

GEOSCIENCE AND REMOTE SENSING

GEOSCIENCE AND REMOTE SENSING

Edited by
PEI-GEE PETER HO

In-Tech
intechweb.org

Published by In-Teh

In-Teh

Olajnica 19/2, 32000 Vukovar, Croatia

Abstracting and non-profit use of the material is permitted with credit to the source. Statements and opinions expressed in the chapters are those of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published articles. Publisher assumes no responsibility liability for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained inside. After this work has been published by the In-Teh, authors have the right to republish it, in whole or part, in any publication of which they are an author or editor, and the make other personal use of the work.

© 2009 In-teh

www.intechweb.org

Additional copies can be obtained from:

publication@intechweb.org

First published October 2009

Printed in India

Technical Editor: Goran Bajac

Geoscience and Remote Sensing,

Edited by Pei-Gee Peter Ho

p. cm.

ISBN 978-953-307-003-2

Preface

Remote Sensing is collecting and interpreting information on targets without being in physical contact with the objects. Aircraft, satellites... etc are the major platforms for remote sensing observations. Unlike electrical, magnetic and gravity surveys that measure force fields, remote sensing technology is commonly referred to methods that employ electromagnetic energy as radio waves, light and heat as the means of detecting and measuring target characteristics.

Geoscience is a study of nature world from the core of the earth, to the depths of oceans and to the outer space. This branch of study can help mitigate volcanic eruptions, floods, landslides... etc terrible human life disaster and help develop ground water, mineral ores, fossil fuels and construction materials. Also, it studies physical, chemical reactions to understand the distribution of the nature resources. Therefore, the geoscience encompass earth, atmospheric, oceanography, pedology, petrology, mineralogy, hydrology and geology.

Normally speaking, remote sensing is a technology that applied on earth as inverse problem. The object around earth may not be directly measured, by using data derived computer model from the equipments on aircraft or satellites, some other information may well be detected and observed. Take one simple example for remote sensing, while it is impossible to directly measure temperatures in the upper atmosphere, it might very well be possible to measure the spectral emissions from carbon dioxide in that region. Through thermodynamics principle, the frequency of the emission can be related to the temperature in that area.

The remote sensing field has experienced rapid growth in recent the years. The advent of satellites has provided the opportunity to acquire global and synoptic information about the environment and its associated phenomena. The wide coverage capability allows for monitoring the rapid changed phenomena on a large scale. The repetitive capability allows the observation of seasonal, annual and long term changes. The global capability allows the viewing of regional and large-scale structures. In addition, remote sensing also allows measurement of hurricanes, bottom of ocean, volcano and mountainous terrain... etc hazardous regions that might not be easily accessible.

This book covers latest and futuristic developments in remote sensing novel theory and applications by numerous scholars, researchers and experts. It is organized into 26 excellent chapters which include optical and infrared modeling, microwave scattering propagation, forests and vegetation, soils, ocean temperature, geographic information, object classification, data mining, image processing, passive optical sensor, multispectral and hyperspectral sensing, lidar, radiometer instruments, calibration, active microwave and SAR processing. With rapid technological advances in both sensors and computing, signal processing and image processing are playing increasingly important roles in remote sensing. In chapter 15, 16 and 21, the state of the art Kernel Learning Machine method, Support Vector Machine and Maximum A Posteriori statistical classification schemes are presented. The use of electromagnetic waves for remote sensing can be separated into active and passive remote sensing. Active remote sensing utilizes an external source to irradiate the object or phenomena. The interaction of the object and radiation is used to extract information about the object. Most active remote sensors such as radars operate at the microwave and millimeter wave frequencies. On the other hand, the optical systems such as lidars are getting more commonly use. In passive remote sensing, the natural radiation properties of the object are utilized to extract information. Sensors that handle illumination from sunlight are generally categorized as passive sensors. They cover the entire electromagnetic spectrum from microwave frequencies to the visible region and beyond. Due to the fact that the strength of natural sources is weak, passive sensors require very sensitive detectors as compared with active sensors. Chapter 4, 12, 13, 18, 20, 25 and 26 contain detailed contents of these particular active/passive sensors' new technologies. In chapter 14, the important and newer remote sensing techniques for land cover change detection are introduced. Nevertheless, in chapter 3, the development of the high resolution wireless sensor network for monitoring volcanic activity is very well introduced.

A picture worth thousand words. In this book, many pictures and graphs has been included in each chapter concisely to convey information about positions, sizes and correlations between objects in the remote sensing. They portray information on things that can be recognized as objects. These objects in turn can deliver deep levels of meaning. We think that humans normally possess a high level of proficiency in deriving information from such images that we might experience less difficult in interpreting even those scenes that are visually complex.

Last but not the least, this book presented chapters that highlight frontier works in remote sensing information processing. I am very pleased to have leaders in the field to prepare and contribute their most current research and development work. Although no attempt is made to cover every topic in remote sensing and geoscience, these entire 26 remote sensing technology chapters shall give readers a good insight. All topics listed are equal important and significant.

Pei-Gee Peter Ho
*DSP Algorithm and Software Design Group,
Naval Undersea Warfare Center
Newport, Rhode Island, USA*

Contents

Preface	V
1. Remote Sensing of the Ecology and Functioning of the Mekong River Basin with Special Reference to the Tonle Sap <i>Benger and Simon Nicholas</i>	001
2. Remote Sensing of Forest Health <i>Jyrki Tuominen, Tarmo Lipping, Viljo Kuosmanen and Reija Haapanen</i>	029
3. Development of a High-Resolution Wireless Sensor Network for Monitoring Volcanic Activity <i>José Chilo, Andreas Schlüter and Thomas Lindblad</i>	053
4. On Position and Attitude Estimation for Remote Sensing with Bistatic SAR <i>Stefan Knedlik, Junchuan Zhou and Otmar Loffeld</i>	075
5. Unmanned Airborne Platforms For Disaster Remote Sensing Support <i>Vincent G. Ambrosia and Steven S. Wegener</i>	091
6. The Geomorphometry of Rainfall-Induced Landslides in Taiwan Obtained by Airborne Lidar and Digital Photography <i>Jin-King Liu, Kuan-Tsung Chang, Jiann-Yeou, Wei-Cheng Hsu, Zu-Yi Liao, Chi-Chung Lau and Tian-Yuan Shih</i>	115
7. Description and Publication of Geospatial Information <i>Arturo Beltran, Laura Díaz, Carlos Granell, Joaquín Huerta and Carlos Abargues</i>	133
8. Application of Real Time GIS, Remote Sensing and IC Tag for Realization of Geospatial Informationsociety <i>Shikada Masaaki, Takeuchi Sayaka, Shimano Sota and Moriya Mitoshi</i>	153
9. Integrated sea surface temperature products within a coastal ocean observing system <i>Nadya T. Vinogradova</i>	181
10. Soil Backgrounds Impact Analysis on Chlorophyll Indices Using Field, Airborne and Satellite Hyperspectral Data <i>A. Bannari and K. Staenz</i>	197

11. Simultaneous Estimation of Optical Properties of Asian Dust and Ground Reflectance by Polarization Measurements 229
Takashi Kusaka and Ryuichi Taniguchi
12. Moving Target Detection and Velocity Estimation in Multi-Channel AT-InSAR Systems from Amplitude and Phase Data 241
Alessandra Budillon
13. Monitoring tropical peat swamp deforestation and hydrological dynamics by ASAR and PALSAR 257
Dirk Hoekman
14. Multivariate Differencing Techniques for Land Cover Change Detection: the Normalized Difference Reflectance Approach 277
Paolo Villa, Giovanmaria Lechi, Mario A. Gomasasca
15. Using Kernel Methods under a Learning Machine Approach for Multispectral Data Classification. An Application in Agriculture 301
Adrián González, José Moreno, Graham Russell and Astrid Márquez
16. Multivariate Time Series Support Vector Machine for Multispectral Remote Sensing Image Classifications 323
Pei-Gee Peter Ho
17. Surface approximation from rapidly varying data: Applications to geophysical surfaces and seafloor surfaces 347
Apprato Dominique, Gout Christian and Le Guyader Carole
18. Three-Dimensional Microwave Imaging using Synthetic Aperture Technique 375
Shi Jun, Zhang Xiaoling, Yang Jianyu, Liao Kefei and Wang Yinbo
19. Corn Monitoring and Crop Yield Using Optical and Microwave Remote Sensing 405
Jesus Soria-Ruiz, Yolanda Fernandez-Ordonez and Heather McNairn
20. Radargrammetric SAR image processing 421
Stéphane Méric, Franck Fayard and Éric Pottier
21. MAP Classification of a Reference Image Using Auxiliaries Images with Different Prevalent Classes 455
Orlando Alves Máximo and David Fernandes
22. Optical Satellite Volcano Monitoring: A Multi-Sensor Rapid Response System 473
Kenneth A. Duda, Michael Ramsey, Rick Wessels and Jonathan Dehn
23. The Extended Integral Equation Model IEM2M for topographically modulated rough surfaces 497
Jose Luis Alvarez-Perez
24. Microwave Remote Sensing of Soil Moisture in Semi-arid Environment 529
A. K. M. Azad Hossain and Greg Easson

25. Ensemble of retrieval algorithms and electromagnetic models for soil and vegetation water content estimation from SAR images 555
Claudia Notarnicola
26. Methodology for investigation of the factors for georadar signals influencing the directional pattern of synthetic aperture radar 579
Zolotarev I.D. and Miller Ya.E.

Remote Sensing of the Ecology and Functioning of the Mekong River Basin with Special Reference to the Tonle Sap

Benger and Simon Nicholas
*Flinders University
Australia*

1. Introduction

The management of large transnational river basins is subject to a range of challenges stemming from differing national priorities, governance of land use activities and resource use, and differences in institutional capacity, data gathering and data sharing. Over vast, often inaccessible areas, remote sensing allows for rapid assessment of ecological resources and hydrological processes. This includes quantification of the extent and ecological functioning of vegetation communities, defining the distribution, duration and timing of flooding, measurement of water quality parameters, groundwater assessment, habitat assessment, and predictive modelling of the ecological impacts of landuse activities and changes to hydrological cycles. Remote Sensing technologies currently allow unparalleled capability for environmental monitoring and management. Data recording and delivery systems, sensor platforms, and sensor technology are constantly improving and each year deliver better remote sensing products for a wide array of applications. Largely independent of geopolitical constraints and boundaries, remote sensing systems allow investigation and analysis of water resources and ecosystem functioning and processes at a range of scales. Large transnational river basins such as the Mekong River basin, can be studied in their entirety or in part.

This chapter examines the use of remote sensing techniques in various investigations in the Mekong River Basin, with particular reference to work on the Tonle Sap (Great Lake) of Cambodia.

1.1 The Mekong River basin

The Mekong is the 10th largest river basin in the world in terms of mean annual outflow, with an annual discharge of 475 billion m³ (Daming, 1997). From its source on the Tibetan Plateau, it flows some 4,800 km south to the Mekong Delta in Vietnam, draining a total catchment area of 795,000 km² (MRC, 2005). The Mekong River Basin spans the six countries of China, Myanmar, Lao PDR, Thailand, Cambodia and Vietnam and forms the major hydrological resource for Southeast Asian. The basin has always faced the challenges of

widespread poverty, increasing demands on water and environmental resources, and conflict throughout the region (Jacobs, 2002). There is lack of coordinated management of the basin, although the Mekong River Commission (MRC), and its predecessors the Mekong Committee and the Mekong Interim Committee have sought to foster dialogue between the member countries since the late 1950s. The main achievement of the MRC, however, has been the development in recent decades of an extensive data gathering and dissemination system, flood forecasting and warning systems, and advancing the understanding of the ecological and physical attributes of the basin (Jacobs, 2002).

Flow and runoff in the Mekong is strongly seasonal, reflecting the influence of the annual monsoon in the lower reaches of the basin. The wet season peaks in September-October with flows in the lower basin of 20,000-30,000 m³s⁻¹, compared to dry season flows of approximately 2,000 m³s⁻¹, which are derived mainly from snow melt in the upper basin (Mekong Secretariat, 1989). The Mekong is subject to natural annual variability which affects the size of the flood peak in any given year and is driven primarily by El Niño Southern Oscillation (ENSO) events (Kiem et al. 2004). Future flood pulse activity may be threatened, however, with significant water resources development occurring throughout the Mekong basin, along with the uncertain effects of climate change on precipitation and river flows. Development and water impoundment and extraction upstream on the Mekong, particularly in southern China but also in Laos, Thailand and Vietnam, is thought to be affecting the size, timing and intensity of the monsoonal flood pulse (Blake, 2001; Osbourne, 2006). Although catchments in China account for approximately one fifth of the flows in the Mekong overall, they can contribute 70-80% of flows during the dry season (MRC, 2005). The two main dams built by China on the upper reaches of the Mekong are the Manwan dam, which was completed in 1993, and the larger Dachaosan dam, which was completed in 2003. Campbell et al. (2006) show a reduction in average flood height and flooded area over the past decade. One of the most significant hydrological features of the lower reaches of the Mekong basin is the Tonle Sap lake in Cambodia, which fills annually and plays an important role in flood attenuation and sediment and nutrient exchange from the Mekong (MRC, 2005). Events occurring in the upper reaches of the Mekong that systematically alter the flood hydrograph or change its timing are likely to have significant effects on the sustainability of the Tonle Sap (Kummu et al. 2004).

1.2 The Tonle Sap

The Tonle Sap or Great Lake of Cambodia (Figure 1) forms part of a unique and ecologically significant sub-system within the Mekong basin. It is the largest freshwater lake in Southeast Asia, covering an area of 250,000-300,000 Ha during the dry season and up to 1.6 million Ha during the wet season (ADB, 2002). Expansion of the lake during the wet season is due primarily to the annual monsoonal flood pulse moving down the Mekong and entering the lake through the Tonle Sap River, which reverses its course as the water level in the Mekong rises above that of the lake. Besides drainage from the Mekong during the monsoonal flood, 13 other catchments drain into the lake. The lake plays an important role in flood peak attenuation and flow control to the Mekong Delta, storing up to 40 km³ of Mekong floodwater each year and releasing it slowly back into the system (MRC, 2005). It was listed as a UNESCO Biosphere reserve in 1997, and is designated as a Protected Area under Cambodian Royal decree and through numerous international agreements. By far the largest

infrastructure has probably been subsumed into more recent schemes, or obliterated by the flooding cycles of the lake, several examples of the largest ancient structures remain. These include the Domdek channel: a 200m wide channel extending approximately 80km through the floodplain with 10-15 m high walls; and the Western Baray at Angkor; a water storage covering 17.5 km².

The inflow from the Mekong accounts for approximately 70 % of flow into the Tonle Sap lake (Penny, 2006), with the remainder coming from local catchments. Some 80 % of the sediments and nutrients entering the lake from the Mekong are retained (MRC, 2005) and this annual process supports floodplain and fisheries productivity. The Tonle Sap lake is therefore highly susceptible to changes in the size, timing and duration of the annual monsoonal flood pulse, whether that occurs as a result of climate change or upstream water resources development. The past decade has seen reductions in flood height and flooded area of the lake (Campbell et al. 2006), although 2008 saw larger than normal floods throughout the Mekong. Kiem et al. (2008) in their latest modelling, suggest that precipitation will increase by 4.2% on average throughout the Mekong basin, concentrated in the upper sections of the basin in China. Chinvanho (2003) suggests that while there will be some shift in the timing of the flood peak, flooding durations will still be adequate for the survival of significant wetland areas on the Tonle Sap.

Most management efforts on the Tonle Sap to date have focussed on maintaining the lake's fisheries, which provide up to 70% of the protein intake for the entire Cambodian population (van Zalinge et al. 2000), and protection of the Ramsar wetlands as bird nesting sites. Natural resource management is severely under-resourced and occurs in a piecemeal manner (Bonhuer and Lane, 2002) in the face of poorly delineated jurisdictions and conflicting economic interests. Despite the importance of the Tonle Sap lake to the Cambodian economy, only in recent years have authorities and research agencies begun to characterise the flooding cycles of the lake or map floodplain vegetation distributions. Some modelling of lake hydrology was completed in 2003 (Koponen et al. 2003) and an Asian Development Bank project is currently underway to produce GIS datasets of lake resources (ABD, 2002). The Cambodian Mekong National River Commission (MNRC) in association with the multi-country Mekong River Commission (MRC) now monitor flood conditions in the Mekong and the Tonle Sap, but data is restricted to a limited number of gauging stations and is often not reliable. For example, the nearest MRC gauging station is located at Kampong Chhnang, on the Tonle Sap tributary (Figure 1).

2. Remote Sensing of Floodplain Structures

Many extensive water impoundment structures as part of irrigation schemes have been built throughout the Tonle Sap floodplain to retain flood waters and support dry season rice cropping. Such anthropogenic modification of the floodplain occurs primarily on the northern margins of the lake in closer proximity to larger settlements. It is likely that these structures have a significant impact on floodwater distribution and movement and will simply serve as flood barriers if peak lake levels are diminished. Floodplain structures result in permanent inundation of large areas that were previously subject to a wetting and drying cycle; essential for the maintenance and survival of many plant and animal species,

including many economically important fish species. In addition, retention and restriction of floodwater movement inhibits nutrient exchange between the floodplain and the lake, and movement of juvenile fish into the lake and the Mekong. The impoundments disrupt the moving littoral of the lake's flood pulse (Junk et al. 1989) where high turnover rates of organic matter and nutrients occur. The gradient of plant species adapted to seasonal degrees of inundation, nutrients and light no longer experiences the conditions under which it evolved.

An aim of the current study was to use remote sensing to determine the extent of floodplain structures around the Tonle Sap and where they lay in relation to flooding extent and duration. Major structures associated with irrigation schemes located within the annually flooded zone of the floodplain were mapped using WAAS corrected GPS to an accuracy of 2-3 m during fieldwork in 2006. High resolution Japanese/NASA ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) imagery was obtained over the floodplain for a range of wet and dry-season dates. ASTER senses in 14 spectral bands in the visible, shortwave and thermal infrared, at 15 m, 30 m and 90 m resolutions respectively (Lillesand et al. 2008). From the 37 ASTER multi-spectral surface reflectance product images obtained for the study, a mosaic of 11 dry-season images covering the Tonle Sap floodplain was constructed with rectification carried out using GCPs (Ground Control Points) collected during fieldwork. The available ASTER coverage over the Tonle Sap is fragmented, both spatially and temporally, due to almost perpetual high levels of cloud cover, but it was possible to generate a near-complete mosaic (Figure 2). As most of the structures occurred on the northern shore of the lake, generally they tended to have an east-west orientation. Horizontal spatial filtering was carried out on the imagery to identify and map the extent of major structures. Spatial filters operate on an image to emphasise or deemphasize image data of varying spatial frequencies. Directional first differencing is a simple directional image enhancement technique which improves the delineation of linear features (Lillesand et al. 2008).

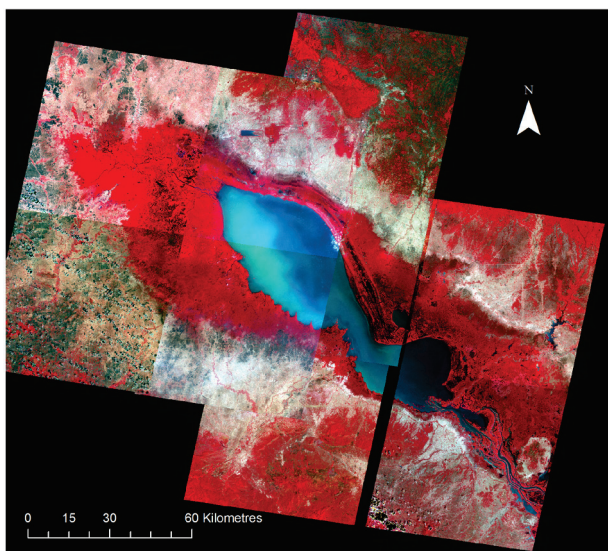


Fig. 2. ASTER Dry Season Image Mosaic of the Tonle Sap Floodplain

Using the filtered images it was possible to identify and map approximately 321 km of major impoundment structures which directly affect water movement across the floodplain (Figure 2). These were generally constructed parallel to the lake shoreline and serve to retain large volumes of water behind them as the lake waters recede after October in any given year. Major structures are defined as being greater than approximately 2m in height, although there are large networks of smaller formal and informal dykes, weirs and regulators which are also used or have been used to modify water movement. Most were built by hand during the Khmer Rouge years using forced labour, and in the absence of any hydrological or engineering knowledge (Kiernan, 1996). Extensive colonisation of these structures with floodplain vegetation has meant that they now form permanent features on the floodplain. According to the flood cycle patterns revealed by the time series analysis described later in this chapter, most of the impoundment structures are built within the zone that would normally be inundated around the end of August in any given year, drying out by mid-December, giving a flood residence time of around 3-4 months (Figure 3). There is also an obvious interaction with floodplain soils. Significant waterlogging occurs around these structures for much of the year, which is a commonly observed phenomenon associated with water storages (Ramireddygarari et al. 2000). This is causing a number of changes to wetlands in these areas. Euphorbiaceae, Fabaceae, and Combretaceae species, which once colonised the mosaic of flooded savannah forest are being replaced by those which can tolerate saturated soils. In the areas behind the dyke walls, which now form permanent water storages, natural wetland species have disappeared completely, due either to blanket infestations of water hyacinth and fringing introduced scrub species.

A secondary impact can also be observed. Irrigated rice fields are present on the lake shore side of most water impoundment structures. Increased nutrient levels associated with the application of fertilisers to the rice fields are likely to be affecting the surrounding wetlands through mobilisation during flooding in the wet season and affecting groundwater quality. Leaching of nutrients into the groundwater from these areas, along with increased utilisation of the groundwater by wetland plants due to higher groundwater levels has created succession towards more nutrient tolerant weeds such as *Mimosa pigra* (Campbell et al. 2006). Similarly, pesticides leaching into groundwater which lies close to the surface are affecting the wetland soils which contain the eggs of hundreds of fish species deposited when the lake is in flood. Changes in predator prey relationships that are important for the ecology of the lake (Scheffer, 1998) and its fisheries are likely to be occurring due to floodwater containment. The impoundments would restrict movement of larger fish into shallow areas of lake for predation during flooding and also form barriers to movement of juveniles out from hatchery zones to the lake and the Mekong system. This undoubtedly contributes to the well-documented reduction in the number of fish species and changes in size of individuals (Puy et al. 1999; Bonheur & Lane, 2002). However, the impoundments are also an important source of protein for the occupants of the floodplain, as they effectively operate as large unmanaged aquaculture sites for much of the year, possibly reducing pressure on lake fish stocks.

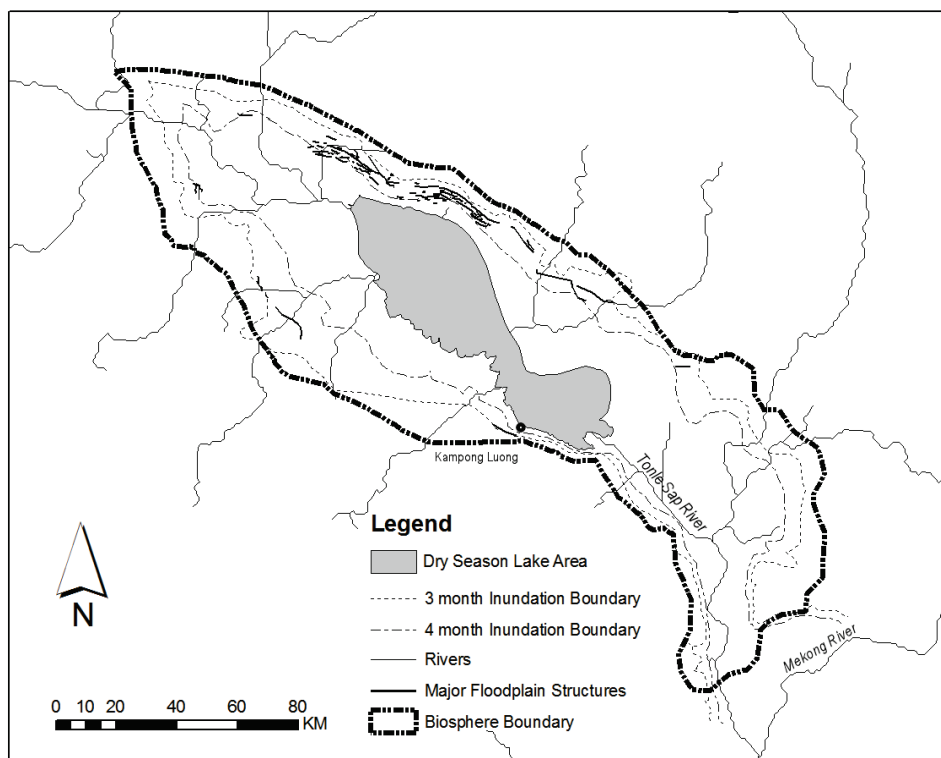


Fig. 3. Major Water Impoundment and Barrier Structures on the Tonle Sap Floodplain

The Khmer Rouge under Pol Pot sought to dramatically increase the areas of land under cultivation on the floodplain, and emptied the cities to provide forced labour for the extensive irrigation schemes that were established (Kiernan, 1996). These structures form by far the largest spatial extent of modifications to the present day floodplain, although many have now been abandoned or are in partial use. Of those surveyed during fieldwork, approximately 40% are now in disuse and others partially used on a seasonally varying basis depending on flooding extent, land availability and population pressures (Bonheur & Lane, 2002). Many of the areas originally modified for rice cultivation have failed to be maintained by the present population because of their inaccessible locations within the floodplain, poor siting, lack of centralised management and maintenance of the schemes, and destruction of infrastructure by flooding. Many of these areas have now reverted to permanent wetlands in areas that would previously have dried out when the floodwaters receded each year. Wright et al. (2004) report on some 570 irrigation schemes existing within the Tonle Sap basin, with only 195 being fully operational today. It is not known how many of these schemes fall within the area of the floodplain, although it is likely that a proportion are located in non-flooded areas. A recent phenomenon on the floodplain is the development of large scale privately owned irrigation schemes which seek to harvest floodwaters for rice production. Substantial areas of floodplain previously utilised by

village communes for lower impact agricultural activities are being modified in this way. The use of ring-dyke structures to harvest flood waters for rice production are seen as the ideal new model for agricultural development of the Tonle Sap floodplain (Someth et al. 2009).

3. Remote Sensing of Groundwater Resources

Remote sensing has been widely used to measure the moisture content of soils (Jensen, 2007), although this often depends on the soil grain size and mineralogy, which will affect the ability of a soil mass to store water. Recently, a number of studies have begun to examine the use of remote sensing for inferring the nature of groundwater resources. Brunner et al. (2007) provide an overview of the potential use of remote sensing in the provision of data to support groundwater modelling in a number of large river basins. Other examples of recent studies include Mutiti et al. (2008), who examined groundwater resource development potential using Landsat imagery, Hendricks Franssen et al. (2008) who inferred groundwater patterning from remotely sensed data, and Milzow et al. (2009) who examined groundwater and hydrology of the large river/wetland system of the Okavango Delta using remote sensing. A range of remote sensing technologies are available to assist in the study of groundwater resources. These include technologies such as radar, LIDAR and digital photogrammetry to derive elevation products, airborne EM (electromagnetics) to examine changes in electrical conductivity in the shallow subsurface, and the remote sensing of vegetation, salt crusts and other surface features as a proxy for subsurface groundwater conditions (Brunner et al. 2007).

Groundwater resources are particularly important for the region in and around the Tonle Sap floodplain, as they form the major water supply for human use (Wright et al. 2004). The sedimentary depression of the Tonle Sap is surrounded by low-lying alluvium, with older coarser ferruginous silts, sands and grits around the perimeter overlain by red-clayey and silty sediments (Stanger et al. 2005). The alluvial deposits of the Tonle Sap floodplain are believed to be very good shallow aquifers, with high recharge rates (5-20 m³/h) and a groundwater table generally within 4-6m of the surface. Groundwater quality is generally good apart from high iron content reducing palatability in some areas, and dangerous levels of arsenic contamination in others (Wright et al. 2004). In response to the large amplitude floods that characterise the hydrological cycle of the Tonle Sap, there is an annual cycle in groundwater levels from depths of around 6 m in riparian areas to a few centimetres in some parts of the floodplain (Stanger et al. 2005).

Loss of vegetation, particularly deep rooted tree species, reduces uptake of water from the soil profile and exacerbates waterlogging problems in the wetlands. A large seasonal population usually migrates from upland areas and the non-flooded areas of the Tonle Sap basin to the floodplain as the floodwaters recede, building temporary settlements on and around the water impoundment structures (Bonheur & Lane, 2002). The temporary settlements facilitate activities such as dry season rice cropping and fishing and informal aquaculture. Human settlement compounds the loss of larger wetland tree species in these areas as they form the primary source of fuelwood and building materials. This occurs on a wide scale despite a complete ban on all forms of timber extraction from the flooded forest

areas. As well as the loss of deep rooted tree species, groundwater levels are also likely to be affected by the permanent and semi-permanent water impoundments, which would have a subsurface connection to the local water table (Ramireddygari et al. 2000). An aim of the current study was to investigate whether these effects existed and were detectable using available optical remotely sensed imagery. Soil moisture absorbs incident radiant energy in the 1.4, 1.9 and 2.7 μm regions, although the spectral response can be complex depending on soil type and soil characteristics (Jensen, 2007).

In three fieldsite locations on the Tonle sap floodplain, the relationship between groundwater and water storages was examined. During fieldwork elevated soil waterlogging adjacent to water storages could be observed through the presence of dark saturated soils along with consequent changes in vegetation type. Trenches were dug adjacent to the structures to ascertain depth to water table, and these confirmed water tables lying at or near the surface. Remotely sensed analysis of Landsat imagery over these areas made it possible to map the extent of waterlogging extending out from these structures. This involved the generation of wetness index maps, using the Kauth-Thomas (KT) transformation (Kauth & Thomas, 1976; Collins & Woodcock, 1996). A wetness index map derived from the KT transformation will indicate not only the level of surface soil moisture, but also the wetness of associated vegetation (Mutiti et al. 2008). A wetness index map for the fieldsite locations examined is presented in Figure 3. While the results do indicate a relationship between the size of the water storage and the area detected, such results are difficult to interpret without further information on the quantity of water stored, the duration of storage, the soil types and localised topography, all of which are unavailable for the Tonle Sap Floodplain. However, they did indicate the potential of remote sensing to detect and quantify these effects, and demonstrate the effects of waterlogging of soils adjacent to water impoundment structures – an important consideration given the rapid agricultural development occurring in some areas of the floodplain utilising water impoundments.

4. Remote Sensing of Floodplain Vegetation

The monsoonal driven flood pulse fills the lake and floods an extensive area of the floodplain, usually for several months from August through to January, creating a unique flooded forest plant community (McDonald et al. 1997). These temporary wetlands serve essential ecosystem processes in terms of nutrient exchange between the lake (and the Mekong system upstream) and the floodplain, and are essential for fish breeding (Puy et al. 1999). Flooded forests are found mainly around the dry-season lake shoreline and comprise about 10% of the floodplain and are dominated by *Barringtonia acutangula*, *Barringtonia micratha* and *Diospyros cambodiana*. At higher elevations are extensive areas of short tree shrubland dominated by species of Euphorbiaceae, Fabaceae, and Combretaceae, together with *Barringtonia acutangula* (Wikramanayake & Dinerstein 2001) and seasonally flooded sedgeland and grasslands occupy the distal margins. Large seasonal contrasts in lake levels affect the characteristics of the wetland vegetation (Penny, 2006), with some forest areas enduring fluctuations of up to 8m and complete canopy submergence for months at a time (McDonald et al. 1997).

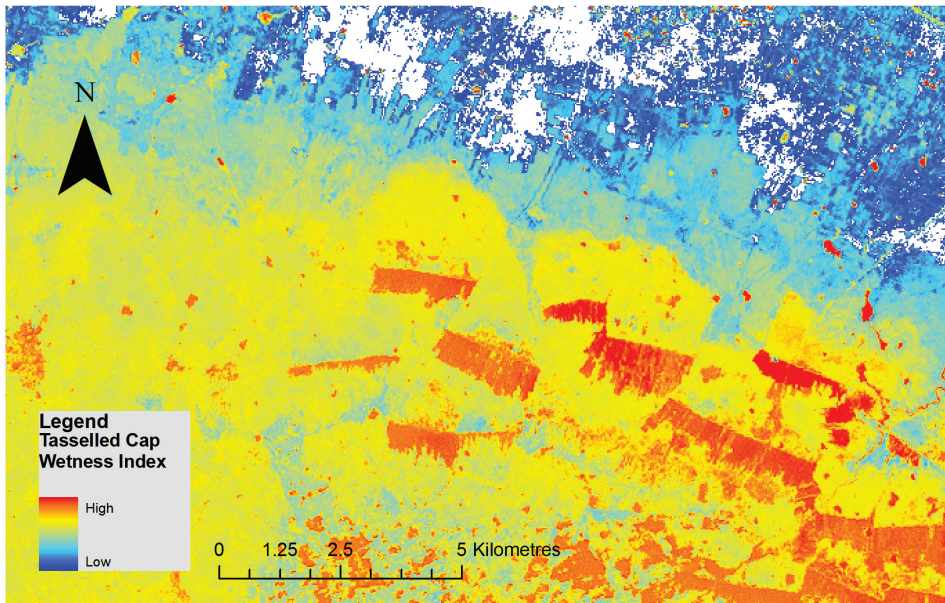


Fig. 4. KT Wetness Index Map of an area of the Tonle Sap Floodplain

Like many ephemeral wetlands around the world, the distribution of the mosaic of flooded forest, scrub and grassland around the lake is determined largely by the duration and depth of flooding (Bonheur & Lane, 2002), and to a lesser degree substrate. The tropical climate, nutrient rich soils and abundant water present on the floodplain mean that vegetative growth occurs rapidly, and forests and wetlands quickly regenerate. It has been suggested that much of the present wetland vegetation is secondary regrowth (McDonald et al. 1997), although this seems unlikely over the majority of the Tonle Sap wetlands, which are highly inaccessible. Only on the northwestern margins of the lake, where the ancient civilizations of Angkor flourished between A.D. 802 and 1431 (Chandler, 1996) is large scale clearing likely to have occurred to facilitate extensive agricultural schemes. Many of these were re-established, often unsuccessfully, during the Khmer Rouge period (Kiernan, 1996). While limited historical data exists on the distribution of plant communities across the floodplain, the majority of the floodplain vegetation is still intact, although often modified in areas closer to settlements, and capable of near normal ecological functioning subject to floodwater availability.

One aim of the current study was to use remotely sensed data to map the current distribution of wetland and floodplain vegetation around the Tonle Sap, so that these could be examined in relation to where they occur spatially on the floodplain in relation to flooding extent and duration. Previous efforts to map wetland vegetation distribution and its relationships to flooding (e.g. Kite 2001) utilised generalised USGS landcover classifications not particularly suited to the Tonle Sap floodplain. The ADB has generated

landuse maps and metrics for all of the Tonle Sap catchments (ADB, 2009), although the landuse categories used are generalised and non-species specific.

Image classification procedures can be used to identify, map and quantify vegetation units of interest in remotely sensed imagery. The overall objective of image classification procedures is to automatically categorise all pixels in an image into land cover classes or themes (Lillesand et al. 2008). Image classification attempts to use the spectral information available in the data for each pixel as the numerical basis for categorisation. Different feature types will manifest different combinations of spectral response in each band (depending on sensor type) based on their inherent spectral reflectance and emittance properties which may also be variant in space and time. Spectral pattern recognition refers to the family of classification procedures that utilise this pixel-by-pixel spectral information as the basis for automated land classification (Lillesand et al. 2008).

Supervised classification is the procedure most often used for quantitative analysis of remote sensing image data. It rests upon using suitable algorithms to label the pixels in an image as representing particular ground cover types or classes (Richards & Jia, 2006). The multidimensional normal distribution of a spectral class is specified completely by its mean vector and its covariance matrix. Consequently, if the mean vectors and the covariance matrices are known for each spectral class then it is possible to compute the set of probabilities that describe the relative likelihoods of a pattern at a particular location belonging to each of those classes (Lillesand et al. 2008). It can then be considered as belonging to the class which indicates the highest probability. Therefore, if the mean vectors and the covariance matrix are known for every spectral class in an image, every pixel in the image can be examined and labelled corresponding to the most likely class on the basis of the probabilities computed for the particular location for a pixel. Before that classification can be performed however, the mean vectors and covariance matrix are estimated for each class from a representative set of pixels, called a training set. These are pixels which the analyst knows as coming from a particular spectral class.

Supervised classification consists therefore of three broad steps. First a set of training pixels is selected for each spectral class using the reference data available in the form of digital vegetation maps. The second step is to determine the mean vectors and covariance matrices for each class from the training data. This completes the learning phase. The third step is the classification stage, in which the relative likelihoods for each pixel in the image are computed and the pixel labelled according to the highest likelihood (Richards & Jia, 2006). Numerous mathematical approaches have been developed for spectral pattern recognition and it is beyond the scope and relevance of this chapter to review them all. Some commonly used classifiers are the Minimum-Distance-to-Means, parallelepiped, Gaussian Maximum Likelihood Classifier (MLC) and the Piecewise Linear Classifier. In the current study, MLC was used for the supervised classification of the ASTER optical imagery. MLC has a demonstrated reliability in achieving accurate classification of land cover types across a range of different environments (Bolstad & Lillesand, 1991; San Miguel-Ayaz & Biging, 1997).

MLC is one of the most commonly used supervised classification methods and it has been demonstrated to be extremely powerful and efficient in a great number of investigations (Maselli et al. 1990). It works most effectively when dealing with normal distribution in the spectral data, although it has also been shown to be relatively resistant to class distribution anomalies (Hixson et al. 1980; Yool et al. 1986). This classifier quantitatively evaluates both the variance and the covariance of the category spectral response patterns. It assumes a Gaussian distribution in the category training data, which is generally a reasonable assumption. Using this assumption the distribution of a category response pattern can be completely described by the mean vector and covariance matrix, it is possible to compute the statistical probability of an unknown pixel belonging to particular land cover class. In essence the maximum likelihood classifier delineates ellipsoidal "equiprobability contours" in the scatter diagram of spectral values which act as the decision regions (Lillesand et al. 2008).

The main limitation of maximum likelihood classification is the large number of computations required to classify each pixel. This is particularly true when either a large number of spectral channels are involved or a large number of spectral classes must be differentiated. Numerous extensions and refinements of the maximum likelihood classifier have been developed (Lillesand et al. 2008). These include the use of lookup tables in which the category identity of all possible combinations of digital numbers is determined prior to classifying the image and each unknown pixel is classified simply by reference to these lookup tables. This avoids the need to carry out complex statistical calculations for each pixel as they have already been determined for each category. Another means of optimizing maximum likelihood classifiers is to use some method to reduce the dimensionality of the dataset used to perform the classification. Procedures such as the principal components, canonical components (Jensen and Waltz, 1979) and tasseled cap (Kauth and Thomas, 1976) transformations achieve this reduction of the dataset by making use only of the significant sections of the data.

Floodplain vegetation type and distribution were observed and mapped in the field in and around the Tonle Sap through a number of fieldwork surveys conducted in 2005 and 2006, in accessible locations. Remote sensing offers the ability to map landcover types over large areas, based on spectral information collected from representative vegetation communities and other landuse and landcovers (Lillesand et al. 2008). On the basis of training sites mapped throughout the floodplain during fieldwork, wetland vegetation and landcover across the floodplain was classified into 20 classes using maximum likelihood classification on the 9 visible/near-infrared and shortwave infrared bands of the ASTER imagery, which were resampled to 30 m. This facilitated determination of the types and extent of wetland vegetation directly affected by the water impoundment structures and their relationship to flooding patterns. Classification accuracy was assessed using standard confusion matrices to generate overall accuracy and Kappa statistics (Congalton & Green, 2008), using one training site for each landcover type not used in the original classification. As a result of the classification carried out over the floodplain using the imagery, it was possible to generate floodplain metrics for the various vegetation and landuse classes, and these are presented in Table 1.

Vegetation/Landuse Class	Area (ha)	Percentage
<i>Barringtonia acutangula</i> dom. Flooded Forest	107928	7.75
<i>Barringtonia acutangula</i> dom. Savannah	765737	55.01
<i>Diospyros cambodiana</i> dom. Savannah	184344	13.24
Euphorbiaceae Shrubland	53794	3.86
Tiliaceae Shrubland	61062	4.39
<i>Mimosa pigra</i>	7551	0.54
Sedge	17810	1.28
Phragmites Reeds	36928	2.65
Thornbush	3380	0.24
Water storage, unvegetated	19257	1.38
Water storage, vegetated	11345	0.81
Agricultural - rice	34972	2.51
Agricultural - fallow	11839	0.85
Legume cropping	250	0.02
Grasslands	67410	4.84
Mudbanks saturated soil	2722	0.20
Bare dry soil	2344	0.17
Firescar	1569	0.11
Rock outcrop	183	0.01
Human settlement	1664	0.12
Total (excluding Lake Area)	1392089	100.00

Kappa = 0.83

Table 1. Landcover classification results for the Tonle Sap floodplain

5. Relationships between Elevation and Floodplain Vegetation

High quality digital elevation data are essential for the assessment of floodplains and spatial arrangement of vegetation communities. Numerous studies of wetland vegetation have suggested that elevation is a primary determinant of vegetation type and location within wetland systems (Scoones, 1981; Hughes, 1990), and substrate to a lesser degree. Analysis of elevation data can yield important information on the spatial arrangement of vegetation in wetland and floodplain environments as it determines the extent and duration of flooding of these areas. In many of the developing countries which comprise the Mekong River basin high quality survey data is simply not available, and over large inaccessible areas such as the Tonle Sap floodplain, ground based survey is logistically impossible. Therefore remote sensing offers the primary means of gathering such data.

There are a range of remote sensing techniques available for the generation of elevation data or digital elevation models (dems). In general, higher precision in these products is

accompanied by higher cost of acquisition and processing. Techniques include digital stereo photogrammetry, radar interferometry and LIDAR. For the Mekong basin the primary dataset that has been utilised is the United States Geological Survey (USGS) GTOPO 30 dem, which is a 30 arc-second resolution product. The Shuttle Radar Topography Mission (SRTM) global product can also be used which has a 3 arc-second (approximately 90 m) resolution with 5 m vertical accuracy (Slater et al. 2006), and more recently, the ASTER GDEM global dem became available in 2009 with 30 m resolution and 15 m vertical accuracy. Of these datasets, only the latter is suitable for use in a low relief environment such as the Tonle Sap floodplain. In the GTOPO 30 and SRTM data, variations in floodplain relief are dominated by data anomalies. In all cases where remotely sensed elevation data are available, finer resolution dem data can be interpolated, but these may lead to a false representation of precision as they will normally retain the errors present in the original data (Longley et al. 2007). Kite (2001) used the USGS GTOPO 30 product for hydrological modelling of the Mekong Basin, and the ADB (2009) show flooded area maps and metrics for the Tonle Sap catchments interpolated from contour maps. In the current study, remotely sensed elevation data was utilised to investigate the relationship between the primary wetland and floodplain vegetation types and elevation, and hence relationship to flooding. For this purpose the ASTER GDEM product was used after processing to remove anomalies, most of which occur over areas of open water and along tile edges, and extracting only elevations below 30m in height. The resultant 30m dem for the Tonle Sap floodplain is shown in 3D in Fig 5.

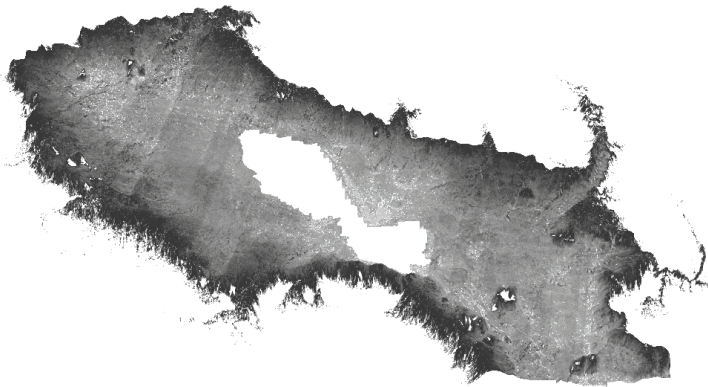


Fig. 5. 3D dem of Tonle Sap floodplain, derived from ASTER GDEM

A simple GIS-based analysis of the location of vegetation communities in relation to elevation yields information on the elevation ranges they occupy within the floodplain. While elevation alone is not the sole determinant of flooding effects on vegetation, in a floodplain such as the Tonle Sap, where overbank flooding from the lake is the primary source of floodwater, it does indicate the sensitivities of various ecological communities to water level ranges. The depth and duration of flooding for these communities is a primary determinant of their evolution in a given location and their ecological functioning (Campbell et al. 2006). The main vegetation classes and their elevation ranges derived from this analysis are shown in Table 2.

Vegetation Class	Minimum Elevation (m)	Maximum Elevation (m)	Elevation Range (m)
<i>Barringtonia acutangula</i> dom. Flooded Forest	0.6	1.8	1.2
<i>Barringtonia acutangula</i> dom. Savannah	1.9	8.2	6.3
<i>Diospyros cambodiana</i> dom. Savannah	2.4	8.7	6.3
Euphorbiaceae Shrubland	5.3	10.4	5.1
Tiliaceae Shrubland	6.2	12.6	6.4
Sedge	8.2	11.3	3.1
Phragmites Reeds	1.0	2.6	1.6
Grassland	9.1	18.4	9.3

Table 2. Elevation ranges for primary vegetation classes

The elevation ranges of the primary vegetation classes of the Tonle Sap confirm that the flooded forest and reed communities occupy lower elevations on the floodplain, with Savannah woodland communities at higher elevations followed by shrubland, sedge and grasslands. Flooded forest, reed and sedge communities occupy the narrowest elevation ranges on the floodplain, while those communities at higher elevations are most likely to be affected by reductions in flood height. This information can then be used with the information on temporal flood extent patterns described below to characterise the horizontal and vertical arrangement of species on the floodplain. This landscape ecology approach to the understanding of floodplain structure provides important information on ecological functioning. Landscape ecology is based on the hypothesis that the interactions among biotic and abiotic components of the landscape are spatially mediated. Not only are the flows of energy material or species from place to place affected by the locations of the places in the landscape, but these flows then determine the interactions among energy, material and species (Malanson, 1993). A central theme of landscape ecology is that spatial structure controls the processes that continuously reproduce the structure. Landscape ecology is an approach to the study of the environment that emphasizes complex spatial relations. The relative locations of phenomena, their overall arrangement in a mosaic and the types of boundaries between them, become the priorities of study (Forman & Godran, 1986; Ingegnoli, 2002).

6. Flood Detection and Mapping

The monsoonal flood pulse is the primary mechanism affecting productivity in the Tonle Sap lake, wetlands and floodplain. The economically important fisheries of the Tonle Sap are strongly influenced by the maximum flooded area and resultant area of fish feeding and breeding habitat (Webby et al. 2005). Remote sensing of the inundation patterns across the study area therefore formed an important part of the current study. Knowledge of the extent and residence time of floodwaters on the floodplains of major rivers is essential for hydrological and biological studies of these systems, and yet for most areas of the Mekong, this remains largely unknown beyond simple maps of flood extent. For the Tonle Sap, the ADB has compiled maps showing minimum and maximum flood extents for the catchments

around the lake as derived from satellite image interpretation (ADB, 2009). For the areas examined in this study, information on flow rates and stream heights may be available, but because of the low relief and complex hydrology of many wetland areas, these data do not correlate well with inundation patterns. Rates of organic matter production, decomposition and export to the river channel are closely linked to floodplain inundation patterns. Primary production rates in inland wetlands are very high and these communities may cover hundreds of thousands of square kilometres (Matthews and Fung, 1987). In many large river systems with associated extensive wetland areas, the difficulty in determining the extent of flooding makes it difficult to accurately estimate wetland area and characterise vegetation relationships. Ground measurement of flooding in forested wetlands is severely limited by the inaccessibility typical of these areas, where mobility is often hampered by flooding and boggy conditions. Remote sensing offers the ability to detect flooded over such areas, and this is typically done using optical or radar imagery.

With regard to optical remote sensing of inundation and the spectral reflectance of water, probably the most distinctive characteristic is the absorption of energy at near-infrared (NIR) wavelengths. Locating and delineating water bodies with remote sensing data is done most easily at NIR wavelengths because of this property (Lillesand et al. 2008). However, various conditions of water bodies manifest themselves primarily in visible wavelengths. Landsat TM imagery has been used to map floodwater distribution and characteristics (e.g Imhoff et al. 1987; Pope et al. 1992; Mertes et al. 1993, 1995; Johnston and Barson 1993), and optical SPOT data has also been used for floodwater mapping (Blasco et al. 1992).

Remote sensing of flooding may also be hampered by forest canopies that render the land/water boundary invisible to infrared and visible wavelength sensors and by frequent cloud cover during periods of rainfall. These limitations are largely overcome by SAR radar sensors which are unaffected by clouds and can significantly penetrate relatively dense forest canopies (Hess et al. 1990). Passive microwave remote sensing has also proved useful for revealing large-scale inundation patterns, even in the presence of cloud cover and dense vegetation (Choudery 1991, Sippel et al. 1994). The bright appearance of flooded forests on radar images results from double-bounce reflections between smooth water surfaces and tree trunks or branches. Enhanced back scattering at L-band has been shown to occur in a wide variety of forest types and is a function of both stand density and branching structure (Hess et al. 1990). Steep incidence angles (20-30°) are optimal for detection of flooding, since some forests exhibit bright returns only at steeper angles. Backscattering from flooded forests is enhanced by underlying water. For forests of moderate density, L-band returns are dominated by corner reflections between trunks and surface and between branches and surface (Richards et al. 1987a). Scattering from a smooth water surface is specular, whereas that from soil includes a significant diffuse component and therefore the amplitude of returns will be higher for standing water beneath forests.

There is a high degree of structural diversity associated with flooded forests, as they occur on numerous substrates, in both saline and fresh water and at a wide range of latitudes (Matthews and Fung 1987). Most frequently studied have been the swamp forests of the coastal plains of the southeastern United States. Relatively bright L-band returns from semi-permanently to permanently flooded stands have been reported in several studies (e.g. Hoffer et al. 1986, Evans et al. 1986, Wu and Sader 1987). Detection of underlying water in

mangrove swamps was demonstrated by Imhoff et al. (1987) in the Sundarbans region of Bangladesh and by Ford and Casey (1988) in East Kalimantan. Bright returns for seasonally inundated temperate forests are described by Richards et al. (1987b) for *Eucalyptus camaldulensis* forests in Australia. Ford et al. (1986) distinguished flooded varzea forest from non-flooded forest using SIR-B scenes of the Rio Japura in the Amazon Basin.

The forest stands cited above have very diverse structures: canopy depth relative to total tree height, dominant branching angle, and crown shape are quite variable. They also encompass a wide range of leaf type and tree heights. It is clear that stands with low stem densities may appear bright at L-band (Hess et al. 1990). Enhancement has also been shown for stands described as dense or thick (Hoffer et al. 1986, Ford and Casey 1988). Enhanced backscattering from flooded forests thus occurs over a broad range of tree species, canopy structures and stand densities. Richards et al. (1987b) demonstrated that brighter returns from flooded forests are not simply a function of vegetation differences between upland and lowland sites. They were able to clearly distinguish between flooded and non-flooded portions of a single forest type.

The accuracy of flood detection using radar imagery is difficult to determine since most studies of flooded forests focus only on those areas which do yield bright L-band returns. Near or complete absence of backscatter from flooded Maryland swamps with dense canopies has been noted by Krohn et al. (1983). It appears that dense undergrowth may significantly affect double-bounce returns. Ford and Casey (1988) found the opposite to be true, however, in flooded mangrove forests of Kalimantan. They found that open stands of low slender trees did not yield bright returns on SIR-B imagery while adjacent denser mangrove stands did. The above examples suggest that for certain forest types, the extent of flooding beneath the canopy would be underestimated using L-band radar. Overestimation would occur if other targets yielding bright returns were mistaken for flooded forests. Other sources of bright returns would normally be able to be visually distinguished from flooded forest based on shape, pattern, associated features and minimal site knowledge. A more serious source of confusion is non-forest vegetation naturally occurring adjacent to flooded forest. Flooded marshes (emergent herbaceous vegetation) typically appear dark at L-band (Krohn et al. 1983, Ormsby et al. 1985). However, marsh vegetation sometimes yields bright returns very similar to those from flooded forests (Krohn et al. 1983).

The magnitude of enhancement associated with double bounce beneath flooded forests can vary significantly. In many studies, variations in magnitude appear to be the result of differences in stand composition as well as flooding (Hess et al. 1990). The problem of separating backscatter variation caused by differences in vegetation from that caused by flooding was minimised in the study by Richards et al. (1987b), because of the virtually monospecific stands of eucalyptus examined. Backscattering from flooded and non-flooded sites within the forest was estimated to vary by 10.8 dB: a substantial difference. Treating the canopy as a uniform layer of small particles, Engheta and Elachi (1982) estimate the enhancement resulting from the presence of a perfectly reflecting surface beneath the canopy to be 3 to 6 dB. It appears from the literature that L-band radar imagery used in the current study should enable accurate delineation of floodwater boundaries.

Aims of this study in relation to flood detection and mapping were to utilise remote sensing methods to (a.) characterise the flood cycles of the lake; (b.) map the spatial distribution of water across the floodplain, and; (c.) determine the relationship between the flooding cycles of the lake and vegetation distributions across the floodplain. The current flood monitoring and mapping efforts of the MNRC and MRC rely on simple linear models of the relationship between river gauge height collected at only a few locations and maximum annual volume and flooded area. Few of the tributaries which drain the 13 catchments around the lake and make significant contributions to lake volume and flooded area have any gauging stations, and hydrological relationships between these tributaries and the lake are complex (Penny, 2006).

For the current study, regional scale MODIS (Moderate resolution Imaging Spectrometer) data was used to determine inundation patterns. MODIS images in 36 spectral bands at 250 m, 500 m and 1 km resolutions, dependent on wavelength, and is widely used for multiple land and ocean applications which require high frequency temporal coverage (Lillesand et al. 2008). A large time-series of MODIS 500m 8 Day Surface Reflectance imagery collected over the period 2001-2005 was used to characterize the flood cycles during the period June to March, at weekly intervals, where the data was of sufficient quality. The MODIS imagery was subsetting to the area of the Tonle Sap and rectified to the ASTER basemap (Figure 2) with its much higher spatial precision using 6 GCPs per image.

The temporal dynamics of the flooded area for the lake are affected by landcover, infiltration rates, and local catchment inputs and cannot be estimated simply from lake gauge height. Inundation mapping in floodplain environments can be problematic due to the presence of high levels of vegetative cover, shallow inundation over large areas and dark organic rich alluvial soils which can appear inundated when they are not (Pearce, 1995). The methods used to map inundation can have a marked effect on the observed patterns (Frazier et al. 2003). On the Tonle Sap, the use of AIRSAR and JERS-1 radar data has been investigated as means of mapping inundation at localised scales (Milne & Tapley, 2005) but this has not been applied at the scale of the entire floodplain. Usual inundation mapping methods using optical imagery involve use of a ratio of mid-infrared reflectance to a visible band reflectance (Lillesand et al. 2008) although this is generally only suitable for relatively deep water. Investigations of techniques for floodplains suggest a combination approach using this ratio and mid-infrared (MIR) change detection is necessary to deal effectively with the shallow water problem (Sims, 2004). Due to the unique nature of the floodplain vegetation and shallow inundation over much of the wet season lake area, a specialised flood detection algorithm was developed for the Tonle Sap using MODIS B6/B4 ratio combined with a B1 threshold, and the accuracy of the technique was verified using the wet-season ASTER imagery. An example of the output from this analysis for a single image date is shown in Figure 6. The MODIS time series was used to determine the extent of flooding and flood duration in conjunction with hydrological data from the CNMC and Mekong River Commission.

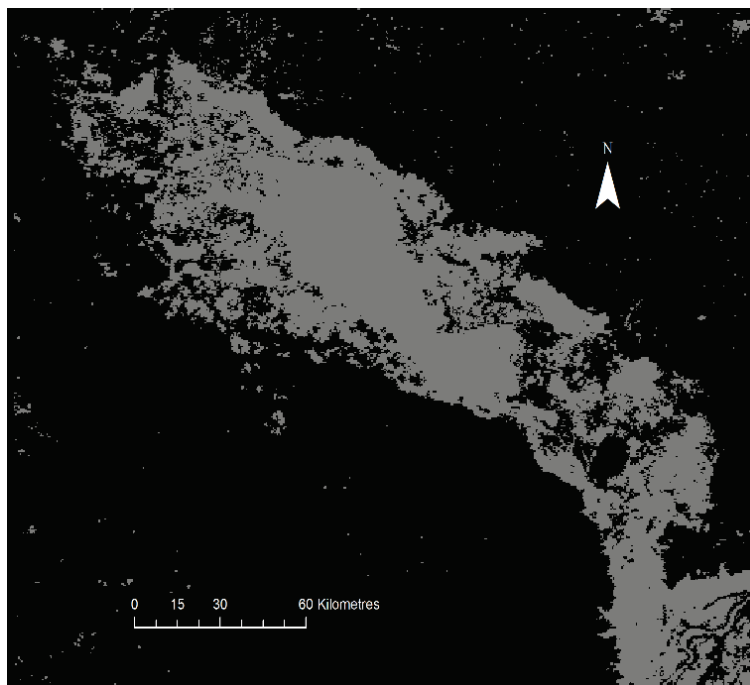


Fig. 6. MODIS derived flood extents for the Tonle Sap – an example

The MODIS derived flood maps indicate a reduction in flooding extent of the Tonle Sap lake since 2000. While the 2000 flood was large by historical standards, and caused widespread damage and loss of life throughout the Mekong Basin (CNMC, 2006), every year since then has been characterised by a reduction in the spatial extent of flooding across the floodplain, apart from the 2008 flood for which the MODIS imagery products are not yet available. This corresponds with MNRC and MRC observations that the flood peaks are now reduced in amplitude and have a much faster fill and drain cycle (CNMC, 2006). Some authors have suggested large dam development throughout the Mekong, and particularly in China, may be responsible (Blake, 2001). The very large Dachaosan dam in southern China began filling in 2003. The monsoons deliver large quantities of water very quickly into the dams where it can be released slowly throughout the year for hydroelectricity generation and for irrigation. The Chinese government currently has another three dams under construction in the upper reaches of the Mekong, with the Xiaowan dam now nearing completion, and another three are at the planning stage (Osbourne, 2006). This will form an 8 dam cascading system capable of retaining very large volumes of water that would otherwise contribute to the monsoonal Mekong flood pulse. With limited fossil fuel reserves and exponential growth in energy demand, the Mekong and other Chinese rivers are seen as offering abundant cheap and clean power. The Chinese dams in the upper reaches of the Mekong are unlikely to be responsible for all reduced flow into the Tonle Sap, as the region may also be experiencing some ongoing effects of drought and climate change (MRC, 2005), and irrigation development is also occurring rapidly on other tributaries which feed the lake. Other current and proposed dams for Laos, Thailand and Vietnam are likely to further ameliorate

the Mekong flood pulse in the future. The output from the analysis of the MODIS time-series was then used to model the effects of inundation variability on the wetland and floodplain vegetation on the Tonle Sap floodplain.

7. GIS Modelling of the Effects of Flooding Changes on Vegetation

The MODIS time-series for the period 2001-2005 shows the area of the Tonle Sap flooded each year and duration of inundation. A goal of the current study was to be able use all the remotely sensed data and derived information on the functioning and spatial arrangement of vegetation and landuse on the floodplain to predict what changes might occur due to interference with the annual flood pulse. This entailed determining the flooding characteristics of floodplain vegetation in terms of depth, timing and duration of flooding and relating these to the spatial distribution of changes in flood patterns. The effects of possible diminished flood peak height and duration on the floodplain were simulated by using an average dry year hydrograph averaged from the four years 1992, 1993, 1999 and 2003 from CNMC data for the Tonle Sap to modify the maximum flood extent model derived from the MODIS imagery. Increasing water use and extraction throughout the Mekong is likely to create move toward dry year conditions with reduced water availability. The average dry year extent was subtracted from the average maximum flood extent derived from the five years of MODIS data for the period 2001-2005, and processed at the resolution of the floodplain dem (30 m). The results show the likely changes in the extent of flooding on the Tonle Sap floodplain if the flooding was likely to be reduced to drier year conditions due to water resource development in the Mekong Basin (Figure 7). When used with the vegetation and landuse cover classifications of the floodplain, this enables GIS modelling of the changes likely to occur in respective landcover types due to reductions in flooded area.

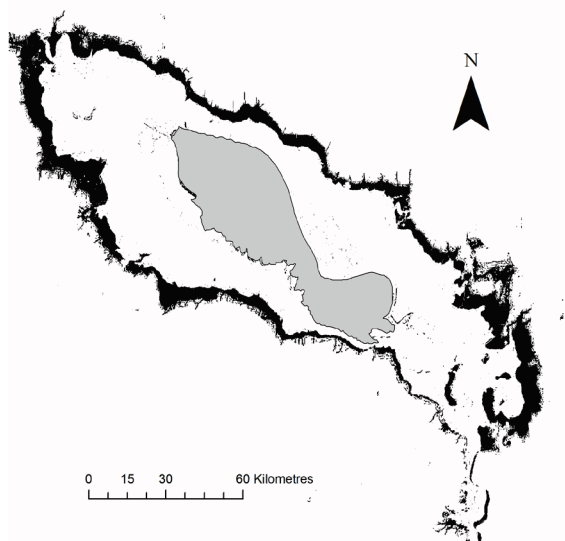


Fig. 7. Modelled reduction in flooded area from MODIS derived flood extents, with change shown in black.

The temporal dynamics of the annual flood event on the Tonle Sap are also revealed in the MODIS data, and show the variability in the duration of the flooding. A cell-based GIS analysis was used to calculate the change in the duration of flooding between the simulated dry year conditions and average conditions. This approach also allows for the determination of how changes occur over temporal cycles, making it possible to develop a dynamic means of estimating changes to vegetation and landuse types. The temporal flood cycle model derived from the MODIS time-series and the dem data were integrated into ArcGIS 9.3 ModelBuilder (Environmental Systems Research Institute, 2009). The time series data provides weekly time steps showing change in flooded area. The mean duration of inundation per cell of landcover type was then generated and this data is summarised in Table 3, showing the change in flood residence time for primary vegetation units between the simulated dry season flood and an average flood. Results indicate that the largest reductions in flood duration will be experienced by the Savannah woodland communities, followed by the shrubland communities, with minor change in the sedge and grassland communities. The core wetland areas of flooded forest and reeds occur at lower elevations and show no reduced flood duration in this analysis. The results of simulating reduced flooding based on the average dry year hydrological data for the Tonle Sap from the GIS-based flood extent model indicate that reductions in flood peak and duration such as those experienced during dry years will have a significant effect on inundation area of the floodplain. This would result in reduction in flooded area reducing from 13,286 km² to 11,134 km², or approximately 16%.

Vegetation Class	Simulated Dry Year Flood Duration (days)	Average Flood Duration (days)	Change (days)
<i>Barringtonia acutangula</i> dom. Flooded Forest	318	318	0
<i>Barringtonia acutangula</i> dom. Savannah	242	311	69
<i>Diospyros cambodiana</i> dom. Savannah	213	288	75
Euphorbiaceae Shrubland	58	79	21
Tiliaceae Shrubland	51	68	17
Sedge	43	52	9
Phragmites Reeds	324	324	0
Grassland	11	23	12

Table 3. Change in flood durations between simulated dry year and average flood conditions for primary vegetation classes

Core areas of wetland on the floodplain including the high conservation value *Barringtonia acutangula* dominated flooded forests are most immune to changes in the flood amplitude as they are subject to greater depths of inundation and these are estimated to decline in area by only 2.5%. However, reduced lake levels and reduced flood duration will mean that normal full canopy submergence may no longer occur or submergence time will be reduced. This may affect productivity and growth characteristics and cause a transition towards shorter trees. Similarly, core emergent reed and grass mat areas will suffer only limited effects and

as they are short rooted and colonise quickly they can more easily make spatial transitions. Flooded woodland savannah, which makes up the majority of the floodplain is likely to be significantly affected, with areas predicted to reduce by some 23%. Grassland and sedge communities on the distal margins of the floodplain will be greatly reduced in area by an estimated 76%, although they are fast disappearing anyway due to human encroachment. In terms of human landuse, dry season cropping area within the flooded zone will be reduced by an estimated 43%, which will displace these activities to other locations, most likely toward lower elevations in the floodplain. The infrastructure associated with dry season cropping will in many cases no longer be viable.

The results of the GIS modelling indicate that a number of habitats within the Tonle Sap floodplain are vulnerable to changes in the monsoonal flood pulse. This will possibly have ramifications throughout the Mekong due to the importance of many areas as fish breeding habitat. These problems will be compounded by the incursion of agricultural activities into core wetland areas as water availability is reduced on the lake margins (Campbell et al. 2006), along with associated land clearance and resource extraction.

8. Conclusion

Remote sensing is able to provide valuable information on the structure, processes and functioning of the Tonle Sap floodplain. Large, inaccessible wetland and floodplain systems such as the Tonle Sap can be studied from space with a range of remote sensing technologies in combination with appropriate fieldwork and reference data. Interference with the natural flood cycles and inundation patterns of the lake and surrounding floodplain are causing changes in vegetation and are likely to be affecting the biological productivity not only on the Tonle Sap but throughout the Mekong system. The myriad impacts occurring in and around the impoundment structures on the floodplain are changing wetland community composition and structure, which in turn will affect fisheries productivity and species biodiversity. Local livelihoods are already affected by fierce (often violent) competition for lake and floodplain resources (Bonheur & Lane, 2002), and as the wetlands and floodplain degrade further this is likely to increase. Historical development of water resources has had significant impacts on the environments and catchments in parts of the floodplain, and caused permanent changes in the hydrology of these areas (Kummu, 2009), and this will continue and accelerate with population growth in the region. Water resource use upstream of the Tonle Sap is potentially reducing and moderating the monsoonal flood pulse which sustains the lake and floodplain system. This may be linked to the timing of large dam construction within the Mekong River basin, although Laos and Thailand are extracting increasing amounts of water from the Mekong as well for use in rapidly expanding rice irrigation schemes (Osbourne, 2006). While social benefits may arise from amelioration of floods which in some years can cause extensive property damage and loss of life, this must be balanced against the need to maintain flood cycles which can sustain the environment of the Tonle Sap, and economic activities such as fishing and agriculture. There is an urgent need to develop effective cross-border management plans and agreements for the water resources of the Mekong system before the unique and economically important Tonle Sap region slips into further decline.

Future events in the Mekong basin, whether related to climate change or human development, will have important ramifications for the Tonle Sap. The annual flood pulse which sustains lake and floodplain ecology is vulnerable to change and as it changes the primary vegetation communities on the Tonle Sap floodplain will most likely face significant declines. In addition, in-situ impacts from upstream developments in the sub-catchments of the lake, as well as further modification of the floodplain will act to reduce water availability and wetland area. The floodplain is already exhibiting signs of over-exploitation (Campbell et al. 2006) and this will increase in line with population and development pressures. It is critical that future basin planning and water resource extraction between the Mekong Basin countries be coordinated in order to preserve the size, duration and timing of the flooding of the Tonle Sap.

9. References

- Asian Development Bank (2002). *Report and Recommendation for the Tonle Sap Environmental Management Project*, ADB Report RRP: Cam 33418.
- Asian Development Bank (2005). *Summary Initial Environmental Examination Report for the Tonle Sap Sustainable Livelihoods Project in Cambodia*, August 2005, ABD.
- Asian Development Bank (2009) *The Tonle Sap Initiative: Future Solutions Now*, Available online: http://www.adb.org/Projects/Tonle_Sap/default.asp
- Benger, S.N. (2006) Groundwater interactions with the wetlands of the Tonle Sap, Cambodia, in *Proc. HydroEco 2006*, Karlovy Vary, Czech Republic, Sept 2006, pp.45-48. ISBN 80-903635-1-2
- Blake, D. (2001). Proposed Mekong Dam scheme in China threatens millions in downstream countries. *World Rivers Review* 4-5, pp.43-51, ISBN 08906211
- Blasco, F.; Bellan, M.F. & Chaudhury, M.U. (1992) Estimating the extent of floods in Bangladesh using SPOT data, *Remote Sensing of Environment* 39, pp.167-178, ISSN: 0034-4257
- Bonheur, N. & Lane, B. D. (2002). Natural resources management for human security in Cambodia's Tonle Sap Biosphere Reserve, *Environmental Science and Policy* 51, pp. 33-42, ISSN: 1462-9011
- Bolstad, P.V. & Lillesand, T.M. (1991) Rapid maximum likelihood classification, *Photogramm. Eng. Remote Sens.* 57, pp.67-74, ISSN: 0034-4257
- Brunner, P.; Hendricks Franssen H.J.; Kgotlhang, L.; Bauer-Gottwein, P. & Kinzelbach, W. (2007) How can remote sensing contribute to groundwater modelling? *Hydrogeology Journal* 15(1), pp.5-18, ISSN 1431-2174
- Cambodia National Mekong Committee (CNMC) (2006). Cambodia Country Report: Flood information in Cambodia, *Proceedings of the 4th Annual Mekong Flood Forum "Improving Flood Forecasting and Warning Systems for Flood Management and Mitigation in the Lower Mekong Basin"*, Siem Reap, Cambodia, May 2006, pp.23-36.
- Campbell, I. C.; Poole, C.; Giesen, W. & Valbo-Jorgensen, J. (2006) Species diversity and ecology of Tonle Sap Great Lake, Cambodia, *Aquatic Sciences - Research Across Boundaries* 68, pp. 355-373, ISSN 1015-1621
- Chandler, D. (1996). *A History of Cambodia*, Westview Press Inc, ISBN 974-7100-65-7, Melbourne.

- Choudery, B.J. (1991) Passive microwave remote sensing contribution to hydrological variables, *Surveys in Geophysics* 12, pp.63-84, ISSN 0169-3298
- Collins, J.B. & Woodcock, C.E. (1996) An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data. *Remote Sensing of Environment* 56, 66-77, ISSN: 0034-4257
- Congalton, R. G. & Green, K. (2008). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, 2nd Edition*, CRC Press, ISBN: 9781420055122, Boca Raton, FL.
- Daming, H. (1997) Facilitating regional sustainable development through integrated multi-objective utilization management of water resources in the Lancang-Mekong river basin, *Journal of Chinese Geography* 7, 4, ISSN 1861-9568
- Engheta, N. & Elachi, C. (1982) Radar scattering from a diffuse vegetation layer over a smooth surface, *IEEE Trans. Geosci. and Remote Sens.* 20, pp.212-216, ISSN: 0196-2892
- Environmental Systems Research Institute (2009) ArcGIS 9.3 ModelBuilder Software, ESRI, Redlands, CA.
- Evans, D.; Pottier, C.; Fletcher, R.; Hensley, S.; Tapley, I.; Milne, A. & Barbetti, M. (2007) A comprehensive archaeological map of the world's largest preindustrial settlement complex at Angkor, Cambodia, *Proceedings of the National Academy of Sciences of the United States of America* 104 (36), pp. 14277-14282, ISSN 1091-6490
- Evans, D.E.; Farr, T.G.; Ford, J.P.; Thompson, T.W. & Werner, C.L. (1986) Multipolarisation radar images for geological mapping and vegetation discrimination, *IEEE Trans. Geosci. and Remote Sens.* 24, pp.246-257, ISSN: 0196-2892
- Ford, J.P. & Casey, D.J. (1988) Shuttle radar mapping with diverse incidence angles in the rainforest of Borneo, *Int. Journal of Remote Sensing* 5, pp.927-943, ISSN: 1366-5901
- Ford, J.P.; Cimano, J.B.; Holt, B. & Ruzek, M.R. (1986) *Shuttle Imaging Radar Views the Earth from Challenger: The SIR-B Experiment*, Jet Propulsion Laboratory publication 86-10, Pasadena, California
- Forman, R.T.T. & Godron, M. (1986) *Landscape Ecology*, Wiley, ISBN: 0471870374, New York.
- Frazier, P., Page, K., Louis, J., Briggs S. & Robertson, A. I. (2003). Relating wetland inundation to river flow using Landsat TM data. *Int. Journal of Remote Sensing* 24(19), pp. 3755-3770, ISSN: 1366-5901
- Hendricks Franssen, H. J.; Brunner, P.; Makobo, P. & Kinzelbach, W. (2008) Equally likely inverse solutions to a groundwater flow problem including pattern information from remote sensing images, *Water Resources Research*, 44, W01419, doi:10.1029/2007WR006097
- Hess, L.L.; Melack, J.M. & Simonett, D.S. (1990) Radar detection of flooding beneath the forest canopy: a review, *Int. Journal of Remote Sensing* 11, pp.1313-1325, ISSN: 1366-5901
- Higham, C. (2001). *The Civilisation of Angkor*, Orion Books, London. ISBN 1 84212 584 2
- Hixson, K., Scholz, D. and Funs, N. (1980) Evaluation of several schemes for classification of remotely sensed data, *Photogramm. Eng. Remote Sens.* 46, pp.1547-1553, ISSN: 0099-1112
- Hoffer, R.M.; Lozano-Garcia, D.F.; Gillespie, D.D.; Mueller, P.W. & Ruzek, M.J. (1986) Analysis of multiple incidence angle SIR-B data for determining forest stand characteristics, *The Second Spaceborne Imaging Radar Symposium*, JPL Publication 86-26, Pasadena CA, pp.159-164

- Hughes, F.M.R. (1990) The influence of flooding regimes on forest distribution and composition in the Tana River Floodplain, Kenya, *Journal of Applied Ecology* 27, pp.475-491, ISSN 1365-2664
- Ingegnoli, V. (2002) *Landscape Ecology: A Widening Foundation*, Springer, ISBN: 978-3-540-42743-8, Amsterdam.
- Imhoff, M.; Vermillion, C.; Story, M.H.; Choudery, A.M. & Gafoor, A. (1987) Monsoon flood boundary delineation and assessment using spaceborne imaging radar and Landsat data, *Photogramm. Eng. Remote Sens.* 53, pp.405-413, ISSN: 0099-1112
- Jacobs, J.W. (2002) The Mekong River Commission: transboundary water resources planning and regional security, *The Geographical Journal* 168, pp. 354-364, ISSN 1861-9568
- Jensen, J.R. (2007) *Remote Sensing of the Environment, 2nd Edition*, Pearson Prentice Hall, ISBN 0-13-188950-8, Upper Saddle River, NJ.
- Jensen, S.K. & Waltz, F.A. (1979) Principal Components Analysis and Canonical Analysis in Remote Sensing, in *Proc. American Photogrammetric Society 45th Annual Meeting*, pp.337-348, ISSN: 0099-1112
- Johnson, R.M. & Barson, M.M. (1993) Remote sensing of Australian wetlands: an evaluation of Landsat TM data for inventory and classification, *Aust. J. Mar. Freshwater Res.* 44, pp.235-252, ISSN: 0067-1940
- Junk, W. J.; Bayley, P. B. & Sparks, R. E. (1989). The Flood Pulse Concept in river-floodplain systems, *Can. Spec. Publ. Fish. Aquat. Sci.* 106, pp. 110-127. ISSN: 1205-7533
- Kauth, R.J. & Thomas, G.S. (1976) The Tasseled Cap - A graphic description of the spectral temporal development of agricultural crops as seen by Landsat, in *Proc. LARS 1976 Symposium on Machine Processing of Remotely Sensed Data*, Purdue University.
- Kiem, A. S.; Ishidaira, H.; Hapuarachchi, H. P.; Zhou, M. C.; Hirabayashi, Y. & Takeuchi, K. (2008) Future hydroclimatology of the Mekong River basin simulated using the high-resolution Japan Meteorological Agency (JMA) AGCM, *Hydrological Processes* 22: 1382-1394, ISSN: 1099-1085
- Kiem A.S., Hapuarachchi H.P. & Takeuchi, K. (2004) Impacts of climate variability on streamflow in the Mekong River: an interesting challenge for hydrological modelling, In: *Proc. River Symposium 2004*, Sept 2004, Brisbane.
- Kiernan, B. (1996). *The Pol Pot Regime – Race, Power and Genocide in Cambodia under the Khmer Rouge, 1975-79*, Yale University Press, ISBN 974 7100 43 6, New Haven
- Kite, G. (2001). Modelling the Mekong: hydrological simulation for environmental impact studies, *Journal of Hydrology* 253, pp. 1-13, ISSN: 0022-1694
- Koponen, J.; Josza, J.; Lauri, H.; Sarkkula, J. & Markku, V. (2003) Modelling Tonle Sap Watershed and Lake Processes for Environmental Change Assessment, Mekong River Commission MRCS/WUP-FIN Model Report.
- Krohn, M.D.; Milton, N.M. & Segal, D.B. (1983) Seasat synthetic aperture radar (SAR) response to lowland vegetation types in eastern Maryland and Virginia, *J. Geophys. Res.* 88, pp.1937-1952, ISSN 0148-0227
- Kummu, M. (2009) Water management in Angkor: Human impacts on hydrology and sediment transportation, *Journal of Environmental Management* 90, 3, pp. 1413-1421, ISSN: 0301-4797
- Lillesand, T.M.; Kiefer, R.W. & Chipman, J.W. (2008). *Principles of Remote Sensing and Image Analysis, 6th Edition*, Wiley, ISBN 978-0-470-05245-7, New York.

- Longley, P.A.; Goodchild, M.F.; Maguire, D.J. & Rhind, D.W. (2005) *Geographic Information Systems and Science, 2nd Edition*, Wiley, ISBN 0-470-87000-1, New York.
- Malanson, G.P. (1993) *Riparian Landscapes*, Cambridge University Press, ISBN-13: 9780521384315, Cambridge.
- Maselli, F.; Conese, C.; Zipoli, G. & Pittau, M.A. (1990) Use of error probabilities to improve area estimates based on maximum likelihood classifications, *Remote Sensing of Environment* 31, pp.155-160, ISSN: 0034-4257
- McDonald, J.; Bunnat, P. & Virak, P. (1997). *Plant Communities of the Tonle Sap Floodplain*, UNESCO/IUCN/WI, Phnom Penh.
- Mekong River Commission (2007) *Annual Mekong Flood Report 2006*, Mekong River Commission, ISSN: 1728 3248, Vientiane.
- Mekong River Commission (2005) *Overview of the Hydrology of the Mekong Basin*, Mekong River Commission, ISSN: 1728 3248, Vientiane, November 2005.
- Mekong Secretariat (1994) *Annual Report 1994*, Mekong Secretariat, ISSN: 1728 3248, Bangkok.
- Mertes, L.A.K.; Daniel, D.L.; Melack, J.M.; Nelson, B.; Martinelli, L.A. & Forsberg, B.R. (1995) Spatial patterns of hydrology, geomorphology, and vegetation on the floodplain of the Amazon River in Brazil from a remote sensing perspective, *Geomorphology* 13, pp.215-232, ISSN: 0169-555X
- Mertes, L.A.K.; Smith, M.O. & Adams, J.B. (1993) Estimating suspended sediment concentrations in surface waters of the Amazon River wetlands from Landsat images, *Remote Sensing of Environment* 43, pp.281-301, ISSN: 0034-4257
- Milne, A.K. & Tapley, I.J. (2005). Change Detection Analysis in the Wetlands Using JERS-1 Radar Data: Tone Sap Great Lake, Cambodia, IEEE doi 0-7803-9119-5/05. pp. 146-150
- Milzow, C.; Kgotlhang, N.; Bauer-Gottwein, P.; Meier, P. & Kinzelbach, W. (2009) Regional review: the hydrology of the Okavango Delta, Botswana - processes, data and modelling, *Hydrogeology Journal*, ISSN 1431-2174, Published Online DOI 10.1007/s10040-009-0436-0
- Mutiti, S.; Levy, J., Mututi, C. & Guturu, N.S. (2008) Assessing Ground Water Development Potential Using Landsat Imagery, *Groundwater*, Published Online DOI 10.1111/j.1745-6584.2008.00524.x
- Ormsby, J.P.; Blanchard, B.J. & Blanchard, A.J. (1985) Detection of lowland flooding using active microwave systems, *Photogramm. Eng. Remote Sens.* 51, pp.317-328, ISSN: 0099-1112
- Osbourne, M. (2006) *River at risk: The Mekong and the water politics of China and Southeast Asia*, Lowy Institute for International Policy, ISBN 1 921004 02 9, New York.
- Pearce, B. (1995). *The compilation of regional flood maps using remote sensing techniques over the Ballonne river catchment and downstream areas*. Technical Report. Queensland Department of Primary Industries, Brisbane, QLD.
- Penny, D. (2006). The Holocene history and development of the Tonle Sap, Cambodia. *Quaternary Science Reviews* 25, pp. 310-322, ISSN: 0277-3791
- Pope, K.O.; Sheffner, E.J.; Linthicum, K.J.; Bailey, C.L.; Logan, T.M.; Kasischke, E.S.; Birney, K.; Nlogu, A.R. & Roberts, C.R. (1992) Identification of the central Kenyan Rift Valley fever virus vector habitats with Landsat TM and evaluation of their flooding

- status with airborne imaging radar, *Remote Sensing of Environment* 40, pp.185-196, ISSN: 0034-4257
- Puy, L.; Lek, S.; Touch, S. T.; Mao, S-O. & Chhouk, B. (1999). Diversity and spatial distribution of freshwater fish in Great Lake and Tonle Sap river Cambodia, Southeast Asia, *Aquatic Living Resources* 126, pp. 379-386, ISSN: 0990-7440
- Ramireddygar, S. R.; Sophocleous, M. A.; Koelliker, J. K.; Perkins, S. P. & Govindaraju, R. S. (2000). Development and application of a comprehensive simulation model to evaluate impacts of watershed structures and irrigation water use on streamflow and groundwater: the case of Wet Walnut Creek Watershed, Kansas, USA. *Journal of Hydrology* 2363-4, pp. 223-246, ISSN: 0022-1694
- Richards, J.A. & Jia, X. (2006) *Remote Sensing Digital Image Analysis - An Introduction, 4th Edition*, Springer-Verlag, Berlin. ISBN: 978-3-540-25128-6
- Richards, J.A.; Sun, G-Q. & Simonett, D.S. (1987a) L-band radar backscatter modelling of forest stands, *IEEE Trans. Geosc. Remote Sens.* 25, pp.487-498, ISSN: 0196-2892
- Richards, J.A.; Woodgate, P.W.; & Skidmore, A.K. (1987b) An explanation of enhanced radar backscattering from flooded forests, *Int. Journal of Remote Sensing* 8, pp.1093-1100, ISSN: 1366-5901
- San Miguel-Ayanz, J. & Biging, G.S. (1997) Comparison of single-stage and multi-stage classification approaches for cover type mapping with TM and SPOT data, *Remote Sensing of Environment* 59, pp.92-104, ISSN: 0034-4257
- Scheffer, M. (1998). *The Ecology of Shallow Lakes*, Chapman and Hill, ISBN: 0-412-74920-3, London.
- Scoones, I. (1991) Wetlands in Drylands: key resources for agricultural and pastoral production in Africa, *Ambio* 20, pp.366-371, ISSN: 0044-7447
- Sims, N. (2004). The Landscape-scale Structure and Functioning of Floodplains, Unpublished PhD Thesis, University of Canberra.
- Sippel, S.J.; Hamilton, S.K.; Melack, J.M. & Choudery, B.J. (1994) Determination of inundation area in the Amazon River floodplain using the SMMR 37 GHz polarisation difference, *Remote Sensing of Environment* 48, pp.70-76, ISSN: 0034-4257
- Slater, J.A.; Garvey, G.; Johnston, C.; Haase, J.; Heady, B.; Kroenung, G. & Little J. (2006) The SRTM data "finishing" process and products. *Photogramm. Eng. Remote Sens.* 72(3), pp.237-247, ISSN: 0099-1112
- Someth, P.; Kubo, N.; Tanji, H. & Lyd, S. (2009) Ring dike system to harness floodwater from the Mekong River for paddy rice cultivation in the Tonle Sap Lake floodplain in Cambodia, *Agricultural Water Management* 96, pp.100-110, ISSN: 0378-3774
- Stanger, G.; VanTruong, T.; Ngoc, K. S.; Luyen, T. V. & Thanh, T. T. (2005). Arsenic in groundwaters of the Lower Mekong, *Environmental Geochemistry and Health* 27, pp. 341-357, ISSN: 1573-2983
- Top, N.; Mizoue, N.; Kai, S. & Nokao, T. (2004). Variation in woodfuel consumption patterns in response to forest availability in Kampong Thom Province, Cambodia, *Biomass and Energy* 27, pp. 57-68, ISSN: 0167-5494
- Van Zalinge, N.; Thouk, N.; Tana, T.C. & Leung, D. (2000). Where there is water, there is fish? Cambodian fisheries issues in a Mekong River Basin perspective. In: Ahmed, M. and Hirsh, P. (Eds) *Common Property in the Mekong: Issues of Sustainability and Subsistence*. ICLARM Study Review.

- Webby, R.; Adamson, P.T.; Boland, J.; Howlett, P.G.; Metcalfe, A.V. & Piantadosi, J. (2005) The Mekong - Applications of Value at Risk (VaR) and Conditional Value at Risk (CVaR) simulation to the benefits, costs and consequences of water resources development in a large river basin. In: *MODSIM 2005 International Congress on Modelling and Simulation*. (ed. by A. Zerger & R.M. Argent), pp.2109-2115, ISBN: 0-9758400-0-2, Modelling and Simulation Society of Australia and New Zealand, December 2005, Brisbane.
- Wikramanayake, E. & Dinerstein, E. (2001) *Terrestrial Ecoregions of the Indo-Pacific*, Island Press, ISBN: 1559639237 Washington DC.
- Wright, G.; Moffatt, D. & Wager, J. (2004). *Establishment of the Tonle Sap Basin Management Organisation: Tonle Sap Basin Profile*, Cambodia National Mekong Committee, Asian Development Bank Report TA2412-CAM.
- Wu, S.T. & Sadler, S.A. (1987) Multipolarisation SAR data for surface feature delineation and forest vegetation characterisation, *IEEE Trans. Geosc. Remote Sens.* 25, pp.67-76, ISSN: 0196-2892
- Yool, S.R.; Star, Y.L.; Estes, J.E.; Botkin, E.B.; Eckardt, D.W. & Davis, F.W. (1986) Performance analysis of image processing algorithms for classification of natural vegetation in the mountains of southern California, *Int. Journal of Remote Sensing* 7, pp.683-702, ISSN: 1366-5901

Remote Sensing of Forest Health

Jyrki Tuominen, Tarmo Lipping, Viljo Kuosmanen* and Reija Haapanen**

*Tampere University of Technology
Finland*

Geological Survey of Finland
Finland*

*HaapanenForestConsulting**
Finland*

1. Introduction

Global forest health is declining. The main reasons for this unfortunate development include climate change, air pollution and increased human activities. There is a need to monitor and quantitatively measure the change in forest health. Forest health can be defined in many different ways depending on the perspective one takes. The relationship between the cause and the symptom of forest health deterioration is complex mainly because the same symptom can often be induced by multiple different stressors.

Modern state of the art remote sensing technologies provide the means for wide coverage measurement of forest health with reasonable accuracy. In the assessment of forest health by means of remote sensing, features called Vegetation Indices (VIs) are usually extracted from the data. VIs are combinations of surface reflectances at two or more wavelengths designed to highlight a particular property of vegetation. In addition to specific VIs some attempts to develop a general forest health index combining the assessments of the various properties have been published.

In this chapter we first discuss the term 'forest health' as well as the various causes and symptoms of forest health deterioration. We then argue about the role of remote sensing in forest health monitoring and present the most common VIs used for the assessment of forest health. The VIs' capability to detect different types of forest damage is assessed in case studies; the results for Ni contaminated and pest inflicted forest areas are presented in sections 7 and 8, respectively. Finally, future possibilities in applying the remote sensing technologies to forest health monitoring are discussed.

2. Forest health

Despite its widespread use, the term "forest health" is vaguely defined in the literature, making its application to forest management difficult (Kolb et al., 1994). The definition of forest health is always a matter of perspective. Social, economic and ecological perspectives

are taken most frequently. Looking from different perspectives the definitions of healthy forest can even appear contradictory.

Social perspective emphasises people's needs for healthy living and recreational environment. In many countries people are becoming more and are more concerned about their environment. Since the 1960s, people have become more concerned about their environment, due to intensified management of natural resources, increased availability of information and structural changes in the society, i.e. decreased importance of primary production and increased leisure time. Healthy forests are needed for aesthetical pleasure and variety of outdoor activities. Economical perspective is quite complex. Historically short-term needs have usually exceeded long-term needs in forest industry. Production of commodities and services has outweighed ecological considerations in the past. Nevertheless, long-term economical benefits can only be obtained by practicing sustainable forest management. Ecological perspective is focused on ecosystems instead of human needs. This perspective emphasises the fact, that counterproductive interaction between ecosystems and humans should be minimized. The potential should exist for all biotic and abiotic elements to be present with sufficient redundancy at appropriate spatial and temporal scales across the landscape. Human intervention should not impact ecosystem sustainability by destroying or significantly degrading components that affect ecosystem capabilities.

Utilitarian perspective versus ecosystem perspective is another categorization presented in literature (Kolb et al., 1994). The utilitarian perspective emphasizes forest conditions that directly satisfy human needs. The ecosystem perspective emphasises the maintenance of sustainable ecosystems over the landscape. Often different perspectives do overlap, however, on occasions they may be contradictory. Depending on the perspective, the condition of the same forest can be viewed as healthy or unhealthy. For example, a common component in ponderosa pine forests is dwarf mistletoe. It reduces the growth of ponderosa pine and increases its mortality. The existence of dwarf mistletoe is harmful from economical perspective. However, abundance and species richness of birds is higher when dwarf mistletoe is present. Thus, the existence of dwarf mistletoe is desirable from the ecological perspective (Kolb et al., 1994).

The first definition of forest health usually cited in the scientific literature is by Aldo Leopold (1949): "Health is the capacity of the land for self-renewal. Conservation is our effort to understand and preserve this capacity." Although Leopold's definition is concerned also with other issues than forest health, it has founded the basis for all later definitions. Since Leopold's definition many new ones have been presented in the scientific literature. Some of these definitions deserve more in-depth consideration.

O'Laughlin et al. (1994) defined forest health in the following way: "Forest health is a condition of forest ecosystems that sustains their complexity while providing for human needs". This definition is an effort to take all the different perspectives of forest health, i.e., social, ecological and economical, into account. The question often asked when discussing forest health is: 'Can forest health be measured?': According to O'Laughlin et al. (1994) the answer is: "Objective indicators of forest condition can be specified and measured, but forest health assessments contain subjective value judgements which must be clearly recognized." Some definitions of forest health like that by Monnig&Byller (1992), for example, emphasize ecological perspective: "A healthy forest is an ecosystem in balance"(Monnig&Byler, 1992). Human social and economical needs are not accounted for and the condition of ecosystems

matters for its own sake. Kolb, Wagner & Covington (1994) defined forest health as follows: "The term forest health should be restricted to the examination of the role of biotic and abiotic agents in ecosystem processes." Several characteristics for such a system were mentioned: resistance to dramatic change in populations of important organisms within the ecosystem not accounted for by predicted successional trends; a functional balance between supply and demand of essential resources; and a diversity of seral stages, cover types and stand structures that provide habitat for many native species and all essential processes. Climate change, in particular increased temperatures and levels of atmospheric carbon dioxide, as well as changes in precipitation and the occurrence of extreme climate events, is having notable impacts on the world's forest health. Increased temperatures may relieve forest stress during colder seasons but increase it during warmer seasons. Impacts of increased temperatures vary widely among different climatic zones. Climate change has both direct and indirect impacts on forest health (Moore&Allard, 2008). For example, climate change has strong influence on forest pests which can be considered as a direct impact. Pests can rapidly react to changing climate because of their short generation times. Drought is a good example of a cause that is influencing forest indirectly. Drought can change tree physiology in a way that it is more vulnerable to certain insect and pest species. Such changes are, e.g., sugar content of foliage, changes in leaf colour, change of leaf thickness and structural changes of foliage.

3. Causes and symptoms of deteriorated forest health

In the scope of this chapter it is only possible to address the complex relationship between causes and symptoms of deteriorated forest health briefly. The relationship is complex mainly because there are often multiple stressors causing a certain symptom (Ferretti, 1997). One of the major causes of deteriorated forest health is air pollution, particularly acid rain and ground-level ozone. There are two major types of pollutant threatening forest health: photochemical oxidants of which ozone is the primary compound, and nitrogen pollutants. Ozone is toxic (plant-killing) to sensitive plant species. Nitrogen is the primary growth-limiting nutrient, yet it is also a pollutant when in excess. The emission levels of these two pollutants are expected to increase significantly globally (Moore&Allard, 2008).

Some causes of deteriorated forest health are controversial. Depending on the perspective they can be seen as beneficial or negative. The role of disease agents (pathogens) and insects, for example, is controversial. They are essential to the function of dynamic ecosystems as they recycle nutrients and create habitats for different species. They can also negatively affect forest health, increase mortality and create growth losses. Diseases and insects influence the health of forests, trees outside forests and other wooded lands. Globally, all ecosystems with tree cover are under increasing threat, as the periods between sequential outbreaks are rapidly decreasing because of a range of factors including climate change and lack of proper forest management (Moore&Allard, 2008).

The impact of forest fires may also be controversial. Although they are usually seen as a threat to forest health, these natural events are key elements in many forest ecosystems as well. Another major cause for deteriorated forest health is posed by droughts. The consequences of a long-lasting drought can be very severe. In addition to direct impacts, they have indirect ones, for example pest outbreaks are associated with droughts. Drought can also increase the risk of forest fires.

Another cause of forest health deterioration is invasive species. Any species non-native to a particular forest ecosystem and whose introduction and spread causes deteriorated forest health can be considered invasive species. A major cause of increasing number of invasive species is increased human activity. Transport vehicles act as carriers of seeds and plants. Sometimes invasive species can be intentionally introduced to an ecosystem to provide economic or environmental benefits. These species have later spread and caused serious problems in forests ecosystems.

There is an abundance of symptoms of deteriorated forest health. Some symptoms are easy to detect and follow while others might be very difficult to monitor. Discoloration is a good example of a symptom, which is relatively easy to monitor. Discoloration is usually also a very useful index of forest health. Growth losses, in turn, are difficult to measure from large forest areas. Usually sample plots are used, but, as the problem may be sporadic and local, it may not be caught with the sampling design. The new remote sensing technologies offer possibilities to large-scale forest growth monitoring.

Needle or leaf loss is a common symptom, which is also difficult to monitor unless the loss is severe. The needle loss symptom is often observed from needle retention. Needle retention provides an index of the number of years that needles are retained. It is only useful as a measure of needle loss if the loss occurs progressively from the oldest to the youngest needles (Innes, 1993). Defoliation can be estimated by observing the form of the tree crown. Mechanical damage to the crown is usually caused by wind. Butt and stem damage is another mechanical damage usually caused by animals like rabbits and squirrels, for example. A common nominator for most of the causes and symptoms of deteriorated forest health is, that almost all are expected to become more frequent.

4. The role of remote sensing in forest health monitoring

First attempts to introduce aerial photographs as a remote sensing tool in forestry were made in 1887. An airborne balloon was used as a photographic platform to produce photographs of forests in the vicinity of Berlin (Van Laar & Acka, 2007). The objective was to examine the possibility of preparing forest maps from aerial photographs. The forest was classified and described on the basis of visual examination of the photographs. Airborne photography from aircraft was introduced during World War 2. After the war the techniques developed for military use became available for civil applications. Since then aerial photography has been widely used in forestry applications. Forest inventory and measurement has been the most widely used application in forestry remote sensing. Earlier, stereo imagery consisting of pairs of oriented aerial images was used to measure individual trees. Today LiDAR based 3D-measurements have made stereo images obsolete (Clement, 2004).

Storm damage studies using aerial photographs were probably the first forest health applications using remotely sensed data. After the introduction of false colour photographs and early multispectral satellites it was possible to study red edge related observables (Barret & Curtis, 1997). By the red edge the difference between the reflectance maximum at near-infrared region and corresponding minimum at visible red region typical for all green vegetation due to chlorophyll absorption is meant. Red edge related indices such as normalized difference vegetation index (NDVI) and leaf area index (LAI) were applied. The introduction of hyperspectral sensors and the increased spatial resolution of multispectral

data available enabled new forest health applications: detection of root diseases (Leckie et al., 2004), detection of pest inflicted damages (Vogelmann & Rock, 1989) and assessment of photosynthetic efficiency (Gamon et al., 1992).

Recent advances in high spectral resolution hyperspectral technology have enabled a whole new approach to forest health studies - the remote chemistry. The most advanced technology to study the health of forest is foliar chemistry. Estimates of the foliar chemistry of canopies allow a better understanding of the functioning of forest ecosystems since many biochemical processes such as photosynthesis, respiration and litter decomposition, are related to the foliar chemistry of trees (Huber et al., 2008). The use of high spectral resolution data has enabled to study many forest health relevant observables, such as the concentration of nitrogen (Fourty et al., 1996), carbon (Daughtry et al., 2001) and leaf pigments (Gitelson et al., 2002).

In the concept of forest health remote sensing nowadays means airborne or satellite imaging of forest areas. Depending on the number of wavelength channels used images obtained by remote sensing are categorized into aerial photographs, multi- and hyperpectral images. Other methodologies such as radar and Light Detection and Ranging (LIDAR) also used in remote sensing of forests. It has been shown that remote sensing can provide useful and relevant forest information (Solberg, 1999). Historically, the potential of remote sensing for forest health studies remained limited for a variety of reasons. Most of the remote sensing data suffered from insufficient spatial, spectral, or temporal resolution. Modern remote sensing instruments have overcome these problems opening new possibilities for forest health assessment. Especially promising is the modality of hyperspectral imaging.

Traditionally, forest health monitoring is performed by means of field studies. In many cases systematically monitored sample plots are used. There is no conflict between field studies and remote sensing. Both techniques are needed and they have their own important role in forest health monitoring. The advantages of remote sensing over conventional field studies include better spatial coverage, shorter sampling intervals, efficiency of data acquisition as well as access to remote or restricted areas. Climate change and increasing pollution set new requirements to forest health monitoring. State of the art remote sensing technology and methods offer an efficient solution to meet those requirements.

Disadvantages of remote sensing in forest health assessment mainly arise from quality and interpretation issues. The quality of modern remote sensing instruments has improved remarkably, however, unfortunately the advancements in algorithm development and verification are not at the same level. Algorithms used to retrieve end-user products such as chlorophyll content have to be validated using field measurements. Algorithms are used worldwide, but the field measurements used in their validation often cover only some specific area. Therefore it is not unusual that algorithms fail to produce reliable results at certain climate zone or geographic location. Remote sensing data products are sometimes difficult to interpret and that makes otherwise reliable data useless. Reliable verification of airborne remote sensed data can only be done using adequate ground truth measurements during the flight operation.

Practically all remote sensing algorithms require the data to be atmospherically corrected. Atmospheric correction is a procedure where the filtering effects of the atmosphere are compensated for. Optical properties of the atmosphere are time and place variant making correction procedure complex and difficult. Atmospheric models used in the correction procedure are not precise enough to ensure correct results in all cases. Atmospheric

correction should be made with utmost care in order to produce quality results. Results should always be verified using known ground targets.

Another remarkable difficulty is the so called "mixed pixel problem". This means that usually in remote sensing data one pixel represents many different materials. In forest areas one pixel often represents tree canopy, soil and some other materials like rock, for example. Then vegetation indices measuring certain forest property can give false results because the percentage of tree canopy is too low. In dense tropical forest the problem is not so serious as in boreal forest areas with low tree density. Because of the mixed pixel problem remote sensing usually provides general estimate of forest health over larger area rather than precise condition of a single tree.

5. Vegetation indices

In the assessment of forest health by means of remote sensing, features called Vegetation Indices are usually extracted from the data. VIs are combinations of surface reflectances at two or more wavelengths designed to highlight a particular property of vegetation. They are derived using the reflectance properties of vegetation. Each of the VIs is designed to accentuate a particular vegetation property. VIs are usually developed by means of empirical laboratory measurements of the property to be studied as well as correlation analysis of remotely sensed data. VIs can be categorized into narrow or broadband ones according to the bandwidth of used wavelength channels. The use of narrowband VIs requires data of high spectral resolution, i.e., hyperspectral data. The use of many VIs is limited because they saturate in dense vegetation areas (Mutanga & Skidmore, 2004). Some narrowband VIs can overcome this problem. Another problem with VIs is nonlinearity (Jiang et al., 2006). The relationship between VI value and measured property is nonlinear which makes the use of VI somewhat difficult. Here we present the most important categories of VIs and some examples of VIs for each category. The ability of the VIs to detect different types of stressed forest areas are tested by means of two case studies presented in sections 7 and 8.

5.1 Greenness (chlorophyll concentration) VIs

Greenness VIs are designed to measure the general quantity and vigor of green vegetation. They measure many different aspects: chlorophyll concentration, canopy area and canopy structure. VI value is always determined by the combination of these different effects. Greenness VIs are based on the measuring of reflectance peak in near-infrared region (NIR). Red wavelength where the chlorophyll absorption is strongest is used as a reference.

Normalized difference vegetation index (NDVI) is the most frequently used and most well know VI. It simply measures the reflectance peak at NIR region. NDVI is a good overall measure of green vegetation, but it has problems with saturation and non-linearity (Jiang et al., 2006). NDVI is defined according to the equation

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}, \quad (1)$$

where ρ_x represents reflectance at wavelength band x (Tucker, 1979).

Enhanced Vegetation Index (EVI) is designed to be used in dense vegetation areas. NDVI saturates in densely vegetated areas (Huete et al., 1997). In order to overcome this problem blue reflectance is used to compensate the effects of background soil and atmospheric scattering effects. EVI is defined according to the equation

$$EVI = 2.5 \left(\frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + 6\rho_{RED} - 7.5\rho_{BLUE}} \right). \quad (2)$$

The Red Edge Normalized Difference Vegetation Index (RENDVI) is a broadband version of the NDVI. While NDVI uses the minimum and maximum reflectances of the red edge region, the RENDVI employs wavelength bands along the red edge (Sims&Gamon, 2002). RENDVI is very sensitive to small changes in canopy chlorophyll content. It can only be calculated based on hyperspectral data. RENDVI is defined according to the equation

$$RENDVI = \frac{\rho_{750} - \rho_{705}}{\rho_{750} + \rho_{705}}. \quad (3)$$

Malenovsky et al. (2006) presented a new approach to canopy chlorophyll content measurement in the form of an index called Area under curve Normalized to Maximal Band depth between 650-725 nm (ANMB). Most of the VIs use simple band ratios to calculate the measured property. In ANMB the surface integral and maximum band depth are calculated. In that way the whole red edge region is estimated instead of a few bands. In the first phase the area under the reflectance curve between 650 and 750 nm is integrated according to the equation

$$AUC_{650-725} = \frac{1}{2} \sum_{j=1}^{n-1} (\lambda_{j+1} - \lambda_j)(\rho_{j+1} + \rho_j), \quad (4)$$

where ρ_j and ρ_{j+1} are reflectances at the j and $j+1$ bands, λ_j and λ_{j+1} are wavelengths of the j and $j+1$ bands, and n is the number of the used spectral bands. The ANMB index is computed according to

$$ANMB_{650-725} = \frac{AUC_{650-725}}{MBD_{650-725}}, \quad (5)$$

where $MBD_{650-725}$ is the maximal band depth of the reflectance, placed at one of the spectrally stable wavelengths of the strongest chlorophyll absorption peaks around 675-680nm.

5.2 Water content VIs

Water content VIs are designed to provide an estimate of canopy water content. Water content is an important vegetation property which correlates with vegetation health. Water content VIs are based on the fact that there are well known water absorption features in the near-infrared and shortwave infrared regions. The use of water content VIs requires high spectral resolution data.

The Water band index (WBI) is a simple reflectance measurement that is sensitive to changes in canopy water content. WBI is utilizing well know water absorption feature at 970nm. The ratio of the reflectance at 970nm to that at 900nm is measured (Penuelas et al., 1995). WBI is defined according to the equation

$$WBI = \frac{\rho_{900}}{\rho_{970}} . \quad (6)$$

Normalized difference water index (NDWI) is sensitive to changes of canopy water content. It uses two different bands, 857 and 1241 nm having similar but slightly different water absorption properties (Gao, 1995). The scattering of light by canopy enhances the weak water absorption at 1241nm. NDWI is defined according to the equation

$$NDWI = \frac{\rho_{857} - \rho_{1241}}{\rho_{857} + \rho_{1241}} . \quad (7)$$

The Moisture Stress Index (MSI) is a simple reflectance measurement that is sensitive to increasing canopy water content. The strength of the absorption at 1599 nm increases when the canopy water content increases. The absorption at 819 nm is used as a reference because it is nearly unaffected by changing water content in the canopy (Ceccato et al., 2001). MSI is defined according to the equation

$$WBI = \frac{\rho_{1599}}{\rho_{819}} . \quad (8)$$

5.3 Leaf pigment VIs

Leaf pigment VIs are measuring the amount of stress-related pigments in vegetation. Carotenoids and anthocyanins are pigments which are present in higher concentrations in stressed vegetation. Carotenoids are pigments that protect vegetation from high light condition. Anthocyanin pigment content is high in senescence and in new leaves.

Anthocyanin reflectance index 700 (ARI_700) is sensitive to anthocyanin amount in vegetation (Gitelson et al., 2001). ARI_700 is defined according to the equation

$$ARI_{700} = \left(\frac{1}{\rho_{550}} \right) - \left(\frac{1}{\rho_{700}} \right) . \quad (9)$$

Anthocyanin reflectance index NIR (ARI_NIR) is similar to ARI_700, but it uses one additional NIR band (Gitelson et al., 2001). ARI_NIR is capable in detecting higher anthocyanin concentrations than ARI_700. ARI_NIR is defined according to the equation

$$ARI_NIR = \rho_{800} \left(\frac{1}{\rho_{550}} \right) - \left(\frac{1}{\rho_{700}} \right). \quad (10)$$

Carotenoid Reflectance Index 550 (CRI_550) measures the amount of carotenoids in canopy. CRI_550 calculates the difference of two bands sensitive to carotenoid amount (Gitelson et al., 2002). CRI_550 is defined according to the equation

$$CRI_550 = \left(\frac{1}{\rho_{510}} \right) - \left(\frac{1}{\rho_{550}} \right). \quad (11)$$

Carotenoid Reflectance Index 700 (CRI_700) is a similar reflectance measurement as CRI_550, but it uses NIR band instead of the green one. CRI_700 is designed to measure higher carotenoid concentrations compared to CRI_550. CRI_700 is defined according to the equation

$$CRI_700 = \left(\frac{1}{\rho_{550}} \right) - \left(\frac{1}{\rho_{700}} \right). \quad (12)$$

5.4 Carbon VIs

Vegetation contains many types of carbon: cellulose, lignin, sugar and starch. Cellulose is used to form cell walls of vegetation tissues. Lignin is used in structurally strong parts in vegetation. There is large amount of carbon in dead or senescent vegetation. These VIs can be used to observe the state of senescence of vegetation.

Normalized Difference Lignin Index (NDLI) is designed to estimate the amount of lignin in vegetation. Reflectance at 1754 nm is primarily determined by lignin concentration of the canopy (Serrano et al., 2002). Reflectance at 1680 nm is used as a reference. NDLI is defined according to the equation

$$NDLI = \frac{\log(1/\rho_{1754}) - \log(1/\rho_{1680})}{\log(1/\rho_{1754}) + \log(1/\rho_{1680})}. \quad (13)$$

The Cellulose Absorption Index (CAI) is a vegetation index indicating surfaces containing dry wood material. Absorptions in the 2000 nm to 2200 nm range are sensitive to cellulose (Daughtry et al., 2004).

CAI is defined according to the equation

$$CAI = 0.5(\rho_{2000} + \rho_{2200}) - \rho_{2100}. \quad (14)$$

5.5 Light Use Efficiency VIs

The light use efficiency VIs are providing a measure of the efficiency with which vegetation is able to use incident light for photosynthesis. Light use efficiency correlates with carbon uptake efficiency and growth rate. Light use efficiency VIs can be used in precision forestry to estimate growth rate and production.

The Photochemical Reflectance Index (PRI) is a reflectance measurement that is sensitive to changes in carotenoid pigments in canopy. The amount of carotenoid pigments indicates photosynthetic light use efficiency and the rate of carbon dioxide uptake (Gamon et al., 1992). It can be used to estimate vegetation productivity and stress. PRI is defined according to the equation

$$PRI = \frac{\rho_{531} - \rho_{570}}{\rho_{531} + \rho_{570}}. \quad (15)$$

The Structure Insensitive Pigment Index (SIPI) is a reflectance measurement designed to maximize the sensitivity to the ratio of bulk carotenoids to chlorophyll while minimizing the affects of variation in canopy structure (Penuelas et al., 1995). SIPI uses three bands: blue, NIR and the maximum chlorophyll peak at 680 nm. SIPI is defined according to the equation

$$SIPI = \frac{\rho_{800} - \rho_{445}}{\rho_{800} - \rho_{680}}. \quad (16)$$

The Red Green Ratio (RG Ratio) index is a reflectance measurement that indicates the relative expression of leaf redness caused by anthocyanin to that of chlorophyll. The Red Green Ratio has been used to estimate the course of foliage development in canopies. The RG Ratio index is an indicator of stress, leaf production and may also indicate flowering in some cases. The ratio is calculated as the mean of all bands in the red range divided by the mean of all bands in the green range. RGRI is defined according to the equation

$$RGRI = \frac{\rho_{RED}}{\rho_{GREEN}}. \quad (17)$$

6. General forest health indices

Forest health is likely to deteriorate due to climate change, pollution and increased human activities. This unfortunate development sets new demands for forest health assessment. There is a need to obtain a quantitative and reliable assessment of the general condition of forest health. Most of the VIs measure a certain property of forest that doesn't necessarily

represent the general condition of the forest. A general forest health index capable of detecting a variety of different damage types would be very desirable. Several methods for obtaining general forest health index have been proposed in the literature.

NDVI index is generally used as a forest health index. Xiao&McPherson (2005) presented a method using NDVI with multispectral data to assess tree health in urban environment. Although NDVI is usually a good VI for general assessment of forest health it leaves opportunities for improvement. NDVI can detect chlorophyll content and defoliation, but there are problems regarding saturation with dense vegetation and non-linearity.

Solberg et al. (2005) presented a method based on combining the LiDAR and hyperspectral data. Variation of canopy chlorophyll mass per area is used to measure forest health. NDVI calculated from hyperspectral data is used as a measure of chlorophyll mass while the Leaf area index (LAI) calculated from the LiDAR data is used as a measure of canopy area. The proposed method produced promising results.

In hyperspectral data analysis the ENVI software package is widely used. ENVI features a forest health tool, which creates a spatial map showing the overall health and vigor of a forested region. The spatial map is showing health index using 9 different classes. Forest health tool uses three different vegetation indices in the classification process. Each VI represents one of the following VI categories: greenness, leaf pigments, canopy water content and light use efficiency. Forest health tool has produced good results in the detection of contaminated forest areas (Tuominen et al., 2008). It constitutes a promising effort to develop a tool to calculate general forest health index using combined VIs.

Forest health can be evaluated using remote sensing by measuring chlorophyll content, defoliation and content of certain pigments (Solberg, 1999). A good general forest health index should measure at least those variables. In addition, water content and photochemical indices could be used. Even advanced foliar chemistry indices yet in experimental phase could be utilized. General forest health index integrating across different damage types is feasible, but a lot of research has to be done. Comprehensive algorithm verification with different forest and damage types would be required.

7. Case study 1: Detection of talc dust contaminated forest areas

The objective of this study was to evaluate how well talc dust contaminated forest areas can be detected using the various vegetation indices. Contamination is one of the major causes of deteriorated forest health. The level of contamination in forest areas can be influenced by related to industrial enterprises. Therefore, it is important to provide reliable forest health information in order to support decision making in all administrative levels.

7.1 Test area

The test area contains Boreal forest around the Lahnaslampi talc mine in North-East Finland. Geographic location as well as the true colour image of the test area are shown in Fig1. Forest in the test area is dominated by Pine, Spruce and Birch trees. Due to large scale mining of crumbly ore, talc, carbonate etc. dust from the rock piles and tailings is carried by the wind to the environment (Kuosmanen et al., 2004). Magnesium and nickel are components of dust emitted from the Lahnaslampi talc plant and mine. Magnesium is bound to talc and magnesite, whereas sources of Ni are both sulphides and Ni bearing Mg-silicates, e.g. talc. Elevated concentrations of Ni and Mg in both moss and humus samples

reflect the amount of precipitated talc dust along the prevailing wind direction (Helminen&Räisänen, 2002). Another source of environmental impact is constituted by the seepage of acid waters through rock piles and tailing ponds. Generally, the level of vegetation stress around the Lahnaslampi mine is low, i.e. the mine does not restrict forestry or other utilization of nature except in the mining area itself. Although talc is not the only source of contamination, it can be assumed that there is strong correlation between Ni concentration and forest stress level.

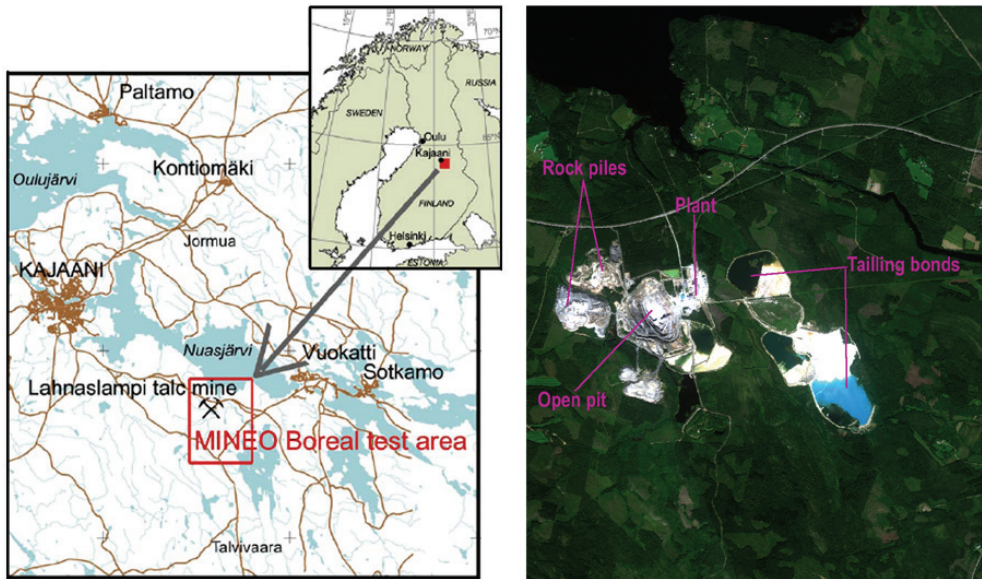


Fig. 1. Geographic location of the test area (Kuosmanen et al., 2004) (left), true colour image of the test area (right).

7.2 Data acquisition

The data acquisition was part of the EU-funded MINEO project, which aimed at the development advanced methods for the extraction of information and knowledge from earth observation data. This study utilized imagery from the HyMap airborne hyperspectral scanner recorded at 28th of July 2000. During the data acquisition cloud cover was non-existent. The HyMap sensor collects reflected solar radiation in 126 bands covering the wavelengths from 420 to 2480 nm. This includes visible, near infrared and short-wave infrared regions of the electromagnetic spectrum. The ground resolution of one pixel was 5*5 meters (Kuosmanen et al., 2004). Hyperspectral data was atmospherically corrected using ATCOR software. ATCOR uses MODTRAN 4 atmospheric model and parameters describing atmospheric type, solar geometry and hyperspectral sensor.

7.3 Methods

In order to remove non-forest pixels from hyperspectral data, a mask image was constructed. Masked pixels were not accounted when test results were calculated. Non-forest pixels were detected by calculating the forest discrimination index (FDI)

$$FDI = \rho_{838} - (\rho_{714} + \rho_{446}) \quad (18)$$

where ρ_x represents reflectance measured in the spectral band centred at x nm (Lucas et al., 2008). All pixels whose FDI-value was under certain threshold were removed from hyperspectral data. It is quite difficult to separate different types of green vegetation reliably. Therefore it was necessary to remove pixels representing green agricultural fields and grasslands manually when they were located near the sites where moss samples were collected. High albedo of the ore minerals causes an atmospheric scattering halo effect around the mining area, a halo which is clearly visible in HyMap imagery especially at VNIR-wavelengths. In order to avoid the influence of halo effect on the test results, 50 meters wide buffer zone was masked around bright ore material. The mask image used to remove non-forest pixels is shown on the left panel of Figure 2.

During the flight campaign 51 moss samples were collected from the test area around Lahnaslampi talc mine. Moss samples were collected by hand using contamination-free gloves. The samples were stored at low temperature and the concentrations of heavy metals were measured in the Geolaboratory of the Geological Survey of Finland (Helminen&Räsänen, 2002). Measurements of nickel concentration were utilized in this case study. Sample sites were divided into six classes according to measured nickel concentration as follows:

- Class 1 Ni concentration 0-5 mg/kg
- Class 2 Ni concentration 6-16 mg/kg
- Class 3 Ni concentration 17-30 mg/kg
- Class 4 Ni concentration 31-60 mg/kg
- Class 5 Ni concentration 61-90 mg/kg
- Class 6 Ni concentration 91-130 mg/kg

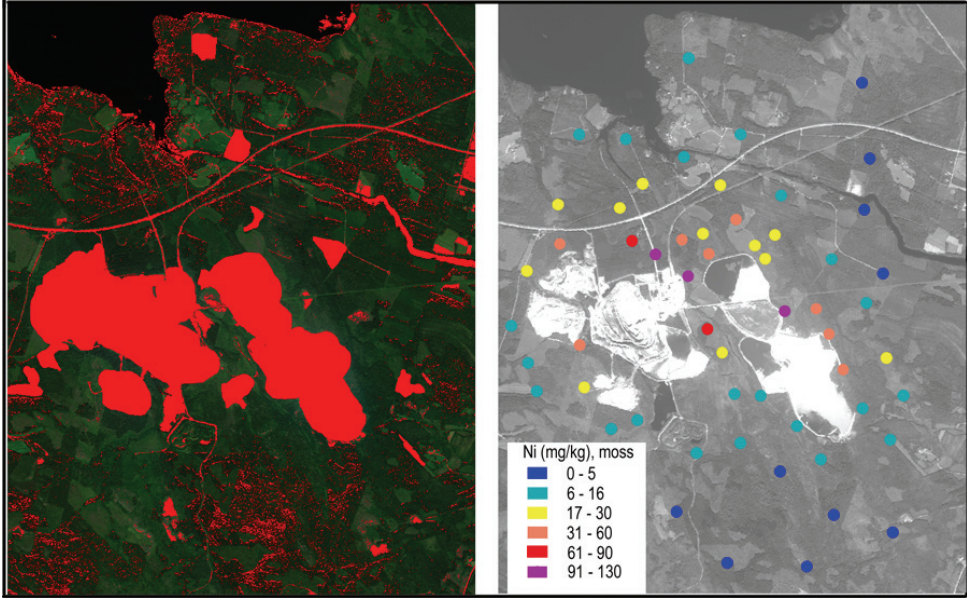


Fig. 2. Mask image used to eliminate non-forest pixels (left), ground truth map indicating Ni-content (mg/kg) of collected moss samples (right).

It is safe to assume that Ni concentration varies slowly within immediate neighborhood of the sampling site. Circular neighborhoods (radius 50 m) were created around each sampling site. These circular areas were color coded according to their Ni concentration levels. Resulting color coded map was used as a ground truth map in the testing process of Vegetation Indices. Color coded ground truth map is shown on the right panel of Fig 2. Correlation between the values of VIs and the measured Ni concentration as well as separability of the classes were considered as the evaluation criteria for Vegetation Indices. Both measures were calculated for each VI using all pixels situated within the 50 m circular neighborhood around the sample sites. Computation of tested VIs is described in section 5.

7.4 Results and discussion

The coefficients of determination r^2 for each VI are shown in table 1. Separability between two classes was calculated using the following simple and robust measure

$$d_{norm} = \frac{|\mu_1 - \mu_2|}{\sigma_1 + \sigma_2} \quad (19)$$

where mu and sigma are the mean and standard deviation of the particular class, respectively (Landgrebe, 2003).

Overall separability between all different class combinations is then

$$S = \frac{1}{15} \sum_{i=1}^5 \sum_{j=i+1}^6 \frac{|\mu_i - \mu_j|}{\sigma_i + \sigma_j} \quad (20)$$

where i and j denote the classes. Separability values for the vegetation indices are shown in table 1.

Vegetation Index	R ²	S
Normalized Difference Vegetation Index (NDVI)	0.9027	0.4866
Enhanced Vegetation Index (EVI)	0.6405	0.2953
Red Edge Normalized Difference Vegetation Index (RENDVI)	0.8135	0.3110
Area Normalized to Maximum Band depth (ANMB)	0.8101	0.3036
Water Band Index	0.4092	0.1514
Normalized Difference Water Index (NDWI)	0.2403	0.0925
Moisture Stress Index (MSI)	0.40141	0.2026
Anthocyanin Reflectance Index 700 (ARI_700)	0.8704	0.3242
Anthocyanin Reflectance Index NIR (ARI_NIR)	0.8098	0.1653
Carotenoid Reflectance Index 550 (CRI_550)	0.7490	0.2868
Carotenoid Reflectance Index 700 (CRI_700)	0.7785	0.3223
Normalized Difference Lignin Index (NDLI)	N/A	N/A
Cellulose Absorption Index (CAI)	N/A	N/A
Photochemical Reflectance Index (PRI)	0.8141	0.2707
Red Green Ratio Index (RGRI)	0.6258	0.1411
Structure Intensive Pigment Index (SIPI)	0.0023	0.1146

Table 1. Coefficients of determination and separability values for different vegetation indices (N/A means not applicable).

In general, correlation and separability values obtained for greenness VIs were relatively high. Highest correlation and separability values were obtained for the NDVI index. Narrow band indices (RENDVI, ANMB) produced better results than the broad band EVI. Test results of water content indices were rather poor. Among these indices best separability of classes was achieved by the MSI index. The separability of NDWI index was the worst of all tested VIs. The value of the WBI index is almost the same for all classes; this can clearly be seen from Figure 3 where the median as well as upper and lower quartiles values are shown for each class.

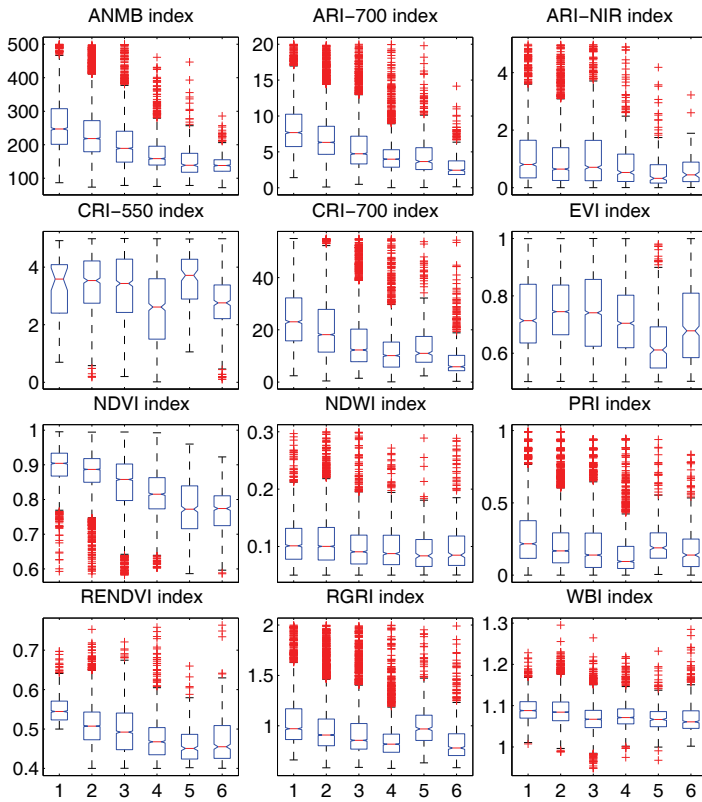


Fig. 3. Median, upper and lower quartile values of classes for tested vegetation indices. Leaf pigment indices that measure the carotenoids and anthocyanins concentration produced quite good test results. Among these indices the best separability and correlation values were obtained by the ARI_700 index. The separability of classes was poor using the ARI_NIR index. VIs designed to provide an estimate of the amount of carbon, i.e., NDVI and CAI were totally useless in the detection of Ni contaminated forest areas. The test results of VIs used to estimate light use efficiency varied a lot. The PRI index produced good results while the results of the SIPI index were very poor.

7.5 Conclusions

Most of the tested VIs could not detect Ni contaminated forest areas with adequate accuracy. However, three VIs, the NDVI, the ARI_700, and the PRI, produced good results. When the spatial distribution of the ARI_700 index is studied from Figure 4, it is easy to see that the more contaminated areas are stretched into North East direction from the mine. From this we can correctly conclude that North East is the prevailing wind direction in the area as the dust emitted from the talc plant and mine is transported by the wind. Poor separability of classes was the common problem for almost all VIs. This can be seen clearly in Figure 3. The major cause of poor separability is the mixed pixel problem. One pixel is

usually representing canopy, shadow and soil. In this Boreal test site the forest is so dense that soil is not present in the pixels normally. The basic problems of the VIs i.e., non-linearity and saturation at high levels can be seen when the test results are analyzed. When Fig 4. is studied, it is easier to detect stressed forest areas by analyzing the spatial distribution of the ARI_700 index, although the NDVI index had better correlation and separability values in the test. This is due to non-linearity and high-end saturation of the NDVI index. The carbon concentration Vis, the CAI and the NDLI failed to produce any meaningful test results. It is understandable because Ni contamination is not assumed to increase carbon content of the tree unless the contamination level is very high.

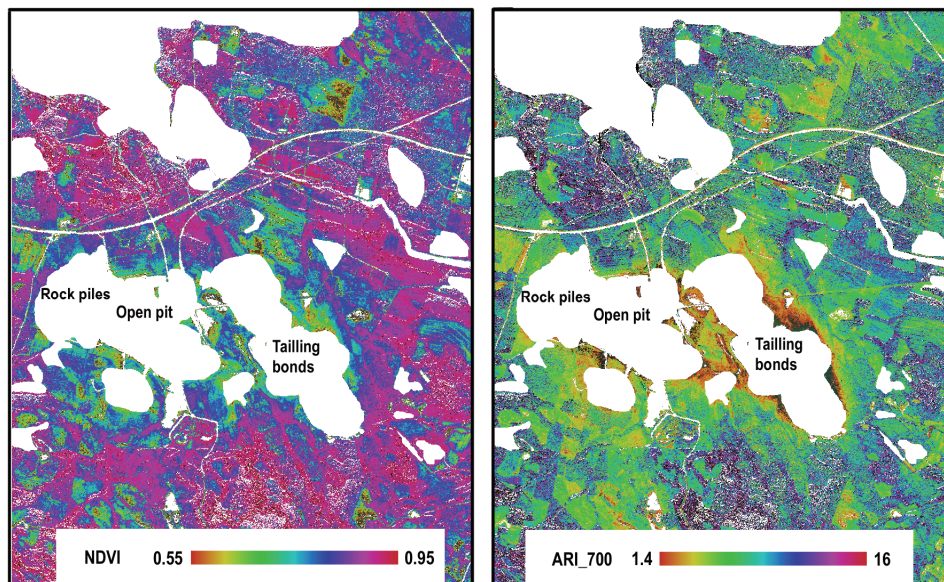


Fig. 4. Spatial distribution of the NDVI and the ARI_700 indices.

VIs estimating canopy water content had only very weak correlation with Ni concentration. Water absorption regions of hyperspectral data are very difficult to measure as often data in those regions is just noise. Light use efficiency indices produced mixed results. PRI had rather good correlation and separation values, but SIPI did not have any correlation with Ni concentration. A probable cause of this is the high noise level in one channel in the blue region of the spectrum. Test results of SIPI index were not excluded from this study, but they should be interpreted with cautious mind. As an overall conclusion from this study it can be said that the deterioration of forest health due to Ni contamination can be detected using carefully selected VIs.

8. Case study 2: Detection of pest inflicted defoliation

The objective of this study was to evaluate how well defoliated tree crowns can be separated from healthy tree crowns using different vegetation indices. Pest inflicted stress is globally one of the most important threats to forest health. Pest inflicted damages can be sometimes

avoided or reduced with proper countermeasures. Therefore it is important to be able to monitor and detect early stage pest damages in forest areas.

8.1 Test area

The test area constituted a sand soiled boreal forest area in South-West Finland. Forest in the test area is dominated by Pine with some occasional Spruce and Birch trees. In the summer of 2006 a seriously damaged pine forest area was found in South-West Finland.

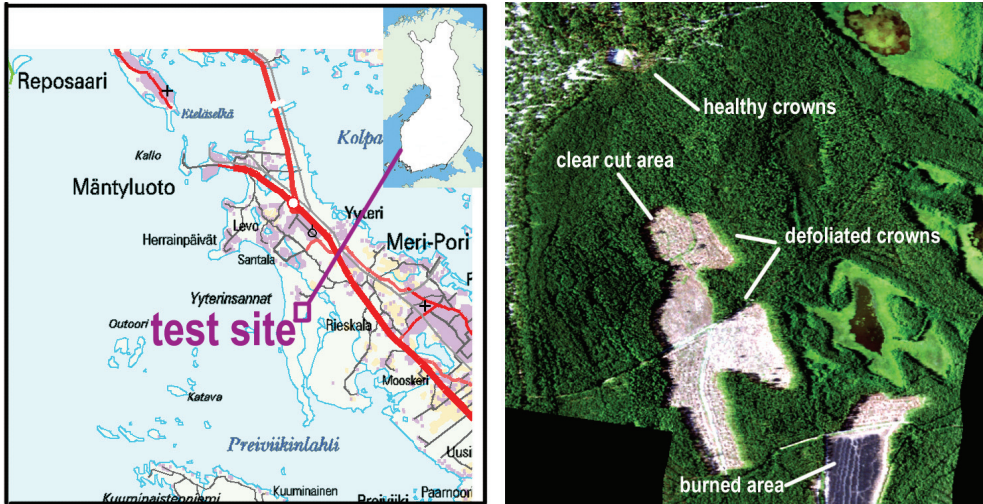


Fig. 5. Geographic location of the test area(left), true colour image of the test area(right).

Studies showed that the cause of the damage was the great web-spinning pine sawfly (*acantholyda posticalis*). Great web-spinning pine sawflies have been found in Finland before, but they have not caused any forest damage before. This time recent warm weathers and sandy soil of the area have made it possible for the sawflies to reproduce and cause damage. It is very possible that forest damage caused by great web-spinning pine sawfly will be more frequent due to climate change. The tree mortality in the area was so high that it was necessary to clear cut 30 hectares of forest. 10 hectares were burned in order to decrease future damage. Most of the damaged trees were successfully clear cut, but there were still some left.

8.2 Data acquisition

This study utilized data from the AISA dual airborne hyperspectral scanner recorded at 13th of July 2008. During the acquisition the cloud cover was non-existent. The AISA dual spectrometer collects reflected solar radiation in 481 bands covering the wavelengths from 399 to 2452 nm. This includes the visible, near infrared and short-wave infrared regions of the electromagnetic spectrum. The ground resolution of one pixel was 2.5*2.5 meters. Hyperspectral imagery was atmospherically corrected using ATCOR software. This study utilized also LIDAR data acquired on May 1st, 2008, by a small aircraft carrying Leica ALS50-II LIDAR scanner. Acquisition was carried out in early spring, so deciduous trees

were still without leaves. The point density of the LIDAR data was 0.76 points per sq. m. Measurement accuracy of elevation data was 0.15 m.

8.3 Methods

Ground truth data was collected by doing an extensive field study in the test area. Moderately defoliated trees were located using visual inspection. Heavily defoliated trees were excluded because in this case most of the reflected radiation actually comes from the soil. Geo-coordinates of promising candidates were recorded using a GPS-instrument. Candidates for healthy trees were also collected. Final selection of the trees included in the study was done using LIDAR data.

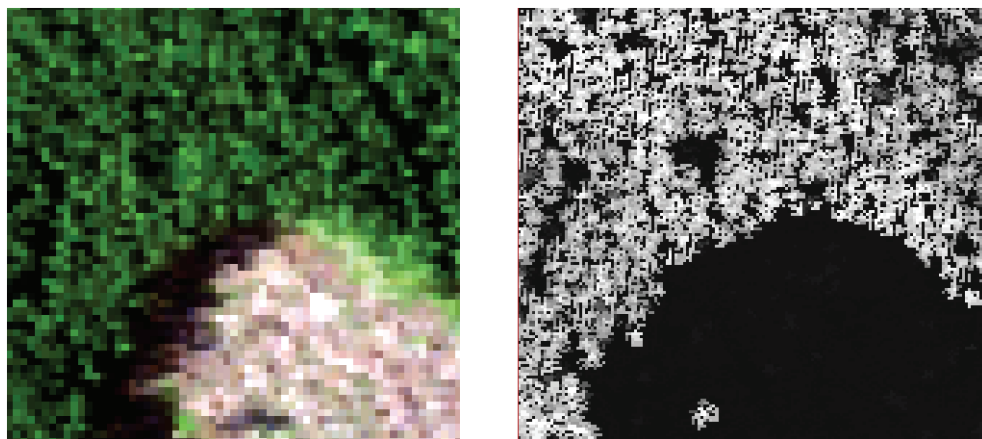


Fig. 6. True colour image of the hyperspectral data(left), DSM image from the same area(right).

The reason for using LIDAR data is mixed pixel problem. Usually spectral signature imaged from the test area represents the mixture of canopy, soil, shadow and maybe some other materials too. In order to test VIs reliably it was necessary to select final tree crown using digital surface model (DSM). DSM was generated by rasterizing and interpolating raw LIDAR data using ENVI software. DSM was geo-referenced to same coordinate system as the hyperspectral data using polynomial triangulation of ERDAS software. Both data sets were linked in ENVI software and the neighborhood of each tree crown candidate were studied. The tree crowns for which the DSM of the tree top neighborhood showed full canopy coverage in the corresponding hyperspectral pixel were chosen for VI testing. Geo-referenced DSM image is shown on the right panel Fig 7. White pixels show laser beam returns from tree tops. As a result altogether 20 hyperspectral pixels representing 10 healthy tree crowns and 10 partly defoliated tree crowns were selected for the analysis.

8.4 Results and discussion

VIs were calculated for all 20 test pixels. Mean and standard deviation values for both healthy and defoliated classes are shown in table 1. Separability S between healthy and defoliated classes was calculated using the formula

$$S = \frac{|\mu_1 - \mu_2|}{\sigma_1 + \sigma_2} \quad (21)$$

where mu and sigma are the mean and standard deviation of the particular class, respectively.

Vegetation index	Healthy		Defoliated		S
	Mean	std.	Mean	std.	
NDVI	0.934	0.0219	0.791	0.035	2.513
EVI	N/A	N/A	N/A	N/A	N/A
RENDVI	0.502	0.0139	0.4823	0.01844	0.5820
ANMB	232.7	50.69	139.8	28.79	1.168
WBI	N/A	N/A	N/A	N/A	N/A
NDWI	N/A	N/A	N/A	N/A	N/A
MSI	0.3865	0.0810	0.2805	0.0872	0.6302
ARI_700	0.0059	0.0011	0.0030	0.0009	1.4500
ARI_NIR	1.654	0.433	3.487	0.611	1.7471
CRI_550	0.0101	0.0144	0.0023	0.0031	0.7936
CRI_700	0.0143	0.0021	0.0177	0.0024	0.7348
NDLI	N/A	N/A	N/A	N/A	N/A
CAI	N/A	N/A	N/A	N/A	N/A
PRI	N/A	N/A	N/A	N/A	N/A
RGRI	0.8574	0.1078	0.70838	0.1540	0.5692
SIPI	1.003	0.0174	1.0280	0.0189	0.6887

Table 2. mean, standard deviation and separability values for each tested Vis (N/A means not applicable).

Many of the tested VIs did not produce any meaningful results. PRI used one wavelength channel that appeared to be noisy and therefore the results for this index are not valid. NDLI and CAI probably failed because they use SWIR channels. Most VIs were calculated using ENVI software VI calculator and some using MATLAB codes. Best separability value was obtained using NDVI. NDVI correlates well with reduced chlorophyll content, so its good performance could be expected the result is understandable. ARI_700 and ARI_NIR produced also rather good results. Recently published ANMB produced clearly better separability than the well established EVI and RENDVI VIs

8.5 Conclusions

The objective of this study was to determine if defoliated tree crowns can be detected using vegetation indices. Widely used and tested NDVI proved to be the best VI in detection of defoliation. Separability was calculated using standard deviation. The number of tested pixels in each class was rather small for accurate statistical analysis. The results show that

some of the VIs can detect defoliated tree areas rather reliably. Results also indicated that reduced chlorophyll content and increased anthocyanoid pigment levels are good indicators of defoliation.

9. The future of forest health related remote sensing

High spatial resolution remote sensing for forestry applications has reached an almost mature phase with wide range of applications available. However, numerous opportunities and challenges remain. The robustness of the data processing methods is one of the issues to be considered. Processing methods currently in use often need extensive calibration and adjustment for each new imaged forest area. Processing methods should be able to handle different types of forest areas in routine fashion despite the variation of climate zone, soil type or forest structure. A lot of research has still to be done to fully utilize the potential of high spatial and spectral resolution data.

In the forest health assessment, most studies are characterized by a limited geographic extent concentrating on test sites where the complexity of forest environment is low. The methods are mostly empirical and require local calibration. Expanding forest health studies over wide geographic regions is a challenge because the added complexity of varying vegetation types can hide the relatively weak signal feature associated with forest stress. One possible solution to overcome this problem is to identify forest health change using data obtained at different times as it may be easier to identify stress from spectral change, rather than from the spectral properties of stress itself (Brandberg & Warner, 2006).

It would be highly desirable to be able to use satellite data, instead of airborne data for remote sensing of individual trees. Airborne remote sensing campaigns are very costly limiting the accessibility of data. Airborne hyperspectral instruments offer superior signal quality and spatial resolution, however, wide coverage multi-temporal forest health monitoring using airborne data is not feasible. Satellite data has the advantage of a relatively uniform illumination and view angle over large regions, thus minimizing problems associated with combining data from individual flight lines. Satellite-borne high spatial resolution hyperspectral data is not available at the moment, but can be anticipated in the future as the development of space technology continues. Spatial resolution of the hyperspectral imager currently in space (EO1 Hyperion), at 30 m, is too coarse for studying individual trees. OrbView-4 which failed to reach orbit after launch in 2001, would have offered 250 band hyperspectral data at 8 m spatial resolution. Specifications of Orbview-4 provide some kind of reference on what can be expected from the capability of future sensors (Castro-Esau & Kalashka, 2008).

Data fusion where several data sources are used together has the potential for revolutionary impact on forest health measurement. For example, with LiDAR data it is possible to directly measure the structural attributes of trees. The data fusion of high spectral resolution LiDAR and high spectral resolution hyperspectral data can raise forest health studies on a new level: precise information on foliar chemistry pin pointed to a single tree. Data fusion of LiDAR and hyperspectral data has already showed promising results (Solberg et al., 2005).

10. References

- Barret, E.C. & Curtis, D.E. (1997). *Introduction to Environmental Remote Sensing*, Chapman & Hall, ISBN 0412371707, London
- Brandtberg, T. & Warner, T. (2006). High-Spatial-Resolution Remote Sensing In: *Computer applications in sustainable forest management*, Shao, G. & Reynolds, K.M., 19-39, Springer, ISBN 9781402043055, Dordrecht
- Castro, K.L. & Kalacska, M. (2008). Tropical Dry Forest Phenology and Discrimination of Tropical Tree Species Using Hyperspectral Data, In: *Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests*, Kalacska, M. & Sanchez-Azofeifa, G.A., (1.), 1-25, CRC Press, ISBN 9781420053418, Boca Raton
- Ceccato, P.; Flasse, S.; Tarantola, S.; Jacquemoud, S. & Gregoire, J.M. (2001). Vegetation Leaf Water Content Using Reflectance in the Optical Domain. *Remote Sensing of Environment*, Vol.77, 22-33
- Clement, J. (2004). *LiDAR Derived 3D Forest Stand Parameters of Dutch Pine*, Wageningen University, Wageningen
- Daughtry, C.S.T., (2001). Discriminating Crop Residues from Soil by Short-Wave Infrared Reflectance. *Agronomy Journal*, Vol.93, 125-131
- Daughtry, C.S.T.; Hunt, E.R. & McMurtrey J.E. (2004). Assessing Crop Residue Cover Using Shortwave Infrared Reflectance. *Remote Sensing of Environment*, Vol.90, 126-134
- Ferretti, M. (2001). Forest health assessment and monitoring - Issues for consideration. *Environmental monitoring and assessment*, Vol.48, 45-72
- Fourty, T.F.; Baret, S.; Jacquemoud; Schmuck, G. & Verdebout, J. (1996). Leaf Optical Properties with Explicit Description of Its Biochemical Composition: Direct and Inverse Problems. *Remote Sensing of Environment*, Vol.56, 104-117
- Gamon, J.A.; J. Penuelas, & Field, C.B. (1992). A Narrow-Waveband Spectral Index That Tracks Diurnal Changes in Photosynthetic Efficiency. *Remote Sensing of Environment* Vol.41, 35-44
- Gao, B.C. (1995). Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. *Proceedings of SPIE*, Vol.2480, 225-236
- Gitelson, A.A.; Merzlyak, M.N. & Chivkunova, O.B. (2001). Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves. *Photochemistry and Photobiology*, Vol.71, 38-45
- Gitelson, A.A.; Zur, Y.; Chivkunova, O.B. & Merzlyak, M.N. (2002). Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy. *Photochemistry and Photobiology*, Vol.75, 272-281.
- Helminen, T.R. & Räisänen, M.L. (2002). Regional atmospheric deposition patterns of dust in the vicinity of the Lahnaslampi talc mine, Sotkamo, Finland, as revealed by moss and humus samples, *Archive report RS/2002/8*, 10 p., 9 appendices. Geological Survey of Finland, Espoo
- Huber, S.; Kneubuhler, M.; Psomas, A.; Itten, K & Zimmermann, N.E. (1997). Estimating foliar biochemistry from hyperspectral data in mixed forest canopy. *Forest Ecology and Management*, Vol. 256, 3, 491-501
- Huete, A.R.; Liu, H.; Batchily, K. & van Leeuwen, W. (1997). A Comparison of Vegetation Indices Over a Global Set of TM Images for EOS-MODIS. *Remote Sensing of Environment*, Vol.59, 440-451

- Innes, J.L. (1993). *Forest health: its assessment and status*, CAB international, ISBN 0851987931, Oxon UK
- Jiang, Z.; Huete, A.R.; Chen, J.; Chen, Y.; Li, J.; Yan, G. & Zhang, X. (2006). Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. *Remote Sensing of Environment*, Vol.101, 366-3
- Kolb, T.E.; Wagner, M.R. & Covington, W.W. (1994). Forest health from different perspectives, *Proceedings of the 1995 national silviculture workshop*, pp. 5-13, Gen. Tech. Rep. RM-267, U.S. Department of Agriculture, Fort Collins, Colorado
- Kuosmanen, V.; Arkimaa, H.; Helminen, T.; Hyvönen, E.; Kuronen, E.; Laitinen, J.; Lerssi, J.; Middleton, M.; Ruohomäki, T.; Räisänen, M.L.; Saarelainen, J. & Sutinen, R. (2002). MINEO Boreal environment test site, Finland. Contamination/impact mapping and modelling - Final report, *Archive report RS/2004/2*, 85 p., 6 appendices. Geological Survey of Finland, Espoo
- Landgrebe, D.A., M. (2003). Signal theory methods in multispectral remote sensing, *John Wiley & sons*, ISBN 047142028-X, New Jersey
- Leckie, D.G.; Jay, C.; Gougeon, F.A.; Sturrock, R.N. & Paradine, D. (2004). Detection and assessment of trees with Phellinus weirii(laminated root) using high resolution multi-spectral imagery. *International Journal of Remote Sensing*, Vol.25, 793-818
- Leopold, A. (1949). *A Sand County Almanac and Sketches Here and There*, *Oxford Univ. Press*, ISBN 0195053052, New York
- Lucas, R.; Mitchell, A. & Bunting, P. (2008). Hyperspectral Data for Assessing Carbon Dynamics and Biodiversity of Forests, In: *Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests*, *Kalacska, M. & Sanchez-Azofeifa, G.A.*, (1.), 50-51, CRC Press, ISBN 9781420053418, Boca Raton
- Malenovsky, Z.; Ufer, C.; Lhotakova, Z.; Clevers, J.; Shaepman, M.E.; Albrechtova, J. & Cudlin, P. (2001). A new hyperspectral index for chlorophyll estimation of forest canopy: area under curve normalized to maximal band depth between 650-725nm. *EARSeL eProceedings*, Vol.5, 2/2006,161-173
- Monnig, E. & Byler, J. (1992). *Forest health and ecological integrity in the Northern Rockies*, USDA For. Ser. FMP Rep. 92-7
- Moore, B. & Allard, G. (2008). Climate change impacts on forest health, *Working Paper FBS/34E02/8*, 38 p. FAO, Rome
- Mutanga, O. & Skidmore, A.K. (2004). Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, Vol.25, 3999-4014
- O'Laughlin, J.; Livingston, R.; Their, R.; Thornton, J; Toweill, D.E. & Morelan, L (1994). Defining and measuring forest health. *Journal of Sustainable Forestry*, Vol.2, 65-85
- Penuelas, J.; Baret, F. & Filella, I. (1995). Semi-Empirical Indices to Assess Carotenoids/Chlorophyll-a Ratio from Leaf Spectral Reflectance. *Photosynthetica*, Vol.31, 221-230
- Penuelas, J.; Filella, I.; Biel, C.; Serrano, L. & Save, R. (1995). The Reflectance at the 950-970 Region as an Indicator of Plant Water Status. *International Journal of Remote Sensing*, Vol.14,887-1905.
- Serrano, L.; Penuelas, J. & Ustin, S.L (2002). Remote Sensing of Nitrogen and Lignin in Mediterranean Vegetation from AVIRIS Data: Decomposing Biochemical from Structural Signals. *Remote Sensing of Environment*, Vol.81, 55-364

- Sims, D.A. & Gamon, J.A (2002). Relationships Between Leaf Pigment Content and Spectral Reflectance Across a Wide Range of Species, Leaf Structures and Developmental Stages. *Remote Sensing of Environment*, Vol.81,337-354.
- Solberg, S. (1999). *Forest health monitoring: Evaluation of methods, trends and causes based on a Norwegian nationwide set of monitoring plots*, Norwegian Forest Research Institute, ISBN 8271698974, Oslo
- Solberg, S. ; Lange, H. ; Aurdal, L. ; Solberg, R. & Naasset, E. (2005). Monitoring forest health by remote sensing of canopy chlorophyll : first results from a pilot project in Norway, *Proceedings of 31st International Symposium on Remote Sensing of Environment 2005*, St. Petersburg, June 20-24 2005
- Tucker, C.J., (1979). Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of the Environment*, Vol.8,127-150
- Tuominen, J.; Lipping, T. & Kuosmanen, V. (2008). Assesment of ENVI forest health tool in detection of dust and seepage contaminated forest areas. *Proceedings of IEEE International Geoscience & Remote Sensing Symposium*, 1358-1361, Boston, July 07-11, IEEE
- Van Laar, A. & Acka, A. (2007). *Forest mensuration (Managing Forest Ecosystems)*, Springer, ISBN 978402059902, Dordrecht
- Vogelmann, J.E. & Rock, B.N (1989). Use of the thematic mapper data for the detection of forest damage caused by the pear thrips. Vol.30, 217-225
- Xiao, Q. & McPherson, E. (2005). Tree health mapping with multispectral remote sensing data at UC Davis, California. *Urban Ecosystems*, Vol.8, December 2005, 349-361

Development of a High-Resolution Wireless Sensor Network for Monitoring Volcanic Activity

José Chilo¹, Andreas Schlüter² and Thomas Lindblad³

¹*University of Gävle S-80176 Gävle, Sweden*

²*Institute of Computer Science, Freie Universität Berlin, Germany*

³*Royal Institute of Technology, S-106 91 Stockholm, Sweden*

1. Introduction

This chapter describes the design of a high-resolution wireless sensor network to monitor infrasonic signals from volcanic activity. A prototype system is constructed and tested. The system is based on the ultra low power microcontroller MSP430 with the requirements of energy-awareness and high sensor node autonomy. The infrasonic signals are measured at 200 Hz using 12 bit resolution and the result is buffered on SD cards in case of a lack of bandwidth. The implementation of a cost-table driven network routing protocol allows a radio sleep schedule of almost 97% when no data has to be transmitted. Furthermore, the sensors need to be time-synchronized for later event localization. This work shows that it is feasible to have a synchronization accuracy of less than 1 ms using a GPS receiver that is powered on only a few seconds per hour.

In recent years the installation of infrasound sensors at seismic measuring stations has become common and now researchers can obtain large and heterogeneous infrasound signal data-sets generated in near real-time. However, most current infrasound stations are still using expensive infrasound microphones and traditional data acquisition systems which limit the deployment of new infrasound stations. To improve on this, we propose in this work a wireless data acquisition system based on FreeWave FGR09CSU 900 MHz radio modem and a Wireless sensor networks (WSN).

Infrasound is defined as the range of frequencies between 0.001-20 Hz. It is generated by a variety of events, both man-made and natural. Among the latter type, active volcanoes are efficient sources of infrasound. Volcanic eruptions are characterized by the acceleration of hot fluids from subsurface reservoirs into the atmosphere generating acoustic waves in the 1-20 Hz frequency range. Infrasonic airwaves produced by active volcanoes provide valuable insight into the eruption dynamics and related phenomena. Infrasound also provides a special opportunity for the comparison of eruptive activity among different volcanoes because atmospheric pressure records are mostly independent of site-specific propagation effects (Chilo 2008).

However, infrasound propagating long distances is a complex phenomenon. It is strongly influenced by the detailed temperature and wind profiles. The infrasonic signal detected by

traditional infrasound systems contains the combination of the source's infrasound power spectrum and the distortions introduced by the atmosphere. In order to extract the source characteristics the data should be collected at close range: from a few meters to a few km distances. At short distances, the atmosphere is a homogeneous medium that preserves the infrasonic waveform as it is generated by the source. We need to seek new ways to enhance the capability of monitoring volcanic activity close to the source. Wireless sensor networks have the potential greatly benefit studies of volcanic activity.

A wireless sensor network is a collection of small devices having sensors, computational processing ability, wireless receiver and transmitter technology and a power supply (Culler, Estrin and Srivastava 2004). Typical WSNs communicate directly with centralized controller or a satellite, thus communication between the sensor and controllers is based on a single hop. Another kind of WSN could be a collection of autonomous nodes that communicate with each other by forming a multi-hop radio network and maintaining connectivity in a decentralized manner by forming an ad hoc network.

The last few years, the WSN has been used by a number of authors for volcanic eruptions monitoring. In references (Werner-Allen, Johnson, Ruiz, Lees and Welsh 2005; Werner-Allen, Welsh, Locrincz, Johnson, Marcillo, Ruiz and Lees 2006) a WSN was used together with infrasound microphones and seismometers for geophysical studies in the area of the volcano Tungurahua, Ecuador. The infrasound signals were sampled by 102 Hz, 10 bits resolution and transmitted over a 9 km wireless link to a remote base station. For the time synchronization a single GPS receiver was used in combination with the *Flooding Time Synchronization Protocol* (FTSP). The archived accuracy was 10 ms with an error of more than six milliseconds. The data transport was controlled by the remote base station.

In this approach we present a high-resolution WSN for long-term monitoring. The infrasonic signals are sampled and converted to digital data at a frequency of 200 Hz and a resolution of 12 bit. The proposed WSN requires a time stamp per infrasonic sample with an accuracy of one millisecond. Therefore, an algorithm was developed and evaluated which synchronizes the sensor nodes under the support of an equipped GPS device. The algorithm needs to be powered on the GPS device only a few seconds per hour. The collected data is handled under the concept of unlimited virtual data memory, whereby the data is swapped out to an SD card or, if the radio is switched on, transmitted towards the observatory. The used and evaluated routing protocol is the first implementation of the data-centric data dissemination protocol (D3). We extended this cost-based ad hoc routing protocol for the usage of radio time slots. This approach allows a sleep scheduled radio of almost 97% of the time.

The proposed WSN is planned to be deployed at University of San Agustín observatory station (ARE) in Southern Perú. The station is located 12.7 km south-east of Arequipa city, latitude S 16° 27'56.67443", longitude W 71° 29'35.23676" and elevation 2450 m. The ARE station provides a unique laboratory for studying regional infrasound and seismic wave propagation. The ARE station is located in the shadow of three giant volcanoes: Chachani (6075m), Misti (5821m) and Picchu Picchu (Chilo, Jabor, Liszka, Lindblad and Persson 2006). Furthermore, the most active volcano of Perú, volcano Ubinas, is situated 65 km from the ARE station which will be the focus for our infrasonic studies.

2. Hardware and requirements

The platform used is the Modular Sensor Board MSB-430 shown in Figure 1 (left). In this figure the top part (MSB-430S) is the sensor module; the middle part (MSB-430) is the core module; and the bottom part (MSB-430T) is the base module. The base module MSB430T carries three AAA batteries, has a JTAG and serial/USB socket and is available with a GPS receiver FALCOM Smart Antenna FSA01. It should be noted that this platform can easily be exchanged to a more suitable platform. The features and capabilities of the MSB430 concerning our proposal are summarized in Table 1.

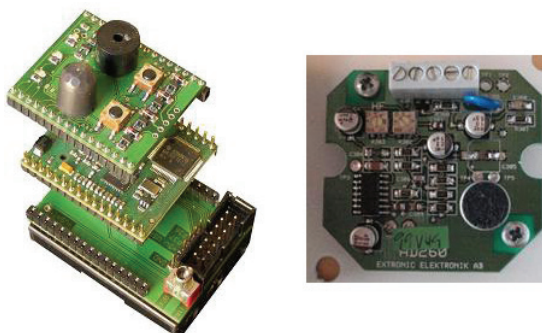


Fig. 1. Title Photo of MSB-430 (left) and the MCE-200 microphone with pertinent preamplifier and filter mounted on a PCB (right)

Microcontroller	MSP430F1612 16bit RISC
	100 kHz - 8MHz
	55 kB Flash-ROM
	5 kB RAM
	256B Infomemory
Transceiver	Chipcon CC1020
	868MHz
	8.6 dBm, max 1km (tuneable to more than 5 km)
	19.2 kbit/s using Manchester encoding
Mass storage	SD card (max 4 GB)
ADC	12 Bit, unipolar [0-3 V]
Sensors on board	humidity and temperature sensor Sensirion SHT11
Supply voltage	2.7-3.6V
Energy	Active Mode: 330 μ A at 1 MHz, 2.2 V
	Standby Mode: 1.1 μ A
	Off Mode (RAM Retention): 0.2 μ A

Table 1. MSB-430 features

For infrasound recording the electret condenser element microphone MCE-200 from Panasonic was used. The details listed in Table 2 are given by the manufacturer. The

mentioned frequency range is peculiar and seems not to cover the infrasonic range. Nevertheless, in contrast with the manufacturer specification, this microphone operates sufficiently in the infrasonic range according to the experiences of the authors (Chilo and Lindblad 2007).

The MCE-200 microphone is also available from Extronic AB (<http://www.extronic.se/>) mounted on a PCB, Fig 1, right part. A specially designed version includes a filter providing two bipolar output signals, one for infrasound and one for audible sound. The power supply is about 3-12 V. The PCB doesn't suit to be used as a sensor caused by high power consumption of more than 1 mA. We planned to use the microphone without this PCB, therefore, a tailor -made circuit needs to be designed including a low pass filter, a gain amplifier and an antialiasing filter.

Frequency range	20-16000 Hz
Sensitivity	7.9mV/Pa/kHz \pm 2 dB
Output impedance	1-2 k Ω
Signal-to-noise	ratio < 58 dB
Couple capacitor	0.1-4.7 μ F
Working temperature	0-40 $^{\circ}$ C
Power supply	1.5-10V DC /0.5mA

Table 2. Nominal MCE-200 Microphone specifications

The following incomplete enumeration briefly sketches the requirements of the application domain.

- 1) An event classification requires the complete signal opening of each infrasonic record.
- 2) The record of an infrasonic event per microphone shall be automatically transported from volcano Ubinas to the University of San Agustin observatory station (ARE). The distance is about 65 km.
- 3) The operating time of the monitoring system is weeks or permanent.
- 4) The system shall automatically handle the failure of sensor nodes. The system will continue at least under a restricted operating mode following the malfunction of multiple nodes. It means that a complete system breakdown should be avoided.

The third requirement is in contradiction to the first one, because it forces a continuous AD conversion. Therefore, the sensor nodes will quickly run out of power. A trade-off is the continuous conversion by just a single node. The remaining nodes could use a more energy conserving comparator.

To analyze the wave propagation and in order to localize the infrasonic source six time synchronized signals recorded from different positions are required:

- 5) At least six microphones shall be deployed close to the volcano. They shall be spatially separated.
- 6) The chronological, accurate to a millisecond, and spatial position, accurate to 10 meters, of all records per microphone of an infrasonic event shall be available.

3. System design

3.1 System overview

The data transport requirements (about 65km distance) affect materially the hardware selection and distribution. The transmission range of the MSB-430 mounted radio goes about 5000 m, but only under special circumstances like a clear line of sight. Hence the node has to be combined in a multi hop network a more powerful transmitting devices must be used. The FreeWave FGR09CSU 900MHz radio modem (http://www.freewave.com) could be used for this task.

According to the manufacturer specifications the FGR09CSU modem reaches a range of 95 km in a clear line of sight. If no line of sight is possible, then either a modem can be used as a repeater or additional intermediates allow a more suited gateway position. An example deployment using the FreeWave modem is sketched in Figure 2. In the example seven infrasonic sensor nodes (green) are used for the data acquisition, one gateway node (magenta) is connected to a workstation located in the ARE observatory and one intermediate node (blue) pass the data from the sources to the sink, the gateway node, which finally delivers the data to a workstation.

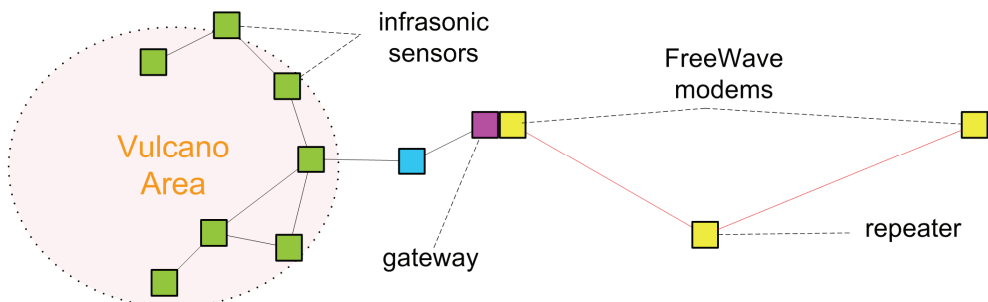


Fig. 2. Spatial system arrangement of a hybrid network; red line represents a long range link; black lines accords a radio coverage between two nodes

As aforementioned, Figure 2 contains seven infrasonic sensors, one more than needed. Not all sensors have to be simultaneously active. So, one can rest and maybe recover energy. Not every infrasonic sensor in the figure is inside radio coverage of an intermediate. In order to transport the data of the outer infrasonic sensors all of these must be intermediate nodes simultaneously. To assure sufficient CPU power for the data acquisition this double role needs to be controlled.

It is also vitally important that a failure of one or two nodes must not impact the whole system, thus the architecture shall actualize highly autonomous behaviour of the nodes not just for the day-to-day business but also for the handling of exceptions like malfunctions of some nodes. Furthermore, low power consumption and energy conservation strategies shall be taken into account, which is essential for a long term operation.

3.2 Architecture

The applied method for the system decomposition covered both: the partitioning of the

system according to the system functionality (distinction of concerns), and to map the components to the required hardware.

Figure 3 shows the component dependencies of a sensor node. Time synchronization using GPS and periodical signalization form the Time component. For instance, the analog-digital-converter (ADC) is controlled by a periodical signal. The converted data is evaluated by the DataProcessing component (DP) in order to detect events of interest. The component gathers all needed data for an infrasonic record including the time and position information provided by the Time component and hands it to the Network component. Both components, DP and Network, require major RAM parts, as they handle the same data. The VirtualDataMemory component (VDM) manages RAM and SD card blocks. The current voltage level is periodically watched by EnergyController. Depending to the remaining energy and in collaboration with the neighboring nodes the component set the node into a low-power mode by stopping the Network, the Time and indirectly the DP component. The hardware abstraction functionality is pooled within the System component.

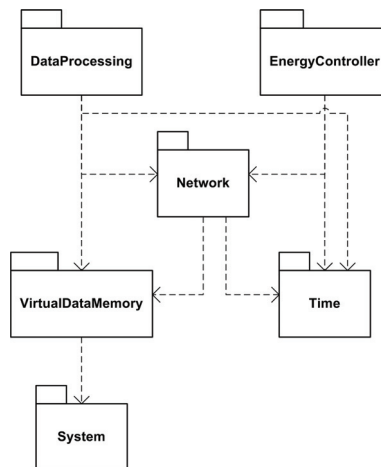


Fig. 3. The diagram gives an overview of the system architecture. Arrows symbolize interface access

4. Time synchronization

4.1 Hardware Mapping

The MSP430 version F1216 provides two timers, *Timer_A* and *Timer_B*. *Timer_A* has three (TACCR0-TACCR2), *Timer_B* has seven independent capture/compare registers (TBCCR0-TBCCR6). A timer register can be used in two different ways. Either the register's content is compared to the current counter value of the timer and creates an interrupt on equality (compare mode), or on an external signal, e.g. the rising edge of the GPS pulse, the current counter value is stored in the register (capture mode). *Timer_A* is configured to count the oscillations of the sub-main clock (SMCLK), which itself will be fed by the digitally controlled oscillator (DCO, Figure 4) and *Timer_B* counts the crystal watch oscillations

(ACLK, 32768 Hz). In order to provide a stable main clock the DCO is watched and controlled according to the more reliable crystal oscillator (dcoChecker).

The second Timer_B register in compare mode is used to create the signal for the ADC. The third register is used to measure the GPS pps. In detail, the GPS signal captures the Timer_B counter value, stores it into the third register and wakes up the CPU, i.e. an interrupt service routine (ISR) of the Time component.

For scheduling tasks like a periodically GPS synchronization the functionality of software timers is needed. Software timers shall run either after a defined delay or at a defined point in time. For an accurate scheduling two values are required, the number of Timer_B cycles and the desired value of Timer_B, i.e. ACLK cycles. When the defined Timer_B cycles of a software timer expires, the desired watch oscillator cycles are written to the sixth compare register, which produces an interrupt if Timer_B reaches the value and the software timer is executed.

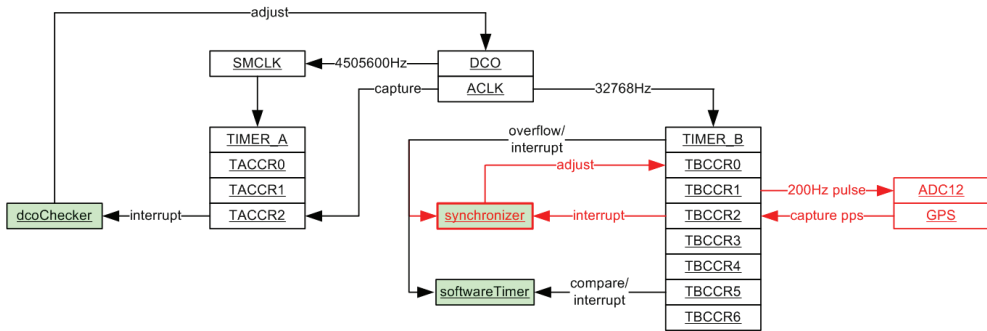


Fig. 4. The diagram shows time synchronization via crystal watch. Green objects are interrupting service routines; red objects are available only on sensor nodes

4.2 GPS-ACLK synchronization

The idea of the ACLK synchronization is to let the timer count from zero to exactly the ACLK frequency decreased by one. Though, the exact ACLK frequency, about 32678 Hz, needs to be measured by using the GPS pulse-per-second (pps) as reference which provides an accuracy of 1 μ s. The real ACLK frequency depends to the environment temperature. The interrupt TBCCR0 (for TBR=32767) and CAP2 (for pps capture) occur simultaneously. This fact complicates the deviation detection. The execution of an interrupt service routine may not follow the real chronological event order. Therefore, it is wise to move the pps capture reference from zero to (TBCCR0+1)/2. Thus, the interrupts are separated by 0.5 s, but only if the system is synchronous. In other words, if the Timer_B is synchronized, the pps is caught exactly by (TBCCR0+1)/2. Moreover, the ACLK ticks between the first and the second pps capture are counted, what should be close to 32768. This is the first synchronization stage.

To illustrate the procedure in detail we sketched a simplified example in Figure 5. In this figure each timer counting step is visible. One relevant detail is the fact that if the timer counts to TBCCR0 it altogether counts TBCCR0+1 ticks; as the timer value is in the range of [0..TBCCR0]. Before the first capture occurred, the timer was initialized by $f_{ACLK_1} = 9$ or

TBCCR0 = 8. After two seconds $capture_1 = 3$ and $capture_2 = 1$ are measured.

$$f_{ACLK_n} = (f_{ACLK_{n-1}} - capture_{n-1}) + capture_n \quad (1)$$

$$shift = capture_n + \left\lfloor \frac{f_{ACLK_n}}{2} \right\rfloor \quad (2)$$

Now, the more exact frequency can be computed by equation 1, with result $f_{ACLK_2} = 7$.

Before the register TBCCR0 is adjusted by f_{ACLK_2} , it is phase shifted, i.e. it is set to $shift = 4$. The shift is determined by the elapsed time since the last timer overflow plus the expected amount of ticks to the desired timer overflow, equation 2. This happens still in the ISR of the second capture.

On the next timer overflow (Figure 5, third TBIFG event) the timer border is set to the calculated value $f_{ACLK_1} - 1 = 6$. The third and fourth capture occurs exactly on TBR=3. The real time clock is synchronous. Furthermore, the local time is adjusted to the UTC time. The UTC time was written by the GPS device to the COM port between the first and second capture and must be mapped to the first capture; i.e. the first pps is the exact point in time given in the following NMEA record. The GPS device could be switched off after the second capture for a time period depending on the allowed time deviation.

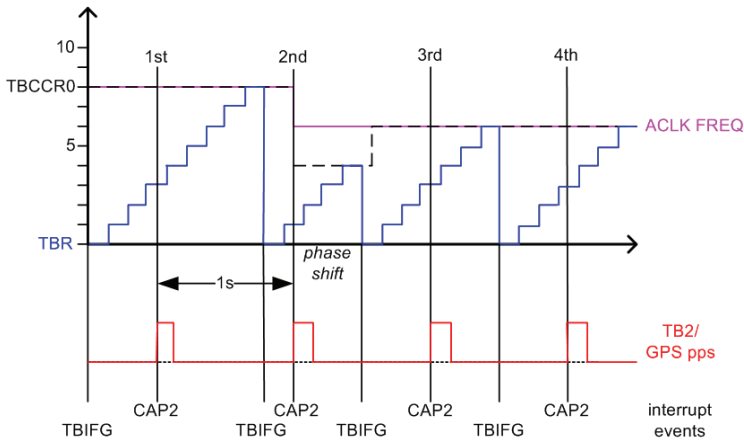


Fig. 5. Example of phase shift and period correction via ACLK

The allowed deviation of $\pm 0.5ms$ accords $\pm 5ms * f_{ACLK} ticks$, which is about 16 ticks. Even by a stable frequency in the worst case it could happen that each second the ACLK deviates almost one tick. It is caused by the fact, that the frequency is calculated as an integer and the fraction is dropped. In case of a fraction close to one the ACLK exceeds the allowed deviation after 16 s.

As a corrective the places after the decimal point for the rational ACLK frequency are approximated by measuring the average deviation for $\Delta seconds$. Through the usage of the integer frequency f_{ACLK} the deviation accords the fractional part of the rational ACLK frequency multiplied $\Delta seconds$. The fraction calculation is done within the second stage of synchronization. First after the third capture the GPS device is switched off and the captured values (captured ticks and local time in seconds) are stored. When for instance tens of minutes elapsed the GPS device is activated again and the fourth capture is awaited. Equation 3 calculates the average fraction, whereby $\Delta ticks_n = capture_n - capture_{n-1}$.

$$fraction_n = \frac{\Delta ticks_n}{\Delta seconds_n} \quad (3)$$

The frequency correction can be simply done by adding uniquely $\Delta ticks_n$ every $\Delta seconds_n$ to TBCCR0 and restoring the original value of TBCCR0 one second later. However, this approach would lead to unwanted steps within the time response. Much better is the method of equally distributing $\Delta ticks_n$ within $\Delta seconds_n$. The resulting value for TBCCR0 is calculated on each timer overflow event like in equation 4. The variable t is the time in seconds after the last phase shift ($t = 0s$).

As $fraction_n$ always is within the range $(-1..1)$, the timer border $TBCCR0(t)$ equals either $f_{ACLK_n} - 1$ or $f_{ACLK_n} - 1 \pm 1$. In other words: the correction of $\Delta ticks_n$ within $\Delta seconds$ is implemented as an equal distribution of atomic corrections for the timer period.

$$TBCCR0(t) = f_{ACLK_n} - 1 + \lfloor (t+1)fraction_n \rfloor - \lfloor tfraction_n \rfloor \quad (4)$$

On the fifth pps the captured values needed for the next GPS adjustment are stored and the device is powered down for maximum $16 * \Delta seconds_n$ (on a constant frequency). One open point is the measuring of $fraction_n$ while concurrently the timer is adjusted by $fraction_{n-1}$. Equation 3 needs to be extended to equation 5 to take a concurrent adjustment into account.

$$fraction_n = \frac{\Delta ticks_n + \lfloor \Delta seconds_n fraction_{n-1} \rfloor}{\Delta seconds_n} \quad (5)$$

First tests showed a non acceptable result. A failure analysis discovered a dependency of the watch crystal on the environment temperature. Within a couple of hours the oscillator slowed down almost linear. The third synchronization stage forecasts the expected ACLK frequency taking the gradient of the frequency response into account.

The gradient is calculated in equation 6. The gradient goes into equation 4 and equation 5 to form equation 7 and equation 8.

$$gradient_n = \frac{(f_{ACLK_n} + fraction_n) - (f_{ACLK_{n-1}} + fraction_{n-1})}{\Delta seconds_n} \tag{6}$$

$$TBCCR0(t) = f_{ACLK_n} - 1 + \lfloor (t+1) fraction_n + (t+1)^2 gradient_n \rfloor - \lfloor t fraction_n + t^2 gradient_n \rfloor \tag{7}$$

$$fraction_n = \frac{\Delta ticks_n + \lfloor \Delta seconds_n fraction_{n-1} + \Delta seconds_n^2 gradient_{n-1} \rfloor}{\Delta seconds_n} \tag{8}$$

The synchronization stages are sketched in the activity diagram in Figure 6. Stage one happens after the second capture but only for the first time a node is activated. Between the third and fourth capture the GPS is powered down for a long time. For the best trade off between energy consumption and accuracy the time duration is determined by test series. On the fourth capture stage one and two are executed. The calculation of the gradient results zero, hence a prior fraction is missing in the current state. The next occurring capture is treated as the third capture again. The second time a *fourth* capture happens, the gradient can be determined.

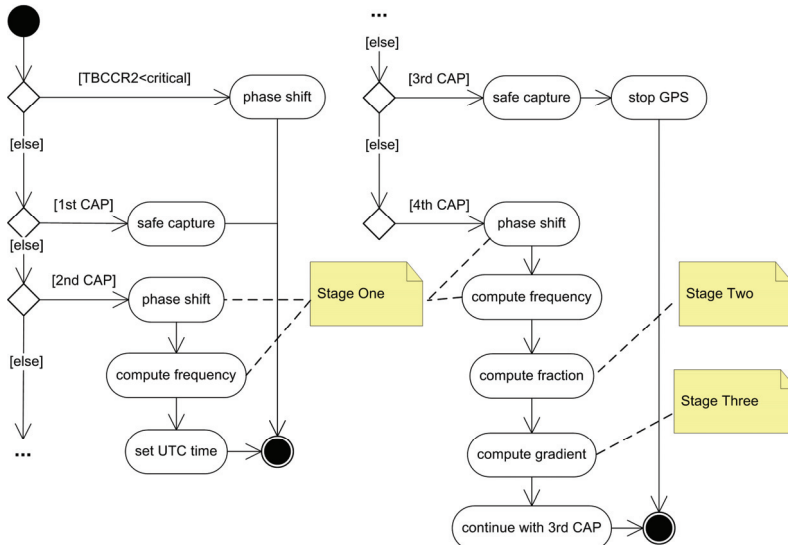


Fig. 6. Final activity diagram of handling CAP2 events

5. Data Acquisition

5.1 Infrasound data

The infrasound sensor is realized by the electret condenser element microphone (ECEM) MCE-200 from Panasonic. A PCB will connect the microphone to the ADC12. Therefore, the signal needs to be transformed to a unipolar signal up to 3V.

The complete data pipeline is arranged in a way that first the analog signal produced by the ECEM is filtered to pass only frequencies below the cutoff frequency and afterwards the

signal is gained to the required voltage, Figure 7. This approach avoids an overmodulation by disturbing frequencies.

To date there are not ideal low-pass filter, i.e. passing the signal below 20 Hz and eliminating the frequencies above. Real filters are approximations, for instance the *Bessel*, *Chebyshev*, and *Butterworth* filter.

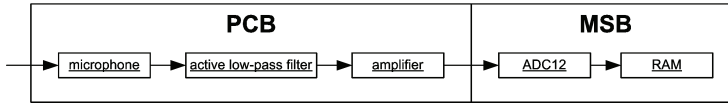


Fig. 7. Data acquisition pipeline

A Butterworth filter was designed in our case. The quality of an active filter, for instance a steeper attenuation, is improved by cascading filter stages by combining multiple single filters. A 4th order low-pass filter was realized which will require two operation amplifiers on the PCB.

5.2 Local event detection

To detect a supposed volcanic eruption a simple approach is watching the signal amplitude. If it passes a threshold one can assume an eruption occurred. The maximum of the amplitude is searched in time window T_w . For the detection of a 1 Hz signal T_w must be at least 500 ms. If the threshold is too high, the amount of false negatives, i.e. missing an event, increases. If it is too low, more false positives will be detected, for instance short-term fluctuations or wind.

Short-term fluctuations can be smoothed out by calculating a moving average. The *EWMA-detector* implemented by (Werner-Allen, Johnson, Ruiz, Lees and Welsh 2005) supplied more reliable events than their implemented threshold based detector. The EWMA (exponential weighted moving average) function

$$average_t = average_{t-1} + \alpha(sample - average_{t-1}) \quad (9)$$

computes the averages by fading older samples. The variable $\alpha \in [0..1]$ determines the fading factor of old samples. A high α fades off more. The elimination of long-term trends can be done by maintenance two averages, by name a short-term $average_s$ and a long-term average $average_l$. If $average_s$ exceeds $average_l$ by the amount of div an event is triggered.

The state machine in Figure 8 maps the distributed event detection. To trigger the data acquisition two independent events must occur within a time window. The elapsing of a timeout without an event always leads back to the initial state.

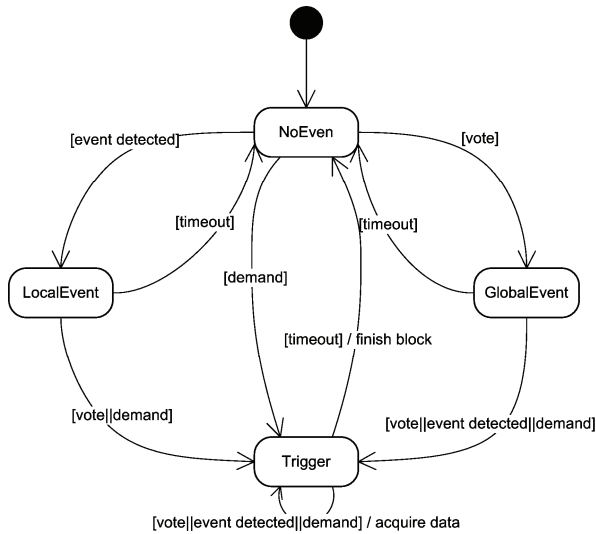


Fig. 8. State machine for distributed event detection

All state transitions are briefly described in Table 3. The demand message is put into brackets because of the following: if a sensor does not receive two votes it is probably too far away, that it could sample meaningful data, hence, the demand message could be discarded. However, the implementation efforts for this functionality are not high and may still finally benefit.

State	Description
<i>local event</i>	local event detected
<i>vote</i>	dialog message received by a neighbour who detected an local event
<i>(demand)</i>	dialog message forces the start of data acquisition
<i>timeout</i>	time without any event elapsed

Table 3. List of state transitions used in Figure 8

6. Data transport

The implemented routing protocol D3 (Ditzel and Langendoen 2005) is a cost table and data centric routing protocol, i.e. no network addresses are used and the data is “floating” down a gradient. Therefore, only broadcasts are used. In respect to energy conservation, the medium is accessed in time slots, small ones for negotiation and large timeslots for the data transfer whereby only the participating nodes are active. The problems of hidden and exposed stations are no more solvable by the classical MACA protocol. Instead we used three different data timeslots: TXDATA for transmission, RXDATA for reception, and IDLE for a powered down radio. An example is given in Figure 9. All nodes with the same hop-

count distance to the data-sink, i.e. cost, need to use the same slots.

Figure 10 shows the general idea of the D3 protocol. First the gateway, the sink, broadcasts an interest message, Figure 10(a). Each node updates its own cost, i.e. it stores the smallest received cost plus one and forwards one-time the INT message with the node's own cost. The data sink labels the interest message by a sequence number. If a node receives an INT message with a new sequence number, it overwrites its cost value even with a higher cost. In this way a changing node topology can be recognized by regular INT messages.

Before a node transmits a data message, it advertises the data by broadcasting an ADV message in the ADV slot before the next TXDATA slot, node E in Figure 10(b). The message contains the cost of the node (here two for node E) and the `segment_key_t` values of all unconfirmed packets in the memory. In the original D3 protocol the message also includes the time of data transmission, i.e. here the duration to the next data slot in ACLK ticks. However, the transmitted duration should be ideally the time between the first byte received by an interested node and the beginning of the data slot. To approach this value as close as possible the time of transmitting the first ADV byte needs to be measured by the initiating node and the duration should be computed according this value. This is only possible by transmitting the duration by a different message immediately after an ADV message; called *Time* message. To map a *Time* message to the responding ADV message, both messages need the node network address.

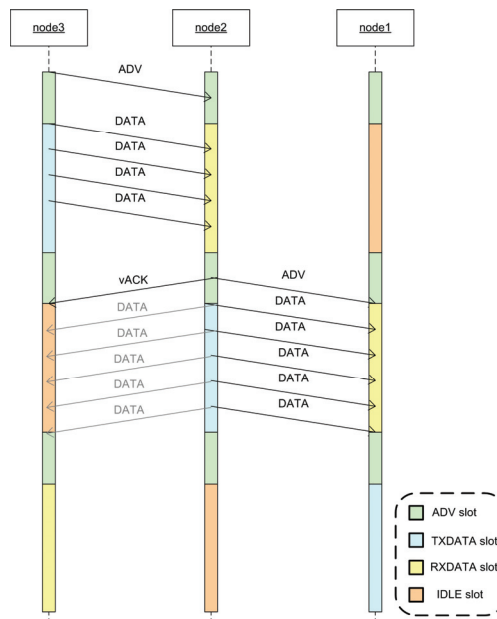


Fig. 9. Sequence diagram of D3 timeslots

Each neighbour with lower costs than the cost of a received ADV message schedules the data reception (node C and D). All other nodes just switch off their radio in the next RXDATA slot (node F and G). The data is transmitted at the beginning of the RXDATA slot, Figure 10(c).

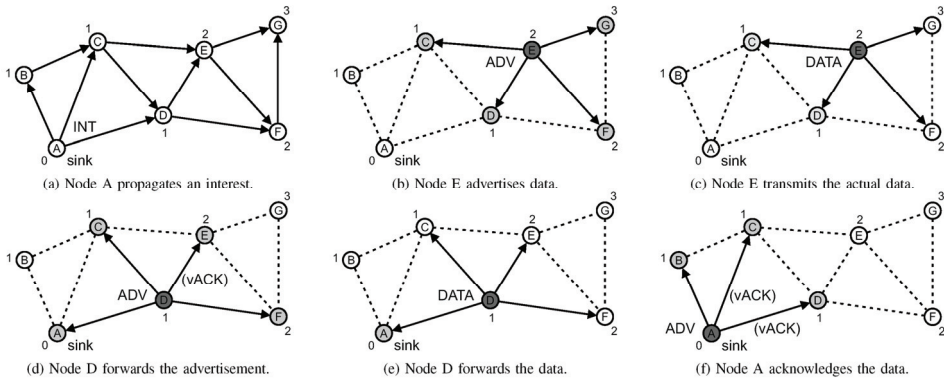


Fig. 10. The D3 routing protocol. Initially the gateway floods an INT message. A node forwards data by advertising the new data and after transmitting the data itself

The next data slot is an IDLE slot for the prior transmitter (node E) and a TXDATA slot for the prior receiving nodes (node C and D). Those nodes initiate a random backoff delay at the beginning of the ADV slot foregoing its TXDATA slot. The node for which the backoff delay expires first, node D in Figure 10(d), immediately broadcasts an ADV, which lets all other nodes with an advertisement intention cancel their backoff delay. Again, the message contains the nodes own cost (one for node D) and all segment identifications of the packets in the memory, which includes also the quite recently received packets. Furthermore the prior forwarding node (node E) takes that ADV message as a virtual acknowledgment (vACK).

Finally, the data message reaches the gateway, Figure 10(e). The gateway also advertises its data in its according ADV slot, Figure 10(f). Of course, in its following TXDATA slot it does not forward the data by the radio, but by the COM port to a gateway client.

The usage of explicit acknowledgments (eACK) is need fully to avoid the retransmission of plenty big data packets. The eACKs are transmitted immediately when a neighbour node advertises already confirmed packets.

Another scenario is the resetting of a node or the adding of a new node to the network. The ADV messages do not suffice for the integration of the node into the topology, because the following could happen: Let's assume a new node has the cost of five, but before the node doesn't receive an ADV message of a neighbour with the cost four, its cost is unknown. Now, this node receives an ADV message with the cost six and assumes a cost seven, i.e. the cost plus one. This lets also confirm all advertised packets in its history. In the case the node receives an ADV message with cost four; it updates its cost to five. If now the node who first advertised data needs to advertise the data twice, the new node assumes the data to be already confirmed and wrongly transmits an eACK. The data packets will be lost. A solution is the usage of interest request messages, which are transmitted the first time a node enters the network or frequently until the node is added to the network. All neighbours answer on an interest request with its last received interest message. However, the requesting node can use the ADV messages together with the time messages to synchronize to the network time slots, so it could transmit the request within an ADV slot.

7. Results and conclusions

7.1 Data processing component

A prototype node is equipped with a specially designed PCB containing the infrasonic microphone, a 4th-order Butterworth active low-pass filter, and an amplifier. For the power supply a 3 V source is used. A function generator is connected to the couple capacitor and ground. It produces a sinus wave with an amplitude of $U_{i-max} = 344$ mV. The PCB output pin and ground is connected to a voltage oscillograph. The amplitude of the PCB output signal U_{o-max} is measured. The output signal swings around 1.5 V.

The gain G_{PCB} of the circuit depending to the input frequency f_i is computed by the equation 10, Figure 11, green curve.

$$G_{PCB}(f_i) = \log\left(\frac{U_{o-max}(f_i) - 1.5V}{U_{i-max}}\right) * 10dB \quad (10)$$

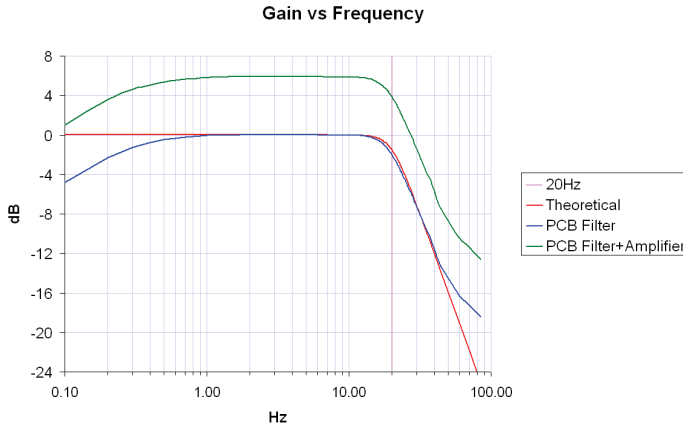


Fig. 11. The response of the amplifier. Red, theoretical 4th-order Butterworth; blue, filter without amplifier and green, filter plus amplifier

To measure the gain of the AC amplifier we used a frequency which is passed by the filter with unity gain. In other words we searched the output voltage maximum depending to the input frequency, what is by $f=2.2$ Hz and $U_{o-max}=2.844$ Hz. The gain at this point is:

$$G_{amp} = \frac{2.844V - 1.5V}{U_{i-max}} = 3.95 \quad (11)$$

The calculated value is $G=4.25$. To get the gain G_{filter} assigned to the filter the amplifier gain must be undone, Figure 11, blue curve:

$$G_{filter}(f_i) = \log\left(\frac{U_{o-max}(f_i) - 1.5V}{3.95U_{i-max}}\right) * 10db \quad (12)$$

The amplitude response of a 4th-order Butterworth filter is plotted for comparison, Figure 11 red curve. Two obvious deviations are visible. The first one below 1 Hz is explained by the high-pass effect of the couple capacitor and maybe even by the high-pass effect of the capacitor of the AC amplifier. The second one close to 100 Hz can be a measurement error caused by the fact that the output voltage in this range is nearly the offset DC of 1.5 V.

The cutoff frequency is defined as the frequency for which the filter returns $\sqrt{1/2}$ of the pass-band voltage. The cutoff frequency f_c can be experimentally determined by finding the frequency which fulfils following equation:

$$\log(\sqrt{1/2}) * 10dB = -1.51dB = G_{filter}(f_c) \quad (13)$$

The realized cutoff frequency is $f_{c-low} = 19$ Hz. The cutoff frequency of the high-pass effect can be determined in the same way and is $f_{c-high} = 0.14$ Hz which is acceptable.

7.2 Time component

The three staged GPS synchronization allows a switched off GPS device for long periods (t_{GPS_OFF}). The time period impacts not only the energy consumption but the RTC accuracy. To find the best trade-off between energy consumption and accuracy test series with different t_{GPS_OFF} values are accomplished. The board is simply made of a single sensor node equipped with a GPS device. It is configured to start the stage one synchronization immediately, stage two after ten minutes, and stage three after further ten minutes. Firstly now, the time period t_{GPS_OFF} is used. Hence, the GPS device needs between 45 s and about 165 s to provide a valid pps signal, the synchronization period is t_{GPS_OFF} plus the fluctuating GPS startup time.

Figure 12 shows the result for $t_{GPS_OFF} = 45$ min. The test runs more than 15 hours. One can extract different assumptions by interpreting the responses. The first fact is that within the first three hours the synchronization is unacceptable. The reason could be the temperature difference between the office and outside, thus the node needs some time to acclimatize to the outside temperature. Furthermore, the outside temperature declined in the evening hours.

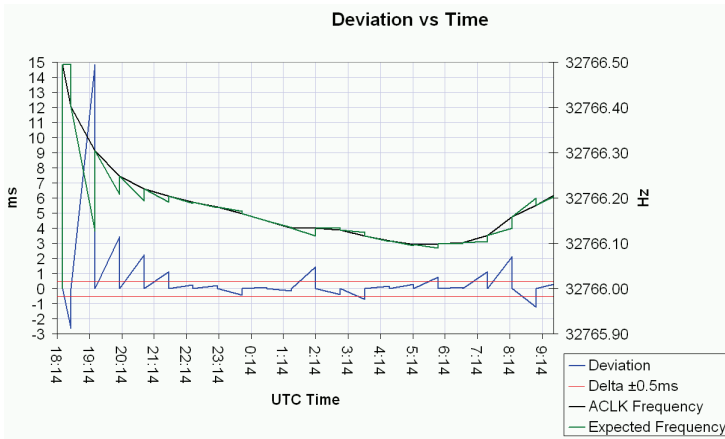


Fig. 12. State 15 hour test result of the RTC deviation; three staged GPS synchronization happened about every 45 minutes

Incontestable is the dependence of the deviation between the expected and the real frequency response to the RTC deviation. A bend in the frequency response leads to a deviation. However, the idea to take the second derivation into account would worsen the prediction of the frequency. One can appreciate this, if the expected frequency response is imaginary extended by the turns of the real curve. The mathematically unpredictable bends can maybe be physically predictable, if for instance the temperature is monitored. A future investigation of dependence of the frequency and temperature could pay off.

Summarized, 33.7% of the time the RTC exceeds the allowed limit (ignoring the first three hours, 22.1%).

For the experimental assurance of the desired accuracy the time t_{GPS_OFF} needs to be defined according to the highest gradient, what is = 648 s, equation 14, ignoring the first three hour.

$$t_{GPS_OFF} = 0.5ms / |gradient_{max}| = 648s \quad (14)$$

The average gradient, equation 15, of the deviation takes both into account: long periods of small deviations and the percentage part of peeks of high deviations. A fixing of the time t_{GPS_OFF} according to the average gradient, equation 16, allows an exceeding of the allowed deviation for a small time. For the average gradient without the first hour the result is $t_{GPS_OFF} = 2150s$.

Very optimistic but energy conserving is to take only 75% of the smallest gradients into account. The resulting value according to the average gradient of 75% of the best values is $t_{GPS_OFF} = 4168s$.

$$\overline{\text{gradient}} = \frac{\sum_{\min < i \leq \max} \frac{|deviation_i|}{\Delta t_i}}{\max - \min} \quad (15)$$

$$t_{GPS_OFF} = \frac{0.5ms}{\overline{\text{gradient}}} \quad (16)$$

An adjustment every 30 minutes, Figure 13, doesn't show a considerable improvement. It needs two hours to acclimatize and to calculate the gradient to the realistic temperature response. After the first two hours the time of exceeding the limit deviation of 0.5 ms is about 19.9%. A shorter time period for the GPS adjustment doesn't solve sufficiently the deviation problem caused by the turns in the frequency response.

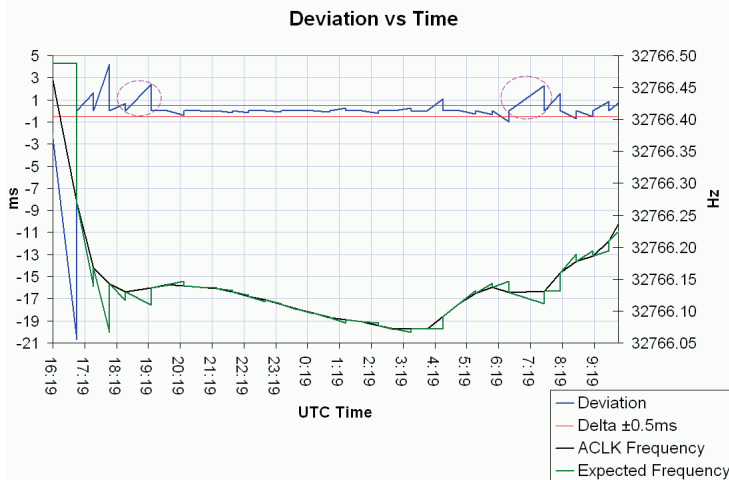


Fig. 13. 18 hour test result of the RTC deviation; three staged GPS synchronization happened about every 30 minutes; pink cycles - GPS timeout caused by rain

For a complete evaluation of the measurement the knowledge of an outstanding fact is required. The pink cycles mark long time periods of a deficient GPS signal, which lasts respectively about an hour. Both GPS timeouts were caused by rain. Rain impedes the synchronization process twice. On the one hand, it induces GPS timeouts and on the other hand, it comes together with an increasing of temperature, which itself speeds up the ACLK. Nevertheless, excluding these values would result a violation of the accuracy requirements for about 12% of the time.

7.3 The network layer

The issues for the network layers are the capacity, the reliability, and the robustness. Two types of test scenarios were accomplished: the transmission of a continuous data stream and a sparse transmission of data records over a long time period.

For both tests the command *cost* was implemented on the gateway node in order to simulate a cost distribution. The first parameter of the command addresses a node. The second one specifies the desired cost of the node. By receiving a dialog message containing the command the destination node adjusts its cost accordingly. While the test application is running, the node ignores messages transmitted by nodes whose cost difference is greater than one.

Unfortunately, a second node equipped with a second GPS device was not timely available. Due to the fact that the network time slots are seeded by time synchronized nodes, the tests can only be done with a single data source. However, the intermediate nodes don't care for the packet originator and multiple data sources are not essentially required to test the functionality.

Four different configurations for the breadboard were realized. Different network depths and different numbers of nodes on the same level were tested. Of course the data sink was always on level zero. The data rates are measured by the GatewayClient.

data rate [kbit/s]	Volume [kB]	cost level 1	cost level 2	cost level 3
2.35	530	one data source		
2.26	502	one intermediate	one data source	
1.37	292	two intermediates	one data source	
2.09	524	one intermediate	one intermediate	one data source

Table 4. Test results of routing a continuous data stream

Table 4 shows the data rates measured for the different network topologies. The packet loss rate (not listed in the table) for all topologies was 0%. The rate for the two node topology (line one in the table) complies exactly the theoretical value of equation 17 (one RAM block of 512 Bytes contains four network packets; one network packet carries a payload of 110 Bytes). The data rate is the measured value for two intermediates on the same cost level. Obviously, the small value is reasoned by collisions. This indicates a small value for the used backoff delay. Anyhow, the results still fulfil the requirements.

$$data\ rate_{MAX} = \frac{4\ blocks * 4\ packets * 110\ Byte}{3 * t_{ADV} + 3 * t_{DAT}} \approx 2.35\ kbit/s \quad (17)$$

By the following test case the reliability of the network synchronizations is measured. During the periods of no network traffic, no synchronization of the time slots happens. The network topology was linear with a depth of three, i.e. the data source was on cost level three. The size of the records was randomly created.

test duration [min]	record separation [min]	volume [kB]	records
124	15	51	8
185	30	33	7

Table 5. Test results of routing periodically single records

A delay of 15 minutes between the record transmissions results a stable time slot behavior of the network, Table 5. However, for the delay of 30 minutes the networks got asynchronous after the 7th record. Another test with a delay of one hour (not listed) failed after the first record. It implies, that the time slot deviation after 30 minutes can exceed at least the duration of the half of the ADV slot duration, i.e. 94 ms or 3072 ACLK ticks. In other words, if the crystal oscillators of the intermediates deviates about 3.4 ticks per second, the ADV slot duration is exceeded after 30 minutes.

In order to increase the allowed deviation the ADV slots could simply be extended. Even, periodical transmissions of ADV messages would synchronize the nodes again. However, both approaches would increase the network energy consumption.

Another solution is the flooding of the measured frequency and its expected trend by the sensor nodes immediately after a sensor node synchronized to the GPS device. Therefore, it is assumed, that the oscillators of all nodes work under the same conditions.

Considering very long time periods without any network traffic, following strategy is promising: If the nodes got asynchronous it firstly matters in the case of transmission purposes. To synchronize again the acting node transmits its slot time multiple times with a delay, which assures that the message is received in at least one ADV slot by a child node. A delay of about the half of the ADV slot duration would suit. The acting node is *scanning* the time slots.

Though the unsatisfying result the implemented D3 routing protocol works in principal. A detailed analyzes and thereby the adjustment of the protocol parameters is the direction of future work, as well as large scale tests.

7.4 Conclusion and outlook

The system presented in this chapter has the potential of a full-fledged application. The analysis of the requirements, as well as the design of the system architecture, was done using common and well proven software engineering techniques. Whereby, the balancing act between the reutilization and the easy interchange ability in opposite to the high application awareness succeeded for the most parts. Though the not entirely satisfying results and the open questions, the design is reasoned and self contained.

In order to connect the microphone to the MSB a sensor module was created. The circuit covers a fourth-order Butterworth low-pass filter and an amplifier. Furthermore, the signal is transformed from bipolar to unipolar. The comparison of the measured frequency response and the expected theoretical response shows acceptable results. A direction for future work is to design a digitally controlled amplifier. This would open the possibility to remotely adjust and fine tune the event detection on hardware level. In the same way, the problem of a missing infrasonic reference source for the best suiting amplifying factor could be avoided.

The time synchronization was one of the big challenges. We showed the feasibility of achieving the required accuracy by a rare use of the GPS device. The implemented approach comprises a high potential to get improved. In view of the measurement results, the prediction of the crystal oscillator frequency by taking the current temperature into account is promising. Even the usage of an external and more stable oscillator would be an enhancement. The maximum achievable accuracy is determined by the GPS receiver and lies in the range of microseconds.

The time synchronization greatly benefits the data acquisition. On the one hand, the synchronized hardware timer controls the sampling frequency through the pulse-width modulation, what assures the desired sampling frequency. On the other hand, the timer determines the beginning of the sampling. So, the mapping of the samples to the UTC time is reliably done.

The conflict of the high CPU demands of both, the data acquisition and the data transmission, is solved by applying a mutex for the CPU. Anyhow, considering a long-term usage it is recommended to implement a hardware trigger and start the data acquisition only if a threshold is passed. Therefore, the requirement of the signal opening needs to be discarded. A trade-off is the continuously sampling by just a single node.

The presented strategies for the local and the distributed event detection can be applied and expediently tested firstly, when several infrasonic sensor nodes are available. Again, the missing of an adequate reference signal complicates a contingently ongoing development. Future work should concentrate on the remote parameterizing of the thresholds. The event detection is an own field of research. An improvement here would payoff by reducing the power consumption of the network.

Another big challenge was the implementation of the routing protocol in respect to the time slots. The procedure was very time consuming and the test results didn't show a sufficient behaviour. Anyhow, we showed the functioning in principle and submitted possible adjustments. The most promising approach is the scanning of the time slots in the case of asynchrony. The benefit of the sleep scheduled radio for almost 97% of the time still reasons such an approach. A detailed analyze and proving of the routing protocol can fill an own study.

The implemented concept of the virtual data memory was highly optimized to the demands of the accessing components. The SD card read/write operations are only performed when it is absolutely required. For the data acquisition a free memory block is always assured. The Network component finds enough free memory in the reception time slot, while in the transmission time slot the memory is filled by the packets to be forwarded. In order to serve gateway requests of missing packets a history maps the packets to SD card addresses. The thereby required trade-off between the history size and the SD card address limitation can be solved by considering the swapping out the entire history to the SD card.

Finally, designing a circuit for the regeneration of chargeable batteries by solar panels could be the essential step towards a real long-term usage.

Wireless sensor networks present many exciting opportunities. The developed system is easily to customize in order to operate in different applications. For instance, the orthogonal usage of the infrasonic microphones can be used to measure the resonance frequency of high buildings. The implementation of a burglar alarm is another example, as well as tracing of

moving objects by the time synchronized nodes. Nevertheless, exchanging the microphones to other sensors opens a broad spectrum of possibilities.

8. References

- Chilo, J. (2008). *Low-Frequency Signal Classification: Filtering and extracting features from infrasound data*, Verlag Dr. Muller, ISBN 978-3639113419.
- Chilo, J.; Jabor A.; Liszka L.; Eide A. J.; Lindblad & Persson L. (2006). *Infrasonic and Seismic Signals from Earthquake and Explosions in Arequipa, Peru*, Proceedings of Western Pacific Geophysics Meeting, Beijing China, 2006.
- Chilo J. & Lindblad Th. (2007). *A Low Cost Digital Data Acquisition System for Infrasonic Records*, International Workshop on Intelligent Data Acquisition and Advanced Computing Systems, pp. 35-37, IEEE Catalog Number 07EX1838C.
- Culler D.; Estrin D. & Srivastava M. (2004). *Overview of wireless sensor networks*, IEEE Computer, Special Issue in Sensor Networks.
- Ditzel M. & Langendoen (2005). *D3: Data-centric Data Dissemination in Wireless Sensor Networks*, IEEE Computer Society, The European Conference on Wireless Technology.
- Werner-Allen G.; Johnson J.; Ruiz M.; Lees J. & Welsh M. (2005). *Monitoring Volcanic Eruptions with a Wireless Sensor Network*, Second European Workshop on Wireless Sensor Networks, EWSN'05.
- Werner-Allen G.; Welsh M.; Locrincz K.; Johnson J.; Marcillo O.; Ruiz M. & Lees J. (2006). *Deploying a Wireless Sensor Network on an Active Volcano*, IEEE Computer, Sensor-Network Applications.

On Position and Attitude Estimation for Remote Sensing with Bistatic SAR

Stefan Knedlik, Junchuan Zhou and Otmar Loffeld
*Center for Sensorsystems (ZESS), University of Siegen
Germany*

1. Introduction

In recent years, there has been an increasing interest in remote sensing with bistatic SAR. Among the advantages bistatic SAR imaging offers in comparison to monostatic SAR imaging are that additional information can be exploited (specific bistatic angles can be chosen, and additional information is obtained from the bistatic reflectivity of targets and because of a reduction of di- and polyhedral effects), that SAR imaging along along-track direction will be feasible, that a cost reduction, as well as reduced size, weight, energy consumption can be achieved for passive receive-only systems, and that passive systems have a reduced vulnerability.

Several experiments have been carried out to prove the feasibility of remote sensing with bistatic SAR. While in the first experiments a stationary receiver or transmitter was involved or two airplanes (with almost parallel flight trajectories) have been used, so called hybrid experiments with Germany's national remote sensing satellite TerraSAR-X as illuminator and with an airborne SAR receiver system have been performed recently, and first processing results have been presented (cf. (Rodriguez-Cassola et al. 2008) or (Ender et al. 2006; Walterscheid et al. 2009) (where a larger bandwidth and a double sliding spotlight mode have been used)).

Among the research challenges in remote sensing with bistatic SAR are the derivation of proper processing algorithms (that yield focused images even in the most general case of arbitrary flight trajectories) and the *position and attitude determination of the SAR antennas* (with the accuracy and real-time ability required, and at relatively low cost).

Accurate position and attitude knowledge of the involved SAR antennas is required at several steps: It is an important issue in *footprint chasing*, that is, to obtain overlapped transmitter and receiver antenna footprints (by appropriate antenna steering) during the mission. In the hybrid experiments, the magnitude of the satellite velocity is about 76 times higher than that of the airplane. The overlap of the antenna footprints will be a few seconds in maximum (the operation in a double sliding spotlight mode is recommended). The aircraft has to fly over the target scene in time. Antenna steering can be used to compensate for smaller errors. The aircraft's trajectory and the attitude of its SAR antenna have to be known in real-time and typically absolute information is required (independent from the satellite (because the related information is not completely available from the satellite)). For

footprint chasing in such hybrid bistatic SAR experiments an antenna pointing error of about 0.5° in all the three angles is acceptable. The positioning requirements are even less strict.

Furthermore, accurate absolute position and attitude information is required for *motion compensation* and for *parameter estimation* (with respect to a_0 = time difference parameter, and a_2 = slant range ratio parameter (required during raw data processing)). Recently, in (Wang et al. 2009), the effect of uncompensated errors on the bistatic point target reference spectrum has been analyzed for the airborne/airborne case.

Absolute position and attitude information with highest accuracy is (in several applications) required for *geo-referencing*.

Moreover, in SAR interferometry the *baseline* (the vector between the antenna phase centers) has to be known. Depending on the parameters (e.g., wavelength, orbit) of illuminator and receiver, the baseline has to be estimated very accurately (e.g., with mm/cm (length) and arcsec (angle) accuracy, respectively) to obtain height errors smaller than 1 m. Here it is relative position and attitude information that is required (relative with respect to the carrier platforms), and here it is often not required in real-time.

The aforementioned position and attitude information can be obtained using inertial navigation or global navigation satellite systems (GNSS). In the following section, some fundamentals regarding these kinds of navigation are introduced, and the data fusion of corresponding measurements is considered. Example data fusion approaches for low cost position and attitude determination are given and analyzed in Section 3.

2. Introduction to GNSS/INS integration

2.1 Inertial navigation, satellite navigation, and its complementary characteristics

In inertial navigation, usually, *gyroscopes* and *accelerometers* are used, and they are mounted in triads so that the sensitive axes of the sensors are mutually orthogonal, setting up a Cartesian reference frame. In an *inertial measurement unit* (IMU), which contains the inertial sensor assembly, the raw data provided by the inertial sensors is converted to angular rates (from gyroscopes) and specific forces (from accelerometers), and typically an integration of the raw data over a certain time and also a calibration is performed. The output of an IMU are angular rates ω_{ib}^b of the body-fixed frame (*b*-frame) with respect to the inertial frame (*i*-frame) (see subscript) given in the *b*-frame (superscript) and specific forces \mathbf{a}_{ib}^b (or, because of the integration, also delta-theta's and delta-V's, respectively). See Figure 1. These data can be processed yielding position, velocity, and attitude, which, in case of systems where the inertial sensor assembly is strapped down to a body frame of the host platform, has been coined *strapdown processing*. That is, starting from a known initial position, velocity, and attitude, dead reckoning is performed using the measurements of the inertial sensors. A system that contains an IMU and an appropriate processing unit has been coined *inertial navigation system* (INS). Strapdown inertial navigation is in detail explained in the literature. With an INS autonomous navigation (almost independent of the environment) at a high data rate (e.g., 100 Hz) can be realized. The full attitude information is available. However, the system has to be initialized, and information about the local gravity is required. Furthermore, due to the dead reckoning principle the errors are growing unbounded with

time. That is, even if expensive systems (e.g., with a price tag of 50.000\$–100.000\$) are used, position errors of about 1 km/h result if no calibration and external aiding is applied.

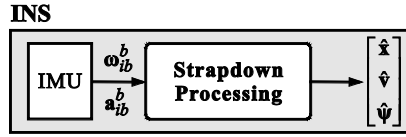


Fig. 1. Simplified block diagram of an inertial navigation system

Global navigation satellite systems (GNSS) such as the Global Positioning System (GPS) can be used to determine position, velocity, and time of a point on a platform. By measuring the time of arrival (TOA) and by using the transmission time, which can be extracted from the received signal, the propagation time and finally the range, the receiver-satellite distance, can be derived. By trilateration the receiver position can be determined. Typically, four or more simultaneous measurements are required to solve for the 3D receiver position and clock bias. The carrier phase/frequency and the code phase of the received signals are tracked by appropriate phase/frequency and delay lock loops. There is a correlation of the inphase (I) and quadrature phase (Q) components with replica signal quadrature components, and the I,Q samples are integrated and dumped. Phase/frequency of the replica carrier and phase of the replica code are the raw measurements of a GNSS receiver; and from these raw measurements pseudorange, delta range (or Doppler, carrier phase) and accumulated delta range (integrated Doppler) observations can be derived, and finally in a navigation processor (typically a Kalman filter) position, velocity, and time can be estimated based on the observations and a model of the dynamics (cf. Figure 2). More information about the systems, the signals, the error sources, the positioning methods (e.g., differential GNSS), and augmentation systems and services can be found in the literature.

GNSS based navigation is non-autonomous. Usually, at least 4 GNSS satellites have to be continuously in view. It is depending on the environment and there is a high vulnerability. On the other hand, one obtains absolute position, velocity and time information which is long-term stable. The output rate is relatively low (e.g., 1 Hz), and regarding the carrier phase measurements, the ambiguity resolution and the cycle slip detection and repair is challenging.

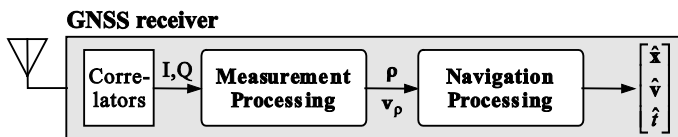


Fig. 2. Simplified block diagram of a GNSS receiver

As indicated above, GNSS based navigation and inertial navigation and the corresponding navigation solutions have *complementary characteristics*. Hence, GNSS/INS integration (in the sense of data fusion) is useful to obtain a complete and continuous navigation solution with high accuracy and high bandwidth at relatively low cost.

2.2 GNSS/INS integration approaches

Raw measurements and derived observations available from inertial navigation and GNSS based navigation, respectively, have been briefly mentioned in the preceding section.

From an information-theoretical point of view, it would be optimal if the raw measurements would be processed in a single centralized filter using, for example, a total state space model for the state space modelling. Especially if the I, Q components at the output of the correlators in the tracking loops are utilized, and if there is a feedback of the estimated Doppler from the fusion filter to the numerically-controlled oscillator (NCO) - which has been coined INS aiding GNSS - it is known as *deeply coupled* (or ultra-tightly coupled) GNSS/INS integration. (However, there is no commonly agreed definition of it). In that case, because of the INS aiding GNSS, only the residuals of the receiver dynamics have to be tracked in the tracking loops. The bandwidth is reduced, accuracy, and robustness can be improved, and the tracking can be faster. In practice, deeply coupled GNSS/INS integration is usually not applied. Access to the tracking loops of the GNSS receiver is usually not given. Moreover, there is a relatively high computational burden (e.g., from theory, a total state space model has to operate at a relatively high data rate) and relatively poor fault-tolerance. However, depending on the application (and specific requirements) it could outperform other approaches. In (Wagner and Wieneke 2003), the incorporation of the strapdown processing into the fusion filter and the use of a total state space filter have been proposed.

On the other hand, a decentralized, a distributed, estimation architecture can be considered which exploits the outputs of a GNSS receiver and of an INS. It is called a *loosely coupled* GNSS/INS integration architecture. Systems off-the-shelf can be used, and with the GNSS receiver and the INS independent and redundant navigation solutions are available. Drawbacks of a loosely coupled GNSS/INS integration are that typically four satellites have to be in view to obtain information from the GNSS receiver and that in case of a Kalman filter in the GNSS receiver (denoted by navigation processor in Figure 2) time correlated estimates of position and velocity are the input for the fusion filter (there are cascaded filters) which has to be accounted for (e.g., by considering this information only every >10 s or by using a federated filter with high computational load). Moreover, cross correlations between position and velocity estimates exist, which can in practice often not be considered - because the belonging measurement noise covariance matrices are often not or not completely provided by the GNSS receiver - yielding a decreased performance.

Finally, one can distinguish a *tightly coupled* integration from the aforementioned approaches where pseudoranges and delta ranges or carrier phases (which are outputted by many GNSS receivers) are exploited. Different definitions of a tightly coupled GNSS/INS integration (e.g., with or without feedback to the GNSS receiver) can be found in the literature. In general, a tightly coupled filter is more complex (the fusion filter) than the loosely coupled filter but the estimation is more robust, and it is especially when signals from less than four satellites can be received superior.

Regarding the state space modelling, an *error state space model* is usually applied, which is also known as *indirect filtering*, cf. (Maybeck 1982). The strapdown processing based on the measurements of the inertial sensors is done separately at a high rate. It provides a reference trajectory. Measurements for the fusion filter are differences between GNSS receiver measurements and predicted measurements based on the INS output. The fusion filter operates at a relatively low rate at which GNSS based measurements are available. The

dynamics are given by the inertial error differential equations which can be well modelled as being linear.

The estimated errors are usually fed back to correct the IMU measurements yielding errors at the output of the INS which do not grow unbounded with time but remain small so that the assumption of a linear system model remains reasonable.

An exemplary GNSS/INS indirect feedback tightly coupled integration architecture is shown in Figure 3.

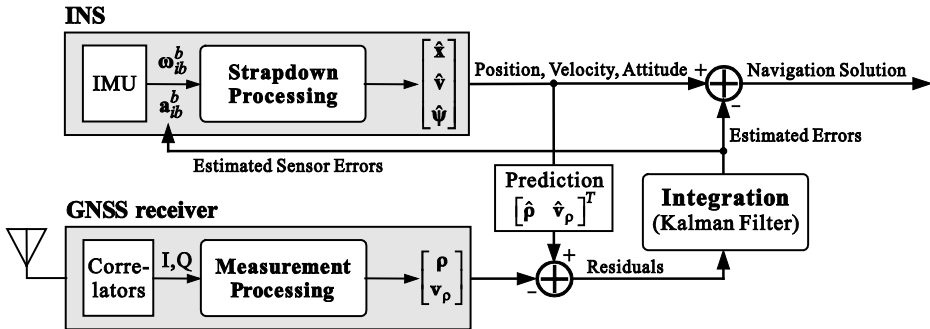


Fig. 3. GNSS/INS tightly coupled indirect feedback integration principle (example)

In general, the development of a proper GNSS/INS integration approach depends on the application. For example, the dynamics of the platform have to be taken into consideration (e.g., orbiting satellite versus unmanned aerial vehicle (UAV)) - not only for a possible modelling of the dynamics but also for a proper derivation and consideration of expected errors (e.g., with respect to multipath), required bandwidth and update rates, possible positioning or initialization methods, etc. Furthermore, among others, the structure of the platform has to be considered, the environment has to be considered, constraints can be derived and incorporated, and the grade of the inertial sensors and the GNSS receivers available has to be taken into consideration.

More details regarding GNSS/INS integration are given in the literature (e.g., (Farrell and Barth 1999)).

3. Low-cost GPS/INS integration with multiple GPS antennas

In this section, an example for GNSS/INS integration is given. GNSS/INS integration for position, velocity, and attitude estimation of an antenna mounted on an aircraft will be considered. The focus is on low cost. That is, a low-cost microelectromechanical system (MEMS) based IMU and L1 GPS receivers (that can output pseudorange, delta range, and carrier phase measurements) are supposed to be available.

Tightly coupled GPS/INS indirect feedback sensor data fusion approaches will be considered. Different proposals to integrate additional, redundant attitude information are compared.

3.1 Preliminary considerations

The data fusion approaches are formulated with respect to the n -frame (with axes pointing locally north, east, down, respectively).

The error is defined as usual, that is, as observed or estimated value – true value. The state vector for the tightly coupled integration is chosen to be

$$\Delta \mathbf{x}_c = \begin{bmatrix} [\Delta x_n^n \quad \Delta x_e^n \quad \Delta x_d^n]^T \\ [\Delta v_{eb,n}^n \quad \Delta v_{eb,e}^n \quad \Delta v_{eb,d}^n]^T \\ [\Delta \alpha \quad \Delta \beta \quad \Delta \gamma]^T \\ [\Delta b_{a,x} \quad \Delta b_{a,y} \quad \Delta b_{a,z}]^T \\ [\Delta b_{\omega,x} \quad \Delta b_{\omega,y} \quad \Delta b_{\omega,z}]^T \\ [c\Delta t_{rb} \quad c\Delta t_{rd}]^T \end{bmatrix} = \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \mathbf{v} \\ \Delta \boldsymbol{\psi} \\ \Delta \mathbf{b}_a \\ \Delta \mathbf{b}_\omega \\ c\Delta t_r \end{bmatrix} \left\{ \begin{array}{l} \text{position error} \\ \text{velocity error} \\ \text{error in misalignment} \\ \text{accelerometer bias error} \\ \text{gyroscope bias error} \\ \text{receiver clock error} \end{array} \right. \quad (1)$$

That is, the state vector comprises 17 states.

Accelerometer levelling can be used to determine the initial bank angle (roll) and elevation angle (pitch) of the platform from the accelerometer measurements as follows

$$\theta = \arctan \left(-a_{ib,x}^b / \sqrt{(a_{ib,y}^b)^2 + (a_{ib,z}^b)^2} \right) \quad (2)$$

$$\varphi = \arctan_2 \left(a_{ib,y}^b / a_{ib,z}^b \right) \quad (3)$$

The MEMS-based IMU with a typical bias instability between several 100°/h and several 10000°/h can not sense Earth's rotation. Hence, gyro-compassing for alignment in azimuth can not be performed. Other sensors, such as a magnetometer, can be used to derive heading. Heading information can, for example, also be derived from GPS velocity measurements according to

$$\psi = \arctan_2(v_E / v_N) \quad (4)$$

or from *a priori* knowledge.

It can be shown by an observability analysis that only with additional redundant attitude information the states are completely observable (independent of the manoeuvre of the platform, dependent on the number of satellites in view).

This redundant attitude information can be provided by a multi-antenna GNSS receiver system. Here we use a non-dedicated system consisting of the master GPS receiver (including antenna) and two more (independent) GPS receivers (with antennas).

Because of an approximately straight and level flight, that can be assumed for a remote sensing experiment, the size effect related to the accelerometer triad can be neglected. Moreover, because of the low-cost IMU, in addition to other approximations (e.g., no Euler acceleration, that is, Earth's rotation rate assumed to be constant), the transport rate and Coriolis terms can be neglected in the strapdown processing and in the system model.

In subsequent sections, continuous-time models are provided. The appropriate discrete-time models can be derived as shown in the literature, e.g., in case of a time-variant system the state transition is described as

$$\begin{aligned}
A(k) &= \phi((k+1)T, kT) = \phi(T) = e^{F((k+1)T-kT)} = e^{(FT)} \\
&= \mathcal{L}^{-1} \left\{ [sI - F]^{-1} \right\} \Big|_{t=T} = I + F \cdot T + \frac{(F \cdot T)^2}{2} + \text{h.o.t.}
\end{aligned} \tag{5}$$

where T is the sampling period. That is, if the continuous-time state transition matrix $F(t)$ is time-invariant or only slightly varying with time, one can approximate

$$A(k) \approx I + F \cdot T \tag{6}$$

3.2 Strapdown mechanization

With aforementioned simplifications due to the characteristic of a low-cost IMU, the mechanization can be expressed as

$$\dot{\mathbf{x}}^n = \mathbf{v}^n \tag{7}$$

$$\dot{\mathbf{v}}^n = R_b^n \mathbf{a}_{ib}^b + \mathbf{g}_l^n \tag{8}$$

$$\dot{q}_b^n = \frac{1}{2} q_b^n \begin{bmatrix} 0 \\ \boldsymbol{\omega}_{nb}^b \end{bmatrix} = \frac{1}{2} q_b^n \begin{bmatrix} 0 \\ \boldsymbol{\omega}_{ib}^b \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 0 \\ \boldsymbol{\omega}_{in}^n \end{bmatrix} q_b^n \approx \frac{1}{2} q_b^n \begin{bmatrix} 0 \\ \boldsymbol{\omega}_{ib}^b \end{bmatrix} \tag{9}$$

It is performed with the specific force vector $\tilde{\mathbf{a}}_{ib}^b$, related to the measurement of the accelerometer triad, with the angular rate vector $\tilde{\boldsymbol{\omega}}_{ib}^b$, related to the measurement of the gyroscopes, with the local gravity vector resolved in the n -frame $\hat{\mathbf{g}}_l^n$, and where \mathbf{x} (or $\hat{\mathbf{x}}$) is (computed) position, \mathbf{v} (or $\hat{\mathbf{v}}$) is (computed) velocity, and where the frame rotation from b -frame to n -frame is described by the computed direction cosine matrix \hat{R}_b^n . The computed platform rotation rate with respect to the inertial frame, $\hat{\boldsymbol{\omega}}_{in}^n$, is the sum of Earth's rotation rate (depending on latitude) and the transport rate (depending on the speed of the platform). Both contributions can be easily computed. Moreover, in some cases, depending on the application (gyroscope bias instability, velocity of the platform), these terms can be neglected. Note that the products in Eq. (9) are quaternion products as defined by Hamilton.

3.3 Error state system model

Because of the low-cost IMU, the uncompensated systematic error $\Delta \mathbf{b}_a$ and $\Delta \mathbf{b}_\omega$ in the measurements of the inertial sensors are considered as states, and they are modelled as random walk processes. The receiver clock drift (related to the frequency error) is modelled as constant plus a random walk process, and the clock bias $c\Delta t_r$ (related to the phase error) is the integral of it.

With the aforementioned constraints and simplifications, the n -frame error state system model for the tightly coupled integration is set up as

$$\underbrace{\begin{bmatrix} \Delta \dot{\mathbf{x}} \\ \Delta \dot{\mathbf{v}} \\ \Delta \dot{\Psi} \\ \Delta \dot{\mathbf{b}}_a \\ \Delta \dot{\mathbf{b}}_\omega \\ c\Delta \dot{t}_r \\ c\Delta \ddot{t}_r \end{bmatrix}}_{\Delta \mathbf{x}_c} = \underbrace{\begin{bmatrix} O & I & O & O & O & \mathbf{0} & \mathbf{0} \\ O & O & F_{23} & -\hat{R}_b^n & O & \mathbf{0} & \mathbf{0} \\ O & O & O & O & -\hat{R}_b^n & \mathbf{0} & \mathbf{0} \\ O & O & O & O & O & \mathbf{0} & \mathbf{0} \\ O & O & O & O & O & \mathbf{0} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T & 0 & 1 \\ \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T & 0 & 0 \end{bmatrix}}_F \cdot \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \mathbf{v} \\ \Delta \Psi \\ \Delta \mathbf{b}_a \\ \Delta \mathbf{b}_\omega \\ c\Delta t_r \\ c\Delta \dot{t}_r \end{bmatrix} + \mathbf{w} \quad (10)$$

where O and I denote a 3×3 identity and 3×3 zero matrix, respectively, and the sub-matrix F_{23} is a skew-symmetric matrix that contains the specific force components transformed to the n -frame

$$F_{23} = F_{a \times \psi} = [\hat{\mathbf{a}}_{ib}^n \times] = \begin{bmatrix} 0 & \hat{a}_{ib,d}^n & -\hat{a}_{ib,e}^n \\ -\hat{a}_{ib,d}^n & 0 & \hat{a}_{ib,n}^n \\ \hat{a}_{ib,e}^n & -\hat{a}_{ib,n}^n & 0 \end{bmatrix} \quad (11)$$

3.4 Observation models

In the tightly coupled integration the measurement vector contains the differences in predicted (based on strapdown solution for position and velocity) and measured pseudorange ρ and delta range v_ρ , respectively. Moreover, if redundant attitude information is available, e.g., derived from an independent GPS multiple-antenna system, the difference between predicted (from INS) and true attitude measurement can be included in the measurement vector. With respect to satellite number m , we have (in the n -frame)

$$\mathbf{y}_{t_1}^{(m)} = \begin{bmatrix} \hat{\rho}^{(m)} - \tilde{\rho}^{(m)} \\ \hat{v}_\rho^{(m)} - \tilde{v}_\rho^{(m)} \\ \hat{\Psi} - \tilde{\Psi} \end{bmatrix} \quad (12)$$

In the tightly coupled integration, the states are *nonlinearly* mapped into the observation space. Hence, an extended Kalman filter can be used for the estimation of the states. The Jacobian has to be computed. The resulting observation matrix, that maps the 17 state vector components (Eq. (1)) into observation space, is

$$H_{t_1}^{(m)} = \begin{bmatrix} -(\mathbf{1}_t^{(m),n})^T & \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T & 1 & 0 \\ \mathbf{0}^T & -(\mathbf{1}_t^{(m),n})^T & \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T & 0 & 1 \\ O & O & R & O & O & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (13)$$

where $\mathbf{1}_t^{(m),n}$ is the unit vector in the line of sight from receiver (master GPS antenna A0) to satellite number m , resolved in the n -frame, and where R is a frame rotation matrix for the transformation from body-axes angular rates to the Euler angle angular rates.

The matrix has the dimension $(3 + \nu \cdot 2) \times 17$, where ν is the number of satellites in view.

Sequential processing has to be performed to avoid inversion of a huge matrix.

If the redundant attitude information is obtained from a multi-antenna GPS receiver system we can, instead of first computing the attitude, directly exploit the double-difference carrier phase measurements of the system. This kind of integration of the attitude information from a GPS multiple-antenna system has been proposed in (Hirokawa and Ebinuma 2009). In the measurement vector we have then rather differences of predicted and present double-difference carrier phase measurements (related to antennas A0, A1, and A2) than the difference in attitude. In case of two satellites, m and n , the measurement vector is given as

$$\mathbf{y}_{i_2}^{(m,n)} = \begin{bmatrix} \hat{\rho}^{(m)} - \tilde{\rho}^{(m)} \\ \hat{\tilde{\nu}}_{\rho}^{(m)} - \tilde{\tilde{\nu}}_{\rho}^{(m)} \\ \hat{\rho}^{(n)} - \tilde{\rho}^{(n)} \\ \hat{\tilde{\nu}}_{\rho}^{(n)} - \tilde{\tilde{\nu}}_{\rho}^{(n)} \\ \nabla\Delta\hat{\varphi}_{A0,A1}^{(n,m)} - \nabla\Delta\tilde{\varphi}_{A0,A1}^{(n,m)} \\ \nabla\Delta\hat{\varphi}_{A0,A2}^{(n,m)} - \nabla\Delta\tilde{\varphi}_{A0,A2}^{(n,m)} \end{bmatrix} \quad (14)$$

where the symbols in parentheses represent satellites (and where the time index has again been neglected for convenience). The true double-difference carrier phase observation can be expressed as shown by the following example:

$$\begin{aligned} \nabla\Delta\varphi_{A0,A1}^{(n,m)} &= \frac{2\pi}{\lambda} \left[\left(\rho_{A1}^{(n)} - \rho_{A1}^{(m)} \right) - \left(\rho_{A0}^{(n)} - \rho_{A0}^{(m)} \right) \right] \\ &= \frac{2\pi}{\lambda} \left[\left(\rho_{A1}^{(n)} - \rho_{A0}^{(n)} \right) - \left(\rho_{A1}^{(m)} - \rho_{A0}^{(m)} \right) \right] \end{aligned} \quad (15)$$

and the measured double-difference carrier phase can be modelled as

$$\nabla\Delta\tilde{\varphi}_{A0,A1}^{(n,m)} = \frac{2\pi}{\lambda} \cdot \nabla\Delta\rho_{A0,A1}^{(n,m)} + \frac{2\pi}{\lambda} \cdot c \cdot \nabla\Delta MP_{\varphi,A0,A1}^{(n,m)} + e_{\varphi\nabla\Delta} - 2\pi \cdot \nabla\Delta N_{A0,A1}^{(n,m)} \quad (16)$$

where besides the integer ambiguity $\nabla\Delta N$, only the non-common mode errors multipath $\nabla\Delta MP$ and receiver noise $e_{\varphi\nabla\Delta}$ have to be considered.

As soon as there is another satellite in view, the measurement vector is augmented by four more rows (pseudorange and delta range differences to the new satellite, respectively, and new double-difference carrier phase measurements related to the master satellite m).

In case of $\nu > 1$ satellites in view, we obtain obviously a $(2 + (\nu - 1) \cdot 4) \times 17$ matrix that relates the 17 state vector components to the observations. The Jacobian, the observation matrix, corresponding to Eq. (14) (2 satellites in view) is

$$H_{i_2}^{(m,n)} = \begin{bmatrix} -(\mathbf{1}_t^{(m,n)})^T & \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T & 1 & 0 \\ \mathbf{0}^T & -(\mathbf{1}_t^{(m,n)})^T & \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T & 0 & 1 \\ -(\mathbf{1}_t^{(n,n)})^T & \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T & 1 & 0 \\ \mathbf{0}^T & -(\mathbf{1}_t^{(n,n)})^T & \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T & 0 & 1 \\ \mathbf{0}^T & \mathbf{0}^T & \frac{2\pi}{\lambda}(\mathbf{1}_t^{(m,n)} - \mathbf{1}_t^{(n,n)})^T [\hat{R}_b^{n_1} \mathbf{1}_{01}^b \times] & \mathbf{0}^T & \mathbf{0}^T & 0 & 0 \\ \mathbf{0}^T & \mathbf{0}^T & \frac{2\pi}{\lambda}(\mathbf{1}_t^{(m,n)} - \mathbf{1}_t^{(n,n)})^T [\hat{R}_b^{n_2} \mathbf{1}_{02}^b \times] & \mathbf{0}^T & \mathbf{0}^T & 0 & 0 \end{bmatrix} \quad (17)$$

where $\mathbf{1}_t^{(m,n)}$ is the unit vector in the line of sight from master antenna (A0) to satellite m , resolved in the n -frame, and $\mathbf{1}_{0i}^b$ is the lever arm (baseline) between the phase centers of master antenna (A0) and antenna number i , known in the b -frame.

3.5 Further remarks to the integration approach

Derivations of the Kalman filter and its augmentations for nonlinear models as well as the resulting algorithms are not repeated here. The reader is for example referred to (Maybeck 1982), (Simon 2006).

The measurements from the IMU or from the inertial navigation system can be expected with a high data rate (e.g., 100 Hz), which is much higher than the data rate of the measurements derived from GPS. Hence, to reduce the computational burden, to obtain a filter that operates at a relatively low rate, the measurements from the inertial sensors are not incorporated as measurements (cf. Section 3.4). Instead, they are directly used in the strapdown processing to predict position, velocity, and attitude. In addition, only the *a priori* error covariance matrix $P^-(k)$ has to be predicted as usual (in the *prediction* step, and not necessarily at the high data rate).

As soon as a new measurement vector can be computed, that is, as soon as information from GPS is available, the error state is estimated, and the state vector is *corrected*. In case of small rotations there is approximately no difference between the components of the rotation vector and the Euler angles. That is, the estimated attitude error can be interpreted as rotation vector, and the correction of the *a priori* attitude quaternion is done by computing the quaternion product $\hat{q}_{b,k}^{n,+} = \hat{q}_{b,k}^{n,-} q_k^n$, where the n -frame rotation quaternion is computed

with $\phi = -\Delta \hat{\psi}$, $\phi = \sqrt{\phi^T \phi}$ and with $q^n = [\cos(\phi/2) \sin(\phi/2)\phi/\phi]^T$ (using a Taylor series approximation). The correction of the other states is straightforward (subtraction of the appropriate estimated error state). After the correction, the errors state is zeroed.

3.6 Simulation setup and results

For the following experiments a *hardware-in-the-loop* system has been used for the GPS part. It consists of the RF GPS signal simulator system NavX-NCS from Ifen GmbH and Novatel DL-4 plus GPS receivers. Synthetic inertial measurements (as obtained from a low-cost IMU)

have been generated using the parameters of a typical consumer grade IMU (cf. Table 1). A straight and level flight with a nominally constant velocity of 110 m/s in east has been modelled. The parameters describing the experiment are given in Table 1.

Trajectory	
Velocity (ENU) [m/s]	$\mathbf{v}^E = [110 \ 0 \ 0]^T$
Initial Position (φ, λ, h)	$\mathbf{x}^e = [51^\circ\text{N} \ 8^\circ\text{E} \ 3 \text{ km}]^T$
Angles (φ, θ, ψ) [$^\circ$]	$\boldsymbol{\psi} = [0 \ 0 \ 90]^T$
Start Time (UTC)	October 29, 2006, 00:11:27
Duration	180 s
GPS measurements	
Update rate	1 Hz
Measurements	Depending on integration approach: code pseudo-range, delta range, (double-differenced carrier phase)
Method	Point positioning, and attitude determination from a non-dedicated multiple antenna system (3 antenna-receiver pairs with baselines of 2 m each)
Error modelling	Tropospheric delay is estimated and corrected for; The small ionospheric delay is estimated and corrected for; No multipath (calibrated out); Ambiguities resolved and cycle slips repaired (assumed)
Satellites in view	2-8 (as depicted in Figs. 4-5)
Elevation angle	$\geq 5^\circ$
IMU measurements	
Update rate	100 Hz
Error modelling	<i>Gyroscope (Angular rates)</i>
	Bias stability [$^\circ/\text{h}$] 360
	Scale factor [ppm] 10000
	Noise (ARW) [$^\circ/\text{h}/\sqrt{\text{Hz}}$] 180
	<i>Accelerometer</i>
	Bias stability [μg] 2400
	Scale factor [ppm] 10000
	Noise (VRW) [$\mu\text{g}/\sqrt{\text{Hz}}$] 1000

Table 1. Nominal parameters describing the setup

An extended Kalman filter has been used for the GPS/INS integration which is based on the state space modelling provided in Sections 3.1-3.5. An ideal time-synchronization has been assumed, time-delayed measurements have not been taken into consideration. An initialization has been performed as described in Section 3.1.

The attitude estimation results are depicted in Figure 4.

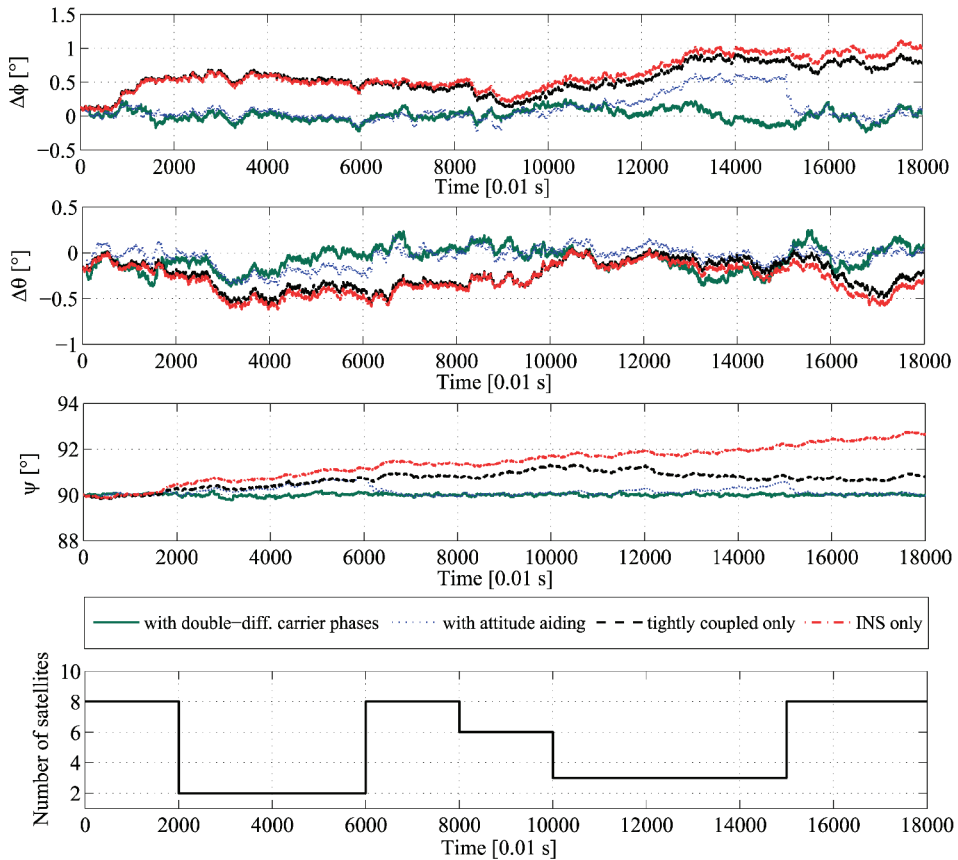


Fig. 4. Attitude estimation result using tightly coupled GPS/INS integration approaches

Because of the proper initialization (including an estimation of the gyroscope bias instability and an appropriate correction) of the IMU, the errors are relatively small, that is, after 3 minutes (180 s) the heading error is in case of the stand-alone IMU still smaller than 3° . Without initialization and calibration the error would be larger than 20° . As expected, in case of the stand-alone IMU, the largest error can be observed in heading.

In the tightly coupled GPS/INS integration without redundant attitude information the attitude errors (and sensor biases) are not directly mapped into observation space. However, these error states will also be updated with every new position and velocity measurement. These quantities are related to each other as shown by the system model, and this is reflected in the predicted error covariance matrix (used for computing the Kalman gain). It is therefore no surprise that the attitude estimation results using a tightly coupled GPS/INS integration (represented by the black, dashed curves in Figure 4) are better than using a similar INS alone.

A much more accurate attitude estimation result is obtained using additional redundant attitude information (shown by blue dotted lines in Figure 4). The measurement vector is

then given by Eq. (12). The redundant attitude information is here obtained from a multiple antenna GPS receiver system. In case of less than 4 satellites in view, the system can not provide attitude information. Hence, the errors become much larger during such periods. Finally, if the double-difference carrier phase measurements from the non-dedicated multi-antenna system are used (with the observation matrix given by Eqs. (17) (for the case of two satellites)) the results are comparable in case of four or more satellites in view, but if only two or three satellites are in view still a robust and accurate attitude information is obtained. In Figure 5, the obtained position, velocity, and attitude errors are shown.

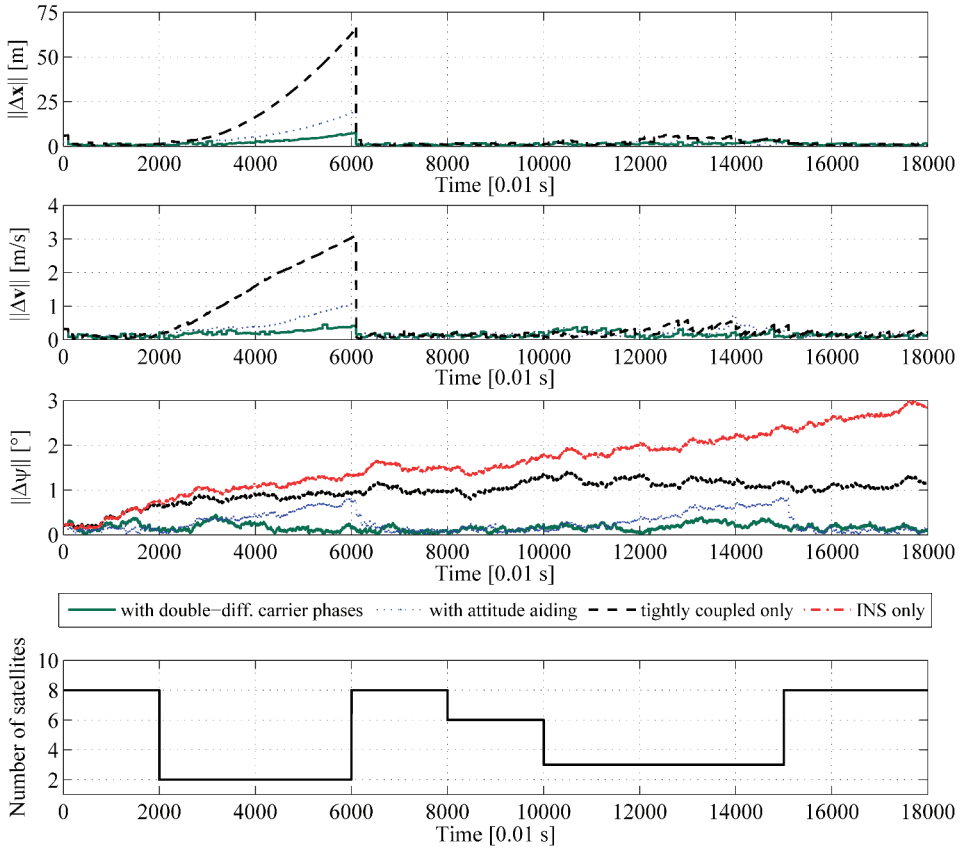


Fig. 5. Navigation solution using tightly coupled GPS/INS integration approaches

The large stand-alone IMU related position and velocity errors (about 1 km and 15 m/s after 3 minutes, respectively) are not included in the figure. In general, if only two satellites are in view, the position and velocity errors grow remarkably. During the period where only three satellites are in view, the position and velocity errors are bounded (which is not the case in a loosely coupled GPS/INS integration). Again, the incorporation of the double-difference carrier phase measurements into the integration approach yields the best results.

3.7 Conclusions

Tightly coupled indirect feedback GNSS/INS integration approaches have been compared in Section 3. Required additional redundant attitude information can be derived from GNSS using three antenna-receiver pairs. In that case, the double-difference carrier phase measurements should be directly incorporated into the integration approach, as has been shown by looking at approaches for low-cost GPS/INS integration and appropriate simulation results. Such approaches and low-cost sensors can for example be used for footprint chasing in bistatic SAR.

Carrier phase measurements have to be exploited to achieve the required attitude estimation accuracy. The detection and repair of cycle slips and the rapid integer ambiguity resolution are challenging. It is easier if shorter baselines are used (because of a reduced search space), however, the longer the baselines the better the expected attitude accuracy.

The lever arm between specific force origin of the IMU and phase center of the main GNSS antenna can be easily incorporated.

Depending on the application, the assumptions have to be proven. For example, delayed measurements have to be considered, and depending on the platform (e.g., UAV) multipath errors that can not be calibrated are an issue, and unconsidered flexure and vibrations can remarkably decrease the accuracy.

Constraints, such as that an airplane is in straight and level flight during a remote sensing experiment, can be easily incorporated into the estimation approach.

A novel GNSS/INS direct integration approach that exploits direction cosine matrix orthogonality constraints and that incorporates a model of the dynamics has been proposed recently (Edwan et al. 2009).

The redundant attitude information required can possibly also be derived from a comparison of a quick-look SAR image with an existing digital map.

To obtain more accurate position and attitude estimates, an IMU of a higher grade and precise point positioning can be exploited.

4. Summary

The importance of accurate and reliable attitude and position information for remote sensing with bistatic SAR has been pointed out in the introduction. In the following section, inertial navigation and navigation based on global navigation satellite systems have been mentioned. Reasons for GNSS/INS integration have been provided, and appropriate data fusion architectures have been briefly introduced and discussed. In the main section, low-cost tightly coupled GPS/INS integration approaches with and without additional redundant attitude information have been presented, and simulation results have been discussed.

5. Acknowledgements

Part of the work reported herein has been funded by the German Research Foundation (DFG), grant number KN 876/1-2, which is gratefully acknowledged.

6. References

- Edwan, E., Knedlik, S., Zhou, J., and Loffeld, O. (2009). "Constrained GPS/INS integration based on rotation angle for attitude update and dynamic models for position update." *ION ITM 2009 (The Institute of Navigation International Technical Meeting)*, Anaheim, CA, USA.
- Ender, J. H. G., Loffeld, O., Brenner, A. R., Wiechert, W., Klare, J., Gebhardt, U., Walterscheid, I., Knedlik, S., Weiss, M., Nies, H., Kirchner, C., Kalkuhl, M., Wilden, H., Natroshvili, K., Medrano-Ortiz, A., Amankwah, A., Kolb, A., and Ige, S. (2006). "Bistatic exploration using spaceborne and airborne SAR sensors - A close collaboration between FGAN, ZESS and FOMAAS." *IGARSS 2006 (IEEE International Geoscience and Remote Sensing Symposium)*, Denver, Colorado, USA.
- Farrell, J. A., and Barth, M. (1999). *The global positioning system and inertial navigation*, McGraw-Hill, New York.
- Hirokawa, R., and Ebinuma, T. (2009). "A low-cost tightly coupled GPS/INS for small UAVs augmented with multiple GPS antennas." *Navigation: Journal of The Institute of Navigation*, 56(1), 35-44.
- Maybeck, P. S. (1982). *Stochastic models, estimation, and control (Vol. I, II)*, Academic Press, New York.
- Rodriguez-Cassola, M., Baumgartner, S. V., Krieger, G., Nottensteiner, A., Horn, R., Steinbrecher, U., Metzger, R., Limbach, M., Prats, P., Fischer, J., Schwerdt, M., and Moreira, A. (2008). "Bistatic spaceborne-airborne experiment TerraSAR-X/F-SAR: data processing and results." *IGARSS 2008 (IEEE International Geoscience and Remote Sensing Symposium)*, III - 451-III - 454.
- Simon, D. (2006). *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*, Wiley-Interscience, Hoboken, N.J.
- Wagner, J. F., and Wieneke, T. (2003). "Integrating satellite and inertial navigation - conventional and new fusion approaches." *Control Engineering Practice*, 11, 543-550.
- Walterscheid, I., Espeter, T., Brenner, A. R., Klare, J., Ender, J. H. G., Nies, H., Wang, R., and Loffeld, O. (2009). "Bistatic SAR experiment with PAMIR and TerraSAR-X - setup, processing, and image results." *IEEE Transactions on Geoscience and Remote Sensing*, under review.
- Wang, R., Loffeld, O., Nies, H., Medrano Ortiz, A., and Knedlik, S. (2009). "Bistatic point target reference spectrum in the presence of trajectories deviations." *IET Radar, Sonar & Navigation*, 3(2), 177-185.

Unmanned Airborne Platforms For Disaster Remote Sensing Support

Vincent G. Ambrosia¹ and Steven S. Wegener²

¹California State University – Monterey Bay;

²Bay Area Environmental Research Institute (BAERI)

United States

1. Introduction

The remote sensing community is increasingly turning to Unmanned Aircraft Systems (UAS) for integration of sensors to support scientific and applications-oriented airborne missions. These UAS platforms are seen as providing support capabilities for applications that require long observation dwell times and/or require operations in regions that are generally too dangerous for manned aircraft to operate efficiently and effectively. One of the most viable utilities for UAS as remote sensing platforms is in supporting rapidly evolving disaster events, be they natural (wildfires, etc) or anthropogenic (chemical releases, etc). Civilian land, resources and disaster management agencies in the United States are critically examining the role of UAS for use in long-duration monitoring over disaster events. Some of the critical elements that must be included in the analysis are the availability or development of autonomous operating sensor systems for integration on these platforms, near-real-time data delivery capabilities from the platform sensor to the ground management teams, and data / information integration into strategic disaster mitigation / management activities. Additionally, there are a number of issues regarding adaptation of UAS in the National Airspace System (NAS) and how those UAS interact with manned airborne assets in an increasingly congested airspace. Advances in UAS platforms, sensor systems, data telemetry capabilities and data manipulation / visualization enhancements have been developed, demonstrated, and evaluated for wildfire situational use in the United States (Ambrosia, et al., 2008). Those capabilities will be described and will form a foundation from which to look towards future improvements to utilizing UAS to support the disaster mitigation / remote sensing community. An assessment of the critical operational and integration challenges are also addressed.

1.1 Background: Current Fire Observation Capabilities

Wildfires are highly dynamic phenomenon, and their progression, consumption rates, and intensity are not easily modeled or predicted. Varying vegetation composition, age class, and moisture content of a fire-prone region are key factors that affect rates of spread. Additionally, terrain, coupled with solar exposure and wind dynamics are key elements to predicting how and where fire will advance. Given the dynamic conditions of these

variables over the course of a wildfire event, it is critical to have current and timely intelligence on the fire location and condition of the fire-front, and unburned vegetation in the fire's path. This information, if provided frequently, allows the fire management team to plan fire attack appropriately, saving resources, time and possibly lives. A wildfire management team cannot attack or manage a fire without "intelligence" about the fire condition, location, speed, vegetation composition, access routes or numerous other factors. A key factor to managing a wildfire event is the ability to access satellite or airborne remote sensing information, at an appropriate temporal and spatial scale. The wildfire management agencies in the United States currently utilize satellite data provided by NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) sensor data to provide synoptic, 2-4 times-daily hot-spot detection of fire at continental scales (Giglio, et. al, 2003; Justice, et. al, 2002; Kaufman, et. al, 1998; Morisette, et. al, 2005; NASA - Goddard Space Flight Center, 2009; U.S. Forest Service 2009). The spatial resolution of MODIS is low / moderate (1000 meters), and is used to derive a regional estimate of fire distribution. Multiple daily observations allow some estimate of fire movement at large scales. Although the temporal frequency of the MODIS data is sufficient for individual incident management uses, the spatial resolution is insufficient for tactical fire management operations.

The U.S. fire management agencies cohesively manage national fire events through the National Interagency Fire Center (NIFC), located in Boise Idaho. The multi-agency operations coordinate the distribution of fire fighting assets for all major wildfire events in the U.S., and will assist in international operations when requested. As part of the NIFC operations, the organization maintains the National Infrared Operations (NIROPS). The NIROPS operate two manned aircraft, a Cessna Citation Bravo II and a Beechcraft King Air B-200, which employ thermal imaging systems onboard for wildfire mapping support nation-wide. During manned mission operations the NIROPS relies on night-time thermal infrared data capture to minimize hot spots false detects from thermally "bright" objects that may be evident during daytime missions. The two NIROPS aircraft operate at similar data capture attitudes (3050 meters (10,000 feet) Above Ground Level (AGL)), and have similar mission endurance capabilities (4-6 hour missions). During extended distance missions, the aircraft and crew will deploy to other bases of operation in the fire vicinity for multiple mission days.

The NIROPS have experimented with various data transfer capabilities including "drop tubes" containing hard copy image maps generated on the aircraft, as well as landing and handing off digital data storage media (USB "thumb" drives", etc) containing the thermal infrared (TIR) fire hot spot detection data. Recently, with the advent of moderate cost over-the-horizon (OTH) telemetry technology, transferring of data from an acquiring sensor on an aircraft is attainable. In 2009, both NIROPS aircraft will have near-real-time data telemetry capability from the thermal infrared sensor, but prior to that installation, fire crews relied on a data handoff following the plane landing after TIR acquisition. The process then required manual spatial data transfer of hot spot detections to incident team map bases. The process, from acquisition over fire to map generation of hot-spots, took over one hour.

The NIROPS aircraft operate with an instrument engineer on-board to maintain the instrument and provide the necessary processing of the data for telemetry distribution to ground incident managers. Each of these processes can be streamlined or automated to

reduce the instrument engineer interaction. Further, automation of instrument collections and on-board image processing can yield significant time savings over “manual” operations of the same functions. UAS, with their long-flight-durations, provide an efficiency improvement over manned aircraft, by allowing either long-term lingering over a single fire event, or by allowing multiple fires to be imaged over the 24-hour duration capability of the platform. The remote / long-duration operations capabilities and tools developed to support those operations are described in the following sections.

1.1.1 NASA Involvement in Fire Observations

- The National Aeronautics and Space Administration (NASA) and the United States Forest Service (USFS) have been collaborating to develop, demonstrate and utilize innovative airborne and satellite remote sensing tools and capabilities for gathering, distributing and analyzing near-real-time wildfire information (Ambrosia, et al., 2009). NASA has been at the forefront of aeronautics research in UAS technologies and has added both small and large UAS platforms to its portfolio of science / research aircraft. UAS platforms, like the Northrop-Grumman *Global Hawk*, and the General Atomics Aeronautical Systems, Inc. (GA-ASI), *Ikhana* (MQ-9 Predator-B), are operated by NASA to support the agency’s science mission objectives, which includes earth and atmospheric research, telecommunications, autonomous sensor operations and applied science missions in support of other partner agency goals. NASA is also at the forefront of satellite-derived autonomous data processing of sensor system information sets, and those same capabilities can be designed for cross-cutting use on UAS. Autonomous processes of sensor-derived spectral information reduces the “data-to-information” lag time common with standard manual processing streams. Autonomous image processing and manipulation allows the derivation of Level II information to be created from baseline spectral data through the development of complex spectral algorithms. Autonomous geo-rectification processes greatly reduce the amount of time for the information to be ingestible in a digital spatial context. Developments and use of both “line-of-sight” (LOS) and OTH data telemetry systems improve space-borne and airborne science mission data-sharing capabilities. On-board data telemetry systems allow the distance-sharing of data sets collected remotely to be transmitted from the UAS to a disparate investigator or science community. These capabilities are critical to support near-real-time image utility by disaster managers on rapidly changing events, such as wildfires and have been integrated and demonstrated and are described in the context of the following sections.

1.2 Wildfire Research and Applications Partnership (WRAP)

The Wildfire Research and Applications Partnership (WRAP) is a joint effort between the National Aeronautics and Space Administration (NASA) and the U.S. Forest Service to explore innovative technologies to improve remote sensing observations of fire events. Since 2003, the WRAP project demonstrates and transitions emerging observation and information technologies to operational utility by wildfire management agencies. Because of this unique partnership between wildfire personnel and the NASA, academia, and industry science and technology community, the wildfire management agencies are better poised to utilize and integrate the demonstrated capabilities to improve wildfire intelligence and reduce wildfire losses and mitigation expenditures. The WRAP project effort focuses on the

fire management community providing the requirements and metrics for improving wildfire observational strategies. The NASA and USFS team members then develop / mature technologies that meet the metrics and requirements defined by the fire community. This process is formalized in the WRAP project's Tactical Fire Remote Sensing Advisory Committee (TFRSAC) creation, which is described in the following sub-section.

1.2.1 Tactical Fire Remote Sensing Advisory Committee (TFRSAC) - The technology transfer successes of the WRAP project are the result of an innovative technical and scientific team structure that marries fire management personnel with science and engineering team members from NASA, academia and industry. The Tactical Fire Remote Sensing Advisory Committee (TFRSAC), chaired by partners from the US Forest Service meet twice annually to discuss and highlight critical wildfire observational technology- and information-gaps. The TFRSAC group engages the NASA / academia / industry members to design new solution sets to fill those gaps within that disaster management community. The partners engage in technology development, enhancement, maturation, demonstration, and technology transfer to that wildfire community to ensure that the capabilities meet the requirements of the fire community. The TFRSAC members become technology enablers, allowing rapid operational integration, meeting the specific requirements of wildfire managers and wildfire technologists. This partnership group has been highly successful in maturing, demonstrating and integrating NASA-derived capabilities in UAS utility, sensor system design, telecommunications systems improvements, image-processing algorithm development, intelligent systems design, inter-sensor systems coordination (sensor-web) and data visualization capabilities.

The objectives of the WRAP and TFRSAC-led efforts were to:

- Demonstrate the efficiency of long-duration observational capabilities of a UAS for disaster management support;
- Develop and demonstrate new sensor design concepts for multi-mission operations on UAS platforms. This includes maturing system architecture to allow long-duration autonomous operations (+24 hours), high altitude operations, and large data collection and storage capabilities;
- Develop and demonstrate new sensor capabilities that utilize increased spectral domains to improve autonomous fire-characterization.
- Demonstrate over-the-horizon data telemetry capabilities that allow efficiency in provision of critical, near-real-time sensor information from a remote UAS platform;
- Provide sensor-derived, GIS-compatible, geo- and terrain-rectified, Level II processed data on wildfire conditions to incident management teams within 15-minutes of acquisition.

2. Western States Fire Mission Configuration Overview

The Western States Fire Mission (WSFM) demonstrations, a major component of the WRAP project, is a multi-agency collaborative effort to explore, develop and evaluate emerging technologies for possible adaptation by fire and other disaster response agencies. This configuration was not developed with a focus on any particular business or cost model, but was driven by scientific and technical needs assessments. The WSFM approach was

requirements-driven as defined by the TFRSAC. The primary focus was to streamline the process and increase the quality and efficiency of wildfire characterization to incident personnel. The project leveraged NASA emerging technologies in UAS, sensor systems, communications, autonomous intelligent systems operations, and sensor-web expertise.

To meet the goals, an operational concept was developed to test and evaluate the following capabilities:

- Broad-area UAS coverage with long-duration day / night capability, and near-real-time broad-band communications telemetry capabilities;
- A calibrated multi-spectral visible through thermal-infrared sensor with onboard processing for near-real-time geometric image correction, geo-location, image analysis, and communications management;
- A Wildfire - Collaborative Decision Environment (W-CDE) for data visualization, mission planning, and situational awareness;
- A real-time, GIS data-base of selected derived wildfire sensor products (GeoTIFF, multi-band imagery, on-board derived hot spot detection products);
- Training and outreach to fire management personnel and data analysts.

The WSM architecture concept is shown in Figure 1.

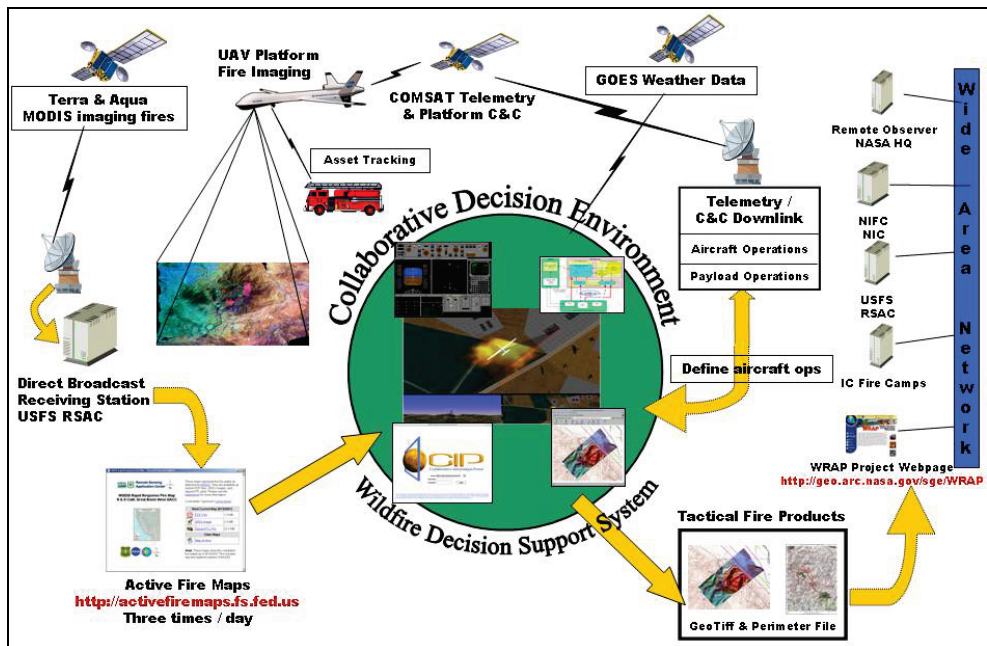


Fig 1. The Western States Fire Mission operational concept highlights the communications architecture for UAS sensor imaging over wildfire disaster events. This configuration integrates additional information sources into a Wildfire - Collaborative Decision Environment (W-CDE), and allows sharing of data elements with a disparate disaster management community.

This configuration allows the WSFM team to observe critical fire components, develop a process flow for autonomous data intelligence management, develop a simplified common operating picture, plan missions within the framework of the COP (the W-CDE), monitor flight operations on UAS aircraft, and provide near real-time fire intelligence to the Incident Command system. The following sections will explore these technologies in greater detail.

2.1 UAS

Our fire mission requirements and expectations for UAS performance included broad-area access to the NAS, with long-duration day / night capability, and near-real-time broad-band satellite telemetry capabilities for remote command / control of the sensor payload and sensor data telemetry. The new NASA *Ikhana* UAS had all these key attributes and made it a capable platform to demonstrate support of disaster incident monitoring.

The NASA *Ikhana* UAV is a modified General Atomics - Aeronautical Systems, Inc. Predator-B (MQ-9) Unmanned Aerial Vehicle (UAV), designed specifically for supporting NASA science missions. "*Ikhana*" is a Native American Choctaw word meaning intelligence, conscious or aware. The name is descriptive of the research goals NASA has established for the aircraft and its related systems. The *Ikhana* is capable of ~24-hour duration, 150-200 knots airspeed, ~13720 meters (45,000 feet) altitude, and flight legs of over 7408 kilometers (4000 nautical miles) (Figure 2). The platform is remotely controlled by a pilot on the ground seated at a console located in the Ground Control Station (GCS). The sensor system operator, seated at a console located in the GCS can monitor and control the AMS-Wildfire sensor payload carried aloft by the *Ikhana*. The *Ikhana* home base is NASA-Dryden Flight Research Center (DFRC) at Edwards Air Force Base (EAFB), California. The *Ikhana* was first put into service for NASA in January 2007, and flew its first science missions in support of wildfire observations in August 2007. The *Ikhana* is ideally suited to support long-endurance / duration missions, where critical observation-time over a dynamic event is required. Special coordination with the Federal Aviation Administration (FAA) was required to safely operate the *Ikhana* UAV in the National Airspace (NAS).



Fig. 2. The NASA *Ikhana* UAS with the sensor pod mounted under the left wing.

2.2 Command / Control / Data Communications Telemetry

The aircraft flight controls, payload system controls, and the payload sensor data are operated through a communications linkage with the GCS. There are two kinds of ground communications to the aircraft: LOS and satellite OTH systems. A portable ground data terminal provides two-way control and sensor communication when the aircraft is within radio line-of-sight (approximately 130 kilometers (70 nautical miles)). A geo-synchronous communications satellite, operating at Ku-band frequencies, provides OTH uplink and downlink between the UAS and GCS. The Ku-band system has bandwidth capacity of 3.0 megabits-per-second (Mbs), where 1.0 Mbs is used for data transmission, and 2.0 Mbs is used for video data transmission, a small bandwidth is required for platform command and control. This telemetry link allows data from the onboard imaging sensor (described in the next section) to be sent from the UAS to the GCS and then redistributed through the Internet to the community.

2.3 Autonomous Modular Scanner – Wildfire (AMS-Wildfire) Sensor

The AMS-Wildfire sensor is an airborne multi-spectral imaging line scanner capable of high-altitude autonomous operations on both manned and unmanned aircraft (Figure 3). The sensor is a highly modified Daedalus AADS-1268 scanning system that can support resolutions of 1.25 milliradian and 2.5 milliradian, with an angular field of view of 43° or 86° respectively. Spatial resolution is determined by altitude and the primary aperture size (1.25 mrad or 2.5 mrad). Operating from an example altitude of 7011 km (23,000 feet) mean sea level (MSL); ~20,000 feet Above Ground Level (AGL)), with a 2.5 mrad setting, the pixel spatial resolution would be 15 meters (50 feet). The system is configured with twelve spectral channels ranging from the visible through short-wave-, mid-, and thermal-infrared (VIS-IR-TIR) (Table 1). The thermal-infrared (TIR) channels have been calibrated for accurate (~0.5° C) temperatures discrimination of hot targets, up to ~1000°C.

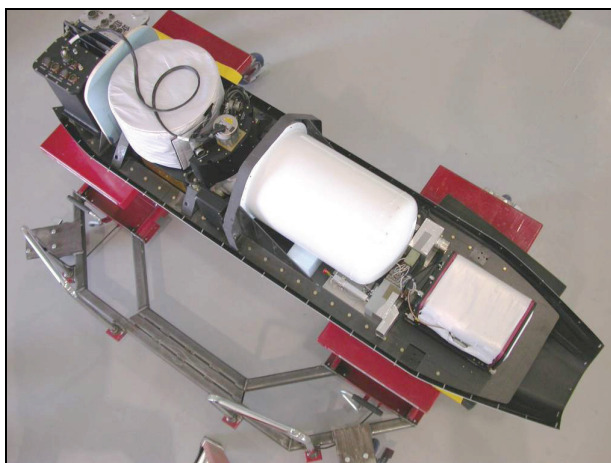


Fig. 3. The NASA AMS-Wildfire instrument arranged in the sensor payload pod tray. The scan-head is located under the cylindrical white thermal blanket at the top-left of the pod tray, while the white pressure vessel in the center contains the digitizer and electronics. Other components include the power supply and various controlling electronics.

Spectral Band	Wavelength, μm
1	0.42- 0.45
2	0.45- 0.52 (TM1)
3	0.52- 0.60 (TM2)
4	0.60- 0.62
5	0.63- 0.69 (TM3)
6	0.69- 0.75
7	0.76- 0.90 (TM4)
8	0.91- 1.05
9	1.55- 1.75 (TM5)
10	2.08- 2.35 (TM7)
11	3.60- 3.79 (VIIRS M12)
12	10.26-11.26 (VIIRS M15)

Total Field of View: 42.5 or 85.9 degrees (selectable)
IFOV: 1.25 mrad or 2.5mrad (selectable)
Spatial Resolution: 3 – 50 meters (variable based on alt)

Table 1. AMS-Wildfire 12-channel scanner specifications.

Major hardware and software modifications to the AMS-Wildfire instrument allow autonomous or remote operations of the sensor aboard a UAS platform during extended mission profiles.

As the AMS line scanner collects a series of scanlines over a wildfire event, the raw spectral data are sent to a computer processor on-board the platform to further process into useful information data sets for delivery to a telemetry system and distribution to receiving nodes on the ground. The on-board autonomous data processing is described in the following section.

2.3.1 On-Board Data Processing – The on-board data processing system was designed to autonomously complete the acquisition, pre-processing, information extraction, and output product generation from the raw spectral data collected by the AMS-Wildfire sensor system. The on-board processing includes fire detection “hot spot” algorithm processing, image generation, and geo-rectification of all data sets. Each of these processes is automated, requiring an initiation by the sensor engineer operating from the GCS. AMS-Wildfire sensor data are first autonomously converted to temperature/radiance data, and the thermal channels are further converted to a “brightness temperature measurement. An appropriate fire detection algorithm is applied to those derived image data sets. The resultant data sets are then autonomously processed to create geo-rectified visual raster products and hot-spot detection vector files. The vector and raster products are transmitted via the *Ikhana* Ku-band SatCom telemetry system to the ground.

2.3.1.1 Fire Hot-Spot Detection Algorithm - For fire hot-spot detection, a multi-channel temperature threshold algorithm, based on that developed by the Canadian Center for Remote Sensing (CCRS), was implemented (Li, et al., 2000a, Li, et al., 2000b, Flasse and

Ceccato, 1996, and Cahoon, et al. 1992). The CCRS algorithm was originally developed for use with satellite (AVHRR) imagery (Li, et al., 2000b), but has been adapted for use on various airborne sensor systems, including the AMS-Wildfire sensor.

The fire hot-spot detection algorithm uses the 3.6 μ m channel of the AMS-Wildfire sensor to define a fire temperature threshold, and two or more additional channels to further refine this classification. Multi-channel thresholding improves commission errors encountered when using a single mid-wave thermal infrared channel-derived temperature value alone. The threshold values used in the algorithm (AMS channels 11 and 12 and, for daytime missions, channel 7; see Table 1) are parameters which can be variably set by the operator during a mission. The fire hot-spot detection algorithm is calculated as:

If:

*Band 11 (3.60- 3.79 μ m) > Band 11 minimum temperature (e.g. 380° K) and
Band 12 (10.26-11.26 μ m) > Band 12 minimum temperature (e.g. 240° K) and
Band 11 – Band 12 > Difference minimum (e.g. 14° K),*

And (if available),

Band 7(0.76- 0.90 μ m) < Reflectance maximum (e.g. 0 .3) (to screen high-reflectance commission errors),

Then,

Pixel is classified as a fire hot-spot

The hot-spot algorithm-defined vector data set is provided as an additional data product transmitted over the telemetry link.

2.3.1.2 Geo-Rectification – Image data sets from the AMS-Wildfire sensor are autonomously geo-rectified on-board the *Ikhana* on a processor. The fully automated geo-rectification processing utilizes meta-data from an Applinix Position and Orientation System for Airborne Vehicles (POS AV) model 310 system. The POS AV-310 integrates precision Global Positioning Satellite (GPS) data with inertial technology to provide near-real-time and post-processed measurements of the position, roll, pitch and heading of airborne sensors. Photogrammetric projective transformation equations are used to determine the position of each pixel in the scanline as projected to the ground, with “ground” being determined by the on-board digital elevation model (DEM) data for the area being over-flown. The onboard DEM consists of a composite data set of one-arc-second Shuttle Radar Topographic Mission (SRTM) 30-meter spatial resolution elevation “tiles” which are mosaiced real-time as needed, creating a seamless DEM for the entire western United States (USGS SRTM website, 2008). The SRTM DEM data are used to define the geospatial context (latitude / longitude, elevation) reference for geo-rectification of the sensor line-scanner data. Each of the AMS-Wildfire data pixels are geo-referenced based upon the relationship between the location of the sensor / platform (which defines the pointing vector of the line-scanner pixel at acquisition time) and the latitude, longitude, and elevation of the terrain (from the SRTM data).

The on-board product generation, algorithm processes, and geo-rectification processes takes approximately 30-seconds (0.5 minutes) per image-file frame (1200 lines of AMS-Wildfire spectral data). With the additional data transmission time (via satellite telemetry) and

ground-based quality control assessment, the total process time (to final delivery to a server for Internet distribution) is within fifteen minutes defined as a metric for near-real-time data delivery.

The geo-rectified data sets and imagery are sent from the on-board link module image-processing computer to the telemetry system. The GeoTIFF files are moderate file sizes (-1 – 3 Mb per frame), allowing for minimal transmittance time through the telemetry link to the GCS, where they are then sent to servers at NASA for redistribution through the Internet.

2.4 Ground Services

The geospatial processing services for the serving of the near-real-time AMS-Wildfire derived data products were implemented utilizing open standards, promulgated primarily by the Open Geospatial Consortium (OGC). The pointers to the image data were of five types:

- A pointer to the raw spectral data availability via anonymous file transfer protocol (FTP);
- A pointer to the data via an OGC-compliant Web Map Server (WMS), used by Geographic Information System (GIS) clients (such as ESRI ARC users);
- A pointer to the data via an OGC-compliant Web Coverage Service (WCS), used primarily by other processing services, including fire and smoke modeling teams;
- A pointer to a GoogleEarth Keyhole Markup Language (*kml*) file; and,
- A pointer to a thumbnail-sized version of the file for quick-look viewing of the data.

The AMS-Wildfire data can therefore be accessed and ingested into a desktop GIS, WMS- or WCS-accessed system, or be visualized using any standard web browser, or Google Earth.

2.5 Wildfire - Collaborative Decision Environment (W-CDE)

A simplified, fire data integration and visualization solution tool was developed using NASA and Commercial-Off-The-Shelf tools. The Wildfire Collaborative Decision Environment (W-CDE) was developed originally to support data and sensor sharing for the NASA's Mars Exploration Rover program, and was modified to allow use as a data- and information-sharing tool for wildfire disaster managers. The W-CDE allows the integration of numerous web-enabled data sources to be collaboratively viewed and implemented to aid in determining appropriate fire management strategies. Simplifying the fire data visualization capabilities, NASA expanded the capabilities of the GoogleEarth® free-ware package as a "front-end" to allow the integration of multiple, pertinent fire-related data elements into a single package. These elements included the integration of real-time satellite weather information, predicted and actual cloud cover, predicted winds, satellite-derived fire "hot-spot" detections, Remote Automated Weather stations (RAWS) throughout the western US, National Weather Service Fire Critical Weather information, *Ikhana* aircraft tracking positional information, sensor information and real-time imagery feeds, Federal Aviation Administration (FAA) flight restrictions data, maps and information, airspace information, National Interagency Fire Center's Large Fire Location data, wildfire management team's Infrared mission support requests, and real-time lightning detection

data (Figure 4). Incident command teams were provided access to the W-CDE through a network link to the data “mash-up” service.

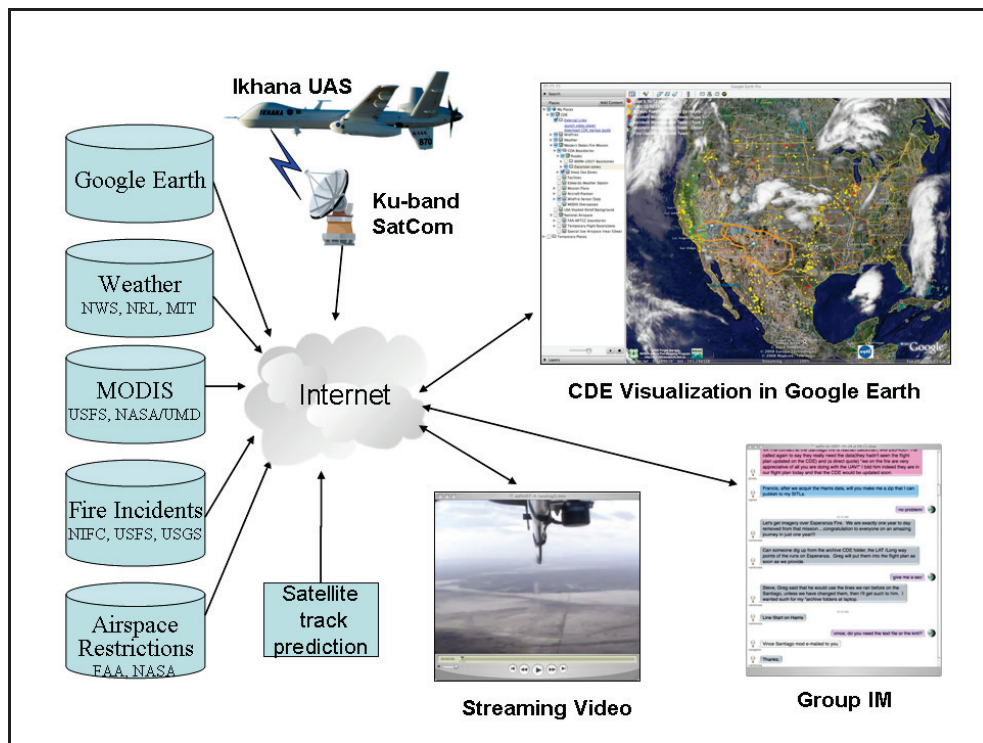


Fig. 4. Components of the Collaborative Decision Environment. The visualization element of the W-CDE employs GoogleEarth. The critical fire data elements (left side) that compose the additional visualization components are a “mash-up” of data from various web served data locations. The W-CDE also allows integration of instant messaging (IM) and provision of streaming video data from the acquiring UAS platform, in addition to the 3-D visualization of the AMS-Wildfire sensor-acquired data.

3. WFSM Missions

The WFSM focus is getting the right information, to the right people, at the right time. Missions planning was done in partnership with the NIFC National Incident Coordination Center (NICC) and the California state fire agency (CalFire) to ensure useful fire data products were generated on-board and transmitted to ground servers for distribution to web-supported Incident command teams in minutes. Highlights of WFSM UAS mission execution steps and results are summarized in the following sections.

3.1 Mission Planning

Mission planning requires knowledge of where the image targets (fires, incident infrared data requests, fire science targets, etc.) are. Additionally, operational constraints (aircraft

performance, FAA- and NASA-imposed keep-out (no fly) zones, FAA Certificate of Authorization (COA) requirements and weather constraints, etc...) impact mission planning. To efficiently perform the WSFM series, it was imperative to assemble and real-time-update the image target information and mission constraint criteria into the W-CDE (Figure 5).

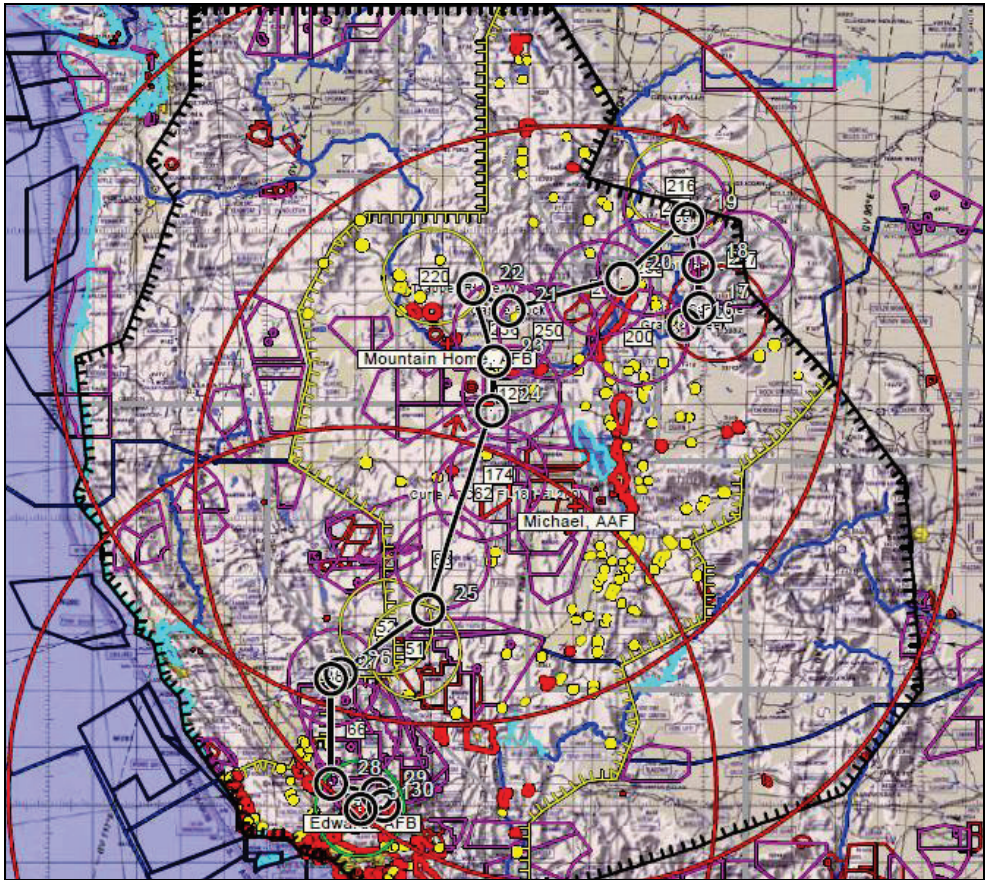


Fig. 5. Graphic of *Ikhana* UAS flight planning for operations over emerging wildfire events. Planned flight route displayed as black line; turn points of UAS defined as small black circles along flight plan route. Large red circles represent 400 nm distance range of potential secondary emergency landing locations (three identified for western U.S. Purple smaller circles represent tertiary emergency landing locations (50-mile radius along flight track). Yellow and red polygon areas on map represent population density centers where certain UAS operations are restricted or not allowed as per the FAA regulations.

The WSFM planning efforts begin two to three days before a mission, and results in two mission plans being developed. A preliminary mission plan filed with the FAA 48-72 hour prior to take off, highlighted the route of flight and denoted imaging targets as various loiter

areas (circular operational areas, typically 15-nm radius). This preliminary plan was used by the FAA to alert FAA flight sector controllers to the NASA *Ikhana* UAS activity in their assigned areas of responsibility. For this preliminary plan, targets are selected from the W-CDE, with coordinated input from the National Interagency Fire Center (NIFC), or other responding agencies such as Cal Fire. Additional requests for imaging (science targets, satellite calibration/validation coincident collections) are also considered. Operational radiuses are assigned to each target, considering extent of imaging, and required maneuvering room. A route of flight is then generated to efficiently transit to each target, while avoiding keep out zones. Take off times are planned to optimize imaging times, and are often modified to address operational constraints (Figure 6).

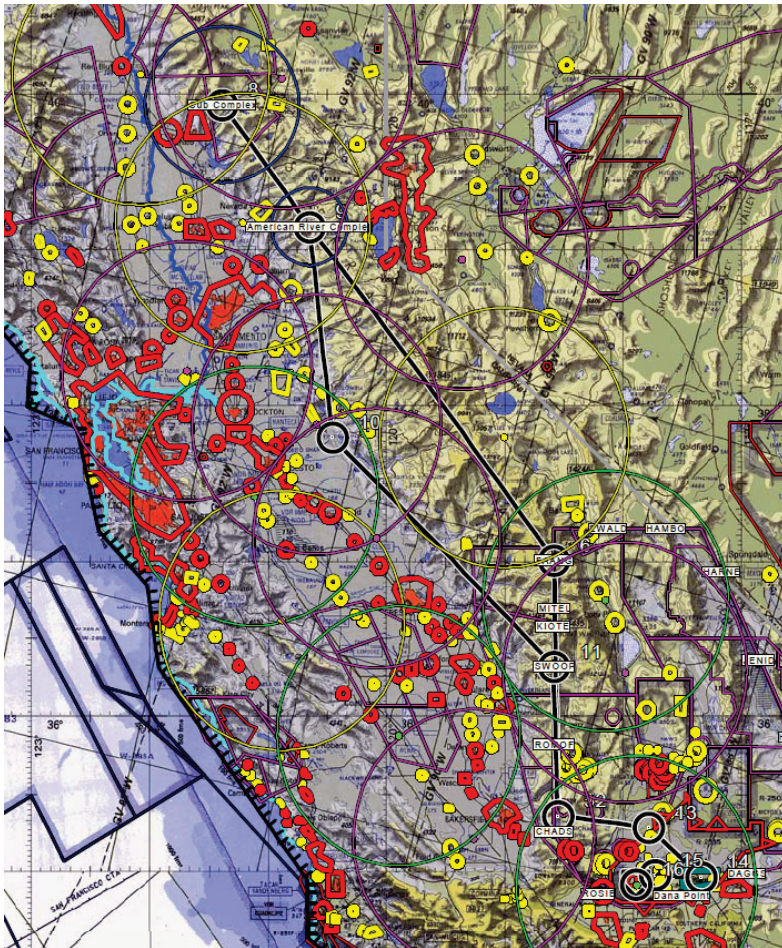


Fig. 6. Flight Plan for FAA defining *Ikhana* fire imaging mission route (black line), UAS turn / linger points (small black circles), tertiary landing locations along flight route (purple circles), restricted-flight / no flight zones (yellow / red polygons), and major lingering imaging locations (fire areas) (blue circles). Data displayed on a digital air navigation chart.

Hours before an *Ikhana* fire imaging mission is initiated, a detailed mission plan, showing the route of the flight (avoiding the FAA / NASA Flight Safety-imposed restricted / keep out areas) and a detailed imaging plan for operations around the imaging targets was completed. The pilots use this detailed plan during the mission to methodically image the target areas. Detailed mission plans build on the preliminary plan, by adding imaging waypoints (start/stop points for imaging runs) in each of the fire imaging areas. For the WSFM, the entire mission planning was done in the W-CDE. Waypoints (latitude/longitude) for all the targets and turning points, are generated, saved as a *kml* file, and compiled into both a text and a Microsoft EXCEL® file highlighting all waypoints, with distance between waypoints and duration of mission through the waypoint series. A completed Mission Plan includes waypoint files and a graphic of the mission plan for visual verification (Figure 7).

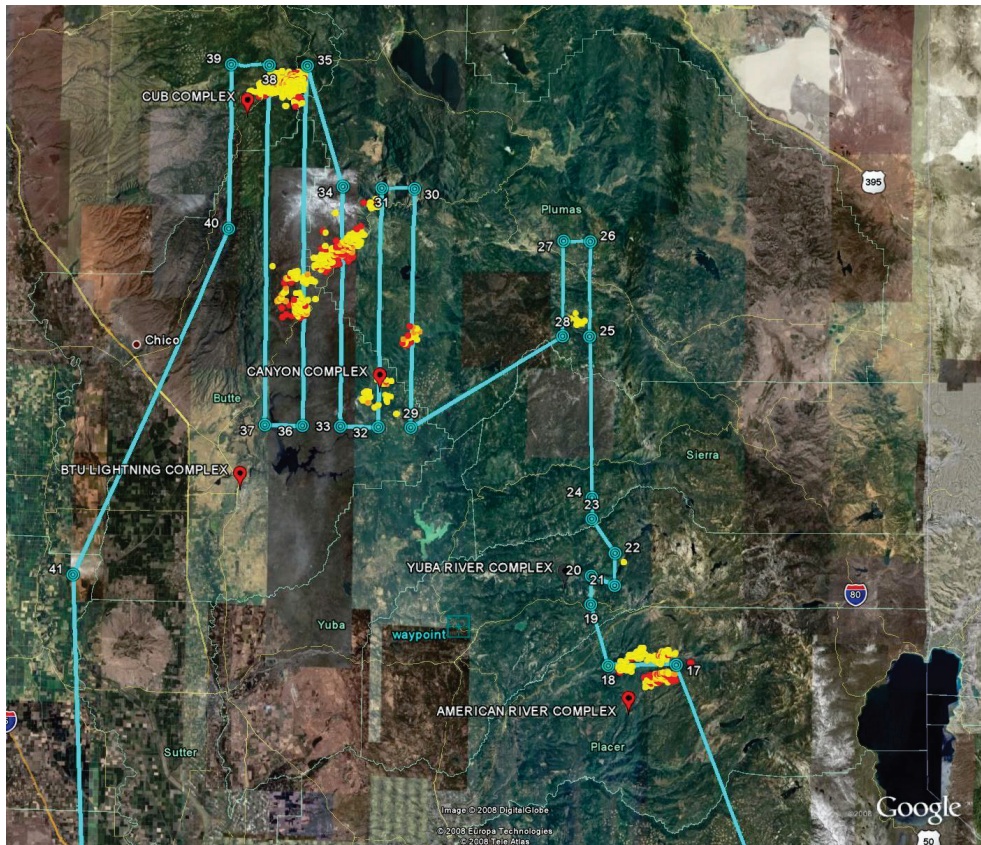


Fig 7. Detailed mission plan indicating sensor data collection flight line locations for an *Ikhana* UAS fire mission. The teal lines represent the flight line plan for regular-spaced flight segments over a complex of fires in Northern California, 8 July 2008. The yellow, orange and red dots represent near-real-time MODIS satellite hot-spot detects, which allow the *Ikhana* sensor team to design flight profiles over the most currently active fire areas for detailed imaging.

During mission operations, fire management personnel and the mission management team can make requests to adjust flight parameters to allow for shifting fire locations or target modifications. The modifications can include both flight parameters and sensor configurations (band combinations, changing algorithms, etc). The mission manager notifies the *Ikhana* pilot of “mid-mission” modifications to the flight parameters, who would request such flight modification from the FAA controller via radio. In almost all cases, the requested flight modifications were allowed. Mid-mission sensor configuration modifications are made through the sensor operator, monitoring the system operations.

3.2 Fire Data Products

The AMS-Wildfire sensor data products delivered from the UAS through the satellite telemetry link to the ground include a geo-rectified 3-band color visual product and a vector file of hot-spot-detect “fire” polygons. Imaging smaller fires may only require the collection of a single “segment” of image data (defined earlier as a section of ~1200 lines of scanner data). An investigator can choose any three channels to form an image data set composite as can be seen in Figure 8.

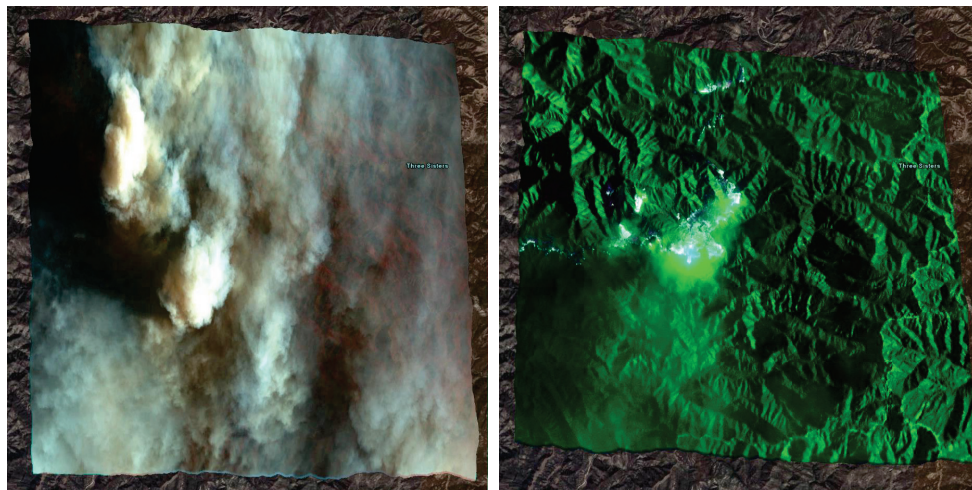


Fig. 8. Image on left represents a single segment image of a geo-rectified three-band composite of AMS-Wildfire visible-band spectral data collected over the Zaca Fire, southern California, 16 August 2007. Fire is obviously not visible in this color composition, but the attenuating smoke clearly is visible and obscures the terrain and fire location. The right-side image is the same region (Zaca Fire), but displays a 3-band color composition that includes both reflected infrared and thermal infrared data. The intense fire locations can be easily seen, even through the dense smoke plume.

For most fire events, a multi-segment / multi-flight line fire data collection mission is flown. Since the imagery and vector files are geo-rectified, the mosaics automatically orient and display correctly into any GIS mapping tool or in the W-CDE (GoogleEarth). The fire management personnel can display the images and overlay the hot-spot detection vector

files on the same data sets. This can provide an indication of both the burned and unburned areas surrounding the hot-spot active fire areas (Figure 9).

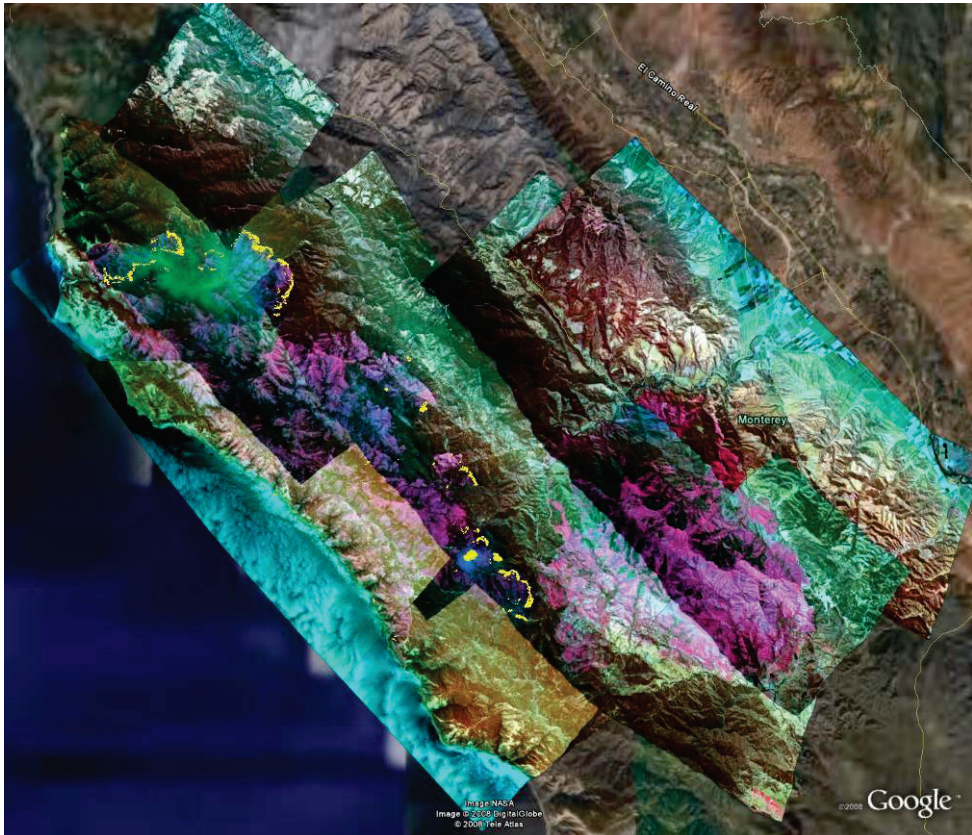


Fig. 9. AMS-Wildfire sensor near-real-time processed image “segment” mosaic for the Basin Fire Complex, Big Sur, California, collected on 8 July 2008. Data was processed and mosaiced in “real-time” from 5 flight lines of 22 segments of AMS data. The fire hot-spot detection algorithm shape-file information, shown draped in yellow, on the 3-channel composite mosaic (which is draped on the GoogleEarth W-CDE).

W-CDE users can “turn-off” the 3-band image files and display just the vector file hot-spot detections, in order to reduce image clutter and focus on identifying critical small hot-spot locations (Figure 10). By having simple functionality built into the data visualization, the wildfire management teams can make effective use of the UAS sensor-derived data. Additionally, the wildfire managers can make sensor data requests for “alternative” band and data sets to be collected and rendered in near-real-time, allowing true sensor-web-enabled functionality.

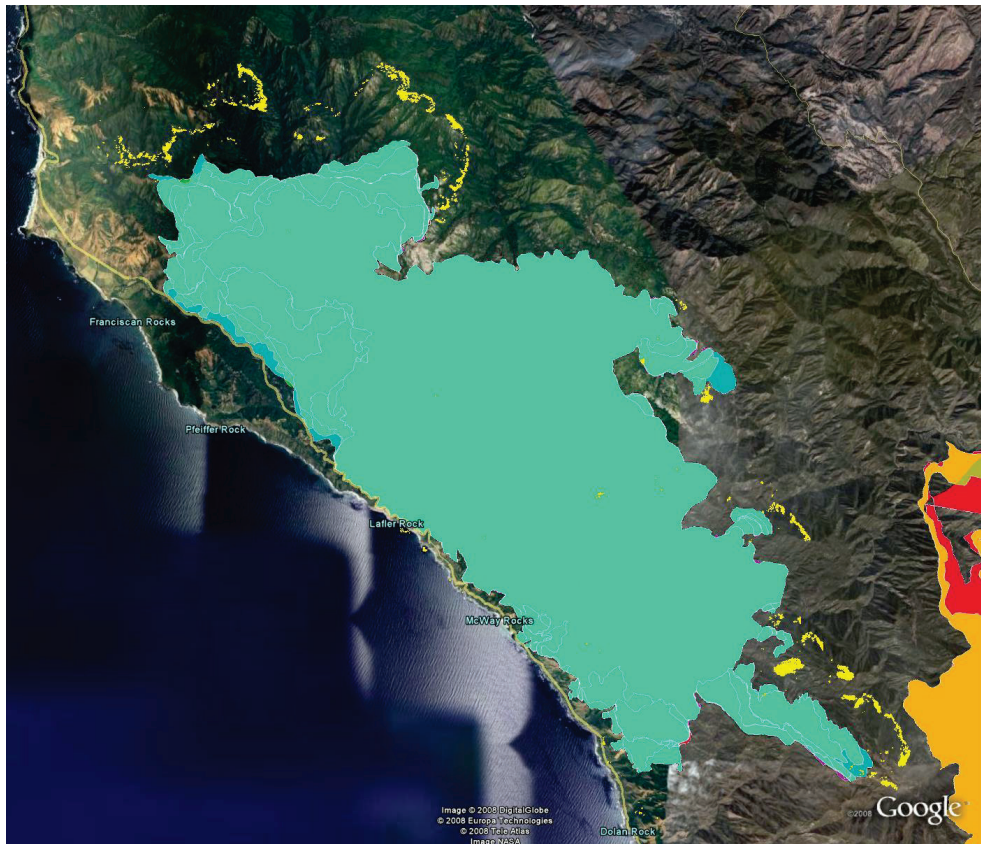


Fig. 10. AMS-Wildfire sensor data collected 8 July 2008 over Basin Fire Complex, California. The hot-spot vector file data (yellow polygons) are shown overlain with the recent fire perimeter polygon (teal). All layers draped in GoogleEarth for visualization. The yellow hot-spot fire fronts can be seen extending outside the fire perimeter.

The AMS-Wildfire sensor data can also be visualized in 3-D within the W-CDE. Since all data are geo-rectified, the mosaic imagery and any additional data layers can be draped on the terrain and various visualization perspectives can be rendered. The 3-D visualization capability can be critical for determine rates of fire spread with various terrain slope or aspect conditions (Figure 11).

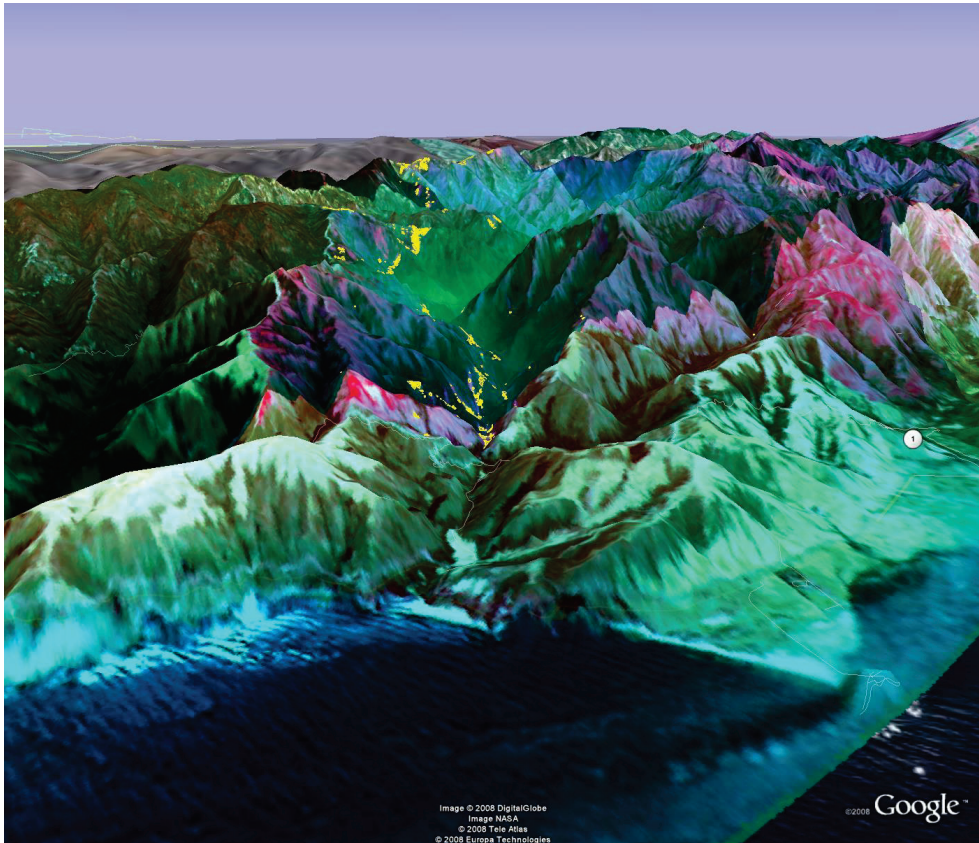


Fig. 11. Three-dimensional view of AMS-Wildfire sensor imagery collected on 8 July 2008 over Basin Fire Complex, California. This eastward looking view shows the hot-spot fire detects (yellow polygons) overlain on the three-band color composite, which is draped on the GoogleEarth background data base. Note the locations of the fire on the south-facing (sun-intense) slopes at this data collection time.

All of these various W-CDE-enabled capabilities were available to fire management teams during the 2006, 2007 and 2008 western U.S. wildfire season. Additionally, project team members were embedded at fire camps and multi-agency and multi-fire coordination centers to provide W-CDE assistance. In 2008, numerous fire personnel were familiar with the UAS, sensor, and W-CDE capabilities to work seamlessly on data utility for wildfire event management. The additional GIS-enabled web-served data sets (GeoTIFF, etc) were used extensively by the various GIS teams to further map and update fire perimeter information on numerous fires in the U.S. using various image processing and GIS software systems. The 2006-2008 mission series flights, data collections and data utility are highlighted in the following section.

3.3 Western States Fire Mission Flight Summary

The Western States Fire Missions (WSFM) in 2006, 2007, and 2008 demonstrated the integration of the technologies detailed in the previous section. The missions, flown over wildfire events in the western United States served as a test-bed and model for improving disaster data delivery to disaster incident management teams. The 2006-2008 Western States Fire Missions are briefly summarized, to provide context for employment of the integrated tools during operational missions.

3.3.1 2006 Mission Series - The Western States Fire Mission Series began on 24 October 2006, when the AMS-Wildfire instrument on the *Altair* UAS (predecessor to *Ikhana*) collected and delivered near-real-time data over prescribed burns on the eastern flanks of the Sierra Nevada Mountain Range in California, USA. The mission demonstrated long duration UAS and sensor operations and was the first National Airspace System (NAS) operations for the UAS. On 28 October 2006, the Altair supported data collection / delivery over the Esperanza Fire in southern California, USA, providing near-real-time information on fire location and progression to the Incident Command Team. During the 16-hour mission, the AMS-Wildfire scanner system provided multiple image data sets of the fire progression. Data were delivered as GeoTIFF files and served in the fire management camp on the W-CDE through GoogleEarth as well. An integration team was embedded at the fire camp to assist in data management and training. In 2006, approximately 40 hours of UAS / sensor operations occurred over fires.

3.3.2 2007 Mission Series - The 2007 mission series were the first flights of the new NASA *Ikhana* UAS, which was delivered in January 2007. The 2007 Western States Fire Mission series began in August following the FAA Certificate of Operation for the *Ikhana*. A total of eight fire data collection missions occurred during the fire season in 2007. The first four missions demonstrated long-duration / range capabilities, with data collection over fires located in eight western states. Mission operations were between 10-22 hours with 2593-5926 kilometers (1400-3200 nautical miles) mission ranges. During those first four flights, a total of 27 fires were flown and imaged with near-real-time geospatial fire data relayed to Incident Command Centers (ICC). To assist in information integration, WRAP team members were embedded at various ICCs.

3.3.2.1 Southern California Firestorm Support Missions – October 2007 - In late October 2007, over eleven major Santa Ana wind-driven fires erupted in the Los Angeles and San Diego areas of Southern California, USA. The NASA *Ikhana* / AMS sensor flew on 24, 25, 26 and 28 October 2007, and provided near-real-time imagery of those eleven complexes to the fire management teams. The FAA facilitated operations through the issuance of an emergency COA for operations in the densely populated area of the fires. Flight endurance each day was between 7-9 hours with ~ 2500 kilometers (1350 nm) mission ranges. Many of the fires were imaged twice a day to provide up-to-date fire progression information. Team members were again embedded in various Incident Command Centers (ICC) and county-level Emergency Operations Centers (EOC). A summary of the 2007 missions is shown in Table 2.

Flight Date	Duration	Fires Flown	Mission Mileage
16 Aug	10 hrs	4	2993 km (1400 nm)
29 Aug	16.1	7	4630 km 2500 nm)
7 Sept	20	12	5926 km (3200 nm)
27 Sept	9.9	4	3333 km (1800 nm)
24 Oct	9	11	2500 km (1350 nm)
25 Oct	8.7	11	2500 km (1350 nm)
26 Oct	7.8	11	2500 km (1350 nm)
28 Oct	7.1	11	2500 km (1350 nm)

Table 2. 2007 Western States Fire Mission Summary.

3.3.3 2008 Mission Series - In late June 2008, lightning ignited hundreds of fires in northern California. When the California Governor (Swartzenegger) declared a State of Emergency, the NASA *Ikhana* and sensor were deployed to support data collection and near-real-time delivery to the embattled fire management teams. The FAA provided an emergency COA-region extension to allow the *Ikhana* unfettered access to the NAS above those fire complexes. Four missions were flown during the mid / late summer in 2008, delivering near-real-time data over 16 wildfire events (Table 3). The missions focused on providing near-real-time fire information to the various ICC and well as to the State Operations Center (SOC), and the Multi-Agency Coordination Center (MACC), where data were integrated into the wildfire management decision process.

Flight Date	Duration	Fires Flown	Mission Mileage
8 July	9.5 hrs	9	2593 km (1400 nm)
19 July	5.0	4	1852 km (1000 nm)
17 Sept	3	1	1482 km (800 nm)
19 Sept	3.5	2	1482 km (800 nm)

Table 3. 2008 Western States Fire Mission Summary.

4. Operational and Integration Challenges

Many of the procedural, operational, and technical challenges were overcome during the Western States Fire Mission series from 2006-2008. Still, it is imperative to highlight those issues, so that substantive efforts towards further improving and enabling the use of UAS and near-real-time sensor collection on emergency-support missions can be realized. Some of the significant operational issues are detailed in the following sub-sections.

4.1 COA Limitations

Over the three fire mission flight years, the FAA issued COAs for the *Ikhana* that ranged in complexity and operational area allowances. In 2008, the *Ikhana* COA allowable mission area was significantly reduced by the FAA. The allowable mission area was limited to flight operations within 50 nm of Restricted Airspace (RA) or a Military Operations Area (MOA). The 50 nm RA / MOA flight operations restriction limited access to airspace and precluded critical data collection over some wildfire events in both 2007 and 2008. The FAA did not provide an explanation for the change in the *Ikhana* COA status. Further refinement to the

COA process and COA allowance are needed. Regulatory guidelines must be established to allow emergency support mission operations in areas of critical need.

4.2 COA Restrictions for UAS to Remain Clear of GPS Testing / Jamming

The WSFM encountered some mission delays and rescheduling due to GPS testing / jamming exercises at military bases in the vicinity of the WSFM routes. These testing / jamming areas were defined as consisting of an inverted cone centered at the GPS test site, with increasing radius with increasing altitude. When flying at 25,000 ft, *Ikhana* could be affected at a range of up to 300 nm from the testing / jamming origin. Since the *Ikhana* is restricted to a specified flight altitude, no deviation could be made to allow the UAS to avoid those regions during potential missions. When testing / jamming occurred (or was even planned), the *Ikhana* was grounded from operations. This had a detrimental effect on supporting some national emergency mission requests over the northern California wildfires in 2008.

4.3 Access to LOS Communications Frequencies

For flights within approximately 70 nm of *Ikhana's* base of operations (Edwards Air Force Base, California), the UAS is controlled via a direct line-of-sight radio link. Significant DOD UAS operations in the same general area required prioritization of the limited number of attainable frequencies. The *Ikhana* was restricted therefore to secondary priority status which postponed or cancelled mission operations. This became a critical issue when emergency data collection flights were requested to support wildfire teams battling conflagrations. This issue is solvable by negotiating a sharing of frequencies and reprioritization of frequency allocation given national or state-level emergency requests for mission support.

4.4 Unexpected Weather Along Flight Route

The *Ikhana* COA restricted flight from areas of adverse turbulence, convection and icing. During the 72-hour advanced flight planning process, it was difficult to predict weather occurrences or timing along a planned mission route with any certainty. Since the FAA required flight planning route information 72-hours prior to UAS launch, weather forecasts for the day of flight were not meaningful. During one mission in 2007, a significant flight deviation was allowed by FAA flight traffic controllers to avoid several rapidly developing convective cells. This was a major breakthrough for the *Ikhana* team, as we would have had to abort the mission and return to base if not granted the near-real-time deviation. Additionally, flight restrictions and mission aborts occurred due to the "potential" for clouds to be at operations altitude during missions or in the vicinity of the Base of Operations (Edwards AF Base) during planned take-off or landings. If cloud cover was predicted, the mission was cancelled. These weather-related mission issues require further refinement to allow a go- / no-go decision to be made much closer to mission take-off than 72-prior to operations.

4.5 Staffing Requirements

Long-duration flights (>10 hours) for the WSFM required multiple crewmembers for all operational positions due to crew duty day limitations (generally 8-hour duty limits). When missions longer than 12-hours were conducted, multiple shifts were implemented and crew-

duty hours were strictly adhered to. In a few instances, General Atomics Aeronautical Systems Inc. provided supplemental engineering staff and pilots when needed on long-duration missions. Non-standard flight schedules, intermittent sleep schedules, and extended on-call status have the potential to fatigue crew members. These staffing requirements will continue to be a significant issue for long-duration, non-scheduled flights that involve supporting and “chasing” dynamic disaster events, such as wildfires. Therefore, when planning such mission concepts, one must be aware of the additional staffing requirements needed to sustain safe operations.

4.6 Air Traffic Control (ATC) Coordination

The WSFM mission team coordinated airspace access for *Ikhana* flights with the FAA Unmanned Aircraft Program Office (UAPO) and Air Route Traffic Control Centers (ARTCC). Clearly, both NASA and FAA worked together as partners to facilitate the success of such demanding mission profiles and objectives. ARTCC personnel were open-minded and receptive to the prospect of *Ikhana*'s flights through their respective airspace. They communicated concerns and suggested resolutions. Fostering and promoting communication between UAS mission operations personnel and the FAA will continue to be required and a key component to successful operations in the future for any entity (Hall, et al., 2008).

5. Summary

We have demonstrated that various platform, sensor, communications, and geospatial technologies can be integrated to provide near-real-time intelligence to support of disaster management entities. In our work with the U.S. wildfire management agencies, we have developed and demonstrated technologies for providing near-real-time emergency geospatial data delivery, a significant advance over current capabilities.

Large-capacity, long-duration, medium-altitude UAS can play a significant role in providing repetitive, lingering operations over disaster events, especially dynamic, evolving events like wildfires. The OTH satellite data telemetry systems on these platforms can be employed to control / command an imaging payload as well as to provide sensor data to ground team members. This telemetry capability allows near-real-time information to be in the hands of Incident Management Teams.

Imaging sensor systems can be designed to collect critical spectral and thermal wavelength information, specifically “tuned” to the phenomenon that is being observed. The use of multispectral data for wildfire observations is necessary to characterize fire locations and movement. The spectral channels defined in this chapter are essential for wildfire observations, and multi-channel capabilities offer clear advantages over single-channel fire detection systems, as we have shown in this chapter. Image processing capabilities, to derive Level II information from sensors, can be automated and included as part of the payload processing package on an UAS platform. Complex algorithms can be integrated into the processing scheme to further reduce those labor / time-consuming, analysis tasks. By integrating sensor / platform IMU and positioning information with terrain DEM data, a fully geo-rectified image product can be developed autonomously on-board an aircraft,

further reducing the critical labor and time requirements for delivery of accurate geospatial information.

The employment of web-enabled GIS tools and systems, such as GoogleEarth, provide a user-friendly “platform” for display of geo-rectified imagery and information. Our goals were to ensure that the information products developed autonomously from the UAS sensor would integrate seamlessly into a multitude of geospatial visualization packages. We achieved that objective by providing autonomously-generated data products in Open Geospatial Consortium (OGC) standard formats. The W-CDE was used extensively as were the access to the various WMS data-formatted holdings.

Following three years of system development and emergency support missions in the western United States, we have demonstrated that current off-the-shelf technologies can be integrated to provide the disaster management community with the data and “intelligence” that they require in near-real-time. We anticipate that the civilian use of UAS will increase dramatically, especially in support of disaster management and disaster relief efforts. The processes and technologies described here for the use of UAS platforms and enabling sensors and technologies should form the foundation for designing future disaster monitoring and observation capabilities. These integrated technologies have obvious cross-cutting application to other disaster events in the United States and the world. As the National Academy of Science has reported, “UAVs provide increased range and flight time and the ability to penetrate environments that might be too hazardous for piloted aircraft. However, issues of cost, reliability, software, and proximity to urban areas have limited the use of UAVs to demonstration missions. For now, conventional aircraft remain more reliable and more cost-effective for Earth sensing, and agencies need to ensure an appropriate balance between these two types of platforms” (Henson 2008).

6. Acknowledgements

The authors acknowledge the support of the National Aeronautics and Space Administration (NASA) through a grant (REASoN-0109-0172) awarded to support this work. We are also grateful for the support of S. Buechel (BAERI), D. Sullivan (NASA), B. Lobitz (CSUMB), F. Enomoto (NASA), S. Johan (NASA), S. Schoenung (BAERI), T. Zajkowski (USFS-RSAC), E. Hinkley (USFS-RSAC), S. Ambrose (NASA), T. Fryberger (NASA), T. Rigney (NASA), B. Cobleigh (NASA), G. Buoni (NASA), J. Myers (UCSC), T. Hildum (UCSC), M. Cooper (GA-ASI), and J. Brass (NASA). We also would like to acknowledge the wildfire management community members who engaged us in defining observation criteria and metrics that allowed us to help improve their wildfire / disaster mitigation capabilities.

7. References

Ambrosia, V.G., Hinkley, E., Zajkowski, T., Wegener, S., Sullivan, D. V., Enomoto, F., Schoenung, S. (2009). Lessons Learned: Experiences in UAS Sensor Operations Supporting Disaster Scenarios (Wildfires) in the United States, *Proceedings of 33rd*

- International Symposium on Remote Sensing of Environment*, CD Proceedings, paper reference # 301, pp. 1-4, Stresa, Italy, May 2009, ISRSE.
- Ambrosia, V.G., Hinkley, E, and Ambrose, S. D. (2008). NASA Science Serving Society: Improving Capabilities for Fire Characterization to Effect Reduction in Disaster Losses. In: *Risk Wise*, pp. 158-161, Tudor Rose Publishing, ISBN 0-9536140-9-3, United Kingdom.
- Cahoon, D. R., Jr., Stocks, B. J., Levine, J. S., Cofer, W. R., III, and Chung, C. C. (1992). Evaluation of a Technique for Satellite-Derived Area Estimation of Forest Fires. *Journal of Geophysical Research*, vol. 97, p. 3805-3814.
- Flasse, S. P., and Ceccato P. S. (1996). A Contextual Algorithm for AVHRR Fire Detection. *International Journal of Remote Sensing*, vol. 17, p. 419-424.
- Giglio, L., Descloitres, J., Justice, C.O., Kaufman, Y. (2003). An Enhanced Contextual Fire Detection Algorithm for MODIS. *Remote Sensing of Environment*, Vol. 87, pp. 273-282.
- Hall, P., Cobleigh, B., Buoni, G., Howell, K. (2008). Operational Experience with Long Duration Wildfire Mapping: UAS Missions Over the Western United States. *Proceedings of Association for Unmanned Vehicle Systems International (AUVSI)*, San Diego, California, June 2008, AUVSI.
- Henson, Robert, (Ed.), (2008). *Satellite Observations to Benefit Science and Society: Recommended Missions for the Next Decade*, National Research Council, The National Academies Press, National Academy of Science, ISBN: 0-309-10904-3, pp. 1-40.
- Justice, C.O., Giglio, L., Korontzi, S., Owens, J., Morisette, J.T., Roy, D.P., Descloitres, J., Alleaume, S., Petitcolin, F., Kaufman, Y. (2002). The MODIS Fire Products. *Remote Sensing of Environment*, Vol. 83, pp. 244-262.
- Kaufman, Y.J., Justice, C.O., Flynn, L.P., Kendall, J.D., Prins, E.M., Giglio, L., Ward, D.E., Menzel, W.P., and Setzer, A.W. (1998). Potential Global Fire Monitoring From EOS-MODIS. *Journal of Geophysical Research*, Vol. 103, pp. 32,215-32,238.
- Li, Z., S. Nadon, J. Cihlar, B. Stocks (2000a). Satellite Mapping of Canadian Boreal Forest Fires: Evaluation and Comparison of Algorithms. *International Journal of Remote Sensing*, vol. 21, pp. 3071-3082.
- Li, Z., S. Nadon, J. Cihlar (2000b). Satellite Detection of Canadian Boreal Forest Fires: Development and Application of an Algorithm. *International Journal of Remote Sensing*, vol. 21, pp. 3057-3069.
- Morisette, J.T., Giglio, L., Csiszar, I., Setzer, A., Schroeder, W., Morton, D., Justice, C.O. (2005). Validation of MODIS Active Fire Detection Products Derived From Two Algorithms. *Earth Interactions*, Vol. 9, pp. 1-23.
- NASA - Goddard Space Flight Center (2009). MODIS Rapid Response System. <http://rapidfire.sci.gsfc.nasa.gov/>.
- U.S. Forest Service, MODIS Active Fire Mapping Program (2009). <http://activefiremaps.fs.fed.us/>.
- U.S. Geological Survey (2008). Shuttle Radar Topographic Mission. <http://srtm.usgs.gov/index.php>, site updated 23 June 2008, site accessed 1 June 2009.
- Wegener, S. S., (2009). UAS for Remote Sensing, Myths and Realities. *Proceedings of 33rd International Symposium on Remote Sensing of Environment*, CD Proceedings, paper reference # 468, pp. 1-4, Stresa, Italy, May 2009, ISRSE.

The Geomorphometry of Rainfall-Induced Landslides in Taiwan Obtained by Airborne Lidar and Digital Photography

Jin-King Liu

*National Chiao-Tung University and Industrial Technology Research Institute
Taiwan*

Kuan-Tsung Chang

*Minghsin University of Science and Technology
Taiwan*

Jiann-Yeou

*National Cheng-Kung University
Taiwan*

Wei-Cheng Hsu, Zu-Yi Liao and Chi-Chung Lau

*Industrial Technology Research Institute
Taiwan*

Tian-Yuan Shih

*National Chiao-Tung University
Taiwan*

1. Introduction

Taiwan has a land area of 36000 m². 26.68% of the land areas are covered by plain region, whereas 27.31% are hilly and 46.01% are mountainous. By official definition for the purpose of land conservation management, hilly lands refer to the area under 100m but with a slope more than 5% or the area between 100m and 1000m. Mountainous lands refer to the area with an altitude above 1000m. Therefore, 73.32% of the areas are under conservation management. The complicated landscape of Taiwan is characterized by small drainage basins, highly fractured rock, high relief, and steep stream gradients. Frequent earthquakes due to the collision of Eurasian Plate and Philippine Sea Plate in eastern Taiwan further loosen the top surface of the land. Rock formations are highly fractured and jointed. Therefore the lands are particularly sensitive to episodic events such as typhoons and earthquakes, and various types of anthropogenic disturbance.

In addition, Taiwan is located in tropical and sub-tropical zones, often suffering from heavy rainfalls, especially in the summer seasons with typhoons. The average annual rainfall of Taiwan is 2500 mm which is about three times the world average. Landslides are easily induced by the heavy rainfall come along with typhoons. These physiographic settings

make Taiwan a fragile land, especially vulnerable to rainfall-induced landslides. The consequence is the sedimentation of the reservoirs. And the turbidity of the water in reservoirs becomes a major factor impacting the sustainable operation of water supply reservoirs in Taiwan. Landslides have to be recovered and their hazards have to be mitigated. The necessity of landslide survey is obvious.

Aerial photo interpretation has long been adopted for landslide inventory (Liu et al., 2001). This conventional method is based on visual perception of colour tone and geomorphometric features of landslides on the aerial photographs. Both manual interpretation and automatic recognition of satellite images are also used. Most of the recent automatic classification methods of landslides using images are based on spectral features other than topographic features. Therefore, landslides cannot be correctly recognized. A recent study is to establish an interactive approach with a software interface for assisting visual interpretation of landslides (Lau et al., 2006). Both spectral and spatial parameters are employed for the inputs of the software to assist the interpreter/operator to correctly recognize and delineate landslides. Automatic recognition of landslides solely on basis of spectral information of digital images is efficient in terms of time consumption, whereas the results usually can not meet the requirements for taking engineering measures (Parise, 2001). Nevertheless, manual interpretation is too slow to meet the requirements for emergency response. A hybrid approach is to combine the advantages of automatic processes with manual interpretation. The extraction of geomorphometric parameters from airborne LiDAR data is thus considered for integrating in the interactive interface to assist the interpreter.

Airborne LiDAR is the state-of-the-art technology for efficiently taking high density and high resolution elevation data for a wide area. This feature is also suitable for emergency response or quick assessment of landslide disasters. Hsiao et al. (2005&2006) shows that the integration of multi-temporal airborne LiDAR and aerial photography can give detailed change information of large-seated deep-seated landslide as demonstrated by the Jiu-fen-er earthquake landslide. For establishing an interactive interface for assisting visual interpretation of landslides, morphometric parameters derived from LiDAR are required for setting the internal defaults (Lau et al., 2006). In this interface, four primary parameters are selected, namely the greenness, the slope angle, the object height model, and surface roughness. Normalized Vegetation Index (NDVI) is taken for denoting the greenness if colour IR digital aerial photography is applied.

For these purposes, surveys were carried out with airborne LiDAR and digital camera to obtain digital terrain models (DTM) and digital surface models (DSM) of 1m grid and colour orthophotos of 50cm grid. DTM, DSM and orthophotos are georeferenced and transformed into the local coordinate system with Taiwan Datum 1997 (TWD97). Subsequently, the geomorphometric features of the landslides are analyzed. In this study, the geomorphometric characteristics of three selected events will be examined and these will be taken as reference values for setting the defaults in the software interface.

2. Conventional API Approach of Landslides and Its Implication

Rainfall-induced landslides are in majority shallow-seated in the high relief terrains of Taiwan. Techniques of stereoscopic airphoto interpretation have been adopted for landslide inventory in Taiwan since 1973 when an aerial survey team was established under Agricultural Council of the government. Though it is labour intensive, it is believed to be

reliable. The core spirit of this approach is the synergy of human perception to include both 2D and 3D features of the target and its environment. Any automated attempt should take this into account. Therefore, geomorphometric features of landslides constitute important ingredients in the automation process.

2.1 Rainfall-induced Landslides in Taiwan

For practical applications in the physiographic environments of Taiwan, the classification scheme of landslides developed by Varnes (1978) is simplified into five major categories, namely rock falls, shallow-seated landslides, deep-seated landslides, dip-slope and wedge slides, and debris flows. Thus, types of landslides can be differentiated by their physical appearance. It is especially useful for practical applications using remotely-sensed images.

		Type of Materials	
		Bed rock	Engineering Soils
		Debris	Soils
Falls	Rock falls		Shallow-seated slide
Topples			
Slide	Translational	Dip-slope and wedge slide	
	Rotational	Deep-seated slide	
Flows		(not applicable)	Debris flow (not applicable)

Table 1. A simplified classification scheme of landslides applied in Taiwan

There are 270 events of natural disasters in Taiwan in 50 years from 1958 to 2007 including categories of typhoons (71.1%), flooding (15%), earthquakes (8.5%), torrential rainfalls (2.2%), wind-storms (1.5%), mountain flooding (0.7%), and landslides (0.7%) (NFA, 2008). As shown in Figure 1, the frequency of natural disasters is in a trend of increasing. In total, 89% of the events are concerning with rainfall hazards and 97% of them are directly or indirectly concerning with landslides. Rainfall-landslides become a critical issue in managing natural disasters.

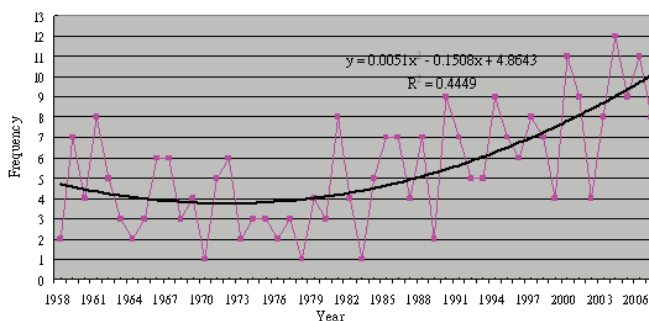


Fig. 1. Statistics of natural disasters in Taiwan from 1958 to 2007

Remote sensing has been an important tool for landslide inventory. The physical appearance of landslides is the basis of the recognition of the boundary and the type of a landslide. However, the displaced materials of a rainfall-induced landslide are usually washed away

from steep slopes. It remains only the fresh scars of the rupture surface. The fresh landslide scars emplacing at various slope gradients and various slope locations would normally include landslide types such as rock falls, debris slides, channel bank failures, and debris flows. In this study, the landslides concerned will cover all these types except debris flows. The exception is due to the reasoning that debris flows are triggered by a different mechanism with more contributions from flowing-water instead of gravity itself. In other words, debris flows can be treated as a transformation of other shallow-seated landslides when high concentration of rainfalls and liquefaction of displaced materials take place.

2.2 Procedures of Air Photo Interpretation

Air photo interpretation (API) is a process of understanding to associate shapes and pattern and other characteristics on vertical images with real features or phenomena on the ground. Interpretation by aerial photographs has been the most efficient and realistic way for identifying landslide topography in a wide area. Currently, researches in automatic extraction of landslides using images and digital elevation data become important topics (Barlow et al., 2003; Chang & Liu, 2004; Fernandes et al., 2004; Parise, 2001; Liu et al., 2008; Mantovani et al., 1996). However, visual interpretation by well-trained personnel is still believed to be more accurate and reliable than by computers. Interpretation process needs high skill and the results largely depend on the expertise of the interpreter. Sense of perception of a specific feature such as landslide can be acquired by practices and by an interpretation key describing visual signature characteristics of the object, including size, shape, pattern, tone, association, and texture. To minimize subjective factors of individual interpreters, cross checks should be implemented for a case covering a wide study area such as a few hundreds of aerial photographs. And, map making should be performed very carefully with, not only aerial photographs, but also site investigation.

The procedures of the conventional API adopted for a wide area of landslide inventory usually include steps as follows:

- (1). Acquisition and preparation of aerial photographs of the study area.
- (2). Aerial photograph interpretation (identifying landslide topography) - A stereoscope is used to pick up accurate landslide topography from aerial photographs. The scale of the panchromatic aerial photographs taken by the Aerial Survey Office of Forestry is about 1:20,000. Since 1976, about 20000 aerial photographs are taken every year. Photo index can be used for choosing the particular cloud-free photographs. Landslides with more than 50m in length were identified and their scarp, moving mass, internal structure, and moving direction are drawn with coloured pencils on the paper-printed photographs. A standard legend should be established.
- (3). Tracing the identified features on the topographic map - Tracing the features of landslides onto the topographic map by comparing identical landforms both on the photographs and the map. An original map of landslides is thus created.
- (4). Digitization and drawing the final map - The landslide features are then digitized. Subsequently, landslide scarps and lineament structures are compiled and printed with a backdrop of conventional contour map in a GIS environment. These maps were examined and revised by the researchers.
- (5). Field check and update the attribute table from field records.
- (6). Ancillary materials for interpretation.
- (7). Final presentation and backups.

The second step of the API procedures is the most critical one where stereoscope is usually used to perceive the sense of 3D features and a well-trained interpreter should be acquainted with interpretation key for the study area.

2.3 Interpretation Key

The perception of landslides from a bird-eye view of aerial photographs is also largely depending on the scale or spatial resolution of the photographs. Landslides can not be mapped properly when they are smaller than a minimum mapping unit such as 5mm on the paper prints. Before 2008, the aerial photographs taken by Aerial Survey Office had been the conventional panchromatic photographs in a scale around 1:20000. Therefore, the minimum mapping unit of the landslides will be larger than 100m in the real ground. In general, four factors affect the quality of the mapping results, namely the scale, the time lag between the landslide event and the aerial photography, the type of film used, and the overall quality of the photographs. Table 2 shows the criteria used for the recognition of landslides on aerial photographs. The general feature of a rainfall-induced landslide is characterized by the fresh landslide scars in elongated shape and located in a relatively steep slope. It takes place in any kind of geology so long as there are some weathered overburdens. Features on aerial photographs include the bright tone, the bare surface, and the features shown in Table 2. Manual interpretation uses both 2D and 3D features of the landslides for recognition. The 2D features include tone, location, and shape. The 3D features include location, direction, slope, and shadow effects. A sound consideration of the automation of landslide recognition should be able to take care of all these aspects.

Feature	Description	Discrimination rule
Tone	Light, grey light	Brightness>Threshold
Location	Near ridges, cut-off slopes, road-sides	Trigger events and buffer zone of the feature
Shape	Spoon-shaped, elongated-oval, dentritic, rectangular, triangular	Location-specific and topography-specific
Direction	The drop direction of the landslide is the gravitational vector on the ground surface.	Roughly perpendicular to the streams and topography-specific
Slope	Depend on types of landslides. E.G. Shallow-seated landslides > 45%; Deep-seated landslides ~40%; Debris flows ~10-20%.	Slope > Threshold
Shadow	Depend on whether the landslides are in shadow-side or sunny-side	Solar azimuth in related to slope aspect

Table 2. The criteria for the recognition of rainfall-induced landslides

2.4 Geomorphometry of Landslides

Obviously, geomorphometry has been applied in manual interpretation. Geomorphometry, the science of quantitative land surface analysis is also known as geomorphological analysis, terrain morphometry, terrain analysis, and land surface analysis (Hengl & Reuter, 2009). The aims of geomorphometry are to extract surface parameters and objects using input digital terrain models. Pike (1988) listed a dozen groups of parameters used as terrain descriptors using manually digitized digital terrain models and he used a resulting "geometric signature or topographic signature" to categorize terrain characteristics and suggested the degree of

danger from landslides. Topographic signature of life and their processes are deemed to be strongly influenced by biota (Dietrich & Perron, 2006). Guth (2001 & 2003) took terrain fabric as measures of a point property of the digital terrain models and the underlying topographic surface. This study is also known as topographic fingerprints (Densmore & Hovius, 2000) for characterizing the location of a landslide on the slope. The state-of-the-art technology of high resolution satellite images, digital aerial photography, and airborne LiDAR opens a new era in the automation of landslide recognition, especially the possibility of applying geomorphometrics. And, the extraction of land surface parameters becomes more and more attractive for both stochastic and process-based modelling, making use all the level of detailed digital terrain models.

It is shown that the topographic-based analyses can be used to objectively delineate landslide features, generate mechanical inferences about landslide behaviour and evaluate relatively the recent activity of slides (McKean & Roering, 2004; Glen et al., 2006). Especially, surface roughness derived from LiDAR DTM allows an objective measurement of landslide topography. Eigenvalues of surface normals can be an effective parameter for differentiating shallow landslides and debris flows (Woodcock, 1977).

For establishing an interactive interpretation software interface to assist the interpreter, it is clear that expert knowledge of the morphometric properties of landslides is required. And, data acquisition with the new sensors of aerial digital camera and LiDAR becomes feasible. Therefore, the general properties of slope angles, OHM and roughness of rainfall-induced landslides are included in this study.

3. The New Interactive Approach and Parameters of Geomorphometry

Figure 2 shows some typical rainfall-induced landslides in Taiwan. Landslides are bare in high relief terrains with densely-vegetated surroundings. Typical modernized aerial survey system nowadays is equipped with a digital camera and a LiDAR sensor. The procedures of landslide inventory are subjected to change to adopt the new types of high resolution digital data. Thus, an interactive system for manual interpretation under a digital environment is required. Standard products generated by the new survey system include orthophoto, DTM and DSM. In addition to the functions for data management and manipulation in the interactive system, algorithms for automatic recognition of landslides are also required to assist or guide the interpreter for improving the efficiency.



Fig. 2. Typical rainfall-induced landslides in Taiwan

3.1 The interactive system for color orthophoto interpretation

On basis of the experiences in airphoto interpretation and national landslide inventory, a man-machine interface is developed using windows software development tools including Visual Studio .NET, Borland C++ Builder, and OpenGL. Figure 3 is the flowchart of the interactive system which includes three data entries and four parameters. The entries and parameters will be modified when more standard products are available. Parameters of roughness, OHM and Slope are derived from LiDAR data. Parameter 4 the greenness is derived from color orthophoto. These four parameters are used for highlighting potential areas of landslides by default settings of threshold for the parameters. Another option is to manually define training areas to obtain the threshold from the training sample.

The visualization on the screen shows both 2D and 3D perspectives of the results (Figure 4). Final setting of parameter thresholds can be optimized visually. And finally, the interpreter can further edit the results of automated detection. Or, the interpreter can even carry out all the interpretation discarding the automated results. Finally, ground truth can be imported to compare with the results for accuracy assessment.

For practical reasons, only four major parameters which can be easily derived from the standard aerial products available by a national agency are used for the automatic back-processing in the interactive system (Figure 3). Simple thresholds are used to highlight the potential landslides. For example, roughness < 5m, OHM < 10m, slope > 40 degrees, and greenness < -0.40. Default settings of thresholds are set on basis of geomorphometric analysis of rainfall-induced landslides for the specific area in related to physiographic conditions and the triggering event. Another option is to obtain the thresholds from the training sample. In this system, a landslide seed is located on the screen by the interpreter. The values of 25 pixels extracted from a 5x5 window centred at the assigned seed are used to calculate statistical means and standard deviations. Three times of the standard deviations are taken as the thresholds. Any pixel with a value within three standard deviations of the means will be assigned as a pixel of landslide. Thus, the omission and commission errors of landslide recognition can be minimized. In addition, the thresholds can be tuned interactively to see the correctness of matching between the landslide feature on the colour orthophoto and the highlighted area (Figure 4).

3.2 The parameter derived from orthophoto

Because rainfall-induced landslides of natural slopes are mostly covered by densely-vegetated surroundings, vegetation index will be critical for indicating the areas of bareness. The most popular one is the NDVI (Normalized Vegetation Index).

$$NDVI = (NIR-R)/(NIR+R) \quad (1)$$

where R stands for the grey value of Red band, and NIR stands for grey value of Near Infrared band. Theoretically, if the image digital values are calibrated to stand for the reflectance of the target, the NDVI can be widely applicable. However, the digital numbers of Red band and NIR band of digital aerial camera are not calibrated for this purpose. Therefore, the NDVI value is a relative indicator of biomass. NDVI can be applied for recent digital aerial cameras which usually includes an NIR band. If the colour aerial photographs include only RGB bands, an alternative of greenness parameter can be used. Greenness is also a relative indicator, of which the radiometric values are not normalized.

$$\text{Greenness} = (G-R)/(G+R) \tag{2}$$

where G is the grey value of Green band, and R is the grey value of Red band. The range of the values of NDVI and Greenness is between -1 and 1. Nevertheless, the range for those of landslides may change with natural weather, terrain conditions and type and settings of the camera sensor. A relative low value implies that the area of the pixel is low-vegetated or bare.

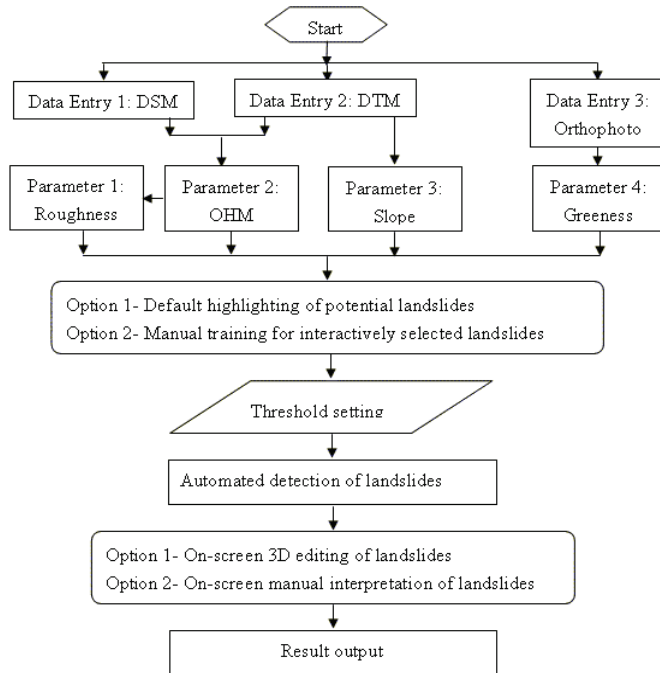


Fig. 3. Flowchart of the interactive system

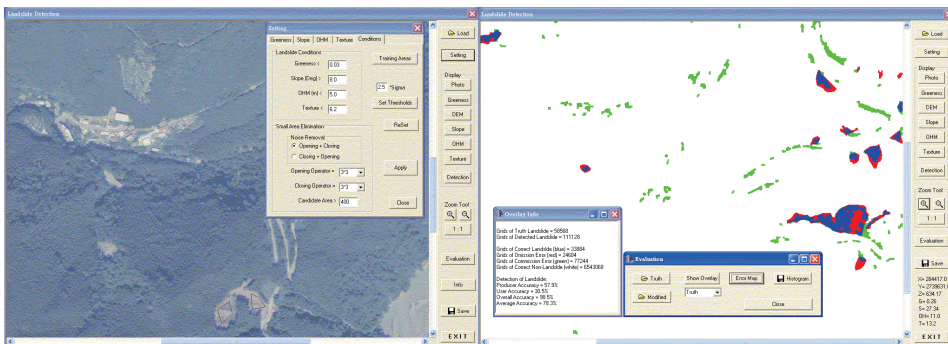


Fig. 4. Screen shots of the interactive system. (Left) Parameter settings; (Right) Accuracy assessment by comparing classified result with ground truth

3.3 The parameters derived from airborne LiDAR

Three parameters are derived from airborne LiDAR DTM and DSM, namely the slope, the object height model (OHM) and the surface roughness. The factors in the mechanism of slope stability usually include slope angle, strength of materials, and pore water pressure (Turner & Schuster, 1996). If the slope gradient is high, the slope can be unstable. Slope is thus selected as the first parameter due to its importance and that it can be easily derived from DTM. There are two surfaces which can be easily defined by LiDAR-derived data. One is the digital terrain model (DTM) standing for the bare ground surface. The other is the digital surface model (DSM) standing for the upper envelope of all the objects above the bare ground surface. For an area of rainfall-induced landslide, the difference between these two well-defined surfaces can be minimal. Therefore, the OHM defined as the difference of these two surfaces can be a good parameter for automatic landslide recognition. It is straightforward that, due to the wash out or sliding, the surface of landslides in nature should be smoother than their surroundings. Surface roughness has been proved to be an objective and useful measurement of landslide topography (McKean & Roering, 2004; Woodcock, 1977; Glen et al., 2006).

(a) Slope

Slope angle of a landslide is the angle between the horizontal and the ground surface of the longitudinal axis of the landslide. Slope angle for each of the landslides can be determined by the slope angles derived from LiDAR DTM. If the surface of the ground is

$$Z=f(X, Y) \tag{3}$$

the slope (in radian) can be defined as.

$$\text{slope} = \tan^{-1} \left(\sqrt{f_x^2 + f_y^2} \right) \tag{4}$$

where $f_x = \frac{\partial Z}{\partial X}$, $f_y = \frac{\partial Z}{\partial Y}$

In common practice, the DTM is stored in grid form. The slope of a grid element such as Z5 in Figure 5 is computed by using a 3x3 moving window.

Z ₇	Z ₈	Z ₉
Z ₄	Z ₅	Z ₆
Z ₁	Z ₂	Z ₃

Fig. 5. Slope calculation by a kernel of 3x3 moving window

If the fluctuation of local height becomes too large due to the nature of LiDAR data or due to the nature of local relief, the resulted slope angles will be subjected to heavy noises with discontinuities of slope angles. It is therefore necessary to introduce an image processing method to resolve the problem. The first order of differentiation is applied for convolution operation with DTM. In the principle of image processing, a 2D (x, y) convolution is equivalent to two passes of 1D convolution of both (x) and (y). This simplification can be implemented more efficiently (Sharpnack & Akin, 1969; Parker, 1997). For example, formula

(5) is a 1D Gaussian function and formula (6) is the first order of its derivative. Therefore, the slope formula in (4) can be implemented by convolution operations in both x and y directions with DTM grid.

$$f(x) = e^{-\frac{x^2}{2\sigma^2}} \quad (5)$$

$$f'_x = f'(x) = \left(-\frac{x}{\sigma^2}\right) e^{-\left(\frac{x^2}{2\sigma^2}\right)} \quad (6)$$

(b) OHM

OHM is obtained by simply subtracting DTM from DSM for describing the height of objects above the bare ground. DTM is also referred to nDSM, i.e. normalized DSM, denoting the significance of the surface is tightly related to DSM. DTM is the bare ground surface excluding all objects above the ground. In forestry land, the difference between DSM and DTM can be referred to CHM (Canopy Height Model), denoting the general heights of the trees. The surface objects especially in forests are generally depleted in areas of landslides. Therefore, a minimal value of OHM can be expected in landslide areas.

(c) Surface Roughness

Surface roughness can be described by either the variance of DSM or OHM in a local window area. In this study, roughness is defined as one standard deviation in a 5x5 moving window on OHM for describing the relief variation in the local area. This can partly diminish the effects of landscape undulation. A 5x5 window is used for extraction the variance of the OHM values in the moving window and then the value of one standard deviation is used to stand for the surface roughness of the central pixel. In the areas of rainfall-induced landslides, the roughness will be lower than other areas due to the depletion of surface materials.

3.4 Scale Effects of Digital Terrain Models

Because slope angle, OHM, and roughness are generated from DTM, they are subject to the change of DTM grid-size. This poses a requirement to understand the possible scale effect due to the change of DTM grid-size for landslide areas (Claessens et al., 2005). A contraction of 1m grid is carried out to obtain grids of 5m, 10m, and 40m for comparison. A pixel on the grid will cover a larger area when the scale is smaller. There are two approaches for the contraction, namely pixel thinning and pixel aggregation. With pixel thinning, every nth pixel is kept. With pixel aggregation, the new pixels represent averages of the n pixels specified by the contracting factor. In Taiwan, DTMs of 5m, 10m and 40m grids are created on bases of photogrammetry. Therefore, pixel aggregation approach is used in this study for its comparability to image matching.

4. Test Areas and Materials

The landslides induced by rainfall events in Shimen, Alishan and Ilan of northern, middle and eastern Taiwan are selected for this study. Figure 6 is the location map of the three

study areas and the landslides of these areas due to the relevant events. Surveys were carried out with both sensors of airborne LiDAR and digital camera to DTM and DSM of 1m grid and orthophotos of 50cm grid. DTM, DSM and orthophotos are georeferenced, co-registered and transformed to the local coordinate system with Taiwan Datum 1997 (TWD97) for the analyses of the induced landslides.

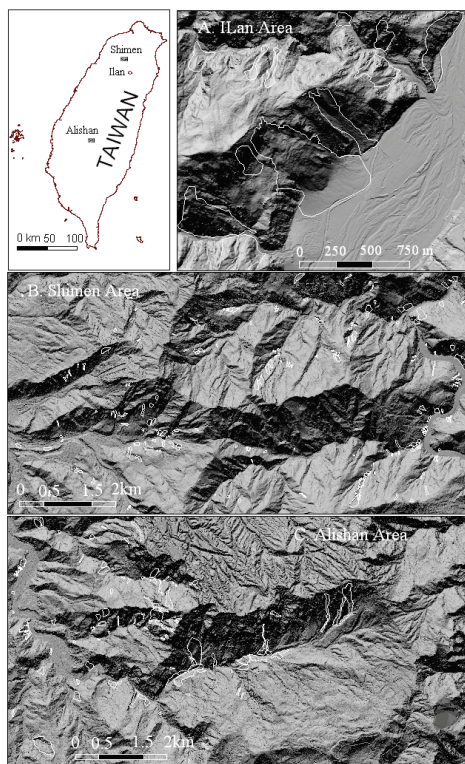


Fig. 6. The location map of the study areas and the landslides of these areas

Aerial surveys were conducted after rainfall events as shown in Table 3. Although the maximum rainfall in the period of Typhoon Longwang in Shimen was as small as 208 mm, this event was the one followed three larger events in three months of the same year, i.e. Haitang (504mm) on July 16, Matsa (818mm) on August 3, and Talim (384mm) on September 1. The event in Alishan was just a concentrated torrential rainfall. On 9th June 2006, the cumulative rainfall had reached 811mm in 24 hours and 1200 mm in 48 hours. Enormous amount of debris flows and slides took places. LiDAR data and aerial photographs were taken right after the event on 22nd June of 2006. There had been no records of heavy rainfall events one year prior to this event. The landslides observed with these datasets can be solely attributed to this rainfall event. Typhoon Kalmaegi on July 17 took place nine month after Typhoon Krosa on October 4 of 2007 in Ilan area. The rainfall took place after a dry and hot summer season. The occurrences of the three selected study areas are different.

Name and size of study area	Date of data acquisition	Rainfall event	Date of the event	Maximum rainfall (mm)
Shimen (48 sq. km)	Jun. 17, 2006	Typhoon Longwang	Sept. 30, 2005	208
Alishan (36 sq. km)	Jun. 22, 2006	Torrential rainfall	Jun. 9, 2006	1200
Ilan (4 sq. km)	Nov. 4, 2008	Typhoon Kalmaegi	Jul. 17, 2008	1100

Table 3. Rainfall events related to the study areas

The orthophotos were then generated by the aerial photographs taken by direct-georeferencing technique and ortho-rectified by LiDAR DSM without using ground control points. Photography and laser scanning are synchronized. Because airborne LiDAR is equipped with GPS and IMU, an event mark is given when photography system triggers a transistor-transistor logic pulse. Thus, the instantaneous GPS and IMU information can be used to resolve the exterior orientation of the photo frame, i.e. x , y , z , ω , ψ , κ . Subsequently, the true-ortho ground surface model, i.e. LiDAR DSM, is used for the ortho-rectification of the central projected photograph.

Leica ALS50 airborne LiDAR system used in this study is consisted of 2 major parts, i.e. a laser scanning assembly and a Position and Orientation System (POS). The former one is for triggering laser pulses, controlling the range, the swath, the FOV, the scan rate and the pulse rate. These parameters decide how fast we can make a complete coverage of the survey area. The second part is critical to the positioning accuracy.

Point density is an important indicator for the spatial resolution of LiDAR DTM and DSM. An understanding of the forest closure and crown density can be obtained by preliminary inspection of the point-density distribution of point clouds (Means et al., 2000; Naesset, 2002). In Alishan study area, the point density in average is around 2.3 points/m² with ground point density of 0.6 points/m². The upper envelope of the point clouds is interpolated to form DSM of 1m grid, whereas the point clouds that hit the bare ground or that are filtered to eliminate off-ground points are interpolated to form DTM. In other words, DTM denotes the bare ground surface. The accuracy of the DTM and DSM can be varied due to the change of land-cover types and density of vegetation. For assuring the accuracy, ground survey with total stations was carried out for 347 selected sample points. The RMSE is 0.82m, and mean error is 0.73m (Table 4). The error actually is a bias verified in the field check because this is due to the dense low bushes underneath the tree-canopies. This over-estimation of DTM is noteworthy especially for tropical and sub-tropical forest. In general, the accuracy of bare grounds is about 0.15m. Similarly, Shimen and Ilan areas were flown with looser point density of 1.5 points/m² with ground point density of 0.45 points/m². Figure 7 is an example of a blown up of 1 square km of the Alishan study area. It is clearly shown that the landslide area can be enhanced on the OHM image where the landslide areas are with low OHM values.

Locations	Sample size	Average error (m)	RMSE (m)	Standard Error (m)
Tree base	219	0.70	0.77	0.33
Open Ground	128	0.79	0.90	0.43
Total	347	0.73	0.82	0.37

Table 4. Accuracy assessment of the DTM in forest lands

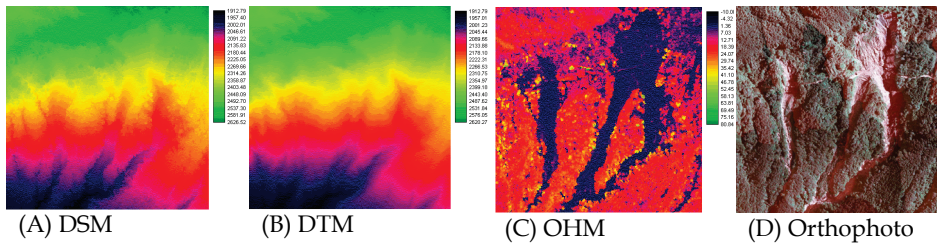


Fig. 7. Blown-up of a 1x1 km area of Alishan study area

5. Results and Discussion

5.1 The resultant images of the four parameters

Greenness can be extracted from RGB orthophoto where the landslide area exhibits lower value (Figure 8C). Local slope can be calculated using 3rd finite difference algorithm (Figure 8D). OHM is a normalized height of objects above the bare ground surface. Because terrain effect has been removed, OHM exhibits a good appearance of landslides (Figure 8E). The roughness of landslide area is obviously lower than that of the environment (Figure 8F). In other words, the smoothness of landslide area is obviously higher than that of the environment.

It also can be observed that the shaded-relief image of DSM gives a better contrast between landslides and their environments than that of DTM due to the contribution of the shading effect of the trees and other above-ground objects (Figure 8A and B). In addition, The DSM-shaded image in nature is a true ortho-image, possessing the advantage of no occlusion of object shading when compared with orthophoto of the same area (Figure 7D). It is costly to process an orthophoto to a true orthophoto which needs to incorporate the correction of objects along with the terrain correction. Therefore, if airborne LiDAR survey is carried out alone without an integrated digital camera, the DSM-shaded image can be a good surrogate of panchromatic photograph for manual interpretation.

5.2 Results of Manual Interpretation of Landslides

Landslides of the study areas (Figure 6) are obtained by manual interpretation of colour orthophotos of 50 cm grid and DSM-shaded images of 1m grid using the criteria of expert knowledge for conventional aerial photo interpretation.

The total number of the rainfall-induced landslides in the 36 km² in Alishan of middle Taiwan is 106 with a total coverage area around 1.29 km². The landslide occurrence rate is around 4%. Statistically, 8% of the landslides have a longitudinal length of less than 30m; 36% between 30~60m; 67% less than 100m; 86% less than 150m. If more than 5 pixels are the minimum mapping unit for visual interpretation, usually more than 36% of the landslides will not be mapped using remotely-sensed images in medium resolution. The total number of landslides in Shimen of northern Taiwan is 200 with landslide coverage of 0.76 km² in 48 km² of study area. The landslide occurrence rate is around 1.4%, which is only one third of the rate in Alishan although the total number of landslides is more than that in Alishan. This implies that smaller consecutive rainfall events in Shimen area trigger more small landslides than that in Alishan area. This assertion can be further supported by the evidence observed in Ilan area of eastern Taiwan. The total number of landslides in Ilan area is 12 in 2 km² of

study area with landslide coverage of 0.14 km². The landslide occurrence rate is around 7.0%. The average area of a landslide in Ilan area is also larger than that of Shimen area, yet comparable with that in Alishan area (Table 5).

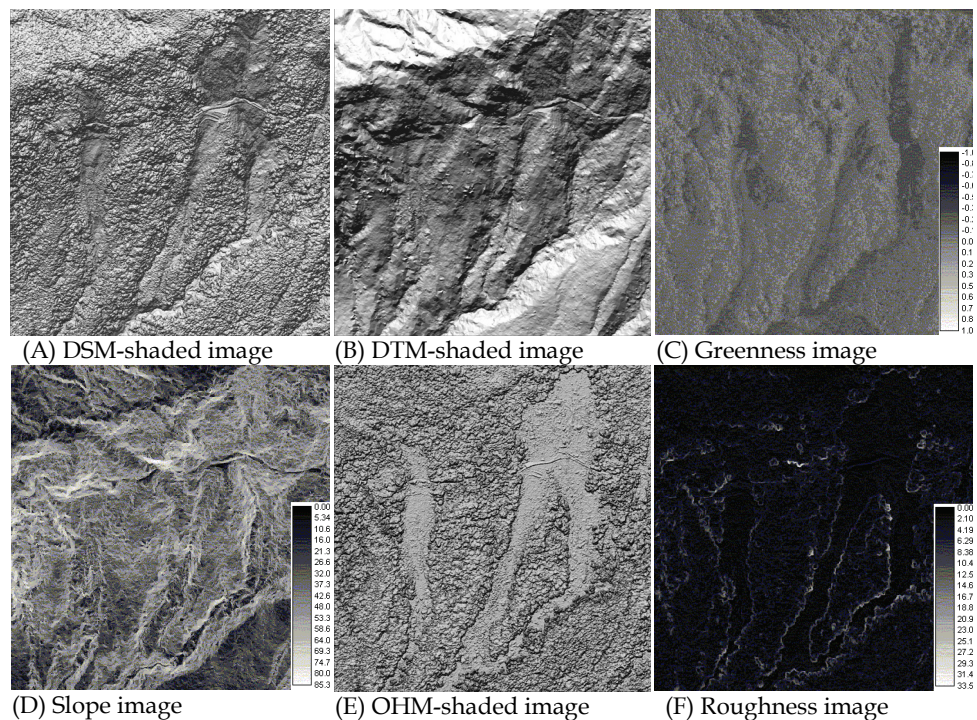


Fig. 8. Resultant images

Study area	Total area (km ²)	Total landslide area (km ²)	Total number of landslides	Landslide occurrence rate (%)	Average area of a landslide (m ²)
Alishan	36	1.29	106	4.0	122
Shihmen	48	0.76	200	1.4	38
Ilan	2	0.14	12	7.0	117
Average	-	-	-	4.1	92

Table 5. Statistics of the landslide distribution of the study areas

5.3 Results of the morphometric analyses of Landslides

Manually-interpreted landslides are overlaid with DTM/DSM derivatives to extract the selected geomorphometric parameters including slope angle of landslides, object height models, and surface roughness. Statistics of the landslides in Alishan area (Table 6) show that the mean slope angle of the areas covered by landslides is 40.99 degrees with one standard deviation of 14.14 degrees. In contrast, the mean slope of the whole study area is 33.97 degrees with a standard deviation of 15.71 degrees. Generally, average slope angle in

landslide areas is higher than that of the whole area. Figure 9 shows that the peak of the curve of slopes of landslide areas is higher and when the slopes are more than 31 degrees the faction of landslide slopes is more than that of the general slopes. This tendency holds true for both Shimen and Ilan areas.

		Slope (deg)		OHM (m)		Roughness (m)	
		Whole area	Slide area	Whole area	Slide area	Whole area	Slide area
Alishan	Mean	33.97	40.99	14.31	4.40	3.25	2.05
	Std. Dev.	15.71	14.14	9.69	6.30	2.69	2.56
Shimen	Mean	35.15	43.79	13.23	2.150	2.37	1.48
	Std. Dev.	14.28	12.95	8.01	4.70	1.87	2.11
Ilan	Mean	29.00	40.48	10.20	6.15	2.55	0.40
	Std. Dev.	20.14	13.14	10.81	8.32	2.82	1.32
Average of the means		32.71	41.75	12.58	4.23	2.72	1.31

Table 6. Statistics of the geomorphometric parameters of the rainfall-induced landslides

The mean value of OHM of the landslide areas in Alishan is 4.40 m with one standard deviation of 6.3 m; whereas for the whole study area, they are 14.31 m and 9.69 m, respectively. OHM of landslide areas are obviously smaller than that of the surroundings where are vegetated with high forests (Figure 8E). Figure 9 shows that the distribution of OHM for the whole study area is bi-modal with one additional peak between 10~31 m. The peak in the right side is a forestry peak representing the concentration of trees. The mean OHM of Shimen area is as small as 2.15m denoting a cleaning ground surface of the sliding areas, whereas the mean OHM of Ilan area is 6.15m denoting the landslide areas remain some tree residues above the ground surface.

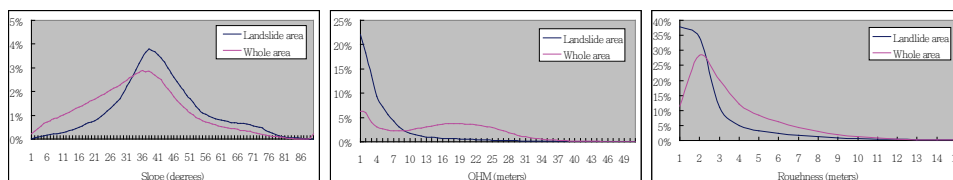


Fig. 9. Statistics of the three selected parameters of Alishan area

The mean roughness of the landslide areas is 2.05 m whereas it is 3.25 m for the whole Alishan area. The cumulative curve of roughness shows that 83% of the landslides have a roughness less than 2m and 88% less than 3m. In general, the means of the landslide areas are less than those of the whole areas. This indicates that ground surface of landslide areas are significantly soother than their surroundings, reflecting the truth of Figure 8(F). The mean surface roughnesses for both of Shimen and Ilan areas are smaller than 2.0m which are even smaller than that of the Alishan area.

The significance of these three morphometric parameters can also be perceived from the average of the means in Table 6 that the differences of the parameters of the whole test area are substantially different from that of the landslide areas.

5.4 Analysis of Scale Effects of Digital Terrain Models

Table 7 shows the statistics of slope angles, OHM, and roughness of landslides in DTM grids of 1 m, 5 m, 10 m, and 40 m, respectively. Two features can be observed in the table: (1) statistics of 40 m grid are obviously different from others; (2) the roughness in four different grids gives quite different values. The former one reflects the unreliability of the statistics when grid-size is comparable to the lengths of landslides (see also Figure 10). The later one shows that there is a significant relationship between surface roughness and grid-size. In other words, there is a scale effect for this parameter. The value of the parameter is changed along with the grid-size. These can be further observed from Figure 10. When the dimension of landslides is similar to or less than the dimension of DTM grid-size, the computed slope angles become unstable, maybe too big or too small. The OHM shows similar phenomena that in 40 m grid, the pixels become mixed cells, i.e. trees nearby the landslide give contribution to the OHM. Surface roughness exhibits changes in all different grid-sizes.

It is noteworthy that there is no cell with a roughness of more than 22m for the curve of 40m grid. 22m is about the half of the 40m-gridsize. This shows that the distribution of roughness is scale-dependent. In short, DTM with a grid size smaller than 40m will not be suitable for analyzing the rainfall-induced landslides which are usually with an area smaller than 40x40m² as demonstrated in this study (Table 5). Therefore, it should be carefully treated when applying DTM with different resolution for geomorphometric studies.

		1m grid	5m grid	10m grid	40m grid
Slope	Mean	40.99	40.69	40.25	37.77
	Std. Dev	14.14	13.77	13.44	13.20
OHM	Mean	4.40	4.86	5.01	6.61
	Std. Dev	6.30	5.99	5.85	5.90
Roughness	Mean	2.05	7.06	13.37	33.96
	Std. Dev	2.56	4.66	7.39	7.38

Table 7. Statistics of slope angles, OHM, and roughness of landslides in four grids

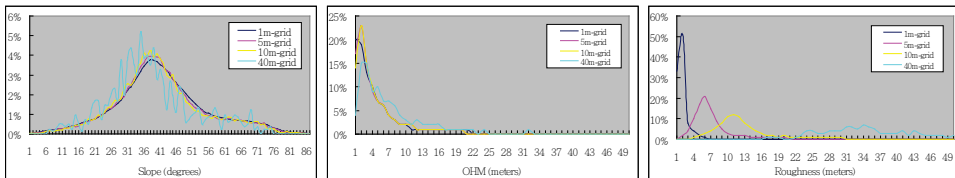


Fig. 10. Scale effects of slope, OHM and roughness derived from various grid-sizes

6. Conclusions

Conventional airphoto interpretation has long been adopted as a standard approach for reliable national mapping of landslides and it is still applied for this purpose in many places of the world including Taiwan. For establishing an interactive interpretation interface to assist the interpreter, expert knowledge of morphometric properties of landslides are required for entries to automatic detection algorithm to highlight the potential areas of landslides in the system. In this study, for understanding these properties, aerial surveys were carried out with airborne LiDAR and digital camera to obtain DTM and DSM of 1m

grid and orthophotos of 50cm grid. The landslides induced after torrential rainfalls in middle, northern and eastern Taiwan are selected for this study. It is proved that the morphometric parameters of rainfall-induced landslides are useful in the automatic detection of landslides for highlighting the potential areas in the interactive system. However, they have to be defined in related to local conditions and the specific events triggering the landslides. It is also observed that scale effects are obvious for roughness but not for slope and OHM. The scale effect takes place when the DTM grid is comparable to the average size of landslides, i.e. 40m in this study.

7. Acknowledgment

This study was sponsored by the grant of Council of Agriculture, Taiwan. Project ID is 95COA-12.1.1-S-a1.

8. References

- Barlow, J.; Martin, Y. & Franklin, S. E. (2003). Detecting translational landslide scars using segmentation of Landsat ETM+ and DEM data in the northern Cascade Mountains, British Columbia. *Can. J. Remote Sensing*, 29(4) 510-517.
- Chang, K. T. & Liu, J. K. (2004). Landslide features interpreted by neural network method. *Proceedings of ISPRS 2004, Istanbul, 2004-7-12~2004-7-22*.
- Claessens, L; Heuvelink, G.; Schoorl, J. & Veldkamp, A. (2005). DEM resolution effects on shallow landslide hazard and soil redistribution modelling. *Earth Surface Processes and Landforms*, 30(4) 461-477. John Wiley & Sons, Ltd., 15 Apr 2005.
- Dietrich, W.E. & Perron, J. T. (2006). The search for a topographic signature of life. *Nature*, 439 (7075) 411-418. 26 January 2006. doi:10.1038/nature04452.
- Densmore, A. L. & Hovius, N. (2000). Topographic fingerprints of bedrock landslides. *Geology*. April 2000. 28(4)371-374.
- Fernandes, N. F.; Guimaraes, R. F.; Gomes, R. A. T.; Vieira, B. C.; Montgomery, D. R. & Greenberg, H. (2004). Topographic controls of landslides in Rio de Janeiro: field evidence and modeling. *Catena*, 55 (2004) 163-181.
- Glenn, N. F.; Streutker, D. R.; Chadwick, D. J.; Thackray, G. D. & Dorsch, S. J. (2006) Analysis of LiDAR-derived topographic information for characterizing and differentiating landslide morphology and activity. *Geomorphology*, 73 (2006) 131-148.
- Guth, P.L. (2001). Quantifying Terrain Fabric in Digital Elevation Models, in Ehlen, J., and Harmon, R.S., eds., *The Environmental Legacy of Military Operations*. Boulder, Colorado, Geological Society of America Reviews in Engineering Geology, 14:13-25.
- Guth, P.L. (2003). Eigenvector Analysis of Digital Elevation Models in a GIS. In: *Geomorphometry and Quality Control, in Concepts and Modelling in Geomorphology: International Perspectives*, Eds. I.S. Evans, R.Dikau, E. Tokunaga, H. Ohmori and M. Hirano, pp.199-220, TerraPub, Tokyo.
- Hengl, T. & Reuter, H. I. (eds.), 2009. *Geomorphometry - Concepts, Software, Applications. Series Developments in Soil Science*, 33, Elsevier, ISBN 9780123743459.
- Hsiao, K. H.; Liu, J. K.; Yu, M. F. & Tseng, Y. H. (2005). Topographic Change Detection Using Aerial Photogrammetric Survey and 3D Laser Data in Jiu-fen-er Mountain. *Journal of Photogrammetry and Remote Sensing*, 10(2), June 2005, p. 191-202.

- Hsiao, K. H.; Liu, J. K.; Yu, M. F.; Chen, T. K.; Hsu, W. C. & Wang, C. L. (2006). Terrain Change Detection Combined Photogrammetric DEM and Airborne LiDAR Data. *Journal of Photogrammetry and Remote Sensing*, (11)3, September 2006, p. 283-295.
- Lau, C. C.; Hsiao, K. H.; Chen, C. T.; Chung, Y. L.; Lin, C. S.; Chiu, C. Z.; Shih, T. Y.; Liu, J. K.; Chen, D. K.; Liao, T. Y.; Shih, C. H.; Weng, S. C.; Lo, S. F. & Pan, L.H. (2006). The development of advanced remote sensing technologies for forestry survey, part 1 of 3. *Report of Council Of Agriculture*, Database number: 950023, Project Number: 95-12.1.1-T-a1, Project ID: 120101a100. 12-15-2006.
- Liu, J. K.; Werng, S. J.; Huang, J. H. & Yang, M. J. (2001). Remote sensing image analyses of rainfall-induced landslides. In *Proc. 21 Century Civil Engineering and Management*, p. C23-33. Minghsin University of Science and Technology, 28 December 2001.
- Mantovani, F.; Soeters, R. & van Westen, C. J. (1996). Remote sensing techniques for landslide studies and hazard zonation in Europe. *Geomorphology*, 15 (1996) 213-225.
- McKean, J. & Roering, J. (2004). Objective landslide detection and surface morphology mapping using high-resolution airborne laser altimetry. *Geomorphology*, 57 (2004) 331-351.
- Means, J. E.; Acker, S. A.; Fitt, B. J.; Renslow, M.; Emerson, L. & Hendrix, C. J. (2000). Predicting forest stand characteristics with airborne scanning LiDAR. *Photogrammetric Engineering & Remote Sensing*, 66(11), p. 1367- 1371.
- Naesset E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80(1) 88-99(12).
- NFA (2008). *Historical records of natural disasters of Taiwan from 1958 to 2007*, National Fire Agency, Ministry of the Interior. Access date : 31 December 2008. <http://www.nfa.gov.tw/Show.aspx?MID=97&UID=827&PID=97>.
- Parise, M. (2001). Landslide mapping techniques and their use in the assessment of the landslide hazard. *Physics and Chemistry of the Earth*, 26(9) 697-703, 2001. doi:10.1016/S1464-1917(01)00069-1.
- Parker, J. R. (1997) *Algorithms for Image Processing and Computer Vision*, Wiley Computer, New York.
- Pike, R. J. (1988). The geometric signature: Quantifying landslide-terrain types from digital elevation models. *Mathematical Geology*. 20(5)491 - 511. ISSN: 0882-8121 (Paper) 1573-8868 (Online). DOI: 10.1007/BF00890333. Publisher: Springer Netherlands.
- Sharpnack, D. A. & Akin, G. (1969). An algorithm for computing slope and aspect from elevations. *Photogrammetric Engineering*, 35(3) 247-248.
- Turner, A K & Schuster, R. L. (eds.)(1996). *Landslide Investigation and Mitigation. Transportation Research Board National Research Council Special Report 247*, National Academy Press, Washington D. C.
- Varnes, D. J. (1978). Slope Movement Types and Processes. In: *Landslides Control, Special Report 176*. eds. Schuster, R. & Krizek, R., National Academy of Sciences, Washington, D.C. P.11-33.
- Woodcock, N.H. (1977). Specification of fabric shapes using an eigenvalue method. *Geol. Soc. Amer. Bull.* 88, 1231-1236.

Description and Publication of Geospatial Information

Arturo Beltran, Laura Díaz, Carlos Granell, Joaquín Huerta
and Carlos Abargues
Universitat Jaume I de Castellón
Spain

1. Introduction

Information systems have evolved to service-oriented architectures (SOA) where dedicated desktop applications have turned into on-line data and services. On one hand this distributed environment let users to share (resources) data and tools, but on the other hand there is a need to develop mechanisms to allow users to find and access to these distributed resources efficiently.

Current trends for discover and access geospatial information are being addressed by deployment of interconnected Spatial Data Infrastructure (SDI) nodes at different scales to build a global spatial information infrastructure (Masser et al., 2008; Rajabifard et al., 2002) being the SOA paradigm in the geospatial domain.

However, current Geographic Information Systems (GIS) and the services provided by the SDIs fail to allow transparent navigation between related geographic information resources. In SDI like in other domains, metadata are a necessary mechanism to describe the information, and together with Catalogue Services are the key elements for discovery and information fusion possibilities (Nogueras et al., 2005; Díaz et al., 2007).

In this context, pointing out this need, there are directives such as INSPIRE¹, that at European level, mandates the creation and maintenance of metadata and related discovery services (Craglia et al., 2007), these elements are, often, the first visible elements of added value in SDIs.

Metadata allow us to describe data and, based on it, we could organize, publicize and facilitate the access to such information. Traditionally, it has been the user or the data provider who creates these metadata that will be published in Catalogue Services, for being discovered and accessed later by different users in a SDI. The fact of generating metadata like who created the data, where are they placed, etc. is a laborious task, fundamentally because the traditional metadata formats are large and complex, the users who are documenting the data usually have no knowledge about some metadata of the original data due to the lack of information supplied by the provider, etc.

¹ <http://inspire.jrc.ec.europa.eu>

In this sense, the production of metadata becomes a laborious job that consumes a large amount of time and effort becoming a task released into the background despite its major importance. This provokes, in reality, a scarcity in metadata availability in SDI and consequently difficulty in data discovery and a miss functioning SDI.

For all the above reasons there is a need to facilitate metadata production to easily create, with minimal user intervention, metadata descriptions when the data are created. In this way, data and metadata can be packed, forming a logical unit, created at the same time and minimizing the inconsistency between data and their metadata.

In this chapter we present a methodology for documenting geospatial information. This methodology provides mechanisms to automate the generation and publication of metadata. For demonstration purposes we describe a prototype implemented within an open-source software GIS/SDI client. This prototype is capable of semi-automatic extraction of explicit metadata from data resources, metadata edition and publication to be catalogued for data discovery in an SDI.

The nature of this integrated workflow that facilitates metadata creation and management, will hopefully contribute to a change in mindset as to the cost/benefit ratio of generating and exploiting metadata, a necessary ingredient for successful SDI.

2. Background

2.1 Geospatial Information

There are studies showing that most of the information (more than 80%) is likely to be linked to a geographic position. When we talk about geospatial information we are talking about data intrinsically related to a geographic position. Although there exists formats specially supporting geospatial data, any other data or information, not considered spatial in nature can be georeferenced and considered as such.

Georeferenced resources are then resources of any nature that have defined their existence in physical space. That is, those that have established their location in terms of map projections or coordinate systems. Nowadays, the act of georeference has gone beyond the fields of geoscience and GIS, thanks to the emergence of new tools which their ease of use has expanded and democratized this task outside of the current technical context.

The use of tools like Google Earth², Flickr³, etc. has meant a qualitative leap in terms of georeferencing, extending the use of georeferencing resources traditionally limited to geodata in geosciences and GIS specialists, and thus accelerating the emergence of a geosemantic web, (Cerda, 2005). In the same way, the overcrowding and constant evolution of the georeferentiation has been boosted by the use of mashups in Web 2.0 sites, allowing the location of digital content (photo, video, news, 3D models, etc.) in digital mapping, nowadays called neogeography (Goodchild, 2007) (Goodchild, 2008).

All this georeferenced content, like geospatial data, can be described by using metadata and published in Catalogue Services in order to be integrated in SDI.

² <http://earth.google.es>

³ <http://www.flickr.com>

2.2 Description of Resources

A description is the explanation, in a detailed and ordered way, of how is certain person, place, object or anything, through the explanation of its various parts, characteristics or circumstances.

As we said earlier, metadata allows us to describe data and, based on it, we could organize, publicize and facilitate the access to such information. Metadata are commonly defined as “structured data about data” or “data that describe the attributes of a resource” or simply “information about data”. In other words, metadata is the information that describes the content, quality, condition, origin, and other characteristics of data. Metadata is the information and the documentation that enable data to be well understood, shared and used effectively by all types of users over time.

These metadata or data description must be generated according to a standard in order to fulfil the minimum requirements for interoperability. One of these metadata standards is DublinCore (DC), this standard was born originally to describe Web resources in a general way proposed by the initiative "Dublin Core Metadata Initiative" (DCMI)⁴. This initiative, created in 1995, promotes the dissemination of interoperable metadata standards and metadata vocabularies to build more intelligent information search systems. The DC standard has been approved as an American standard (ANSI/NISO Z39.85), in the technical European committee CEN/ISSS (European Committee for Standardization / Information Society Standardization System) and since 2003 also as an international standard by ISO (ISO 15836:2003 “Information and Documentation - The Dublin Core Metadata Element Set”).

The need of this kind of metadata standards is pointed out by organizations like World Wide Web Consortium (W3C)⁵. There are many other standards utilized for specific domains, for example, we can find various metadata formats for multimedia resources, like: Apple ITUNES XML, Yahoo MediaRSS, Cablelabs VOD Metadata Content Specification 2.0, MPEG-7 standard, W3C SMIL Standard, etc. W3C tries to standarize all these metadata formats and provide a way to work efficiently. (Toebes, 2007).

The W3C proposal is to develop metadata extending languages based in XML (eXtensible Markup Language) (Bray et al., 2000) or RDF (Resource Description Framework) (Manola y Miller, 2004). In this way in the geospatial domain there is a general consensus.

As we mentioned in the section before, potentially, any resource could be georeferenced and be integrated with other geospatial information in or outside SDI environments. As we focus on a methodology for description of geospatial information, we describe next the goals of the geographic metadata creation and the standards used in this domain.

Geographic metadata help people involved in the use of geographic information to find the information that they need and determine how best to use them (Nebert, 2004). In (FGDC, 2000) it is stated that the creation of geographic metadata has three major goals (which are also benefits):

- Organize and maintain investments in data made by an organization: Metadata seek to promote the reusability of data without having to turn to the team that was responsible for its initial creation.

⁴ <http://www.dublincore.org>

⁵ <http://www.w3.org/>

- Publicize the existence of geographic information through catalog systems: Metadata records are usually published through catalog systems, sometimes also referred as directories. Electronic catalogs not differ too much from the traditional library catalogs except for the fact that it offers a standardized interface for search services. Thus, these catalogs are the tool that put consumers in touch with the producers of information. By means of the publication of geographic information resources through a catalog, organizations can find data to use, other organizations with who share data and maintenance efforts and customers for these data.
- Facilitate the access to the data, their acquisition and a better utilisation of the data achieving information interoperability when it comes from various sources: Metadata help receiving users or organizations in the processing, interpretation and storage of data in internal repositories.

Within the world of geographic information have been defined recommendations for the creation of metadata, whose main purpose is to provide a “hierarchical and concrete” structure to describe fully each of the data to which they refer. These recommendations have been created and approved by standardization bodies according to opinions of experts in the area. These recommendations, in form of standards or metadata schemas, provide criteria to characterize their geographic data properly.

Throughout the years have emerged, at national or European level, even within a specific domain, a set of initiatives to standardize the creation of metadata. However, these initiatives have been repealed for harmonization with the international standard ISO19115:2003⁶. Even the new version of the American standard CSDGM⁷ will converge with the international standard.

Regardless of the metadata standard used, it is usual to classify the elements of metadata respect on their role within the paradigm “discovery, evaluation and access” established in (Nebert, 2004):

- Discovery metadata elements are those that allow minimally describe the nature and content of a resource. These elements usually respond to the questions “What, Why, When, Who, Where and How”. Typical elements in this category would be the title, the description of the data set or its geographic extension.
- Exploration metadata provide information that allow verify that the data are in accordance with the desired purpose, assess their properties or contact with the organization that will provide further information.
- Exploitation metadata include those necessary descriptions for access, transfer, load, interpret and use the data in the final application in order to be exploited.

Another important aspect related to the metadata schemas is their level of detail, which is defined by the choice of the standard itself and the creation of special extensions and profiles. First, the chosen standard defines a more or less large set of elements with different condition: mandatory, optional and mandatory if applicable or conditional. An extension of the standard usually consists on adding new constraints (e.g. conversion of optional elements to mandatory), extension of code lists and the creation of new elements and

⁶ http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020

⁷ <http://www.fgdc.gov/metadata/csdgm>

entities. Some standards such as ISO19115:2003 and CSDGM provide methods for the extension of the metadata within their specification. And if there are a big number of these extra features (they involve the creation of a considerable number of elements), ISO19115:2003 recommends making a formal request for the creation of a specific application profile for that community of users who require it.

However, although the specific profiles and the optional and conditional elements facilitate certain flexibility to the geographic metadata, most of the common used standards like CSDGM and ISO19115:2003 are too complex (Nebert, 2004), both define more than 350 elements distributed into multiple hierarchical sections. This complexity means that, to complete the geographic metadata, it is necessary to devote a big amount of time and highly qualified human resources.

Automatic mechanisms for generating metadata in standard format would be a helpful way to assist user to increase the number of available metadata in distributed environments improving the discovery of the data in an efficient way.

2.3 Generation of Metadata

Metadata is usually created by data providers, generated manually and stored (separated from the resource) in catalogs, according to digital libraries tradition, to be found later for informational purposes. However, practical problems with their creation and maintenance are limiting their effectiveness for tasks such as discovery and evaluation of the usefulness of a given resource.

Some authors emphasize as causes of this low effectiveness the complexity of the rules and standards in the geospatial context or the low automation and synchronization between the creation of data and metadata. In terms of complexity, (Bodoff et al., 2005) regret the overhead of planned uses for some metadata: according to certain rules some metadata must provide at the same time the documentation, the configuration and the access point to the resource. Other authors point out the necessity to automatize data generation, (Bulterman, 2004; Manso et al., 2004).

Nowadays metadata are usually created manually, and only few of them are extracted automatically by software, for example, geographical extent or the date of creation. Although theoretically only must be introduced by hand subjective descriptors such as the abstract, but the complexity and variety of formats limit the application of automated techniques.

Due to the increasing need of metadata to find the great amount of data available in distributed environment, especially in geospatial information systems, being deployed as SDI, there are numerous software applications that try to facilitate this metadata generation. Most of these application started supporting CSDGM standard and ISO 19115. There is a good survey on these applications available in the FGDC metadata working group⁸.

The purpose of this section is to make a small state of the art of existing proposals to improve the automated generation of metadata. The hypothesis is that the automatic generation of meta-information permits decrease human interaction in the creation of metadata, reducing the associated workload and the obstacles arising from the complexity of the metadata schemas that metadata creators must face.

⁸ <http://www.fgdc.gov/metadata/iso-metadata-editor-review>

2.3.1 Methods aimed at the extraction

Actual models of representation of georeferenced information, especially the raster and vector spatial representation models, are characterized by being highly structured and are manifested in multiple exchange formats. Due to the complex nature of digital resources, it is not possible to effectively reuse methods for automatic generation of metadata already existing in the context of information retrieval on textual type documents (e.g. Web browsers). On the other hand, the few existing GIS tools that offer automatic deduction of metadata for raster and vector formats are based on the analysis of these specific formats and the implementation of ad-hoc mechanisms that process the data in these formats to extract information which is used later to populate the metadata elements (Manso et al., 2004).

Among the applications that perform an automatic extraction of metadata from certain geospatial data exchange formats is the free software tool CatMDEdit⁹ (Zarazaga-Soria et al., 2003). As reflected in the work done by (Manso et al., 2004), the amount of information that can be extracted depend fundamentally on the representation model used, and its own file format. In this way, there are elements that can only be extracted from certain types of data and files, while others, such as the size of the data, could be obtained in any circumstances.

Another well-known tool and widely used that includes automatic metadata generation functionality from geographic data is ESRI¹⁰ ArcCatalog, available from version 8 ESRI ArcGIS. This tool allows the automatic loading of a number of basic fields and the synchronized update of data and metadata. To improve this tool have been created some extensions, such as the Metadata Editor of the *Núcleo Español de Metadatos* (NEM), it is a fully integrated tool with the ArcCatalog application, capable of generate a metadata record that meets the standard ISO19115:2003 and NEM v1.0¹¹. Metadata created with this editor will be integrated with the ArcCatalog metadata search functionality, as having been generated by the application (Sanchidrian & Calle, 2005).

Regarding to other data formats, such as text, sound or video documents, content creation software, that is, the range of programs used to create these resources, usually support some automatic metadata generation from the content that they generate. For example, MS Office attaches to the document a title based in the text of its first line, apart from other technical metadata such as dates of creation or modification and the author information. These metadata created by the content creation software are often used by the file system to index and sort the contents. All this kind of metadata can be collected during the creation process (Greenberg et al., 2005), but during the creation of digital content there are other metadata that can be automatic generated, they are usually used in various visualization applications, but not usually taken into account for the description of the resource for future discovery.

A couple of examples of the type of work that is being developed in this area can be the report of (Greenberg et al., 2005) on the generation of metadata for MS Word, Acrobat, Dreamweaver, CityDesk, WinAmp... file formats or the DCS (Dublin Core Services) project, which develops a set of services and applications for the automatic metadata extraction from more than 10 types of digital formats (XML, BibTex, XHTML, PNG, etc.), this project aims to support the development and widespread application of the Dublin Core standard format.

⁹ <http://catmdedit.sourceforge.net>

¹⁰ <http://www.esri.com>

¹¹ <http://www.idee.es/resources/recomendacionesCSG/NEM.pdf>

2.3.2 Methods aimed at the inference

We can infer metadata from other metadata or from geodata, using various techniques of data mining, data recovery, using the context surrounding the data, reasoning techniques and so on. We could know the administrative limits of a geo-spatial data from the knowledge of their bounding box using a gazetteer, or maybe we can infer a more or less adequate abstract from the information of the name, the legend of a layer, its geographical position, etc.

In this sense, (Taussi, 2007) proposes a metadata extraction based on three fundamental steps. The first step consist on apply some metadata extraction techniques largely based on the specific exchange format of the geographic data. Next step is the automatic deduction of the information regarding data quality, using brute force, stochastic or comparison techniques of the analyzed data with other reference data. Finally, the last step to apply is based on the utilization of data mining techniques that lead to obtaining a higher degree of knowledge about the data. Among the proposed data mining techniques, the following can be highlighted (Hand et al., 2001):

- Exploratory data analysis: Goal is to explore data without clear ideas of what we are looking for.
- Descriptive modelling: Idea is to describe all of the data. For example, showing the distribution of the data, partitioning of the data into groups or making models that show relationships between variables.
- Predictive modelling: Goal is to build a model that permits one variable to be predicted from the known values of other variables.
- Discovery methods: These methods are based on pattern detection, and idea here is to identify patterns, rules, outliers or combinations of items that occur frequently in data.
- Retrieval by content: It is based in the comparison of the contents of the dataset according to the pattern of interest to find similar patterns.

A work that tries to take a further step in the induction of metadata from the analysis of data is that developed by (Klien & Lutz, 2005). They propose a method for automatic annotation of geodata that consists of two main steps. In a first step, ontologies are defined from the definition of concepts (eg. floodplain) for a possible dataset. Depending on the spatial relationships that exist and should be verified with a reference dataset, it is checked whether the dataset corresponds to a floodplain, they check their connection to a nearby river, the altitude with respect to this, and if it is a flat terrain. In a second step, existing topological relationships are verified by a spatial processing for each type of relationship included in the concept, and if it meets all the dataset is semantically annotated with the concept. This approach requires a previous readiness to define concepts based on spatial relationships that makes the method is not directly applicable to any set of data. But in any case, it is helpful to specify formally the spatial analysis that allows checking whether a dataset meets certain characteristics, for annotate semantically.

Outside the geographical scope, there are also other works that exploit the idea of extracting information about other resources using techniques related to data mining. In this line we can mention the work done by (Kawtrakul & Yingsaeree, 2005) which provides a framework for extracting metadata from electronic documents, such as text documents or images; the study

of (Day et al., 2007) for extracting metadata of publications from the bibliographic references, or the articles of (Boutell & Luo, 2005) and (Suh & Bederson, 2007) where photographs analysis techniques (based on clusters) are presented to identify, for example, if the photograph was taken during the day or night, in a natural reserve or in a city, inside a place or outside...

2.4 Publication of Georeferenced Resources

The publication is the effect of revealing or expressing to the public some information, that is, the activity of making information available for public view.

In the world of geographic information is widely accepted that publishing means to make available to users some information of the data in a catalog service, as it is driven by Open Geospatial Consortium (OGC)¹² and its standards. However, OGC standards are not the only mechanism for publishing and searching geographic information or georeferenced resources. In the case of entities of medium or small size it might be appropriated to turn to simpler mechanisms that allow the content availability online. Z3950 standard (ISO23950¹³), widely used in digital library environments, includes a GEO5 profile that allows to extract Z3950 metadata as XML whose content is based on FGDC standard.

A more general mechanism, but whose philosophy and operation can be adapted to the field of the georeferenced resources is the Open Archives Initiative (OAI)¹⁴. This initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. OAI has its roots in the open access and institutional repository movements. Over time, however, the work of OAI has expanded to promote broad access to digital resources for eScholarship, eLearning, and eScience.

OAI provides us with the specification Open Archives Initiative Object Reuse and Exchange (OAI-ORE) that defines standards for the description and exchange of aggregations of Web resources. These aggregations, sometimes called compound digital objects, may combine distributed resources with multiple media types including text, images, data, and video. The goal of these standards is to expose the rich content in these aggregations to applications that support authoring, deposit, exchange, visualization, reuse, and preservation. Although a motivating use case for the work is the changing nature of scholarship and scholarly communication, and the need for cyberinfrastructure to support that scholarship, the intent of the effort is to develop standards that generalize across all web-based information including the increasing popular social networks of "Web 2.0".

The specification of standards for the publication of georeferenced information, whose implementation is feasible from a technical and economic point of view, results essential to the progress of technology and services.

¹² <http://www.opengeospatial.org>

¹³ <http://www.loc.gov/z3950/agency>

¹⁴ <http://www.openarchives.org>

3 Proposed Methodology

3.1 Metadata generation

The proposed methodology, to automatically generate complete metadata and of reasonable quality, avoiding as much as possible the participation of the user, is a combination of the methods described before orchestrated efficiently.

Initially, we start by obtaining all relevant information that can be extracted from the data resource itself, for example, the data size or the creation and modification data. Later, we try to extract as much information as possible from the data content. We must emphasize that this is one of the most important sources of information, so we must pay special attention on it. Moreover, how to analyze the resource and the amount of information available will depend entirely on the nature of the resource and the format in which they are represented. By this method we can extract explicit information in the data, for instance, in an email is easy to find information such as the sender, the recipient or the date that it was sent.

Now, we will add the common information pertaining to the creation and the exploitation context. From the creation context we can obtain relevant information as the organization or the company responsible of the data and the theme of the data. We can operate in a similar manner with the exploitation of context, obtaining information such as the theme or the resource quality offered by certain company. All these information can be previously set or revised by the user or automatically obtained exploring the data set and their context.

The next step is to consider collecting information from the process of creation of the data, obviously if it exists. It should be noted that this source of information is "volatile" due to the fact that it is only available at the time that data is created and for this reason we must collect and store all possible information at that moment. We consider that the information that can be obtained from the process of creation of the data is very important and rarely taken into account. By this method we can accurately find out relevant information such as the creation process in order to replicate the results later, costs (computational, temporal, economic, etc.) or the author of the data. In addition to the information that we had commented just now, during the process of creation of the data, a sensor or other measuring mechanism can provide relevant information. Some magnitudes could be measured such as elevation, position or temperature, and incorporate this values to the metadata automatically. We can find a good example of this method in some digital cameras that use it to add, among others, the information that provides their integrated global positioning system (GPS)¹⁵ device to the images in the form of EXIF¹⁶ tags.

Having reached this point we already have a base of information, and applying to it some deductive methods we will try to extend it. One way to deduce new metadata is that an element of metadata is created through a direct correspondence with another existing metadata element. For example, you can get the place name corresponding to the data using the 4 coordinates of their bounding box, using a gazetteer service. Another way to deduce new metadata is based on the calculation of a new element of metadata through a computation process of the data themselves. In this sense there are many lines of investigation open that cover a wide range of possibilities. We can find from different techniques to analyze/process text documents or web pages to find out its main theme or

¹⁵ http://en.wikipedia.org/wiki/Global_Positioning_System

¹⁶ <http://www.exif.org>

keywords, to other techniques that employ the geodata themselves, for example, to determine the province of a town by topological calculations. The last deductive method is the inference of metadata from other existing metadata or from the data content. It represents the best method, in fact in some situations the only one applicable, for the post-hoc creation of metadata, that is, document existing geodata. An example would be to infer the season of the data using the temperature metadata element, perhaps obtained by measurements such as we have explained above, so a rule would establish for a temperature below 15 degrees in Tenerife that we can suppose winter. Today it is obvious that the creation of this inferred metadata overlaps extensively the research fields of data mining and data recovery (Goodchild, 2007).

Finally, we must never forget to offer the user the possibility of modify or introduce the information, although the idea is that users increase their confidence in the methodology in base of the observation of its acceptable results and tend to not participate in the metadata generation process.

While most methods are applicable throughout the lifecycle of data, other methods are only applicable at the moment that data are created. We must pay special attention on them since most times the information that is not collected at that time is lost forever, and some of this information can be essential for a correct description of the resource.

The proposed methodology will allow the improvement of the automatic generation of metadata and their quality, collecting information that is currently ignored, such as that one that is coming from the creation process that can be very important to describe resources properly. Consequently, the result of applying this methodology will obtain more metadata, of higher quality and correction, with reduced participation of the user.

3.2 Metadata publication

Metadata publication is the second step of the proposed methodology. Once the metadata has been generated, users will be assisted in the automatic publication of this metadata. We can give users the possibility of publish these metadata automatically in an integrated way in the workflow. Hopefully, this will lead to increase the amount of published metadata since we are drastically reducing the necessary effort to correctly describe resources (by generating metadata manually) and to publish them.

In our methodology, metadata publication means to publish metadata in a Catalog Service, we will use the CSW¹⁷ protocol, specifically the transactional profile (CSW-T) according to the OpenGIS Catalog Services Specification (Nerbert & Whiteside, 2004). This protocol supports the ability to publish and search collections of descriptive information (metadata) about geospatial data, services and related resources, so it covers our needs perfectly. Furthermore, it had become an OGC standard so it will be widely adopted by GIS applications.

However, we want to explore other ways to publish metadata and resources in order to obtain better levels of the capacity to be found. One way is publish the information resources directly in servers or social networks that support this kind of information using the metadata to document it properly in the server. For example, we can publish maps in MapServer¹⁸, photographs in Flickr or GPS tracks in WikiLoc¹⁹. Other way to explore is put the resources

¹⁷ <http://www.opengeospatial.org/standards/cat>

¹⁸ <http://mapserver.org>

¹⁹ <http://www.wikiloc.com>

available in order to be indexed by Internet search engine's bots, in this way we can try some techniques from simply put the resources available to build an associated KML with the metadata. Other way can be the use of peer-to-peer (P2P)²⁰ (Rüdiger, 2002; Antoniadis & Le Grand, 2007) networks to share data inside an organization network or globally. The main idea is to explore and test these alternatives and others in combination with various levels of metadata generation to measure the capacity to be found of the published resources.

4. Architecture

This section describes the architecture of our proposal to develop this methodology. The following architecture shows the modules and the connections to design an application that we have called *GeoCrawler*.

A crawler is an application that explores the content of a system in a methodical and automated way. This kind of applications is used to build an index the resources found in the system, in basis to the information extracted of each resource in its processing. In this way, *GeoCrawler* will explore the local machine (in the future we can consider to modify it to allow network exploration) and will try to generate metadata of the available geospatial information resources and later publish them according to their respective metadata.

To implement and fulfil the requirements of the proposed automatic metadata generation methodology we have decided to use a three-tier architecture (Eckerson & Wayne, 1995), to place some modules designed to implement the required functionality in each tier. This architecture is a client-server architecture in which the user interface, functional process logic ("business rules"), computer data storage and data access are developed and maintained as independent modules, often on separate platforms.

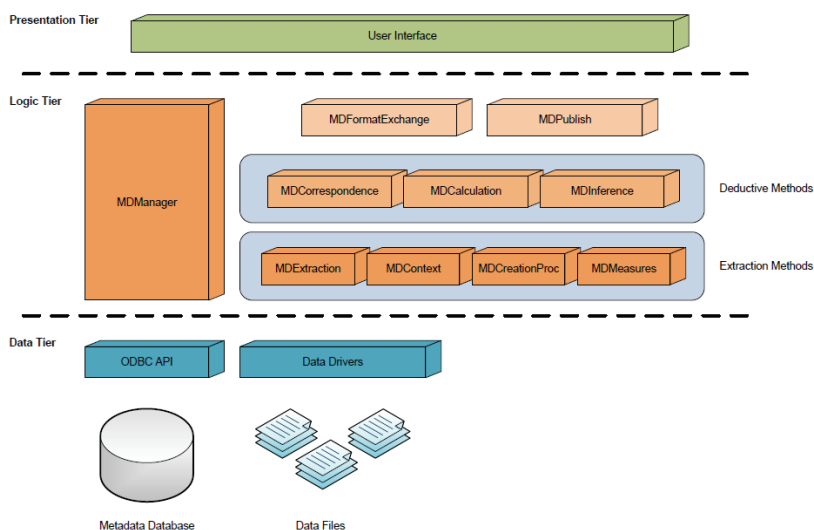


Fig. 1. General Architecture

²⁰ <http://en.wikipedia.org/wiki/Peer-to-peer>

As we can see in Figure 1, at the bottom of the figure and the lowest level of the application is the Data Tier, this tier consists on a database to store the generated metadata and the data files themselves. In this tier the information is stored and retrieved, so it must provide well defined interfaces to manage the data. In our case, this access is provided by data drivers to access to the data files and ODBC²¹ to access to the database. This kind of design, keeps data neutral and independent from business logic, and also improves scalability and performance.

The next tier, which lies just above the Data Tier, is the Logic Tier. It controls the application's functionality by performing detailed processing. In the bottom level of this tier we can find the components implementing the metadata generation methods aimed at extraction. These components correspond to the methods described in the proposed methodology section. Thus, the *MDExtraction* module implements the extraction of all relevant information that can be extracted from the data resource and its content. The *MDContext* module will try to obtain all the information pertaining to the creation and the exploitation context. In a similar manner, the function of the *MDCreationProc* module is collect information from the process of creation of the data. Additionally, the *MDMeasures* module can acquire relevant information from sensors or other measuring mechanisms. At a higher level, based on their previous results, we can find the components implementing the deductive metadata generation methods. These methods will be the deduction of new metadata based on a direct correspondence with another existing metadata element (*MDCorrespondence*), the calculation of a new element of metadata through a computation process (*MDCalculation*) and the inference of metadata (*MDInference*) that includes data mining and data recovery techniques. We have to emphasize that new modules implementing new automatic metadata generation methods could be added.

According to the methodology, and as we can see on the architecture, on the top level of this tier and connecting to the modules which generate metadata, we have the *MDFormatExchange* responsible of generate standard formats and handles the transformation between them. At the same level of this module we have the *MDPublish* module that using, normally the metadata in any standard format, implements the publication business logic, publishing the data in a Catalogue Service or in any other way decided by the user. Finally, in this tier, and covering the whole layer scope, we can see the *Metadata Manager* component whose functionality is to orchestrate the metadata generation efficiently, provide the generated metadata to other components and offer the visible interface to the upper tier.

In the top of the Figure 1 we have the highest application level, where we find the Presentation Tier, this tier displays the information provided by the lower tiers through a graphical user interface. This user interface, moreover, allows users to interact, configure and operate with the application.

This kind of architecture, benefits from the advantages of modularized software containing well-defined module interfaces, it intends to allow any of the three tiers to be upgraded or replaced independently. This is very useful if we want to reuse some components (even the module containing the two lower tiers) and integrate it in other system, to incorporate the functionality of automatic metadata generation and management to any new or existing application.

²¹ http://en.wikipedia.org/wiki/Open_Database_Connectivity

5. Case Study: Metadata Management Platform in gvSIG

We describe next a case of study in which we have implemented a proof of concept for our methodology and architecture. In this sense, we have implemented a prototype of the metadata manager, using the functionality and the extension possibilities that offers an open-source software GIS/SDI client called gvSIG²².

We have extended gvSIG to facilitate, with an integrated workflow, metadata creation, management and publication. This prototype interacts with the gvSIG core to handle the metadata associated to all the resources pointed out to be described with metadata, and provide automatic extraction of explicit metadata from data resources for both internal metadata for user efficiency purposes and external metadata to be catalogued for data discovery in an SDI. In this case, with this integrated solution, we could get lots of information available in the process of data creation. The metadata manager will be working in the background annotating all the metadata while gvSIG users are working with their geospatial data, when it is required the metadata manager will use the implementation of the proposed metadata generation methodology to obtain as much information of the resource as possible, thus, without user interaction. As an added value gvSIG will be using these metadata, as internal metadata to avoid task duplication or recalculations and to visualize the resources properly. On the other hand, when the user wants to share data in a distributed environment like an SDI, he or she can use this metadata to publish metadata. The metadata manager will allow the user to visualize and edit the metadata according to one chosen standard format, and warn about the status of the metadata, for example to fulfil a minimum required set of elements of a certain standard format. Finally, included in this workflow, the prototype includes a user-friendly wizard to guide the user to publish these metadata in a Catalogue Service.

This prototype became available in October 2008 as a pilot plug-in of gvSIG. To sum-up it includes the metadata manager capable of semi-automatic extraction of explicit metadata from data resources for internal use or for being exported to a standard format and/or published in a catalogue service.

In this prototype we support GeoNetwork Opensource²³, as an implementation of the OGC CS-W because is one of the most popular and extended open source Catalog Service implementation. The prototype architecture is shown in the next Figure 2.

²² <http://www.gvsig.gva.es>

²³ <http://geonetwork-opensource.org>

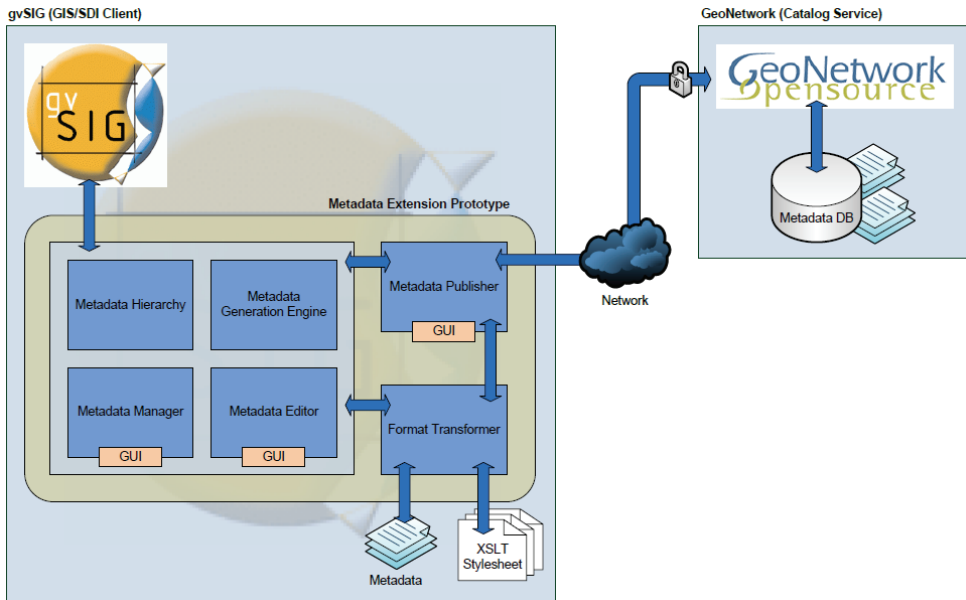


Fig. 2. Metadata Manager Prototype Architecture

As we can see in the Figure 2, in this prototype we define, within the central structure (or core) of gvSIG, an internal metadata dynamic object which would keep all types of metadata associated with their respective resource, as we can see reflected in the *Metadata Hierarchy* module .

The various metadata elements collected will be stored in an XML format file that will be saved together with the data for future uses. When a resource is created, thus it does not have associated metadata yet, the *Metadata Generation Engine* module will be used to generate all the possible metadata according to the proposed metadata generation methodology. In this prototype, we automatically extract so-called explicit metadata of the resource (format, resolution, spatial reference system, creation date, etc.) using the operating system information, and data drivers of gvSIG which are able to read file format headers and other information to collect metadata. As future work we will include inference and information retrieval techniques to create metadata according to the proposed metadata generation methodology, so the user will hardly have to edit or add metadata to publish it in an SDI, thus facilitating the proliferation of metadata and thus the resource discovery in distributed information platforms.

The Figure 3 shows a screen shoot where we can see part of the modules containing GUI (Graphical user interface) shown in the prototype architecture. These modules let the user to visualize and edit these associated metadata by using the metadata editor, he or she can add additional information (such as an adequate title or abstract) that might be required by the standard metadata formats. In this case, when a user wishes to edit the metadata to export it or publish it in a catalogue service, he will choose one of the supported standard formats for this purpose. Once it has been chosen, the metadata manager will start a wizard that will guide the user to view and edit the metadata according to the selected format and validating

the metadata fields. This wizard will guide users to import and export metadata too, validating it according to a standard format to be shared by multiple users without having to publish it in a catalogue service. As we see in the figure the *Metadata Editor* allows users to complete and verify the metadata record according to the selected metadata standard format.

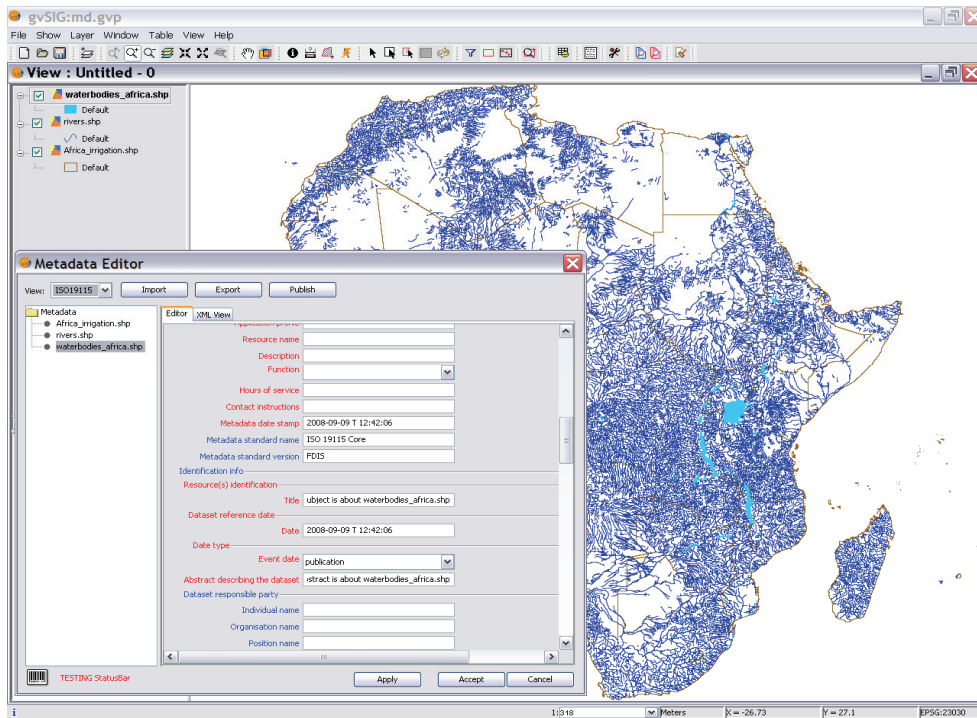


Fig. 3. Screenshot of the Metadata Extension Prototype (Metadata Editor)

As we see in Figure 3. This user interface also links with the *Metadata Publisher* module that will assist the user with a Publish wizard to publish this metadata in a Catalog Service to share the data in an SDI. In the next figure we can observe a screenshot of this Publication Wizard after having finished successfully.

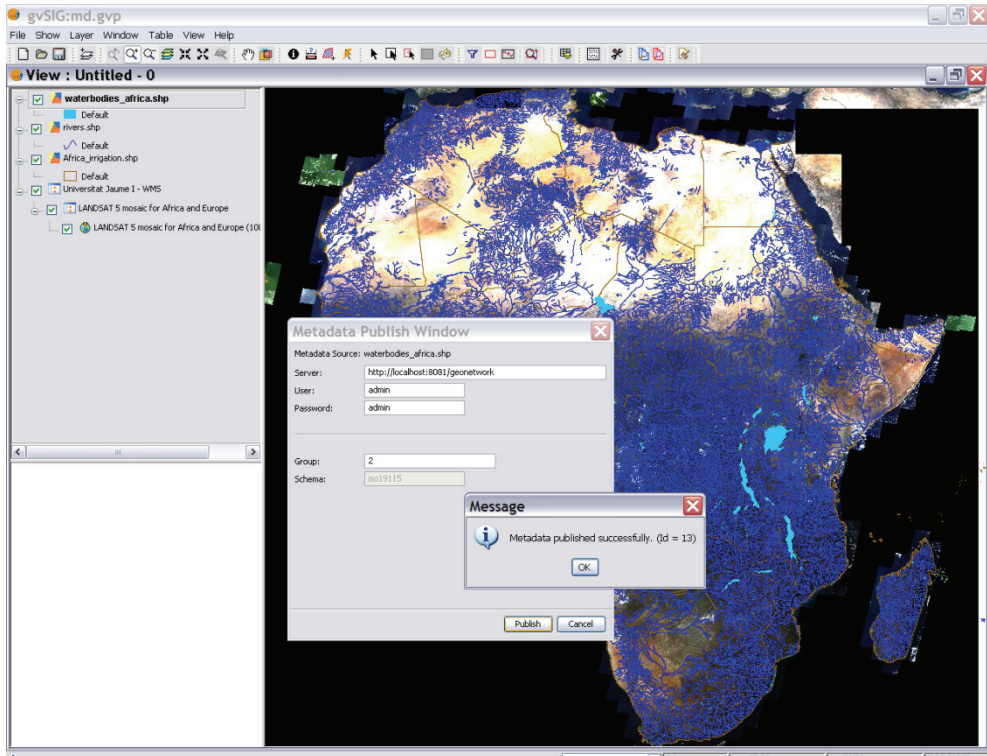


Fig. 4. Screenshot of the Metadata Extension Prototype (Publication Wizard)

Given the above information, let us look at a typical use case. A technician using gvSIG has combined basic geospatial data including terrain data such as slope and aspect, with vegetation data, to create a rough forest fire risk map. Assuming he or she has permission to share this new dataset, he then undergoes the process of publishing the risk map to a map server, and would also like to (or should be required to) publish its description to a metadata catalogue service such as that currently available at the European Commission INSPIRE Geoportal²⁴.

In our use case, the resulting dataset, risk map, will have associated a metadata object that will be created by the process described above. The final step in the workflow is when the user decides to publish the metadata record to a catalogue service the metadata manager checks the validity of the resource associated metadata, the validation will depend on the metadata standard that has been chosen to publish, thus the standards that the Catalogue Service supports. If the metadata conform to minimum requirements according to the selected output metadata standard format, then the metadata manager uses stylesheets to generate an XML string compatible with the catalogue service and carries out the pertinent interaction with the server to establish the connection and publish the metadata.

²⁴ <http://www.inspire-geoportal.eu>

This prototype is only capable of working with shapefile²⁵ vector layer file format. The implemented and supported metadata standard format is the core of the ISO19115:2003 standard. The implementation of the transformation templates for this format has been made based on the standard specification document published by ISO (ISO/FDIS19115). However, its architecture has been designed to support all the desired functionality. So, somehow, this is a proof of the concept of the complete functionality that will be captured in the future Metadata Extension.

Using this integrated solution, the user can close the life cycle of metadata (Baca, 2008) within the same application. So we could create, modify and publish metadata using the metadata manager, and later discover and recover metadata using the integrated catalog client in gvSIG that allow us to recover the linked data from the Catalog Service.

6. Conclusions and Future Work

Metadata descriptions are critical to enhance the discovery, access and use of GI data and therefore are a key element in achieving good data integration and smooth functioning of Spatial Data Infrastructures, as a basic infrastructure to discover, share and use heterogeneous GI data. This points out the need to facilitate metadata production to easily create, with minimal user intervention, metadata descriptions in standard formats.

The presented methodology includes mechanisms capable of automatic generation and publication of metadata in Open Catalogues as means of improving geospatial information sharing in distributed environments like SDI.

As a proof of concept the implemented prototype allows the extraction of explicit metadata from data resources to be catalogued for data discovery in a Spatial Data Infrastructure.

This implementation of the concept of semiautomatic extraction and management of metadata facilitates the creation and edition of images and geospatial data to be published in a Spatial Data Infrastructure. The integrated nature of this solution within the user workflow hopefully will lead to a proliferation of metadata creation, thus improving the functionality and value of SDIs. Furthermore, this development completely supports the philosophy of total integration of data and metadata that we are trying to promote in order to all data generated are easily found and accessible.

In the early future we will complete the development of the metadata manager for documenting well-known imagery and cartographic data sources. The work includes document more types of resources and file formats, add new standard formats and expand the possibilities of publication, but the major effort will be done to continue implementing and improving each of the methods that compose the proposed methodology to automatically generate metadata included in the *Metadata Generation Engine*. Furthermore, this metadata generation engine is a generic approach, so it may be extended to include new data types and multimedia content.

As we had said, the automatic metadata generation methodology includes more intelligent methods to extract metadata by using inferential reasoning techniques from other metadata and data associated. Intuitive extraction of intrinsic (context-based) metadata of the data source in Google-like techniques, including deductive methods to create well formed free text.

²⁵ <http://en.wikipedia.org/wiki/Shapefile>

Another interesting future development is the *GeoCrawler*, a massive metadata generation application, that using crawler techniques and the proposed automatic metadata generation methodology, will allow us to automatically describe the resources available in old data collections currently without documenting, or simply in the user local machine. Subsequently, these data may be published or indexed with respect to the information contained in their metadata to be easily found and accessible by other users. We also consider very interesting the possibility of use this kind of crawler applications in user's local machines, allowing them to share their multimedia resources automatically, for example, in the current social networks.

On the other hand, we will continue to investigate and develop new techniques that allow us the complete integration of data and their metadata. This will greatly facilitate the management, reuse and sharing of resources. Additionally, we will explore other lines of investigation about georeferenced resources publication, for example, the use of indexing techniques that allows us to find the data using simple metadata sets, rather than creating complex formats such as those stored in the current catalogs.

7. Acknowledgements

This work was partially funded by the project "CENIT España Virtual", funded by the CDTI in the program "Ingenio 2010" through *Centro Nacional de Información Geográfica* (CNIG).

8. References

- Antoniadis, P., Le Grand, B., 2007. Incentives for resource sharing in self-organized communities: From economics to social psychology. In *Digital Information Management*, 2007. ICDIM '07
- Baca, M., 2008. "Introduction to Metadata: Pathways to Digital Information (version 3.0)". In Getty Research Institute.
- Bodoff, D., Hung, P.C.K, Ben-Menchem, M., 2005. Web metadata standards: observations and prescriptions. In *IEEE Software*, January-February 2005, pp. 78-85.
- Boutell M., Luo J. , 2005. Beyond pixels: Exploiting camera metadata for photo classification. In *Pattern Recognition*, v. 38, n. 6, pp. 935-946.
- Bray, T., Paoli, J., Sperberg-McQueen, C.M., Maler, E., 2000. Extensible Markup Language (XML) 1.0 (Second Edition). W3C Recommendation 6 October 2000. <http://www.w3.org/TR/2000/REC-xml-20001006>. (last accessed in July 2009)
- Bulterman D. Is it Time for a Moratorium on Metadata? *IEEE Multimedia*, October-December (2004) 10-17.
- Cantan-Casbas, O., López-Pellicer, F. J., Noguerras-Iso, J., Zarazaga-Soria, F. J. 2008. Issues hampering the widespread adoption of catalogues based on the OGC Catalogue Services Specification. In *Computers & Geoscience* 2008.
- Cerda, D. (2005). El mundo según Google: Google Earth y la creación del dispositivo GeoSemántico global. <http://geosemantica.earth.googlepages.com> (last accessed in July 2009)
- Craglia, M., Kanellopoulos, I., Smits, P. Metadata: where we are now, and where we should be going. Proceedings of 10th AGILE International Conference on Geographic Information Science 2007. Aalborg University, Denmark

- Day, M., Tzong-Han Tsai, R., Sung, C., Hsieh, C., Lee, C., Wu, S., Wu, K., Ong, C., Hsu, W., 2007. Reference metadata extraction using a hierarchical knowledge representation framework. In *Decision Support Systems*, v. 43, pp. 152-167.
- Díaz, L., Martín, C., Gould, M., Granell, C., Manso, M.A. Semi-automatic Metadata Extraction from Imagery and Cartographic data, International Geoscience and Remote Sensing Symposium (IGARSS 2007). Barcelona, Julio 2007. IEEE CS Press, pp. 3051-3052.
- Eckerson, Wayne W., 1995. "Three Tier Client/Server Architecture: Achieving Scalability, Performance, and Efficiency in Client Server Applications." *Open Information Systems* 10, 1 (January 1995): 3(20)
- FGDC, 2000. Content Standard for Digital Geospatial Metadata Workbook, version 2.0. Federal Geographic Data Committee (FGDC), Metadata Ad Hoc Working Group.
- Goodchild, M. (2008). Assertion and authority: the science of user-generated geographic content. <http://www.geog.ucsb.edu/%7Egood/papers/454.pdf> (last accessed in July 2009)
- Goodchild, M. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal* 69 (4): 10. 0343-2521.
- Greenberg, J., Spurgin, K., Crystal, A., 2005. Final Report for the AMeGA (Automatic Metadata Generation Applications) Project. UNC, School of Information and Library Science University of North Carolina.
- Hand D., Mannila H., Smyth P, 2001. *Principles of Data Mining*, Cambridge. The MIT Press.
- Hill, L. (2006). Georeferencing. In The MIT Press. ISBN 0-262-08354-6.
- Kawtrakul, A., Yingsaeree, C.A., 2005. Unified Framework for Automatic Metadata Extraction from Electronic Document. In *Proceedings of IADLC2005 (The International Advanced Digital Library Conference)*, pp. 71-77, Nagoya, Japan.
- Klien, E., Lutz, M., 2005. The Role of Spatial Relations in Automating the Semantic Annotation of Geodata. In *Proceedings of the Conference of Spatial Information Theory (COSIT'05)*, Lecture Notes in Computer Science, v. 3693, pp. 133-148, Ellicottville, NY, USA.
- Manola F, Miller E, (eds) (2004). *RDF Primer*. W3C, W3C Recommendation 10 February 2004. <http://www.w3.org/TR/2004/REC-rdf-primer-20040210>. (last accessed in July 2009)
- Manso, M.A., Noguerras-Iso, J., Bernabé, M.A., Zarazaga-Soria, F, 2004. Automatic metadata extraction from geographic information. In *Proceedings of the 7th AGILE conference on Geographic Information Science*, pp. 379-385, Heraklion, Greece.
- Masser, I., Rajabifard, A., Williamson, I. Spatially enabling governments through SDI implementation. *International Journal of Geographical Information Science*. Vol. 22, No. 1, (2008) 5-20
- Nebert, D., 2004. Developing Spatial Data Infrastructures: The SDI Cookbook v.2.0. In *Global Spatial Data Infrastructure (GSDI)*. <http://www.gsdi.org/gsdicookbookindex.asp> (last accessed in July 2009)
- Nebert, D., Whiteside, A., 2004. OpenGIS - catalogue services specification (version 2.0). OpenGIS Project Document 04-021r2, Open GIS Consortium Inc.
- Noguerras-Iso, J., Zarazaga-Soria, F.J., Béjar, R., Álvarez, P.J., Muro-Medrano, P.R. OGC Catalog Services: a Key element for the development of Spatial Data Infrastructures, *Computers and Geosciences*, vol. 31/2, (2005) 199-209.

- Rajabifard, A., Feeney, M-E.F., Williamson, I. P. Future directions for SDI development. *International Journal of Applied Earth Observation and Geoinformation* 4 (2002) 11-22
- Rüdiger Schollmeier, 2002. A Definition of Peer-to-Peer Networking for the Classification of Peer-to-Peer Architectures and Applications. In *Proceedings of the First International Conference on Peer-to-Peer Computing, IEEE*.
- Sanchidrian Cano, N., Calle González, J.V., 2005. Editor de Metadatos NEM v 1.0 para ArcGis. In *Jornadas Técnicas de la IDEE de España (JIDEE05)*, Madrid.
- Stewart, C and Kowaltzke, A. 1997, *Media: New Ways and Meanings* (second edition), JACARANDa, Milton, Sydney. pp. 102.
- Suh, B., Bederson, B.B., 2007. Semi-Automatic Photo Annotation Strategies Using Event Based Clustering and Clothing Based Person Recognition. In *Interacting With Computers*, v. 19, n. 4, pp. 524-544. Elsevier.
- Taussi, M., 2007. Automatic production of metadata out of geographic datasets (master's thesis). University of Technology, Department of Surveying. Helsinki, Espoo. http://www.tkk.fi/Units/Cartography/theses/master/2007/Diplomityo_Taussi_M.pdf (last accessed in July 2009)
- Toebe John, 2007. Enabling a Richer Video Experience With Metadata. A position paper for the W3C Video on the Web Workshop. 12-13 December 2007, San Jose, California and Brussels, Belgium. Chief Architect, Cisco Media Solutions Group. Available in http://www.w3.org/2007/08/video/positions/Cisco_MSG.html (last accessed in July 2009)
- Zarazaga-Soria, F.J. Lacasta, J., Noguera-Isso, J., Torres, M.P., Muro-Medrano, P.R., 2003. A Java Tool for Creating ISO/FGDC Geographic Metadata. In *Geodaten- und Geodienste-Infrastrukturen - von der Forschung zur praktischen Anwendung. Beiträge zu den Münsteraner GI-Tagen*. IfGI prints. 2003, vol. 18, pp. 17-30.

Application of Real Time GIS, Remote Sensing and IC Tag for Realization of Geospatial Information Society

Shikada Masaaki*, Takeuchi Sayaka*, Shimano Sota**
and Moriya Mitoshi***
Kanazawa Institute of Technology (KIT)*
KOKUSAI KOGYO CO., LTD.**
ASIA AIR SURVEY CO., LTD.***
JAPAN

1. Introduction

Japan is now experiencing an aging society and every person should be safe and feel relieved. The Japanese Government executed a new law NSDI for a spatial information society on May 30, 2007. (NSDI: National Spatial Data Infrastructure) The society needs to obtain absolute position for realizing seamless positioning by ubiquitous network technology. However, the technology has not been established yet. An experiment was performed on whether Real-Time GIS (Figure 1), GPS, and the IC tag could obtain the absolute position.

The research is to confirm whether absolute positions can be obtained accurately by Real-Time Kinematic-GPS (RTK-GPS), Virtual Reference System-GPS (VRS-GPS), and Differential-GPS (D-GPS). In addition, Integrated Circuit Tag (IC tag) was used where GPS signals could not be received to obtain information on the absolute position. The IC tag is used in distribution systems, but the method for using geoinformatics has not been established yet. The experiment was conducted to verify the reading rate of IC tag on different types and conditions. The kinds of the IC tag are passive and active types. For example, the IC tag has many advantages of transmitting and receiving the information, and obtaining the absolute position without any contact. A passive and active IC tags made of different materials were experimented to verify the reading rate at the outside and inside of a laboratory. As a result, passive type IC tag become accurate, however, the active type is in a stage of growth because it is no clear method for using by various affect. Therefore it is necessary to do additional experiment of indoor positioning.

Second big purpose of the research is to establish Universal Map (UM). The basis of UM is a surveying and it consists of geoinformatics which is the latest survey technology. Anyone can use UM regardless of the physically challenged, healthy person, age, man and woman. Figure 2 shows the essential and minimum requirements of UM and the concept of UM.UM means the newest map which can display the latest road condition and other information on

a mobile PC. This map includes the recent information by which all users can understand the condition of absolute position anytime, anywhere. Additionally, measures to utilize spatial information become extremely important and the society might need UM with a lot of advantages in the future. Our research suggested that we have to establish a method for acquiring seamless positioning information in the advanced spatial information society by using UM.

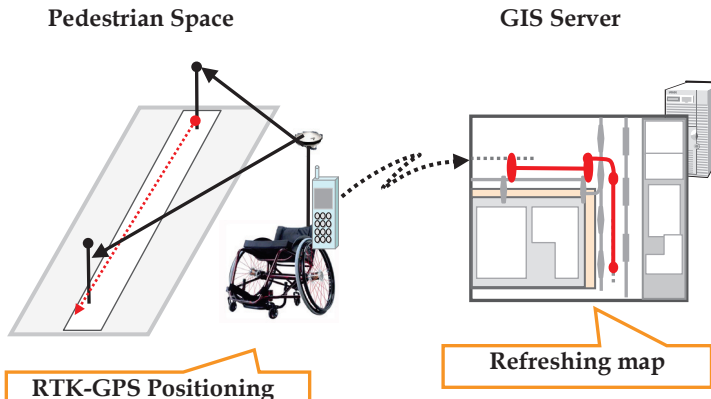


Fig. 1. Concept of REAL TIME GIS

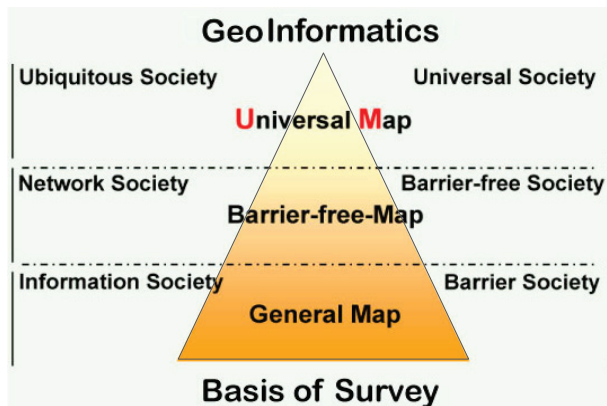


Fig. 2. Concept of Universal Map

2. Background of Study

“The map should be fresh” is a big concept of our laboratory. One of the backgrounds of study is to establish a method for updating a large-scale digital map for local government using a Real-Time GIS (Figure 1). The Real-Time GIS which was defined by our laboratory can be used to renew the new BM. The Real-Time GIS is a technique that updates the new BM instantly by the Real-Time Kinematic Global Positioning System (RTK-GPS), GIS, and mobile phones.

Japan has been adopting a new standard for map geometry since April 1, 2002. Ellipsoid of a new geodetic system in Japan is almost equal to WGS-84 of GPS but most of the Digital Map (DM) of local government is still Tokyo Datum of an old geodetic system. To correspond with two kinds of data which have different geodetic systems, it is necessary to transform coordinates.

On the other hand, much local government has been utilizing a large scale (1/500 or 1/1000) DM with GIS. Government promulgated the law of National Spatial Data Infrastructure (NSDI) to construct the advanced spatial information society May 2007. GIS will be able to efficiently help many workers who are managers and city planners in government and researchers. As an example, it is possible to improve the service to a citizen including elderly people and the physically challenged by sharing those data in local government. However, the maintenance and renewal of UM database need much labor and time and updating a map has not been established yet and, there are only a few successful examples. This is a specific problem of a large scale map to achieve the spatial information society now.

In the master's thesis of Ms. Naoko Matsuda who graduated 2003 from Kanazawa Institute of Technology (KIT), these problems were solved by using Real Time GIS and the achieved result is listed below.

- (1) Position data of latitude and longitude had a high accuracy within 3 cm.
- (2) It is difficult to acquire high-accuracy data because geoids may influence the accuracy of height.
- (3) Tracks were not displayed well though she tried to display tracks which moved by using a RTK-GPS in GIS because canopies interrupt wavelengths from satellites.

As a result, at that stage, it was very difficult to solve problems by using a GPS only. Interruption of signals in course includes very important problems. If anyone is able to receive the positioning data ubiquitously, people will obtain safe, relieved, and comfortable service.

The purpose of this study at the first stage was to establish a method for updating a large scale map for local government, and to propose UM by using RS, GPS, and GIS. On the second stage, we adopted new concepts of collaboration of geoinformatics technologies. On the present stage, the important purpose of study is how to acquire the accurate positioning information without interrupting satellites information.

3. Experiment by GPS Positioning

An experiment was performed inside the KIT campus at Kanazawa district in central Japan by using a D-GPS and a VRS-GPS. The purpose of the experiment is to acquire positional data accurately and to understand the feature of places where accuracy is poor. Accuracy of positioning D-GPS and VRS-GPS was verified.

Figure 3 shows the experiment field and the route. Background of Figure 3 is Base Map of the Nonoichi town on a scale of 1/2500 (Accept: Nonoichi town office). ArcGIS 9.1 which is one of the general software of GIS was used to display the map and analyze the data.

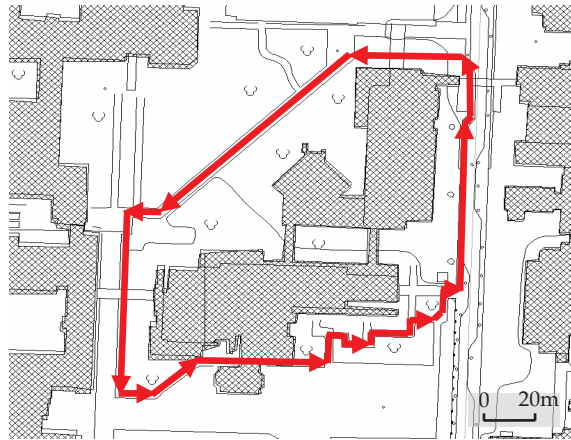


Fig. 3. Experiment field and route in KIT campus

3.1 RTK-GPS Positioning

In this paper, RTK-GPS means that a reference station and a rover station were required for real-time GPS positioning. The reference station was made just on a control point. This control point is a leveling point which has the absolute position because the accurate positioning data by using static positioning were needed. The specified low power radio broadcast is used on the communication from the reference station to the rover station for the RTK-GPS receiver.

The experiment was conducted by moving around by a wheelchair which is equipped with RTK-GPS at uniform speed inside KIT campus. Additionally, we have carried out the experiment by holding the RTK-GPS equipment with hands. In that condition, we could not obtain positioning data because a GPS antenna was swung on moving. This is the reason why a GPS was attached to the wheelchair or a hand truck for reducing an error. However, many errors happened in the most of fields by using a RTK-GPS. One of the reasons for causing errors was the buildings between the positioning points because the communication area of radio broadcasts was narrow in such a situation.

3.2 VRS-GPS Positioning

A second experiment was performed by using VRS-GPS that is a kind of RTK-GPS positioning. In VRS-GPS, the reference station needs not be set. A virtual reference station was made virtually around the positioning point. The distance of a virtual point to an actual point is about 3m to 5m. A rover station received correction information from a mobile phone by using a wireless system. This system consists of GPS-based control stations. GSI made it about 1200 stations in Japan. In the wireless system, only one person is capable of positioning with a light baggage.



Fig. 4. Positioning of D-GPS and VRS-GPS

The method of measurement is similar to that of RTK-GPS. Therefore, VRS-GPS has more advantages. More specifically, it allows only one person to make real-time positioning, has a simple configuration, and has high accuracy.

3.3 D-GPS Positioning

D-GPS positioning is a method that sends the corrected value of a pseudo distance from each satellite, and calculates precision again in the rover station. Additionally, a rover station received correction data from a mobile phone by using GPS-based control station. Position accuracy is from 0.5m to 2.0m. This system is similar to the technique of VRS-GPS and it only needs a rover station. Moreover, the positioning cost is cheaper than that of VRS-GPS because of its simple system. The experiment performed by using a wheelchair with D-GPS and VRS-GPS simultaneously. The reason is to make the experimental environment almost the same and to acquire position data. Figure 4 shows the positioning of D-GPS and VRS-GPS by a wheelchair.

4. Verification of Overlapping

Positioning data of D-GPS and VRS-GPS were displayed on the DM (Base Map of the Nonoichi town on a scale of 1/2500) by using GIS software (Figure 5). Background of Figure 5 is an aerial photograph.

4.1 Experiment Result

As a result of experiments, GPS was able to receive high-accuracy data at almost all places. However, the data were not accurate at any places though the measured place had wide

open sky. It seemed that other reasons affect a receiver. Data were intercepted at three spots, and a lot of measurement errors occurred there.

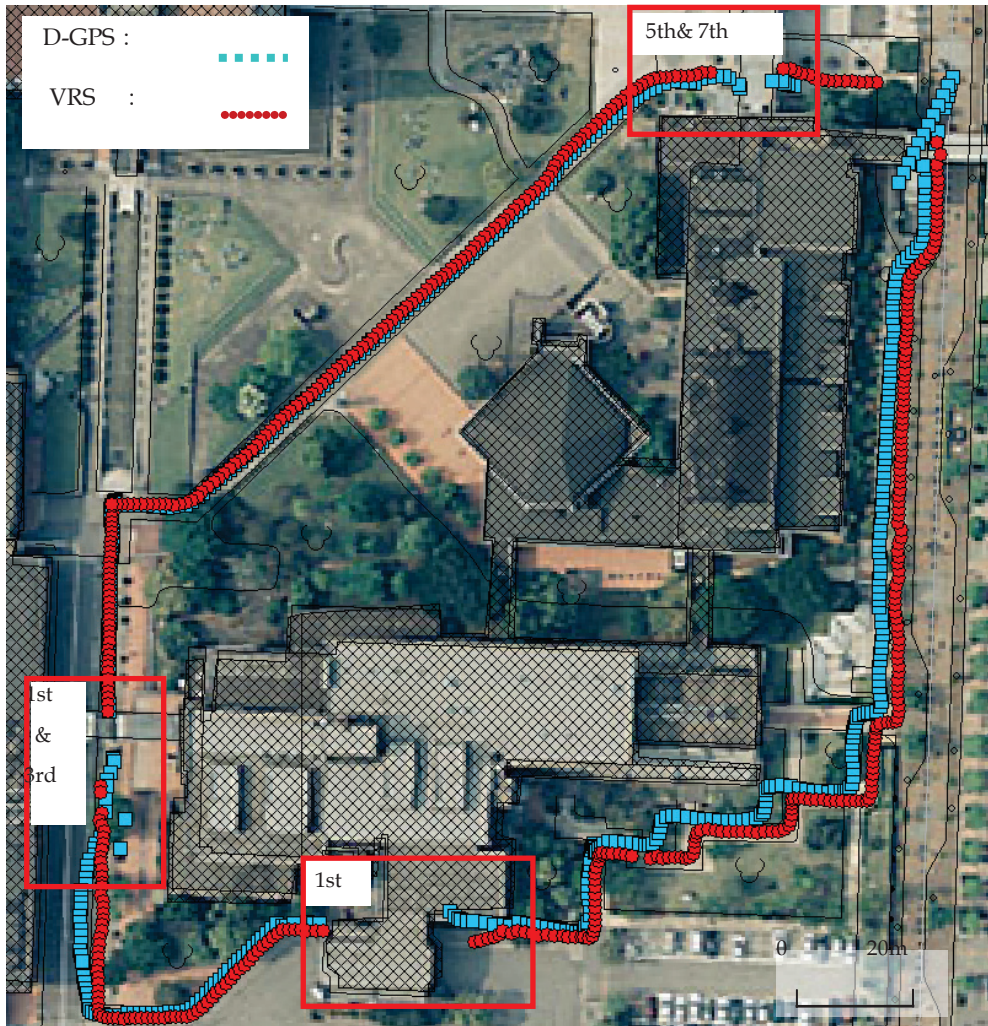


Fig. 5. Overlapping of experiment data

In the next stage, we confirmed how long a signal is received around a canopy by a D-GPS and a VRS-GPS. Figure 6 (a), (b) and (c) shows one of enlarged canopy areas in Figure 5.

Table 1 shows the length of interruption of raw data that were displayed on GIS. Table 2 shows the length of interruption of analytical data that include positional accuracy and error. For example, we considered dilution of precision (DOP), number of satellites, and standard deviation.



Fig. 6 .(a) Verification around each canopy (Front stoop of 1st building)



Fig. 6 .(b) Verification around each canopy (between 1st and 3rd building)

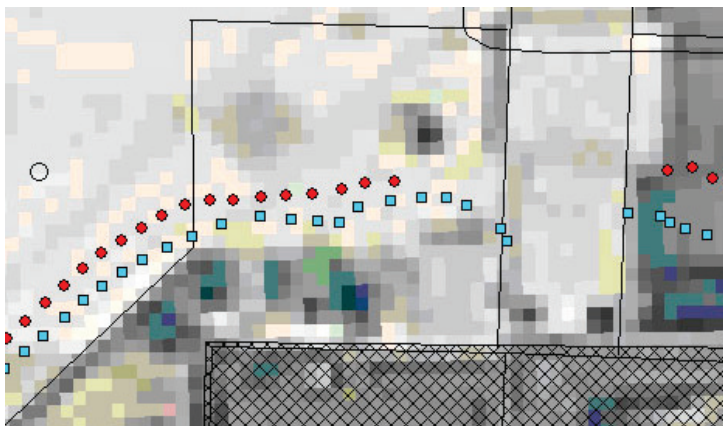


Fig. 6 .(c) Verification around each canopy (Between 5th and 7th building)

Building No.	Length of interrupted tracks	
	D-GPS(m)	VRS-GPS(m)
1st	4.624	7.908
1st and 3rd	5.119	7.396
5th and 7th	2.374	5.419

Table 1. Length of interruption in raw data

Building No.	Length of interrupted tracks	
	D-GPS(m)	VRS-GPS(m)
1st	15.313	12.574
1st and 3rd	null	12.130
5th and 7th	12.952	5.419

Table 2. Length of interruption in analytical data

4.2 Consideration

Important condition of GPS positioning is to receive 4 satellites or more at open sky. As a result, I acquired the absolute position accurately at open sky. However GPS receiver was not able to receive signals from GPS satellites at an area surrounded by canopies and buildings. Additionally, one could not obtain continuous GPS signals at districts overgrown with trees. In such a place, correction data received by a mobile phone might not give acceptable data. An area surrounded by canopies and buildings causes that cycle slip and multipath to badly influence the DOP and GPS signals. If GPS positioning is conducted near buildings, it is necessary to consider satellites situation and multi-path. Because those areas had poor signal conditions, remarkable differences were seen between D-GPS and VRS-GPS. Interrupted signals of the D-GPS were shorter than those of the VRS-GPS as shown in Table 1, but VRS-GPS showed higher accuracy than that of D-GPS as shown in Table 2. As for the reasons for difference, VRS-GPS have the problem of initialization and D-GPS of simple system don't have one. Therefore, it appears D-GPS had high continuousness and VRS-GPS had reliability of positioning accuracy.

In the next chapter, we will show how to obtain an absolute position and other information at a place where the GPS signal does not reach.

5. Solution of Problems by IC Tag and Geoinformatics

Pedestrian space will become safer and securer if people are able to receive absolute positions and other information by future spatial information technology. In this technology, we will adopt an Integrated Circuit tag (IC tag) to assist people including the physically challenged. Major advantages of IC tag are listed below.

The IC tag can

- Receive the data without contact.
- Memorize a lot of data and be very small size.
- Easily add information and update data.
- Have high durability is better than that of its paper type.

If IC tag's powerful advantage is used fully, everyone will be able to receive the positioning data anytime and anywhere. For example, GPS positioning is used at open sky, and IC tags are used at closed sky, which are good ideas because one can continuously receive the signal from satellites and IC tags. In this system, the positioning information will be imagined ubiquitously (Figure 7).

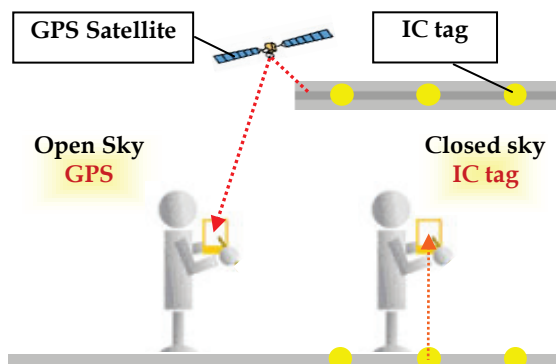


Fig. 7. Utilization of IC tag

6. Experiment by IC Tags for Spatial Information Society

Preliminary experiments were performed by using reader of middle-range and several IC tags which frequency band is 13.56 MHz. The purpose of experiments is to obtain basic data of IC tags for realization of seamless positioning. The experiment was conducted by using hand truck which attached the IC tag reader. IC tags were lineally set out and Unique Identifier (UID) of them was read by moving on the straight line (Figure 8). Additionally, the ratio of UID reading was indicated that how many pieces are able to read among 1000 tags.

As a result of experiments, the maximum reading distance was about 40 cm though the specification of it is 50 cm. Therefore, the height of IC tag's reader fixed from 15cm to 40 cm in consideration for the condition of hand truck. After the experiments, we found that the error factor of IC tags and good measurement environment of reading rate.



Fig. 8. Verification experiment of IC tag

Several experiments were conducted to confirm influence of reading rate. Each condition for research and experimental overview are listed below.

- (1)Material of the ground: It can easily influence that the metal, the water and other materials for to IC tag. First experiment was performed at any places such as cement concrete, asphalt concrete, earthenware tile and fireclay brick.
- (2)Interval of IC tags: As a result of preliminary experiment, IC tag could not be read when the interval was too narrow. Second experiment has the space in the interval of IC tags.
- (3)Material and size of IC tags: Material and antenna size of IC tag is variety. Third experiment was performed by using ceramics type of IC tags which has high durability, general card type and paper type.

6.1 Experiment Result

It was confirmed that IC tags be influenced by the differences the interval of setting and the material on the reading rate. Figure 9 (a) and (b) shows the results of each research as mentioned above.

As a result, the reading rate of IC tags has decreased greatly by 30cm or more on cement concrete. The reading rate at other places is higher than the concrete area (Figure 9 (a)). Additionally, the reading rate was rapidly decreased when the height is above 25 cm (Figure 9 (b)). Furthermore, the reader could not read UID in ceramics type of one when height is more than 20 cm. The reading rate has decreased from 25 cm up in the height of the reader from all results.

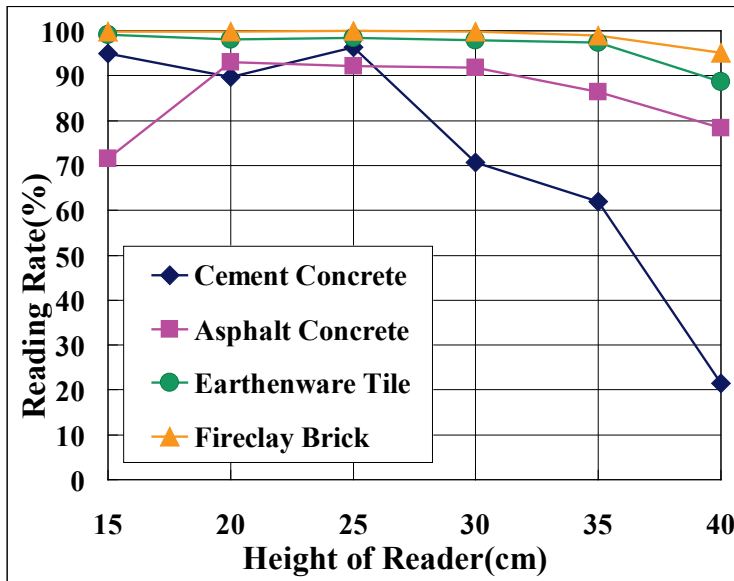


Fig. 9. (a) .Results of IC tag experiment (Reading rate by material of the ground)

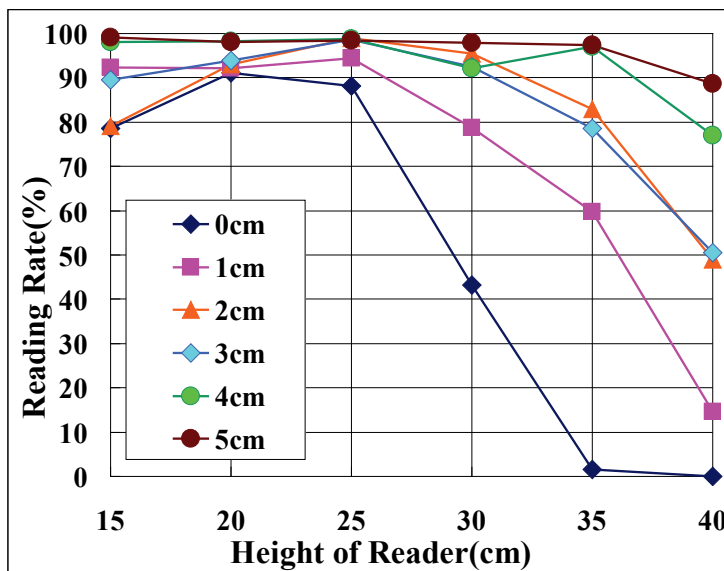


Fig. 9. (b) .Results of IC tag experiment (Reading rate by interval of IC tags)

6.2 Consideration

These results show that the decrease of reading rate was influenced by the moisture included in cement concrete. Installation interval of IC tags should leave space from about 5 cm to 10 cm to avoid anti-collision. Reading rate of ceramics type has narrow area because the ceramics type of IC tag uses alumina to increase durability and is thicker than other kinds of one. And IC tag which has small size of the antenna was low reading rate. Therefore, IC tag should be selected a large size as much as possible and the material should be considered of the environmental condition.

An appropriate reading distance was 20 cm depending on the material of IC tag. However, it is necessary to perform more detailed experiment under various conditions for the realization of proposed UM in my research.

7. VERIFICATION EXPERIMENT OF IC TAG

7.1 Read Experiment by Shielding Material

(1) Experimental Overview

It is necessary to clarify that the height of IC tag's reader and conditions of use. An experiment was performed where the IC tag was buried under the shielding material and was read when reader move through the material. The experiment was conducted by a hand truck to which the IC tags reader was attached. The UID of the IC tag was recognized and the reading rate was investigated. The reading rate was expressed for 1,000 IC tags. Figure 10. shows read experiment by shielding material on wood.



Fig. 10. Read experiment by shielding material (wood)

(2) Equipment in Use

Made by Welcat Inc.

IC tag's reader: EFG-400-01

An antenna of exclusive IC tag's reader writer: ANU-100-01

IC tag: card type(ISO15693, 13.56MHz)

(3) Setting conditions

(a) Used shielding material and its thickness

Wood: 3, 6, 9cm

Concrete: 6, 12cm

Soil: 5, 10, 15cm

Shielding materials made of wood, concrete, and soil were used. These materials are used for general buildings and roads. The thickness of the shielding material has not been unified acquisition conditions

(b)The height of reader

The height of IC tag's reader was set to 15, 20, 25, and 30cm. The reason was shown by Mr. Shimano Co. Author who graduated from Kanazawa Institute of Technology (K.I.T.) where the reading rate was high when the height of the reader ranged form 15cm to 30cm in his research. In this experiment, the height was set to a maximum of 30cm and a minimum of 15cm according to the result.

(c) Setting intervals

A setting interval of the IC tag is 10cm because it was the best interval by his research.

(d)The kinds of IC tags used

In the experiment, passive type was used.

7.2 Result and Considerations

As the thickness of shielding materials is increased, the reading rates tend to decrease. However, each material shows a high reading rate. Therefore, this height is the best suited. The following shows appropriate conditions obtained by the experiment.

- (a) Interval of passive type: Over 10cm
- (b) Size of passive type used: Large size
- (c) Moving speed to read: Normal walking speed
- (d) Height of reader: below 15cm
- (e) Thickness of shielding materials: below 10cm

7.3 Indoor Positioning Experiment

An experiment was used the IC tag of active type. An advance of this type is to transmit radio waves at regular intervals automatically, detect and specify IDs existing in a wide maximum range of 20m. And it can control information as a person, an object, a position, time, and condition at real time. An experiment was investigated to discuss a method for setting active tag in a room and how to set in environment. By changing material and the height, a change in the member of need times and RSSI was verified. RSSI means a sensitivity to receive tag.

(1) Equipment in use

Made by Kyusyu Ten Co.

Wireless reader: TGS-R300W

Wireless tag: YGS-T300

Wireless router: WIN-G54/R4-M(Made by I·O DATA Co.)

(2) Setting conditions

- (a) The height of reader: 240cm

The height of reader is shown from the floor to the ceiling in experimentation area. It easy to receive the electric wave from a tag by reader is attached to the ceiling.

- (b) The distance from a tag to a reader: 10m
- (c) The time to read tag: 60 seconds
- (d) Interval time of electric wave automatic transmission: 3 seconds
- (e) Attached the material: person, wood and iron

According to statistics, an average height of Japanese is 170cm for the past 5 years. So the height of the tag is set 170cm at the maximum.

(3) Experimental Overview

An experiment was performed where a wireless tag is attached to wood, iron, and people. By changing material and the height, a change in the member of need times and RSSI was verified. The active tag can automatically received control position information from attached object. Therefore, it was assumed that a tag was attached to a human body in a basis pattern. Wood and iron were used in comparison with a person from the result of experiment.

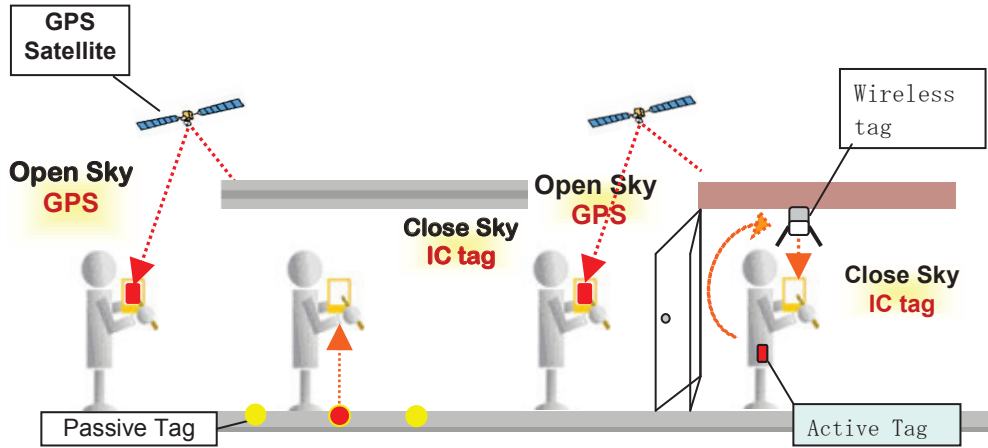


Fig. 11. Concept of indoor and outdoor seamless survey



Fig. 12. Outdoor and indoor experiments



Fig. 13. IC Tags of Passive type

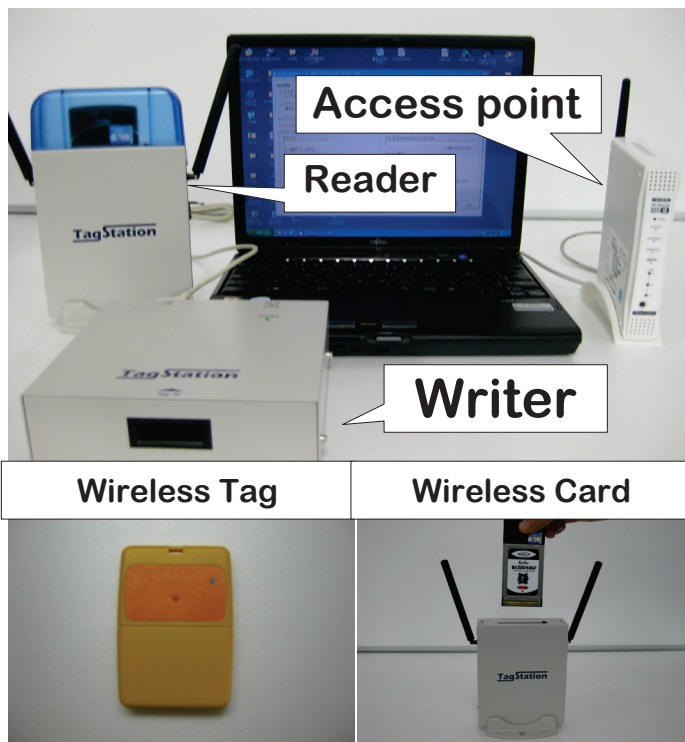


Fig. 14. IC Tag System of Active type

7.4 Results and Considerations

Table 3 shows RSSI and read times. The times and RSSI show higher values when the tag is attached to the iron as compared with a person. It seemed that tag served as an antenna when it was attached to the iron, and electric wave to reader were amplified. In addition, each values of attaching to wood and person were similar to basis. Therefore, the tag does not affect reception sensitivity if it is attached to the wood or person.

	0cm		70cm		100cm		170cm	
	RSSI	times	RSSI	times	RSSI	times	RSSI	times
Basis	4	17	6	21	4	21	6	21
Wood	4	18	6	19	5	19	4	21
Iron	8	21	8	22	8	22	8	22
Person	4	17	5	21	4	19	5	19

Table 3. RSSI and times to read tag

8. Update base map by using regional parameter

“The map should be fresh” is a concept of our laboratory mentioned above. The purpose of study is to establish a method for updating a large-scale digital map for local government using a Real-Time GIS. The Real-Time GIS which was defined by our laboratory can be used to renew the new BM. The Real-Time GIS is a technique that updates the new BM instantly by the Real-Time Kinematic Global Positioning System (RTK-GPS), GIS, and mobile phones. These techniques have been called “Geoinformatics” that is a new field of survey.

Japan has adopted a new general standard for map geometry since April 1, 2002. Ellipsoid of a new geodetic system in Japan is almost equal to WGS-84 of GPS. However, most of the digital maps of local government are still Tokyo Datum of an old geodetic system. To cope with two kinds of data which have different geodetic systems, it is necessary to transform coordinates.

In the master's thesis of Ms. Aki Okuno who graduated from Kanazawa Institute of Technology (KIT), she tried to solve the problem between the old and new geodetic systems by TKY2JGD and Affine Transformation. The result is listed below (Okuno, 2006).

1. The control point of transformation has to be located at four corners in the map. The exact point (national control point and public control point) of the control point could not be found in the field of survey.
2. A 1/500 scale area is desirably converted.

However, the method of making control points was not adopted because it is difficult to obtain the coordinates on a map and to find the points at the field. Therefore, the control points of town planning group data and cadastral data were used for coordinate transformation. Many control points are in a narrow area.

As a result, the transformed old map will be allowed to overlap to a new map measured by RTK-GPS.

9. Control Point for Transformation

To transform the old BM, the control point was used at the field. In the research, the control point means a point for showing both exact coordinates of Tokyo Datum and JGD2000. A verification area has accurate data of Tokyo Datum (based on BESSEL ellipsoid and rectangular plane coordinate system), and the data are managed by the Town Planning Group and the Cadastral Section in Kanazawa City. Experiment areas are "Area A," "Area B," "Area C," and "Area D," (they are marked as A, B, C and D in Figure 17 at next page). Figure 15 shows Sample of verification area (Area D).

To obtain the coordinates of JGD2000, this study used static positioning of GPS and Virtual Reference Station-GPS (VRS-GPS) at test fields (Figure 15). A and B were measured over 2 hours using static positioning of GPS. C and D were measured for one minute using VRS-GPS. The measurement time is decided by the law of Japan.

The control points of the Town Planning Group data and Cadastral Section data were installed simply. Especially, the control point number could be discerned on a road. Therefore, everyone can easily confirm the control point at the field. However, in the future, marking ink of the control point number will disappear. A better method for maintaining and managing the control point for coordinate transformation is to adopt the control point of town blocks and an IC tag.

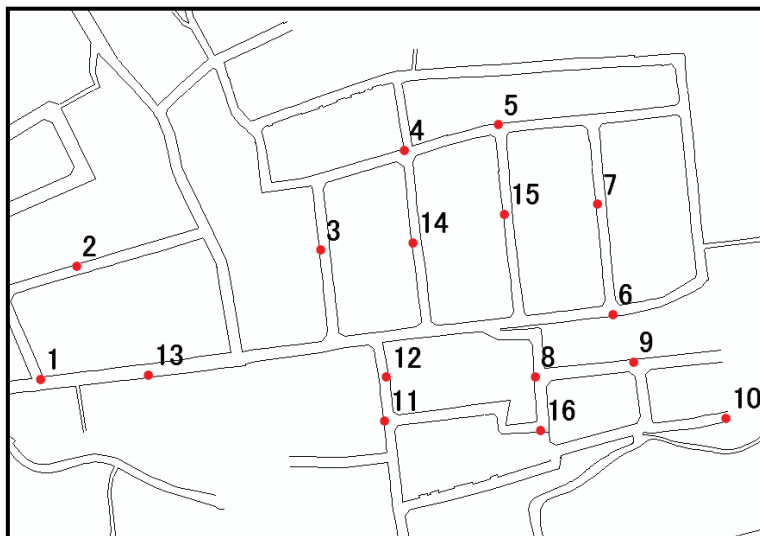


Fig. 15. Sample of verification area (Area D)



Fig. 16. GPS measuring at the field point

10. Verification of TKY2JGD for Japanese Standard Conversion

Geographical Survey Institute of Japan (GSI) opened a website for conversion parameters and programs (TKY2JGD). First, the coordinates of Tokyo Datum of the old geodetic system were transformed to new ones by TKY2JGD. The differences between calculation results and GPS measurement data were verified. The detail results are not shown in this paper by page limitation. The average differences at A, B, C, and D were about 11.3, 31.9, 11, and 14 cm respectively. In addition, A, B, C and D were rotary, parallel, south-east, and south-southeast respectively.

The areas A, B, C and D have a regular accident error. Converted data almost all corresponded to digital BM data in a small-scale map. However, the parameter could not be adopted in a large-scale map. Because the parameter area of TKY2JGD is too large, it is necessary to make the parameter in a narrow area.

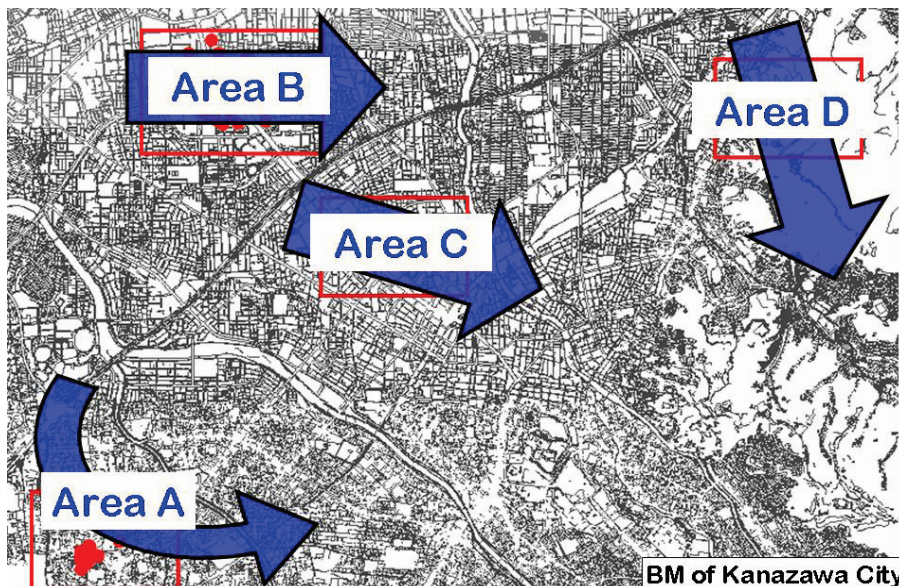


Fig. 17. Differences of vector in each test fields

11. Verification of High-Accuracy regional Parameter Using Affin Transformation in Narrow Area

On the verification, Affine Transformation is the most general and simple method in various geometric conversions. Affine Transformation makes the three parameters. Elements were rotation, scale and parallel. These parameters transform the geodetic system (x, y) of old geodetic system to (x', y') of new one. The conversion formula is as follows.

$$x' = x_0 + k_x x - \theta_y y \quad (1)$$

$$y' = y_0 + \theta_x x + k_y y \quad (2)$$

Where (x, y) = coordinates of Tokyo Datum
 (x', y') = coordinates of JGD2000
 (x_0, y_0) = parallel transformation
 k = scale
 k_x = scale of X axis
 k_y = scale of Y axis
 θ = rotation
 θ_x = rotation of X axis
 θ_y = rotation of Y axis

Parameters obtained by Affine Transformation are called "High-Accuracy Regional Parameter (HARP)". HARP was calculated by the coordinates of Tokyo Datum and GPS data. In the master's thesis of Ms. Aki Okuno, she performed calculation by 11 methods having a different number of control points and different places of control points, and

standard deviation of Affine Transformation had only 3cm errors when using control points at four corners. In this paper, four control points were located at four corners of the area. In addition, calculated parameters by the coordinates of A, B, C and D were named parameter A, parameter B, parameter C, and parameter D respectively.

Transformation methods are as follows.

Areas A, B, C, D were transformed by using parameter A, B, C, D.

Areas B, C, D were transformed by using parameter B.

Experiment Area	Error(m)	
	σ_x	σ_y
Area A (parameter A)	0.002	0.003
Area B (parameter B)	0.001	0.002
Area C (parameter C)	0.001	0.001
Area D (parameter D)	0.005	0.003

Table 4. Error calculated by the same parameter

Experiment Area	Error(m)	
	σ_x	σ_y
Area B (parameter B)	0.026	0.010
Area C (parameter B)	0.008	0.050
Area D (parameter B)	0.111	0.029

Table 5. Error calculated by different parameter

Table 4 shows that result of adapting the parameter of the same area. As a result of verification, the error was not more than 1cm in X and Y.

Table 5 shows that result of adapting the parameter of a different area. The error was large in areas D. However, the error of area A and area C are not more than 3cm in X and Y. There is a possibility that the large area can be converted by one parameter. In addition, the overlaying of New BM and transformed Map should be verified. Then, the accuracy of coordinate transformation will become clear.

12. Characteristics of Error Margin in Large Area Using Triangulation Point

12.1 Triangular point for coordinate transformation

The character of the error was clarified from the result of experiment. However, the regularity of the error in the large area was not able to be clarified. Therefore, the regularity of the error was comprehended by using a triangulation point in large area. Triangulation

point is managed by GSI and, the point is set up in large area. This experiment used triangular points of 53 places in the Kanazawa city.

12.2 Conversion result of triangulation point

On the result of the coordinate transformation, the regularity in the large area indicated the direction of the southeast (Figure 18). In addition, the research area was verified separately plain field and along the mountain. As a result, the character of plain field indicated the southeast pattern, and mountain indicated the south-southeast pattern. Characteristic of triangulation point was compared with to area of verification. As a result, feature of error was similar in the various areas (Figure 19). Therefore, the error in the verification area has the possibility that a triangulation point influences, and the error of a triangulation point has the possibility that diastrophism influenced.

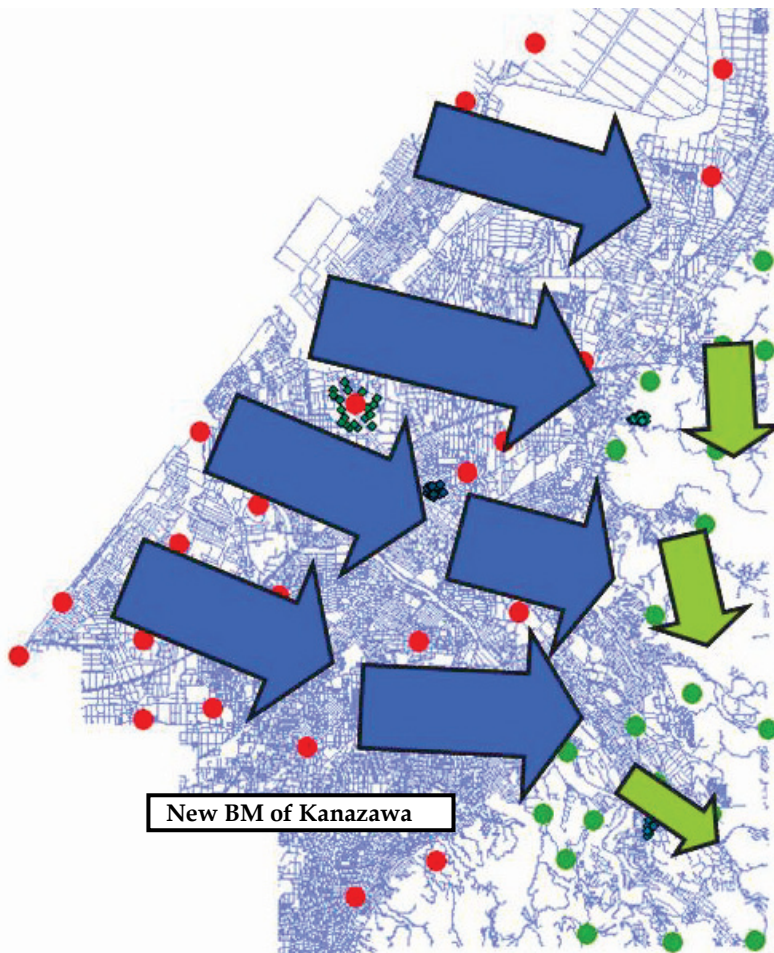


Fig. 18. Character of error in large area (Kanazawa City)

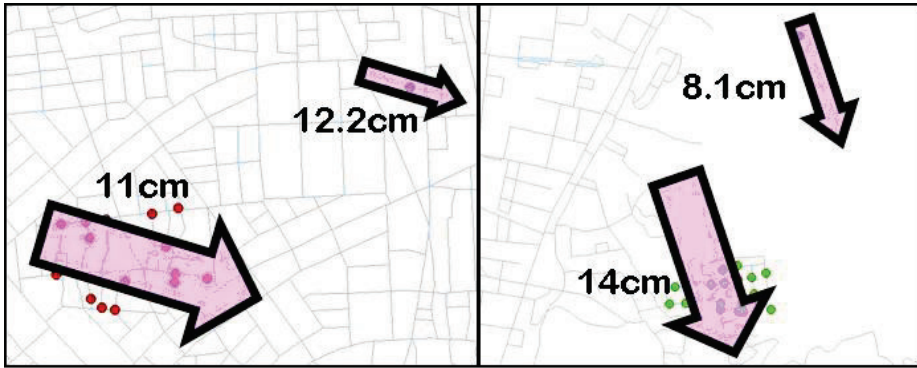


Fig. 19. Comparison between triangulation point and verification area

13. Overly of Transformed Map and New BM of Kanazawa City

Large scale Base Maps used in local government was transformed by using the parameter. Parameter was calculated in Chapter 10 and transformed Base Map was overlapped with New BM. An original program for transformation was made by our laboratory. Transformation parameters were inputted to the program and dot-line from Figure 20((a)-(e)) shows transformed Map using same area parameter. Heavy line in Figure 20((a)-(e)) shows the transformed Map by using parameter of Area A, and narrow line is road data of New BM in the city of Kanazawa's new base map (Kiban Chizu)

Result of applying same area parameter to same area, the error margin of area A, B, C, and D were about 10cm, 4.53m, 12.64m, and 5.36m respectively. On the result of overlapping New BM and converted map with area of A, those maps were accurately overlapped (Figure 20(a)), however, New BM and the converted map of area B, C and D were not overlapped (Figure 20(b),(c),(d)). Result of applying parameter B, the error margin of area A, C, and D were about 6.58m, 4.66m, and 4.69m respectively (Figure 20 (a), (c) and (d)). Therefore, all area did not overlap. Applying the parameter of another area to the area was very difficult.

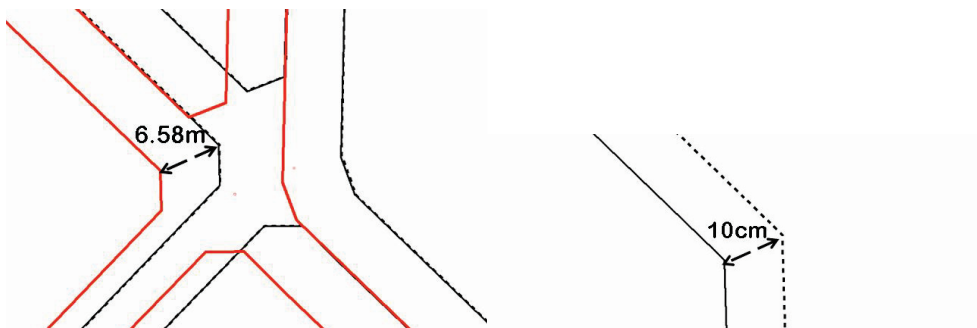


Fig. 20. (a) Overly of area A and New BM (b) Overly of area A and New BM (Amplifier)

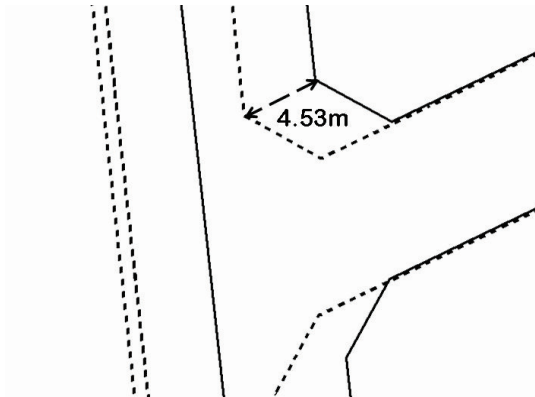


Fig. 20. (c) Overlay of area B and New BM

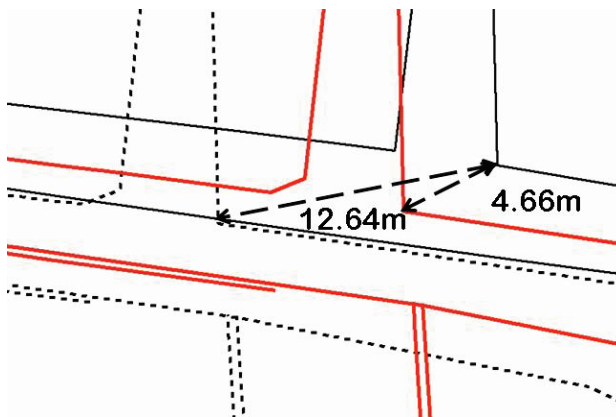


Fig. 20. (d) Overlay of area C and New BM

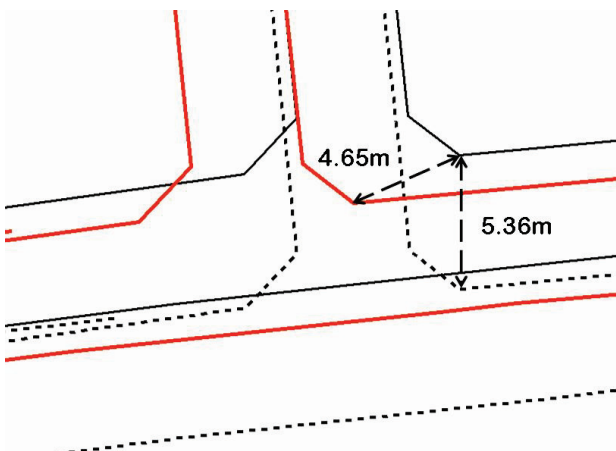


Fig. 20. (e) Overlay of area D and New BM

14. Overly of GPS Data and New BM

New BM was overlapped with GPS data acquired by VRS-GPS. Actual experiment was performed by using RTK-GPS of Virtual Reference Station (VRS-GPS). VRS-GPS does not need to set the reference station. Virtual reference station was made virtually around the measuring point. Distance of virtual point to actual point is about 3m to 5m. Revision information of rover station was sent to mobile phone by using wireless system. This system is possible to measure by only one person with light baggage (Figure 21).

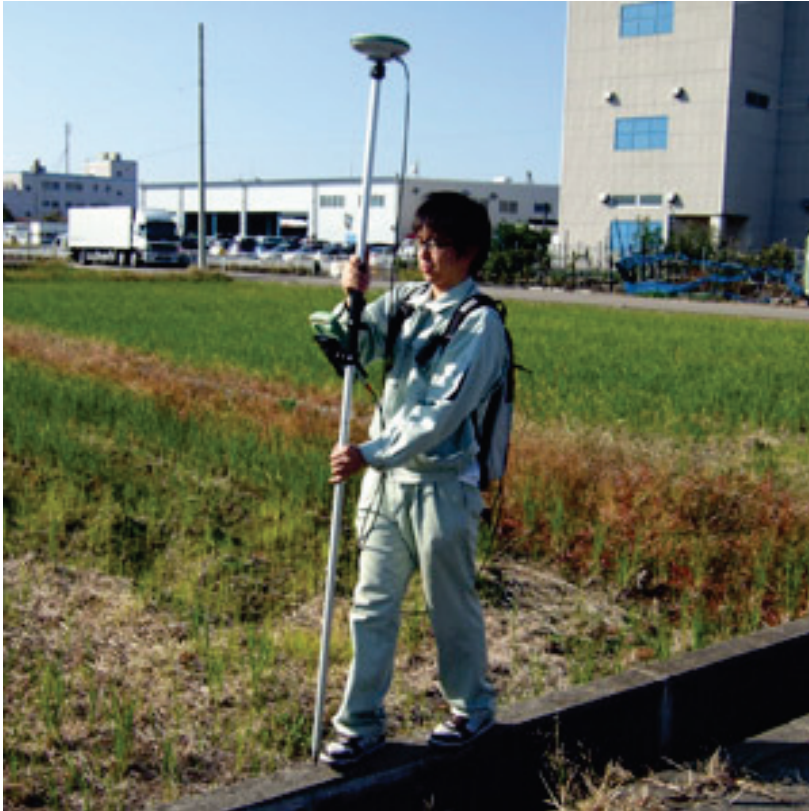


Fig. 21. VRS-GPS on the road line

A circle symbol and solid line in Figure 22((a)-(f)) means VRS-GPS and road line of New BM in the Kanazawa city respectively. Actual experiment was performed at district in A, B and D. GPS was not able to be observed, because area C was a residential area.

Difference between GPS data and New BM were 22cm in Area A, 32cm in Area B and 28cm in Area D (Figure 22). The GPS data was partially overlapped to the map, therefore the update of the New BM using VRS-GPS is effective at that area. However, a part of GPS data and the map did not overlap because GPS is influenced easily by the measurement environment. It is the reason why the update region is limited.

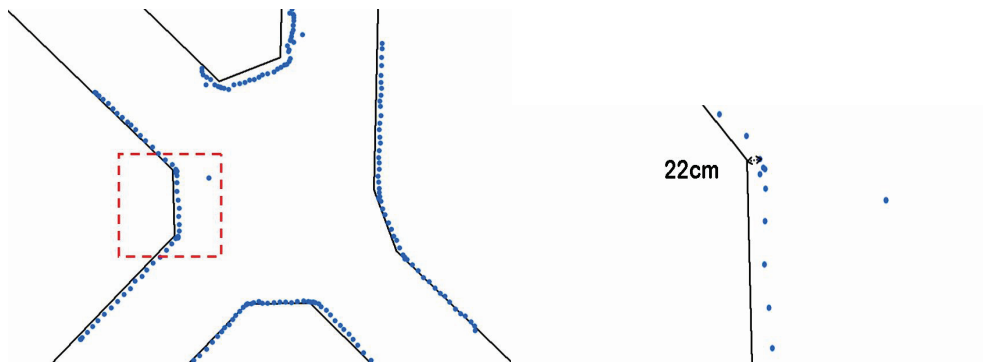


Fig. 22. (a) Overlay of VRS-GPS and Area A (b) Overlay of VRS-GPS and Area A (Amplifier)

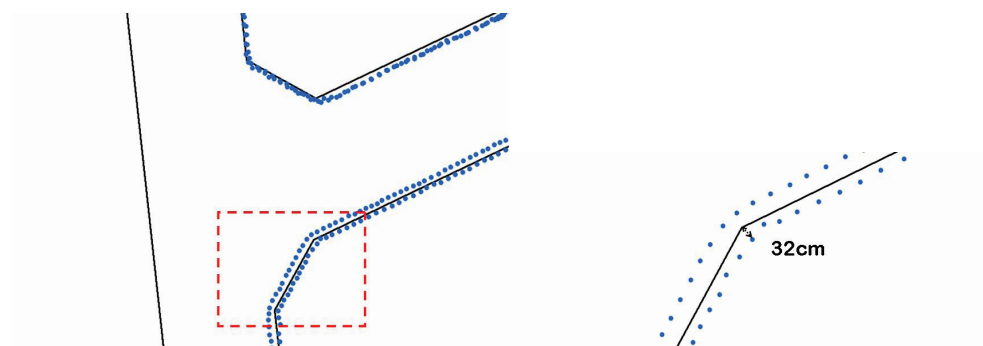


Fig. 22. (c) Overlay of VRS-GPS and Area B (d) Overlay of VRS-GPS and Area B (Amplifier)

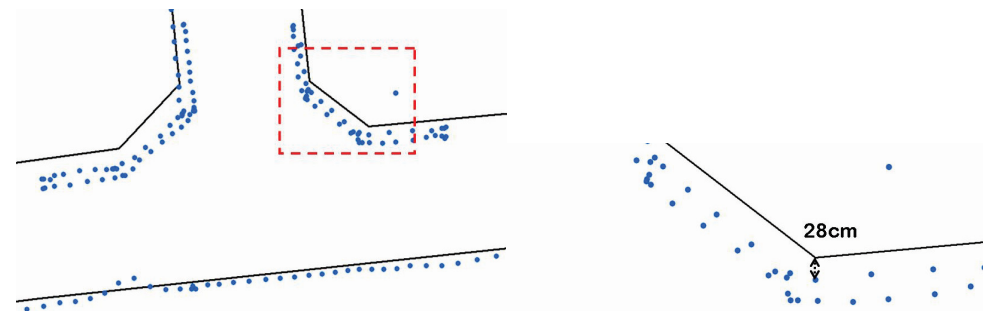


Fig. 22. (e) Overlay of VRS-GPS and Area D (f) Overlay of VRS-GPS and Area D (Amplifier)

15. Conclusion

The experiment was conducted by using a D-GPS and a VRS-GPS attached to wheelchairs. As a result, the positioning data with high-accuracy was obtained under the open sky. However, there were measurement errors including cycle slip and multipath at some places under closed sky. It showed that it is very difficult to solve these problems by using a GPS only. In addition, safe UM cannot be made only by the current technology. Therefore, the

study proposed the method by using geoinformatics and IC tags. Next step is to perform a spatial simulation and it is expected that the proposed method will establish useful UM in the near future.

If the government adopt high-resolution satellite imageries as a background of new BM, the government and general users can easily recognize the urban conditions. We recommend that the government introduce the system of "REAL TIME GIS" and Remote Sensing imageries for their work. The simplification of the mapping process, reduction of mapping and updating costs, and understanding of accurate urban conditions are connected with the improvement of improved service to the citizens. In the experiment, the coordinate transformation of the large-scale map for local government was successfully conducted at a part of test field. However, old maps were not accurately transformed to the new geodetic system. Therefore, it is necessary to investigate the method of the highly accurate conversion.

It is suggested that the town block control point is very useful tool to transform the map. The town block control point was made by GSI on "Basic Survey of Town Block for Renewal of Urban Areas." The town block control points has accurate data because they are managed by the nation, and they were set up at a short interval (every 200m).

Therefore, accurate reference point data can be acquired. If the town block control point is used for geometrical transform, the characteristic of the difference between the old geodetic system and the new one understand easily, and coordinate transformation will perform more efficiently. VRS-GPS and New BM was overlapped. As a result, the effectiveness of VRS-GIS was confirmed. Collaboration of Remote Sensing, Real Time GIS will help local government renew a large-scale map and new BM, and this research will contribute to update of new BM.

Finally, I would like to recommend strongly that local governments have to establish the conference to revise large scale map by using NSDI which include country, prefecture, and cities. It is most important point that reduces much labor money and time.

Previous research confirmed whether high-accuracy positioning information continuously by a D-GPS and VRS-GPS. However, GPS can be obtained positioning was not performed in closed sky. It was proposed to obtain positioning information by a high-accuracy GPS positioning technology like an IC tag utilized at intelligent for control points. Kind of IC tag has passive type and active type. The method is different by a purpose of use. In this study, it became clear how to use appropriately passive type IC tags. Active type IC tags are in a stage of growth on territory of geoinformatics technology. In the future, simulation experiments will be conducted to verify whether positioning information obtained by GPS and IC tags can be shown on the GIS.

16. Acknowledgement

The authors wish to special thank to Program Director Ms.Naoko Matsuda of Nonoichi Foundation for Information and Cultural Promotion, and Ms.Aki Okuno of KOKUSAI KOUGYO Co., LTD. Survey, KokudoKaihatu Center, Ltd., and Nihonkai consultant, Ltd, that offered data. Moreover, we would liketo thank to Nonoichi town office and the cooperation of Mr. Uchida and Mr. Fujita, Leica Geosystems Co., Ltd. which offered much data and suggestions of analysis. Moreover, I would like to express our heartfelt gratitude to

the cooperation of Mr. Fukumori, Toppan Printing Co., Ltd. for supplying IC tag's information and samples.

17. References

- Alfred L.(2004),*GPS Satellite Surveying*, Third Edition,ISBN0-471-05930-7, John Wiley & Sons Ltd.
- Arai C. et al.(2001) *Research on Real-Time Revision of Base Map using Remote Sensing and RTK-GPS*, IGARSS2001,IEEE International Geoscience and Remote Sensing Symposium, 0-7803-7033-3/1,1-3.
- Arai C. et al.(2002) *Management of Mapping in Local Government using Remote Sensing and the REAL TIME GIS*, IEEE International Geoscience and Remote Sensing Symposium 2002, 0-7803-7537-8/02,1-3.
- Arai C. et al.(2003) *An Application of Remote Sensing and REAL TIME GIS to Digital Map for Local Government*, IEEE 2003 International Geoscience and Remote Sensing Symposium, 0-7803-7930-6/3,1-3.
- Christian H,Peter A.W & Markus G.(2008).*Updating geospatial database from images*, Advanced in Photogrammetry,Remote Sensing and Spatial Information Science 2008 ISPRS Congress Book,Li,Chen & Baltasavias(eds),pp.355-362,CRC Press/Balkema,ISBN 978-0-415-47805-2, John Wiley & Sons Ltd.
- Helmut M,Stefan H & Uwe S,(2008).*Automated extraction of road,buildings and vegetation from multi-source data*, Advanced in Photogrammetry,Remote Sensing and Spatial Information Science,2008 ISPRS Congress Book,Li,Chen & Baltasavias(eds), pp.213-226,CRC Press/Balkema,ISBN 978-0-415-47805-2, John Wiley & Sons Ltd.
- Haigans S,Qiming Z,Jianya G & Guorui M,(2008).*Processing of multitemporal data and change detection*, Advanced in Photogrammetry,Remote Sensing and Spatial Information Science 2008 ISPRS Congress Book,Li,Chen & Baltasavias(eds),pp.227-247,CRC Press/Balkema,ISBN 978-0-415-47805-2, John Wiley & Sons Ltd.
- Journal of JACIC (Japan Construction Information Center) Information, (2007). *Feature of Spatial Information Society in the National Spatial Data Infrastructure of Japan (NSDI)*, Geographical Survey Institute of Japan (GSI), <http://law.e-gov.go.jp/htmldata/H19/H19HO063.html> (accessed 6 Dec. 2007)
- JACIC (Japan Construction Information Center), (2005).*Journal of JACIC Information, Feature of IC tag*, Vol.20, No.1
- Moriya M, Shimano S, Shikada M (2007). *Map renewal technique by using collaboration of GPS, GIS and Remote Sensing*, *GPS and Remote Sensing*, IEEE 2006 International Geoscience and Remote Sensing Symposium, p255, 1-3
- Michael N.D,(2005),*Foundamentals of Geographic Information System*,Third Edition,ISBN 0-471-45149-5, John Wiley & Sons Ltd
- Matsuda N et al.(2002) *Proposal of Renewal System for Barrier-free Map by using Remote Sensing and RTK-GPS*, IEEE International Geoscience and Remote Sensing Symposium 2002, 0-7803-7537-8/02,1-3.
- Matsuda N et al.(2003) *Actual experiment of renewal system for barrier-free map by using Remote Sensing and RTK-GPS*, IEEE 2003 International Geoscience and Remote Sensing Symposium, 0-7803-7930-6/3,1-3,2003.

- Matsuda N. (2003). *Application of Real Time GIS using Remote Sensing and RTK-GPS for Local Government*, Resume of Master's thesis Open Hearing Conference at KIT
- Okuno A. (2004) *Application of REAL TIME GIS using Remote Sensing and RTK-GPS for Local Government*, IEEE 2004 International Geoscience and Remote Sensing Symposium, 0-7803-874,1-3.
- Okuno A. (2006).*Coordinate Transformation of Large Scale Map and Establishment of Real Time GIS*, Resume of Master's thesis Open Hearing Conference at KIT,
- Paul A.Longlay et al. (2005), *Geographic Information System and Science*, 2nd edition, pp.385-403, ISBN 0-470-87000-1, John Wiley & Sons Ltd
- Shikada M, et al.(2004).*Coordinate Transformation of Large-scale Map Data for Real-time Update of Maps using RTK-GPS*, The Journal of Survey, Vol.54, No.8, pp.10-13
- Shikada M, et al. (2004).*Coordinate Transformation of Large-scale Map Data for Real-time Update future*, JACIC, Vol.22, No.3
- Shiratori K. (2005). *Change the Business by IC tag*, Printed in Paru Printing Co., Ltd.
- Shunlin L, Michael S & Mathias K, (2008). *Remote Sensing signatures: Measurements, modelling and applications*, *Advanced in Photogrammetry, Remote Sensing and Spatial Information Science* ISPRS Congress Book, Li, Chen & Baltasavias (eds), pp.127-143, CRC Press/Balkema, ISBN 978-0-415-47805-2, John Wiley & Sons Ltd.

Integrated sea surface temperature products within a coastal ocean observing system

Nadya T. Vinogradova

Atmospheric and Environmental Research, Inc. (AER)
USA

1. Introduction

Integration of regional information from existing ocean observing platforms, such as satellite and in situ observations, and from data assimilation and modelling systems is essential for our understanding and prediction of regional environments and ecosystems. The strategy of such integration is to link existing modelling and observing systems - both in situ and space-borne, - and to collect new atmospheric and ocean observations to better understand the Earth system, to monitor the climate, to predict environmental changes and mitigate natural disasters. Remarkable progress has been made in recent years toward the establishment of a global Earth observing system. As a result of an international June 2003 G8 Heads of State meeting, the U.S. Integrated Ocean Observing System (IOOS) was created, as a part of the Global Earth Observing System of Systems (GEOSS). The system is a pioneering architecture that provides new observational capabilities to advance informed decision making on national, regional, and local levels. The IOOS development plan (<http://www.ocean.us/ioospln.jsp>) called for both global and regional components. The coastal component consists of regional coastal observing systems that engage a broad spectrum of data providers and users who can depend on operational systems with the capacity to rapidly detect and provide timely predictions of changes occurring in the coastal environments.

One of the most essential variables in ocean dynamics that is used to monitor climate change is sea surface temperature (SST). Variations in SST are important indicators of climate variability, and can be related to other climate variables, such as sea level change, hurricane intensity, and air-sea fluxes of CO₂. In addition, SSTs are widely used in ocean modelling efforts by providing surface boundary conditions and/or observational constraints for atmospheric and oceanic hindcasts and forecasts. To increase resolution and to improve quality of analysis, SST products are often constructed by combining measurements from a variety of sources. Examples of global, operational, satellite SST products include the Global Ocean Data Assimilation Experiment (GODAE) High-Resolution SST Pilot Project (GHRSSST-PP) (Donlon et al., 2007), NOAA/NASA Advanced Very High Resolution Radiometer (AVHRR) Pathfinder SST analyses (Reynolds et al., 2002), Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) and NASA Advanced Microwave

Scanning Radiometer (AMSR) SST products (Chelton & Wentz, 2005). A list of available operational GHRSSST products is given in Table 1.

SST product name	Satellite sensors used	Agency Identifier	Grid Spacing	Period
OYSSEA	AVHRR, AMSRE, TMI, AATST, SEVIRI, GOES	CNES IFREMER	6 km	Day 274 2007 - present
AVHRR_OI	AVHRR, in situ	NOAA	0.25°	1985 - present
AVHRR_AMSR_OI	AVHRR, AMSRE, in situ	NOAA	0.25°	Day 152 2002 - present
OSTIA	AVHRR, AMSRE, TMI, AATSR, SEVIRI, in situ	UK Met Office	5 km	April 2006 - present
MW_IR_OI	AMSRE, TMI, MODIS	Remote Sensing Systems	9 km	Day 233 2005 - present
NAVO K10	AVHRR, GOES, AMSRE	NAVOVEAN O (NAVY)	10 km	Day 92 2008 - present

Table 1. Examples of available daily global SST analyses provided by GHRSSST group

The products are typically based on merged, optimally-interpolated multi-sensor SST data sets. All these products are high-quality, high-resolution, daily global analyses that also provide the errors associated with their interpolation procedure. For example, Ocean Surface Temperature and Ice Analysis (OSTIA), provided by the UK Met Office (see example in Figure 1), is generated globally in near-real time on a $1/20^\circ$ (~ 5 -km) grid and is routinely validated using independent observations from Marine-Atmospheric Emitted Radiance Interferometer (M-AERI). The system combines satellite microwave and infrared measurements with in situ observations from ships and buoys using optimal interpolation with correlation length of 700 km, and it has a root-mean-square error within 0.8°C . (Stark et al., 2007). Another example of the real-time global SST analyses (RIG_SST) is shown in Figure 2. The fields are developed at the National Centres for Environmental Prediction (NCEP) on a $1/12^\circ$ (~ 9 -km) grid as a blend of in situ and AVHRR observations using variational analysis with isotropic correlation scales that vary from 100 km in areas of high temperature gradients to 450 km in areas of low SST gradients (Thiébaux et al., 2003). Comparison with buoy data resulted in average root-mean-square error within 1°C .

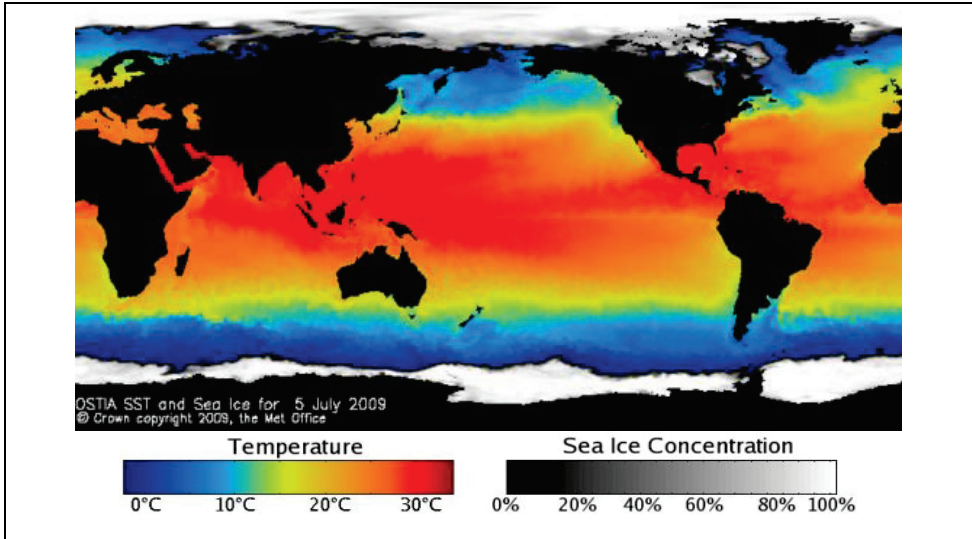


Fig. 1. OSTIA analysis: An example of integrated SST product based on synthesis of infrared and microwave satellite-derived SSTs with in situ data on July 5 2009 in °C. Spatial resolution ~5 km. Source: UK Met Office website

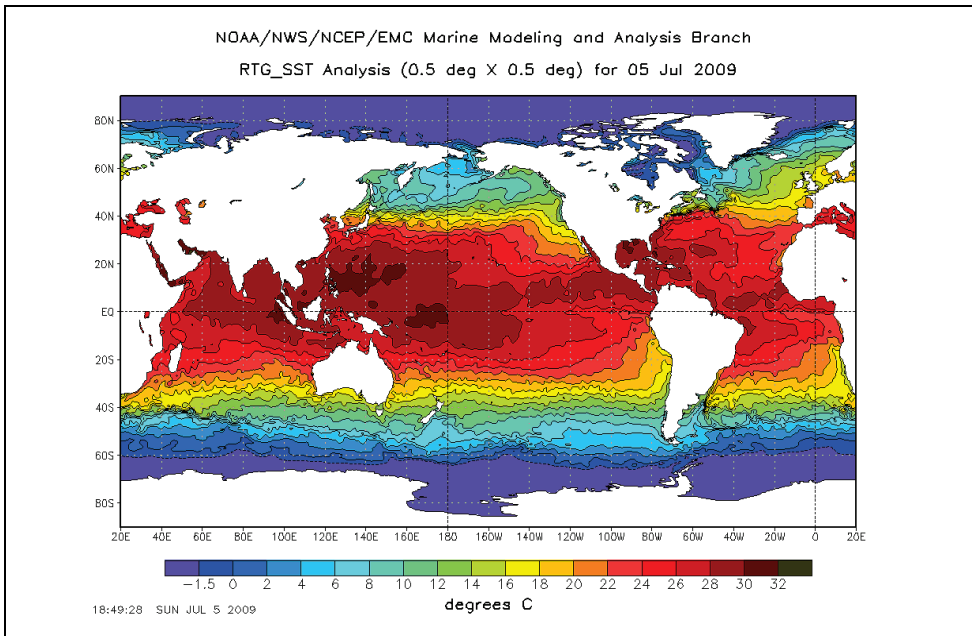


Fig. 2. RIG_SST product: An example of integrated SST product based on synthesis of infrared satellite-derived SSTs with in situ data on July 5 2009 in °C. Spatial resolution ~9 km. Source: NCEP website.

Another source of integrated high-quality SST analyses are solutions provided by general ocean circulation models. Ocean models provide an estimate of the ocean state, including SST, that could be constrained by observations and the model physics and that approximates the time evolution according to the model's equations and parameterizations of the fluxes. Observational constraints range from satellite-derived SST fields (e.g., GoMOOS model, Xue et al., 2005) to almost all available ocean datasets including in situ and satellite-derived observations such as altimetry, Argo, CTD, XBT, scatterometer, SST, SSS, etc. (e.g., ECCO-GODAE solution, Wunsch et al., 2009). Ocean models are typically fit in a least-squares sense to each datasets, each weighted according to the best existing estimate of the data and model errors (see Section 2 for more details). Such combinations provide optimal estimates, given model physics and knowledge of the data. Evaluation of how well model solution fits the data is usually done in terms of cost ratio, which is defined as the variance of model-data differences divided by data error variance:

$$\text{cost} = \frac{\langle M - D \rangle^2}{\sigma_{DATA}^2} \quad (1)$$

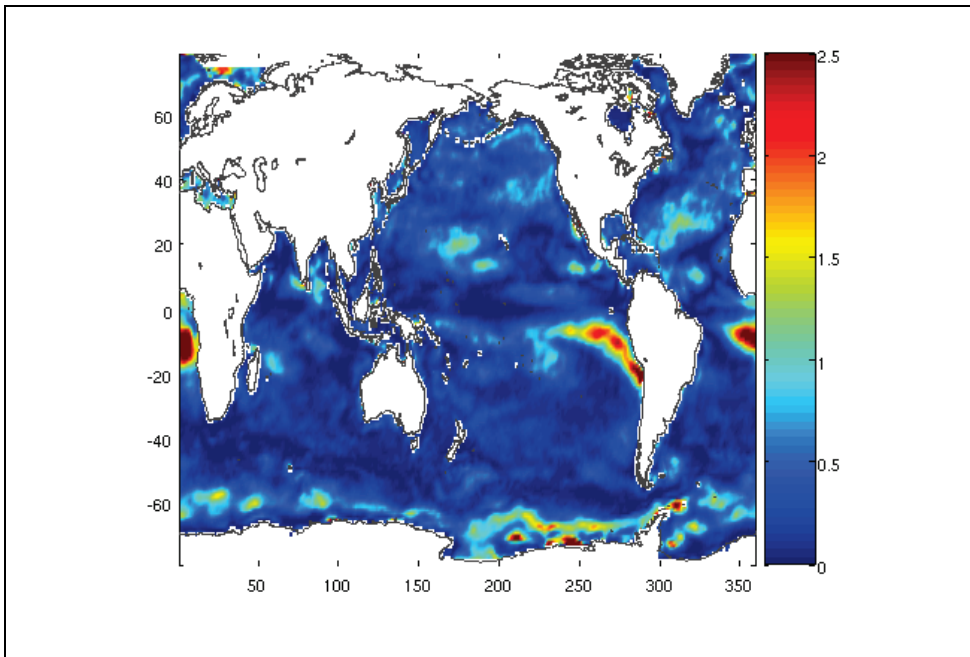


Fig. 3. Example of the cost calculated according to equation (1) for the global ECCO-GODAE model solution and GHRSSST-AVHRR_OI product at annual frequency. ECCO-GODAE optimization is achieved in terms of minimizing the model/data differences. The example above shows consistency between the model and data seasonal cycles, which is the dominant signal in SST variability. Values of the cost that are close or less than one indicate good model/data agreement within the noise level of the data. The optimization of the

ECCO products is still ongoing, but overall there is a good agreement between the solution and SST data within the expected noise level.

A well-constrained model solution typically has values of the cost that are close to one, indicating a good model/data agreement within the noise level of the observations (see example in Figure 3).

2. Integration methods

Each of the two integration approaches mentioned above, i.e. pure data synthesis and model/data integration, has both advantages and limitations. For example, infrared measurements (AVHRR, MODIS) allow higher spatial resolution, especially in coastal areas, but they can be less accurate than microwave data (AMSR, TMI) due to cloud contaminations. In addition, blended analyses tend to over-smooth ocean fine spatial structures, such as fronts and eddies (Donlon et al., 2004). The products are also sensitive to methods that are chosen to blend SST datasets from different sensors, with differences between analysis reaching 2 °C. Ocean models can provide SST estimates that are physically consistent with the dynamical and thermodynamical constraints, but can also introduce additional model errors related to unresolved physics (e.g., sub-grid parameterizations).

When integrating SSTs from different sources, including blended analyses or data assimilation into numerical models, one has to estimate the total error necessary to compute the weights for synthesis algorithms. Consider the problem of estimating the model field M from data D , which measures the ocean variable with some error:

$$D = HM + \sigma_{DATA} \quad (2)$$

where H is mapping matrix that establishes the relationship between M and D . The Bayesian maximum likelihood approach (Vinogradova et al., 2005) allows one to build the optimal field that maximizes the conditional probability of the field M :

$$p(M|D) \rightarrow \max \quad (3)$$

Maximization of conditional probability can be expressed in terms of cost function J with respect to M :

$$J(M) = -\log p(M|D) = -\log p(D|M) - \log p(M) \rightarrow \min \quad (4)$$

Assuming that the errors are uncorrelated and the statistics are Gaussian, the probability density $p(D|M)$ can be expressed in terms of the data error statistics:

$$-\log p(D|M) = -\log p(\sigma_{DATA}) = -\log p(D - HM) = [D - HM]^T W [D - HM] \quad (5)$$

where T denotes transposed matrix, and W is a weight matrix, which is inversely proportional to the error covariance of data error:

$$W = \frac{1}{\sigma_{DATA}^2} \quad (6)$$

Assuming that observational errors are uncorrelated, one can estimate the data error as the difference between the data variance and model/data covariance:

$$\sigma_{DATA}^2 = \langle D^2 \rangle - \langle MD \rangle \quad (7)$$

As seen from procedure (2)-(7), integration occurs within expected uncertainties of each dataset. Accurate characterization of uncertainties, or data errors σ_{DATA}^2 in equation (7), is important when fitting the data to model or blending data from different sources. If the uncertainties are overestimated, one discards and loses information stored in the data. If, on the other hand, the errors are underestimated, the model is fitting noise. Figure 4 shows an example of SST error calculations that follow the procedure (2)-(7) for the North Atlantic region (Vinogradova et al., 2008). These SST errors are computed from the ECCO-GODAE ocean state estimates and global blended Reynolds OI.v2 SST analysis over the last decade. Annual signal has been removed from both model and observations. As seen from Figure 4, areas of high variability, such as western boundary currents (Gulf Stream in Figure 4), are characterized by large errors, indicating larger uncertainties in model and data. In these regions, the integration procedure will assign smaller weight to avoid imposing erroneous variability and noise fitting.

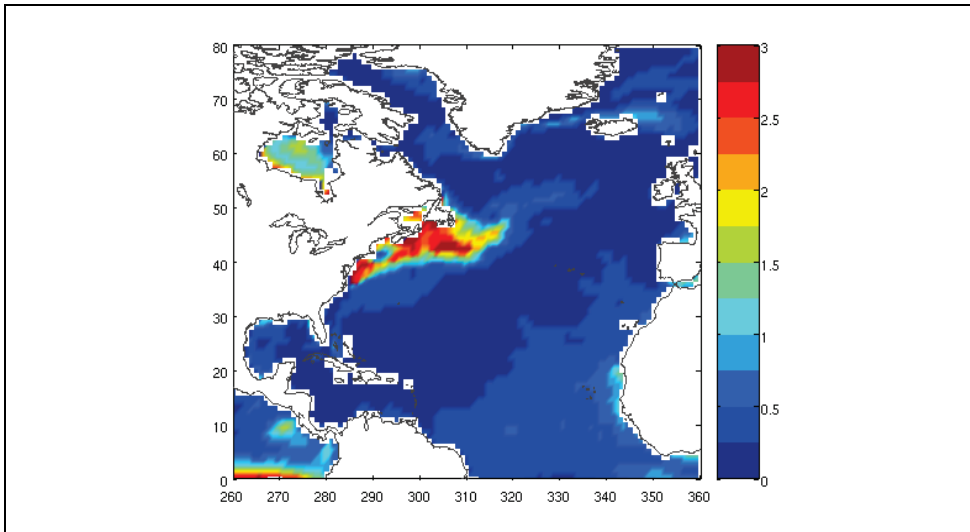


Fig. 4. SST errors (in °C) estimated from model/data difference based on ECCO-GODAE ocean state estimate and Reynolds SST analyses following the procedure (2)-(7). Annual

signal is removed. SST values in the areas with large errors will be integrated with smaller weights to avoid noise fitting.

3. Integrated SST products in coastal oceans

3.1 Integration of observations from various ocean-observing platforms.

Traditional blended SST products, described in the previous sections, are typically available as daily fields as they are based on measurements from polar-orbiting satellites which do not resolve high-frequency signals. However, high-frequency variations, and the diurnal cycle in particular, are important characteristics of the coupled atmosphere-ocean dynamics. The diurnal cycle has substantial implications in numerical weather prediction (NWP) and ocean models. Driven by solar forcing, it directly affects SST variations, the air-sea heat transfer regime, and variations in depth of the upper ocean mixed layer (Stuart-Menteth et al., 2003). Geostationary satellites, such as NOAA Geostationary Operational Environmental Satellite (GOES), provide a continuous stream of environmental data, which can be used to retrieve SST fields with high frequency. Combining observations from geostationary and polar-orbiting satellites allows one to produce a synthesized product with high spatial and temporal resolution. Presented here is an example of such regional application for the North Eastern US coast and Atlantic Canada, which is typically referred to as the Gulf of Maine region (see Figure 5).

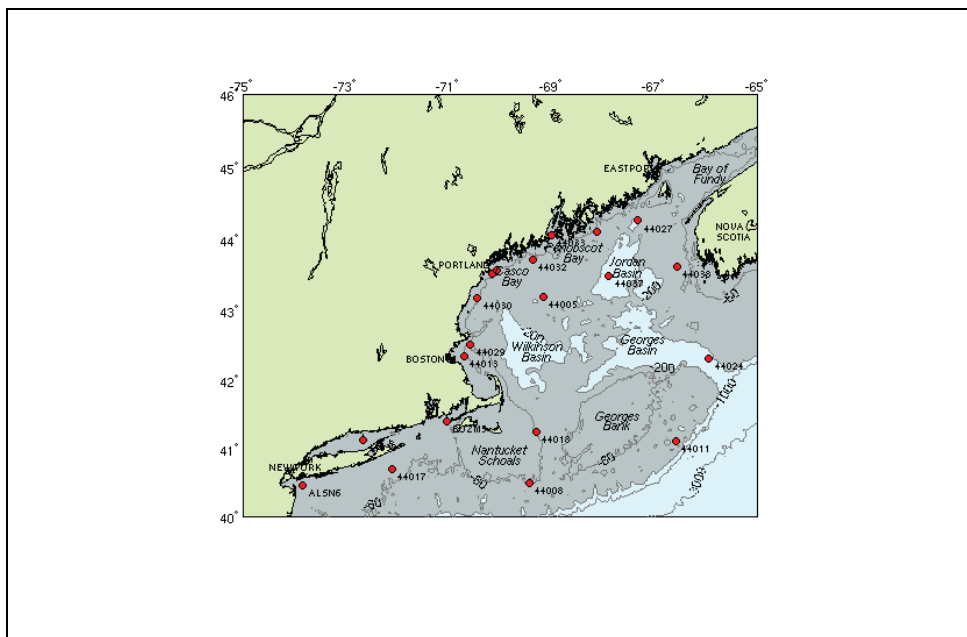


Fig. 5. The Gulf of Maine area, stretching along the coastline of New England and Atlantic Canada. The locations of the buoy stations transmitting in situ SST measurements in real time are shown as red circles. Bottom topography featuring banks and ridges is shown as contour lines.

This part of the ocean is a biologically productive and economically important area that covers about 92,000 km² of the ocean surface. It has a complex bottom topography, which includes banks, ridges, and basins, and extends up to 500 m deep and about 300 km offshore. From a modeling standpoint, the Gulf of Maine is a challenging area due to highly variable surface forcing, strong shelf and open ocean fluxes, and large tidal signal. The ocean circulation is mostly cyclonic and it is predominantly controlled by atmospheric heating and cooling, wind, river runoff, Scotian Shelf inflow, Gulf Stream warm-core ring intrusion and tidal mixing (Xue et al., 2000). For a successful operational forecast of the Gulf of Maine circulation, it is crucial to introduce accurate, high-resolution SST fields into a model, through data assimilation and surface boundary conditions. A synthesized AER-SST product with high spatial and temporal resolution has been developed (Vinogradova et al., 2009) to meet the demand in precise SST forcing. The fused SST product combines three sources: global temperatures estimates from (i) OSTIA and (ii) RTG_SST, described in the previous sections, and (iii) GOES radiances.

GOES SST retrievals are derived from the brightness temperatures of the GOES imager mid-wavelength infrared channel 2 at 3.9 μm and long-wavelength thermal infrared channel 4 at 10.8 μm using the AER cloud-mask detection algorithm (Gustafson et al., 2000). The cloud detection algorithm constructs a binary cloud mask using a set of multispectral tests to detect the presence of various cloud signatures. The cloud mask algorithm detects highly reflective terrain, including sun glint from water surfaces during daytime conditions using visible and thermal IR channels. The thermally-distinct cloud test compares brightness temperatures with the GFS NWP surface temperatures to detect obvious mid- and high-level clouds. The algorithm combines the results of the individual background and cloud tests to create the cloud mask. The algorithm assigns a binary cloudy/clear determination value if at least one of the cloud tests returns a positive result and all of the background tests return a negative result. After cloud mask has been applied, GOES radiances are converted to SST values following the current NOAA GOES operational SST equation:

$$SST = a_0 + a'_0 S + (a_2 + a'_2 S)T_2 + (a_4 + a'_4 S)T_4 \quad (8)$$

where S is the satellite zenith angle,

$$S = \sec(\theta) - 1 \quad (9)$$

T_2 and T_4 are brightness temperature of channel 2 and channel 4, respectively (Maturi et al., 2007). The NOAA retrieval coefficients a_i and a'_i , are listed in Table 2 and are derived from the regression analysis by matching satellite measurements with global drifting buoy observations from the Global Telecommunication System (GTS).

Imager channel	Wavelength, μm	a_i	a'_i
0	-	-2.1000	-1.1500
2	3.78-4.03	1.1177	0.0073
4	10.2-11.2	-0.1620	-0.0690

Table 2. Retrieval coefficients for the NOAA-GOES-12 SST algorithm (from Maturi et al., 2007). The coefficients are used to convert GOES radiances to SST values.

Although GOES radiances are collected every 30 minutes, individual GOES-SST retrievals are usually contaminated by clouds. To increase spatial coverage of the retrieved GOES SSTs, the fields are averaged over the set of eight analyses. The four-hour average provides a better spatial coverage and still resolves the diurnal cycle that could be significant in the coastal dynamics. The example of the input data sources is shown in Figure 6.

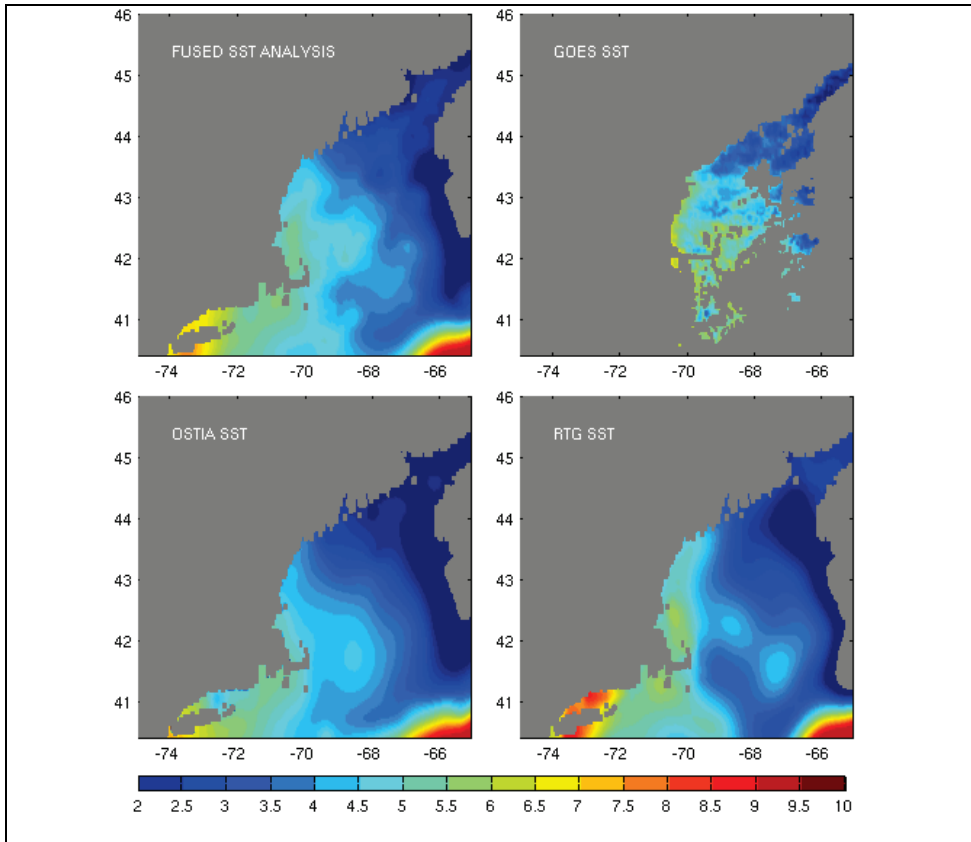


Fig. 6. Example of the SST data sources (GOES, OSTIA and RTG_SST) and fused SST analyses on May 2, 2007 in °C. Notice over-smoothed frontal features in OSTIA and RTG analyses, and unresolved cloud-contaminated coastal regions by the GOES.

Before blending, all data sets are mapped into the GOES domain (4 km) and are quality controlled to avoid artificial errors. The synthesized SST is computed as a weighted average of the three datasets with the weights being inversely proportional to the errors of each data constraint. The errors for each data sets E_i are determined by using the network of in situ SST measurements, that includes about 20 real-time buoys in that area. That allows one to assign time-varying weights, which are recomputed at every run of the system. For each data source, the errors are estimated as a mean difference between the in situ measurement and collocated satellite SST:

$$E_i(t) = \overline{(D_i(t) - B(t))_p} \quad (10)$$

where $i = 1, 2, 3$ represent each data source, D_i , $p = 1..20$ are locations of the buoys B_p , and $\overline{(\cdot)}$ denotes spatial averaging.

To resolve diurnal variability, the analysis runs four times per day and ingests observations received in the preceding (?) four hours, for both in situ and satellite data. Finally, the blended solution is smoothed by its variance with the correlation scale of 10 km, which is close to the Rossby radius in this area (Xue et al., 2000). The algorithm has been implemented into a prototype near-real time production grid system that, since May 2007, has been producing SST fields four times a day on a 4-km grid (see example in Figure 7).

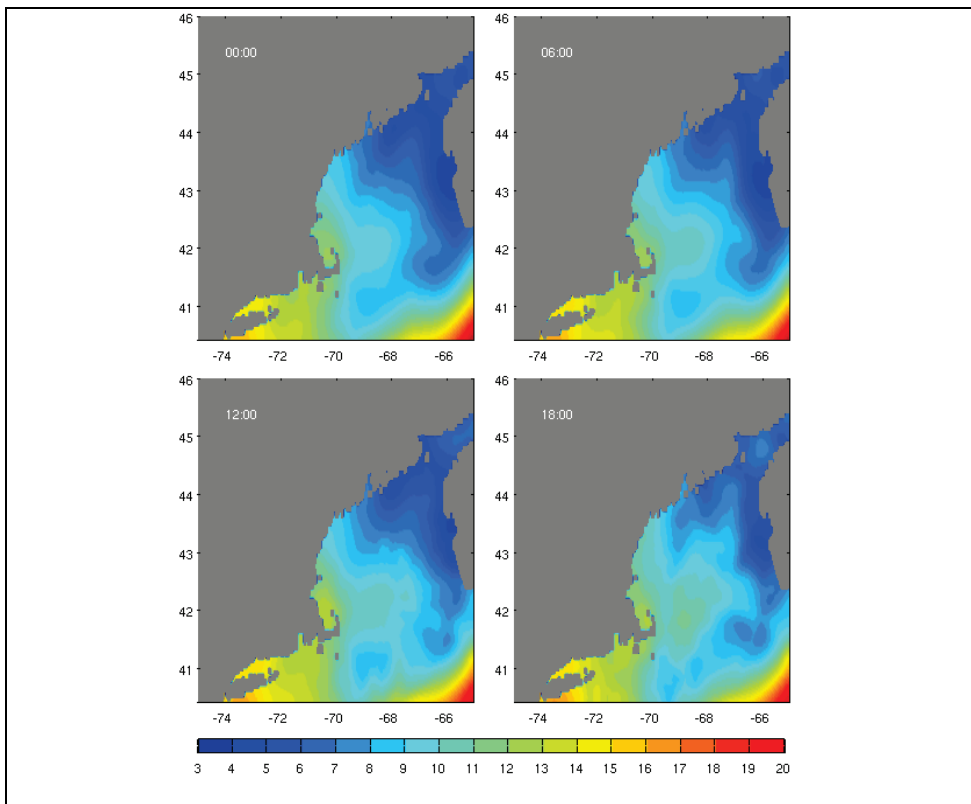


Fig. 7. Example of the fused AER-SST analysis at 00, 06, 12, and 18 UTC on June 8, 2007 in °C. The system resolves typical summer-time patterns with warmer water in the west and cooler water in the east. Other distinctive features include a strong tidal mixing front in the Georges Bank area; warm slope water intrusion in the southwest; cold water in the northwest and Bay of Fundy due to vigorous mixing in response to nearly resonant semi-diurnal tides; and cold temperature along the Maine coast resulting from summer upwelling.

The system is routinely validated by comparing the values of the blended SST analysis with in situ measurements from buoys. The average bias in the domain is found to be 0.02 ± 0.8 °C. SST bias is attributed to bulk correction and partly to the bias of the retrieved GOES SSTs. One way to reduce the bias would be to fine-tune the regression coefficients that are used in the retrieval algorithm in the equation (1) using regional observations instead of the global match-ups. Another way to enhance system performance might be the use of the preceding improved SST synthesis as a background field instead of the daily OSTIA and RTG_SST analyses.

The ability of the AER-SST product to resolve diurnal variations has important implications in ocean studies. Many coastal regions, especially Gulf of Maine area, are well-known for strong high-frequency signals, including diurnal and nearly resonant semi-diurnal tidal responses. Tides affect not only the high-frequency spectrum, but also variability at longer periods though tidal mixing and nonlinear rectification (Xue et al., 2000). One of many possible applications of the AER-SST system is to evaluate the spatial pattern of a diurnal cycle. An example of the monthly mean diurnal cycle is shown in Figure 8. Amplitudes of diurnal variations can reach up to 2-4 °C in summer time and are attributed to high variability of the shortwave insolation that can range from 0 at night to over 900 W/m² at noon (Chen et al., 2003). Large-amplitude diurnal fluctuations suggest that the diurnal forcing is significant and it should be accounted for in ocean models that currently assimilate daily SSTs and do not resolve higher frequency variability. Improved high-resolution regional SSTs would also enhance estimation of the surface heat flux, which is known to play a dominant role in seasonal coastal circulation. Furthermore, knowing diurnal amplitudes will allow one to validate empirical models that are used to retrieve daytime and nighttime satellite sea-surface temperatures (Gentemann et al., 2003).

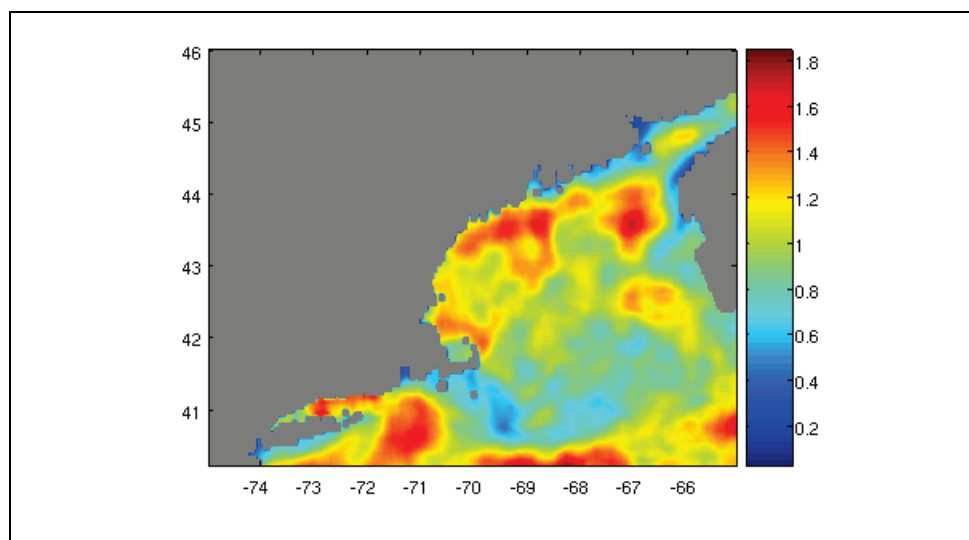


Fig. 8. Spatial distribution of the monthly mean diurnal variability during summer in the Gulf of Maine based on the AER-SST analysis (in °C).

Area-integrated value is about 2 °C, but the extreme amplitude can reach 4 °C. Complex spatial patterns are influenced by geometry of the domain and physical processes that control ocean circulation. Large values of diurnal variation are found along the coast, and in stratified and frontal regions. In the mixed regions, the diurnal cycle is weaker (e.g., the interior of the Georges Bank; western Nova Scotia). Even though solar heating increases during summertime, the buoyancy input is not strong enough compared to vigorous tidal mixing, which keeps the water well mixed and relatively cool in summer. In the stratified and frontal regions, heat is not transferred vertically as efficiently, and surface waters are more prone to diurnal warming, yielding large values of diurnal variability.

3.2 Integration of observations with ocean models

As shown in the previous section, integrated datasets based on pure observations, both satellite-derived and in situ, are useful tools to characterize upper ocean variability as a function of time and space. These products, however, do not explain the mechanisms controlling the observed variability. To interpret the observations, it is useful to analyze the equations defining an evolution of the sea surface temperature, which are also referred to as budget equations. Analysis of ocean surface heat budgets has been addressed both locally and regionally using in situ observations (Wang and McPhaden, 2001; Kim et al., 2006), and globally using theoretical calculations (Gill and Niiler, 1973). A prerequisite of a complete budget analysis is the closure of the budget, meaning that the sum of the budget components exactly matches the property tendency. Such a prerequisite is very difficult to fulfil when using raw observations and, in many cases, even numerical models (Qui, 2002). ECCO-GODAE is one of a few integrated data/model systems that allows computing closed budgets for any prognostic variable, due to consistency of the solution with the model equations and atmospheric forcing as the optimization is achieved through adjustments of the forcing fields and initial conditions. Closed property balances can be used for interpretation of the observed signals and in diagnostics of the SST tendencies, as they relate to advective and diffusive heat fluxes, or atmospheric forcing.

To characterize SST variations in terms of dynamic and thermodynamic processes that drive SST tendencies, it is useful to evaluate the strength of different terms in SST (or surface heat) balance. According to the heat content equation, the rate of change of heat storage in the surface layer occurs due to the advection and diffusive fluxes of heat, and the absorption and radiation of energy through the ocean surface:

$$\rho_0 C_p h \frac{\partial T}{\partial t} + \nabla \cdot (\rho C_p h T \vec{u}) = \nabla \cdot (\rho C_p h \vec{K}) + \frac{\partial Q}{\partial \xi} \quad (11)$$

where T is the temperature, \vec{u} is the velocity vector, \vec{K} is the diffusive heat flux vector, ξ is the vertical coordinate, ρ_0 is constant density of seawater, C_p is the specific heat capacity, and h is the thickness of the surface layer. Analysing the budget (11) provides information about the contribution of each term into surface heat or temperature tendency, where and when one regime is dominant over the other, and how it can be linked to ocean-

atmosphere coupling and predictability. Comparing ratios of each budget term to the total tendency can determine the extent to which each process affects surface heat content. Budget analysis by Vinogradova et al. (2008a) suggests that, in the Northern Atlantic, and in particular in the Gulf of Maine area, seasonal SST tendencies are one of the largest over the globe and can reach the values of 80 W/m^2 (see Figure 9). These studies also indicate that overall SST tendency due to advection is several times smaller compared to other terms, and except large scales, total tendency is a balance between surface mixing and heat fluxes across the air-sea interface.

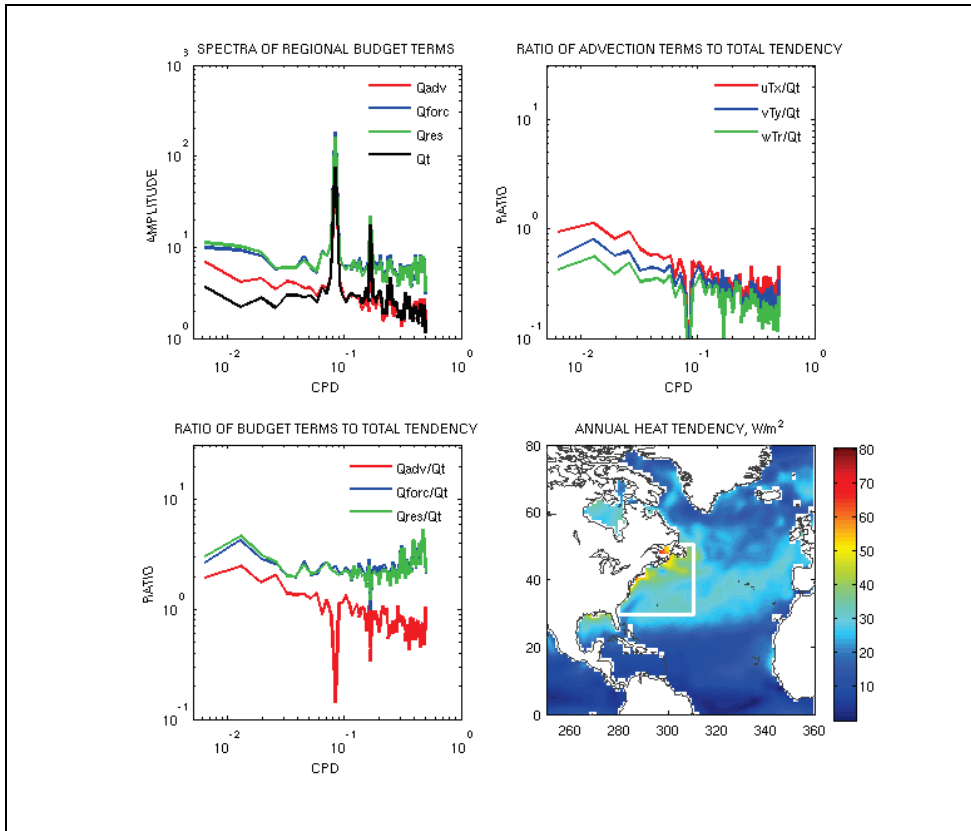


Fig. 9. Analysis of the surface heat budget based on the ECCO-GODAE 13-year ocean estimate for the North Atlantic and the Gulf of Maine area. The terms are computed according to equation (11). The ratios illustrate importance of forcing, advective and diffusive fluxes at different timescales. To analyse dynamical causes of the SST variation, advective fluxes are further evaluated in terms of meridional, zonal and vertical contributions. Along the western boundary, ratios of zonal and meridional components closely follow each other due to angular direction of the Gulf Stream. Vertical advection is generally the smallest term, but it becomes important on the decadal scales. Overall, SST tendency in this region is a balance between the diffusive fluxes and fluxes of heat across the air-sea interface.

4. Conclusion

Combinations of the advantages of the existing observing systems as well as integration of the observations into data assimilating systems provide invaluable tools for environmental management and control in the coastal regions. The need of high-quality, high-resolution SST fields, as one of the essential variables describing climate variability, has been receiving considerable attention in oceanic and atmospheric studies. With the abundance of the SST measurements, many products are based on synthesis of various data sources to produce an analysis with improved resolution and quality. Integration of SST measurements from polar-orbiting and geostationary satellites as well as in situ measurements from oceanic buoys, such as AER-SST product (Vinogradova et al., 2009), produces a new blended estimate with high spatial and temporal resolution. High temporal and spatial resolution of the system allows one to monitor fine ocean structures such as coastal fronts, describe high-frequency oceanic variability and improve regional numerical weather prediction systems. Such systems represent new coastal oceanic and atmospheric products to gain understanding and to improve prediction of the regional environment.

Integrating SST observations with data assimilating ocean numerical models provides best estimates of ocean state, given model physics and knowledge of the data. Integrated systems such as ECCO-GODAE typically continue to evolve and improve as new information (including new data and associated error estimates) becomes available. A considerable advantage of model/data integration is the ability to interpret observed variability as a function of space and time. In particular, understanding the dynamics and forcing of the SST variability is important to determine coupling mechanisms and predictability of the ocean-atmosphere system. By evaluating the prognostic equation for temperature, one can characterize the SST dynamics as a function of timescale and as function of season, which can both affect what physical mechanisms are most relevant (Vinogradova et al., 2008a). One can also differentiate between the regimes where the ocean is mostly affected by local atmospheric forcing, and those where response involve geostrophic and ageostrophic advection processes (Vinogradova & Ponte, 2009).

Identifying and understanding ocean coastal dynamics is one of the major tasks of the Integrated Ocean Observing System. Coastal oceans are highly populated areas of considerable interest to marine commerce, human recreation, oil and gas exploration and development, and are an integral part of the national and international economies. Operational integrated SST products that combine data and/or numerical models help to address important oceanographic problems of air-sea interaction and ultimately improve coastal monitoring and predictability.

5. References

- Chelton, D. B. & Wentz, F. J. (2005). Global microwave satellite observations of sea surface temperature for numerical weather prediction and climate research, *Bull. Amer. Meteor. Soc.*, 86, 1097-1115.
- Chen, C.; Beardsley, R. C., Franks, P. J. S., & Keuren, J. V. (2003). Influence of diurnal heating on stratification and residual circulation of Georges Bank. *J. Geophys. Res.*, 108, 8008, doi:10.1029/2001JC001245

- Donlon et al. (2007), The GODAE High Resolution Sea Surface Temperature Pilot Project (GHRSS-PP), *Bull. Amer. Meteor. Society*, 88 (8).
- Donlon, C. J.; Nykjaer, L., & Gentemann, C. L. (2004). Using seas surface temperature measurements from microwave and infrared satellite measurements. *International J. Remote Sensing*, 25, 1331-1336.
- Gentemann, C. L.; Donlon, C. J., Stuart-Menteth, A., & Wentz, F. J. (2003). Diurnal signals in satellite sea surface temperature measurements. *Geophys. Res. Lett.*, 30, 1140, doi:10.1029/2002GL016291
- Gustafson, G. B.; D'Entremont, R., Collins, J., & Cady-Pereira, K. (2000). VIIRS Algorithm Theoretical Basis Document (ATBD) for the Cloud Cover/Layer EDR, version 1.3
- Kim, S-B.; Lee, T., & Fukumori, I. (2007). Mechanisms controlling the interannual variation of mixed layer temperature averaged over the Niño-3 region. *J. Clim.*, 20, 3822-3843.
- Maturi, E.; Merchant, C., Harris, A., Li, X., & Potash, B. (2007). Geostationary sea surface temperature product validation and methodology. *Proceedings of 23th International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, San Antonio TX, January 2007
- Qiu, B. (2002). The Kuroshio Extension system: its large-scale variability and role in the midlatitude ocean-atmosphere interaction. *J. Oceanography*, 58, 57-75
- Reynolds, R. W.; Rayner, N. A., Smith, T. M., Stokes, D. C. & Wang, W. (2002). An improved in situ and satellite SST analysis for climate, *J. Clim.*, 15, 1609-1625.
- Stark J. D.; Donlon, C. J., Martin, M. J. & McCulloch, M. E. (2007). OSTIA: And operational high resolution, real time, global sea surface temperature analysis system. *Proceedings of Oceans MTS/IEEE Conference, Vancouver, Canada*. September-October 2007.
- Stuart-Menteth A. C.; Robinson, I.S., & Challenor, P. G. (2003). A global study of diurnal warming using satellite-derived sea surface temperature, *J. Geophys. Res.*, 108, 3155, doi:1029/2002JC001534.
- Thiébaux, J.; Rogers, E., Wang, W., & Katz, B. (2003). A new high-resolution blended real-time sea surface temperature analyses. *Bull. Amer. Meteor. Soc.*, 84, 645-656.
- Vinogradova N. T.; Zaccheo, T. S., Alcala, C. M. & Vandemark, D. (2009). Operational high-resolution sea surface temperature product in the Gulf of Maine. *Journal of Operational Oceanography*, in press.
- Vinogradova N. T.; Ponte, R. M. & Heimback, P. (2008a). Sea surface temperature budgets and mechanisms controlling upper ocean climate variability. *Proceedings of ASLO/AGU/TOS Ocean Science Meeting*, Orlando, FL, March 2008.
- Vinogradova N. T.; Zaccheo, T. S., Alcala, C. M. & Vandemark, D. (2008b). Regional fused sea surface temperature system for the Gulf of Maine. *Proceedings of IEEE International Geoscience & Remote Sensing Symposium (IGARSS)*, Boston, MA, July 2008.
- Vinogradova, N. T. & Ponte, R. M. (2009). Dynamics and forcing of upper ocean heat balance on climate scales. In preparation.
- Vinogradova, N. T.; Vinogradov S. V., Nechaev D. A., Kamenkovich V.M., Blumberg, A. F., Ahsan Q. & Li, H. (2005). Evaluation of the Northern Gulf of Mexico Littoral Initiative (NGLI) model based on the observed temperature and salinity in the Mississippi Bight, *MTS Journal*, 38(2), 25-38.

- Wang, W. & McPhaden, M. J. (1999). The surface-layer heat balance in the Equatorial Pacific ocean. Part I: mean seasonal cycle. *J. Phys. Oceanogr.*, 29, 1812-1831
- Wunsch, C.; Hemibach, P., Ponte, R.M., & Fukumori, I. (2009). The global general circulation of the oceans estimated by the ECCO-consortium. *Oceanography*, 22, 88-103.
- Xue H.; Chei, F. & Pettigrew, N. R. (2000). A model study of the seasonal circulation in the Gulf of Maine. *J. Phys. Oceanogr.*, 30, 1111-1135
- Xue, H.; Li, L., Cousins, S., & Pettigrew, N. R. (2005). The GoMOOS nowcast/forecast system. *Cont. Shelf. Res.*, 25. 2122-2146.

Soil Backgrounds Impact Analysis on Chlorophyll Indices Using Field, Airborne and Satellite Hyperspectral Data

A. Bannari ¹ and K. Staenz ²

¹ Remote Sensing and Geomatics of Environment Laboratory
Department of Geography, University of Ottawa, Ottawa (Ontario), Canada K1N 6N5

² Alberta Terrestrial Imaging Center (ATIC)/ Department of Geography, University of
Lethbridge, 400, 817 - 4th Avenue South, Lethbridge, Alberta, Canada T1J 0P3

Abstract

In precision agriculture, crop nitrogen status can be estimated based on the measurement of leaf chlorophyll content at specific stages of crop development. Over the last decade, several spectral chlorophyll indices have been developed to estimate chlorophyll content both at the leaf and the canopy level from different crop types using hyperspectral remote sensing data. For an accurate interpretation of chlorophyll indices derived from hyperspectral data, a “true” chlorophyll content value attributed only to the crop cover signal and free from any contribution of non-photosynthetic elements is required. However, in remote sensing, in spite of the correction and the standardization of the various radiometric distortions such as due to topography, atmosphere, sensor drift, and Bidirectional Reflectance Distribution Function (BRDF) effects, the chlorophyll indices remain sensitive up to a certain degree to the artifacts caused by the soil optical properties particularly in an earlier stage of crop growth. This chapter focuses on the evaluation and comparison of the sensitivity of several chlorophyll indices to bare soils optical property variations. In order to achieve the goal of this investigation, field spectroradiometric measurements were used as well as hyperspectral data acquired with the Probe-1 airborne and Hyperion Earth Observing -1 (EO-1) satellite sensors. The field-based reflectance measurements were acquired above 90 bare soil plots with various optical properties and selected from different agricultural lands. Probe-1 and Hyperion EO-1 data were acquired over the study site on June 28, 2000 and June 30, 2002, respectively. Imagery data were spectrally and radiometrically calibrated, as well as atmospherically corrected. After these pre-processing steps, sixty spectral signatures of different bare soils with various optical properties were extracted from each set of data for use in the analysis. The obtained results show an excellent agreement between the accuracies estimated from field, airborne and satellite data. Indices SIPI, PSSRa and MTCI show a very high root mean square error (RMSE) related to optical background variation. The indices SIPI, SRPI, NDPI, NPCI, GNDVI, CAI and HNDVI have a non-negligible RMSE related to the optical properties of bare soils, and will be very difficult to interpret at a low leaf area index (LAI). PSNDa, hNDVI and PRI show an RMSE less than 20%. The most insensitive

index of this group is the PRI with an RMSE less than 6 %. However, these errors remain significant. Independently from the data source and from the bare soil background, CARI, MCARI and TCARI indices are basically not sensitive to changes in soil optical properties with a RMSE less than 1.2 % and will permit a better estimation of chlorophyll content in sparse crop cover environment.

Keywords: chlorophyll indices, soil optical properties, precision agriculture, hyperspectral remote sensing, field spectroradiometric measurements, Probe-1 and Hyperion EO-1 sensors.

1. Introduction

Nitrogen concentration in crop cover is related to chlorophyll content and, therefore, indirectly to one of the basic plant physiological processes. When nitrogen supply surpasses the vegetation's nutritional needs, the excess is eliminated by runoff and water infiltration leading to pollution of aquatic ecosystems (Daughtry *et al.*, 2000; Wood *et al.*, 1993). This nitrogen loss to the environment represents an economic loss for farmers. However, inappropriate reduction of nitrogen supply could result in reduced yields and, subsequently, substantial economic losses. With this impasse, the optimal solution is an adequate assessment of the nitrogen status and its variability in agricultural landscapes. Since yield is determined by crop condition at the earlier stages of growth, it is mandatory to provide farmers with nitrogen status at those stages in order to supply appropriate rates based upon an accurate assessment of plant growth requirements and deficiencies (Haboudane *et al.*, 2002). For this purpose, remote sensing techniques have been used to assess crop conditions relative to nitrogen status and effects. Foliage spectral properties, reflectance and transmittance, were found to be affected by nitrogen deficiency (Blackmer *et al.*, 1996). Nitrogen shortage reduces leaf chlorophyll content and, therefore, increases its transmittance at visible wavelengths. Thus, reflected radiation from crop leaves and canopies has been used both to estimate chlorophyll content of crop canopies (Daughtry *et al.*, 2000; Filella *et al.*, 1995) and to assess nitrogen variability and stress (Blackmer *et al.*, 1994 and 1996). Over the last decade, several spectral chlorophyll indices have been developed to estimate chlorophyll content both at the leaf and at the canopy levels using hyperspectral remote sensing data of different crop types (Haboudane *et al.*, 2002; Blackburn, 1998a and 1998b; Chappelle *et al.*, 1992). Theoretically, the "ideal" chlorophyll index should be sensitive only to chlorophyll content in crop cover, but insensitive to soil background (colour, brightness, etc.), less sensitive to leaf area index (LAI) variations, independent of the spatial resolution of the sensors, and little affected by atmospheric and environmental effects, the drift of the sensor radiometric calibration, as well as solar illumination geometry and sensor viewing conditions, and not saturate rapidly. These effects intervene simultaneously during *in-situ* measurements and at the time of the airborne and/or satellite data acquisition. Consequently, it is impossible to design an index which is sensitive only to the desired variable and totally insensitive to all other parameters (Daughtry *et al.*, 2000; Bannari *et al.*, 1995; Kim *et al.*, 1994). However, in remote sensing, in spite of the correction and the standardization of the various radiometric distortions (topography, atmosphere, sensor drift, BRDF, etc.) chlorophyll indices remain always sensitive to the artifacts caused by soil optical properties particularly in an earlier stage of crop growth (sparse or fairly dense crop cover). Indeed, the effects of the underlying soil optical properties are very difficult to correct because not all soils are similar in the scene. Different soils have different spectral

reflectance behaviour because many factors influence the soil reflectance (Huete, 1989; Irons *et al.*, 1989). These are mineral composition, colour, brightness, moisture, organic matter content, salt and sodium content, roughness, and texture. In addition, size and shape of soil aggregates also influence the soil reflectance. These soil property variations affect the spectral response of soil and crop canopies and induce noise to the relationships between reflectance data and crop characteristics, such as LAI, absorbed photosynthetically active radiation (APAR), and chlorophyll content (Bannari *et al.*, 1996). These artifacts are likely to increase the chlorophyll index due to the spectral variations of the soils and not to an increase of the chlorophyll content. This chapter focuses on the evaluation and comparison of the sensitivity of several hyperspectral chlorophyll indices (PRI, NDPI, GNDVI, hNDVI, SIPI, SRPI, NPCL, PSSRa, PSNDa, MTCL, CAL, CARI, MCARI, and TCARI) to bare soils optical property variations using field spectroradiometric measurements as well as hyperspectral data acquired with the Probe-1 airborne and Hyperion EO-1 satellite hyperspectral sensors in the beginning of the growing season. With a sparse vegetation cover, the soil background becomes very important to consider using these indices

2. A Review of Spectral Chlorophyll Indices

Hyperspectral remote sensing is very often used to quantify plant photosynthetic pigment content. Plant pigments have a distinct spectral absorption characteristic, which means potential discrimination between them. Numerous studies and experiments have been undertaken in the search for spectral chlorophyll indices for accurate chlorophyll estimation at the leaf or the canopy level using hyperspectral remote sensing (i.e., laboratory and ground spectroradiometric measurements, model simulations and/or image data). Empirical approaches are based on simple relations established between chlorophyll content and spectral data, such as simple spectral analysis (Blackmer *et al.*, 1996; Mariotti *et al.*, 1996) and the analysis of the red-infrared spectral transition, the red-edge (Horler *et al.*, 1983; Guyot and Baret, 1988; Curran *et al.*, 1990 and 1991; Munden *et al.*, 1994; Pinar and Curran, 1996; Gitelson *et al.*, 1996; Filella and Peñuelas, 1994; Jago *et al.*, 1999; Zaroc-Tejada and Miller, 1999; Zarco-Tejada, 2000). In order to minimize the effects of soil background optical properties or the acquisition geometry (view/illumination) on the red-edge parameters, scientists have analyzed the potential of the first and second spectrum derivatives (Peñuelas *et al.*, 1994; Elvidge and Chen, 1995; Peñuelas and Filella, 1998; Jago *et al.*, 1999; Daughtry *et al.*, 2000; Gao, 2006). Other methods indicate that the logarithm of the inverse of the reflectance at specific wavelengths is well correlated with chlorophyll content (Peñuelas *et al.*, 1994; Balckburn, 1999). Semi-empirical approaches have a physical basis, but their mathematical formulation is related empirically to spectral data. In the literature, different indices for detecting and predicting chlorophyll status were developed (Baret *et al.*, 1988; Gamon *et al.*, 1992; Chappel *et al.*, 1992; Carter, 1994; Filella *et al.*, 1995; Jacquemoud *et al.*, 1996; Rollin and Milton, 1998; Blackburn, 1998a and 1998b; Haboudane *et al.*, 2002; Zhang *et al.*, 2008). A review of the spectral chlorophyll indices used in this chapter is given below, and their equations are presented in Table 1.

The *Photochemical Reflectance Index* (PRI) was developed to estimate the photosynthetic activity of canopies (Gamon *et al.*, 1992). It is a physiological reflectance index, which correlates (coefficient of determination $R^2 > 0.91$) with the epoxidation state of the xanthophylls cycle pigments (i.e., a particular group of carotenoids, violaxanthin,

antheraxanthin, and zeaxanthin), and with the efficiency of the plant canopy's photosynthesis. The epoxidation state is the content of xanthophylls cycle pigments. This xanthophylls cycle may be associated with a diurnal reduction in photosynthetic efficiency (Gamon *et al.*, 1992). Therefore, the epoxidation state of the xanthophylls cycle pigments may be a useful indicator of short-term changes in photosynthetic activity. In several studies, this index showed its usefulness in the assessment of radiation use efficiency at the canopy-level. Filella *et al.* (1996) showed that PRI is significantly correlated ($R^2 = 0.88$) with epoxidation, zeaxanthin, and photosynthetic radiation use efficiency for a cereal canopy. Penuelas *et al.* (1997) found significant results to assess photosynthetic radiation use efficiency at the leaf-level in Mediterranean trees, *Quercus ilux* and *Phillyrea latifolia*. Zarco-Tejada *et al.* (2005) showed that PRI is more sensitive to Chl-ab / carotenoid ratios ($R^2 = 0.50$) than to Chl-ab alone ($R^2 = 0.45$) or the carotenoid content for *Vitis vinifera* L ($R^2 = 0.27$) leaves. The PRI was found not affected as much by changing viewing angles for wheat chlorophyll content prediction using the *Compact High Resolution Imaging Spectrometer* (CHRIS) on the platform *PRoject for On-Board Autonomy* (PROBA). This is because the PRI is in the visible part of the spectrum and, therefore, is not influenced by anisotropy in the near infrared. However, this index performs better for wheat chlorophyll content estimation at the canopy level (biomass per unit ground area) than for wheat chlorophyll contents per leaf (Oppelt and Mauser, 2007).

Spectral Chlorophyll Indices	Authors
$PRI = (r_{550} - r_{531}) / (r_{550} + r_{531})$	Gamon <i>et al.</i> (1992)
$SRPI = r_{430} / r_{680}$	Penuelas <i>et al.</i> (1993)
$NDPI = (r_{430} - r_{680}) / (r_{430} + r_{680})$	Penuelas <i>et al.</i> (1993)
$NPCI = (r_{680} - r_{430}) / (r_{680} + r_{430})$	Penuelas <i>et al.</i> (1994)
$SIPi = (r_{800} - r_{445}) / (r_{800} - r_{680})$	Penuelas <i>et al.</i> (1995)
$GNDVI = (r_{801} - r_{550}) / (r_{801} + r_{550})$	Gitelson <i>et al.</i> (1996)
$PSND_a = (r_{800} - r_{680}) / (r_{800} + r_{680})$	Blackburn (1998a)
$PSSR(a) = r_{800} / r_{680}$	Blackburn (1998a)
$CARI = [(r_{700} - r_{670}) - 0.2 * (r_{700} - r_{550})]$	Kim <i>et al.</i> (1994)
$MCARI = [(r_{700} - r_{670}) - 0.2 * (r_{700} - r_{550})] * (r_{700} / r_{670})$	Daughtry <i>et al.</i> (2000)
$TCARI = 3 * [(r_{700} - r_{670}) - 0.2 * (r_{700} - r_{550}) * (r_{700} / r_{670})]$	Haboudane <i>et al.</i> (2002)
$CAI = \int_{r_{600}}^{r_{735}} r_{EQ} dx, \quad \text{where} \quad r_{EQ} = r_{Sc} / r_{ec} \cdot$	Oppelt & Mauser (2001)
where r_{Sc} is the reflectance of the vegetation spectrum at band c, r_{ec} is the reflectance of the envelope at band c, and r_{EQ} is the envelope quotient (see Oppelt and Mauser (2001 and 2004) for provision of more details about this index and for the calculation and extraction of all these parameters).	
$HNDVI = (R_{827} - R_{668}) / (R_{827} + R_{668})$	Oppelt & Mauser (2004)
$MTCI = (r_{753.75} - r_{708.75}) / (r_{708.75} - r_{681.25})$	Dash and Curran (2004)

Table 1. Equations of chlorophyll indices (r_λ indicates the reflectance in a band centered at a specific wavelength λ).

The *Simple Ratio Pigment Index* (SRPI) based on the ratio of the carotenoid and Chl-a content was proposed by Penuelas *et al.* (1993). Penuelas *et al.* (1993 and 1994) found that SRPI correlates well ($R^2 > 0.95$) with different levels of mite attacks in apple trees, as the carotenoid / Chl-a ratio increases with increasing level of mite attack. Similar performances were observed for the same ratio from a wide range of leaves from different species (maize, wheat, tomato, soybean, sunflower, sugar beet, and maple) when the SRPI was highly correlated ($R^2 > 0.95$) with carotenoid / Chl-a ratio (Penuelas *et al.*, 1995). The SRPI was found to be slightly sensitive to low chlorophyll content ($< 50 \text{ mg} / \text{cm}^2$). Penuelas *et al.* (1995) also demonstrated that this index is very sensitive to the leaf structure.

The *Normalized Difference Pigment Index* (NDPI) was proposed by Penuelas *et al.* (1993) in the same way as SRPI to evaluate the ratio of total pigments to Chl-a. Penuelas *et al.* (1993, 1994 and 1995) found that in maize, wheat, tomato, soybean, sunflower, sugar beet, maple, and aquatic plants the NDPI was highly correlated ($R^2 \geq 0.91$) with the ratio of total carotenoids and Chl-a measured at the leaf and plant levels. This index was found sensitive to the leaf surface and structure (Araus *et al.*, 2001). For wheat chlorophyll content estimation in intermediate stage development (approximately 70% of the fields were covered by wheat crop) using the Hyperion EO-1 data against those derived from the SPAD-502 measurements and chemical laboratory analysis, the NDPI showed satisfactory results with an index of agreement of 0.66 and a root mean square error (RMSE) of $2.89 \text{ } \mu\text{g}/\text{cm}^2$ (Bannari *et al.*, 2008).

In a study related to nitrogen (N) and water in sunflower leaves, Penuelas *et al.* (1994) proposed the *Normalized Pigment Chlorophyll Ratio Index* (NPCI). This index varies with total pigment and chlorophyll content and is associated with plant physiological state. This index is sensitive to the proportion of total photosynthetic pigments to chlorophyll, particularly applicable to N stress (Penuelas *et al.*, 1994). For wheat crop (*Triticum aestivum* L.), NPCI was significantly correlated ($R^2 = 0.84$) with total chlorophyll content using field-based reflectance measurements (Riedell and Blackmer, 1999). Exploring a wide range of hyperspectral chlorophyll indices, laboratory based-reflectance data and wheat leaf chlorophyll content estimated from chemical laboratory analysis, Bannari *et al.* (2007a) found that NPCI is significantly correlated ($R^2 = 0.84$) with Chl-ab / Chl-a ratio than with the Chl-ab content only.

Considering a wide range of leaves from several species (corn, wheat, tomato, soybean and sunflower) with the aim of assessing the pigment ratio, Penuelas *et al.* (1995) proposed the *Structure Insensitive Pigment Index* (SIPI). They established an empirical estimation of the carotenoid / Chl-a ratio and found that SIPI provided the best estimate for a range of individual leaves of different species (maize, wheat, tomato, soybean, sunflower, sugar beet, and maple) and conditions ($R^2 \geq 0.95$). Blackburn (1998a) confirmed that this index has a curvilinear relationship with the carotenoid / Chl-a ratio, which is best described using a logarithmic model (i.e., this model gives the highest coefficient of determination: $R^2 = 0.86$). The SIPI lacks sensitivity for low values of the carotenoid / Chl-a ratio and becomes more sensitive for higher values. Using physical simulation on *Vitis vinifera* L. leaves, Zarco-Tejada *et al.* (2005) demonstrated that SIPI is more sensitive to carotenoids and Chl-ab / carotenoid ratio than to chlorophyll Chl-ab content alone.

The *Green Normalized Difference Vegetation Index* (GNDVI) was developed by Gitelson *et al.* (1996) using a green band in a study related to the remote sensing of global vegetation and EOS-MODIS (*Earth Observing System - Moderate Resolution Imaging Spectroradiometer*) data. The development of the GNDVI is based on the idea that an index for chlorophyll estimation should be invariant with respect to pigments other than chlorophyll and should not be influenced by factors including background and atmosphere. Blackburn (1999) reported that there is a curvilinear relationship between GNDVI and the total chlorophyll content ($R^2 = 0.82$). He used this index in a laboratory experiment using stacks of leaves, obtained from four species of deciduous trees at various stages of senescence. He observed that over the wide range of chlorophyll contents, which can be experienced at the canopy scale, GNDVI was found to be sensitive to low content (~ 500 mg / cm²) only. The use of reflectance in a narrow green band, r_{550} , rather than r_{Green} (reflectance in the range of 540-570 nm) in the formulation of this index did not improve the relationship with total chlorophyll content. In addition, no relationship was found between GNDVI and matorral vegetation canopy chlorophyll content per unit ground area (Blackburn, 1999).

The *Pigment Specific Simple Ratio* (PSSR) was proposed by Blackburn (1998a) for the estimation of Chl-a (PSSR_a), Chl-b (PSSR_b), and carotenoids (PSSR_c) contents at the leaf level using samples from deciduous trees at various stages of senescence. Blackburn (1998b) found that PSSR_a has a strong relationship with Chl-a ($R^2 = 0.97$), and McNairn *et al.* (2001) reported the same conclusion for Chl-a estimation in corn and beans. However, the PSSR_c failed to predict carotenoid content in individual leaves of four deciduous tree species at various stages of senescence (Blackburn, 1998b). In another experiment on matorral vegetation canopy, Blackburn and Steele (1999) reported a reasonably linear relationship between PSSR_a and Chl-a content per unit ground area ($R^2 = 0.71$). However, the relationship was much weaker for low Chl-a contents (0 to 500 mg / cm²). Lower linear relationships were also found between PSSR_b and Chl-b ($R^2 = 0.68$) and PSSR_c and carotenoid ($R^2 = 0.50$) content per unit ground area. In this study, only the PSSR_a was considered.

Blackburn (1998a) also developed the *Pigment Specific Normalized Difference* (PSND) index to estimate individual pigment Chl-a (PSND_a), Chl-b (PSND_b), and carotenoid (PSND_c) contents. Like the PSSR indices, the PSND_a and PSND_b were found to have a strong exponential relationship with Chl-a and Chl-b ($R^2 > 0.91$), respectively. However, PSND_c failed to predict carotenoid content in individual leaves of four deciduous tree species at various stages of senescence (Blackburn, 1998b). Blackburn and Steele (1999) found a lower correlation ($R^2 < 0.51$) between these indices and the pigment contents of matorral vegetation canopies per unit ground area than in his previous studies (Blackburn, 1998a and b). The author suggested that this is due to variable background conditions and the structural/spectral complexity of the study sites. As for the PSSR_a, only the PSND_a is used in this study.

The *Chlorophyll Absorption in Reflectance Index* (CARI) was developed by Kim *et al.* (1994) and was designed to reduce the variability of photosynthetically active radiance due to the presence of diverse non-photosynthetic materials. Due to the sensitivity of CARI to soil background, Daughtry *et al.* (2000) presented the *Modified Chlorophyll Absorption in*

Reflectance Index (MCARI). The main change from CARI is the introduction of the ratio (r_{700} / r_{670}) to minimize the combined effect of the underlying soil reflectance and the canopy non-photosynthetic materials. Even though this index was developed to be both responsive to chlorophyll variations and resistant to non-photosynthetic material effects, Daughtry *et al.* (2000) reported that the MCARI is still influenced by the optical properties of the soil background. In order to minimize the underlying soil contribution, they suggested that the MCARI be normalized with a soil line vegetation index like the *Optimized Soil-Adjusted Vegetation Index* (OSAVI; Rondeaux *et al.*, 1996). Combining these spectral indices will further reduce the background contributions and enhance the sensitivity to leaf chlorophyll content variability at the same time. Daughtry *et al.* (2000) found that the MCARI / OSAVI ratio was linearly related to leaf chlorophyll contents ($R^2 = 0.87$) over a wide range of foliage cover of corn (*Zea mays* L.) and soil backgrounds. The combined use of the spectral indices MCARI and OSAVI was successful in producing an accurate assessment of crop chlorophyll contents from remote sensing data (Daughtry *et al.*, 2000). However, this normalization combination was not implemented for predictive purposes, nor have further developments dealt with LAI effects on pigment content estimation from canopy reflectance measurements. In addition, Haboudane *et al.* (2002) noted the limited sensitivity of MCARI for low pigment contents ($> 5 \mu\text{g} / \text{cm}^2$).

Haboudane *et al.* (2002) presented another variation of the MCARI, the *Transformed Chlorophyll Absorption in Reflectance Index* (TCARI). The main reason for developing TCARI was to improve sensitivity for low chlorophyll values of corn. However, according to these authors, this index is sensitive to the underlying soil properties, particularly for low LAIs (< 2.5). In the same study of integrated narrow-band indices for corn-crop chlorophyll prediction, Haboudane *et al.* (2002) proposed the TCARI / OSAVI ratio. The use of this ratio enabled accurate prediction of corn chlorophyll content from hyperspectral remote sensing imagery. These authors established a scaling-up relationship to make chlorophyll estimations as a function of the TCARI / OSAVI ratio derived from above canopy reflectance using *Compact Airborne Spectrographic Imager* (CASI) data. The ratio was found to be relatively insensitive to canopy cover variations, even for very low LAIs (< 1.5). The best fits were obtained for a logarithmic and polynomial function with R^2 values exceeding 0.98. Zarco-Tejada *et al.* (2005) showed that the TCARI / OSAVI was successfully correlated with the Chl-ab content at the canopy scale of *Vitis vinifera* L. using *Reflective Optics Spectrometric Imaging System* (ROSIS) and CASI data ($R^2 = 0.67$). Huang *et al.* (2004) found a significant logarithmic relationship ($R^2 = 0.78$) between TCARI / OSAVI derived from field reflectance measurements and wheat chlorophyll arbitrary values measured with the SPAD (Soil-Plant Analyses Development)-502 meter. Exploring CHRIS-PROBA data for wheat crop chlorophyll content prediction in shaded and sunlit portions of the field, Oppelt and Mauser (2007) showed that using the off-forward-looking angle ($+ 36^\circ$ from the nadir) chlorophyll content per leaf area is weakly correlated with TCARI / OSAVI ($R^2 = 0.56$ and $R^2 = 0.49$ for the sunlit and shaded sides, respectively). However, this index showed a relatively high correlation ($R^2 = 0.71$) with chlorophyll content per biomass of the sunlit portion in the nadir viewing direction. They concluded that the TCARI / OSAVI ratio is significantly sensitive to the viewing angle geometry. In another study for the Chl-ab content estimation at the canopy scale of *Vitis vinifera* L. using CASI reflectance spectra, Meggio *et al.* (2008) demonstrated that BRDF effects significantly affect the TCARI / OSAVI ratio. According to

Wu *et al.* (2008), if different disturbances sources such as shadow, soil background, and non-photosynthetic materials, were considered, the integrated indices TCARI / OSAVI and MCARI / OSAVI are appropriate for chlorophyll estimation of different types of corns with high R^2 of 0.88 and 0.94, respectively. In addition, these authors indicated that these two indices could be used to estimate the chlorophyll of wheat using Hyperion data (R^2 of 0.68 and 0.76 for TCARI / OSAVI and MCARI / OSAVI, respectively). However, Kneubühler (2002) and Bannari *et al.* (2007 and 2008) showed that these indices, developed specifically for corn, performed poorly at the wheat canopy-level using Hyperion hyperspectral data. As well, Haboudane *et al.* (2008), the developer of the TCARI / OSAVI ratio, found that this ratio seems to be a good estimator of leaf chlorophyll content for corn canopies using CASI hyperspectral reflectance data ($R^2 = 0.73$), but weak for wheat chlorophyll content estimation ($R^2 = 0.29$).

The *Normalized Difference Vegetation Index* (NDVI) was proposed by Rouse *et al.* (1974) and was used in various regional and global applications for studying the state of vegetation using multispectral remote sensing (Bannari *et al.*, 1995). When hyperspectral data are used, the name of this index becomes the *Hyperspectral Normalized Difference Vegetation Index* (HNDVI). Oppelt and Mauser (2004) used the HNDVI in a study for monitoring physiological parameters of wheat. They found that HNDVI and OSAVI become insensitive at chlorophyll contents below 0.3 g/m^2 as well as above 1.5 g/m^2 . Nevertheless, it important to mention that these indices were developed for biomass and yield estimation, but not for chlorophyll content prediction.

The *Chlorophyll Absorption Integral* (CAI) was proposed by Oppelt and Mauser (2001) for chlorophyll content of maize (*Zea mays*) derived from *Airborne Visible/near-infrared Imaging Spectrometer* (AVIS) data. This index involves the position of the red edge as well as the depth of chlorophyll absorption at 680 nm (Chl-a) and 650 nm (Chl-b). According to these authors, CAI shows a very good correlation with the maize chlorophyll content per unit area or per unit mass ($R^2 \geq 0.92$), and it could be a good predictor of wheat canopy chlorophyll content provided the dependence of the chlorophyll level on the wheat crop variety is taken into consideration. Oppelt and Mauser (2007) demonstrated that the CAI is the appropriate index to assess the chlorophyll content of both sunlit and shaded layers of wheat canopies using the data acquired with multi-angle CHRIS-PROBA sensor. However, even if the CAI was recommended for wheat crop chlorophyll content prediction (Oppelt and Mauser, 2004 and 2007), this index provide very poor results for wheat using Hyperion EO-1 data (Bannari *et al.*, 2008). Oppelt and Mauser (2001 and 2004) and Khurshid (2004) provide more details on the calculation and extraction of this index.

Using a ratio of the difference in reflectance between bands 10 and 9 and the difference in reflectance between bands 9 and 8 of the *Medium Resolution Imaging Spectrometer* (MERIS), Dash and Curran (2004) developed the *MERIS Terrestrial Chlorophyll Index* (MTCI). This index is used by the European Space Agency to produce the land surface biomass, a MERIS level-2 product. Due to its moderate spatial resolution ($300 \text{ m} \times 300 \text{ m}$) and three-day re-visit time, MERIS is a potentially valuable sensor for the measurement and monitoring of terrestrial environment at regional to global scales (Verstraete *et al.*, 1999). However, using laboratory reflectance measurements for wheat crop chlorophyll content estimation, Bannari

et al. (2007a) showed that it would be very difficult or impossible to interpret the MTCI values at low LAI in the precision agriculture context. Additionally, Haboudane *et al.* (2008) found that this index is very sensitive to the LAI, predicting a weak correlation with wheat chlorophyll content ($R^2 = 0.35$), but a higher one with corn chlorophyll ($R^2 = 0.81$) using CASI airborne data.

3. Material and Methods

3.1. Study Site and field data collection

The field, airborne and satellite data were collected in an agricultural region near Indian Head (50°N, 104°W), approximately 70 km east of Regina, Saskatchewan, Canada. The principal economic activities in this area are based on agriculture. Major crops grown are wheat, pea, canola and corn. This region was used in the context of a large project to investigate the potential of hyperspectral remote sensing in precision agriculture. In this project, the laboratory and the field measurements, and airborne and satellite hyperspectral data were investigated for plant water content estimation (Champagne *et al.*, 2003), nitrogen stress detection (Karimi *et al.*, 2005a and 2005b), LAI modelling and percent crop cover type mapping (Pacheco *et al.*, 2001, 2002, and 2008), crop residues estimation (Bannari *et al.*, 2006 and 2007b), and chlorophyll content prediction (McNairn *et al.*, 2001; Bannari *et al.*, 2007a and 2008).

The soils in the Indian Head and Regina regions are developed on lacustrine, alluvial lacustrine, alluvial glacio-fluvial, and glacial till parent materials in the brown, dark-brown, and black soil zone (Thie, 2006). Dark-brown chernozemic soils, which are the most productive of this group, occupy approximately 75% of the area. Thin black chernozemic soils (about 15 %) are developed on moderately fine and fine textured lacustrine parent materials. The brown soils (less than 2 %) are characterized by moderate-fine texture, and are generally rated as soils whereas the moderately coarse and coarse deposits. Gleysolic soils are generally most widely distributed, but are less represented in the study area. Azonal soils characteristic of the alluvium and hillwash complexer (about 10 %) occur in association with the main drainage channels or their adjacent floodplains. The objective of this study is not the characterization of each soil class using hyperspectral remote sensing, but the evaluation and comparison of the sensitivity of several chlorophyll indices to bare soils optical property variations. In order to achieve the goal of this investigation, different soils were selected based on the spatial representativeness of the major soil types from different agricultural lands with various optical and physico-chemical properties (colour, brightness, roughness, moisture, mineralogical composition, etc.). In the field, samples were taken from the soils upper layer (5 cm depth). Observations and remarks about each sample were noted and photographed using a 35 mm digital camera equipped with a 28 mm lens.

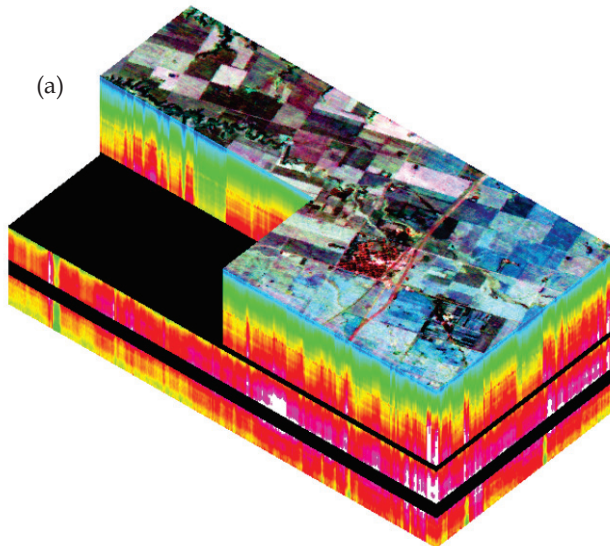
3.2. Hyperion data acquisition

Satellite hyperspectral data were acquired on June 30, 2002 using the Hyperion hyperspectral sensor on NASA's Earth Observer-1 (EO-1) platform over the Indian Head site (Figure 1a). The launch of the Hyperion sensor in November 2000 marked the first operational test of a space-borne hyperspectral sensor covering both the visible and near-infrared (VNIR) and the short-wave infrared (SWIR) spectral regions (Beck, 2003). This

sensor is a pushbroom imaging spectrometer that acquires data in the along-track direction. It collects the upwelling radiance in 242 spectral bands, each approximately 10 nm wide at full width half maximum (FWHM) with an average spectral sampling interval of 10 nm. Hyperion has a single telescope and consists of two spectrographs, one covering the VNIR wavelengths range from 357 to 1055 nm, the other, the SWIR from 851 to 2576 nm. Since Hyperion is a pushbroom sensor, the entire swath is obtained in a single frame with a ground sampling distance of 30 m. Its telescope images the Earth onto a slit with a field-of-view (FOV) of 0.624° , resulting in a swath width of 7.65 km from a 705 km altitude. Each data set acquired by this sensor covers a nominal along-track length of 40 km. Figure 1a illustrates the 3D cube of the used Hyperion EO-1 data.

3.3. Probe-1 data acquisition

The airborne hyperspectral data were acquired using the Probe-1 sensor (Earth Search Sciences Inc., 2001) on Jun 28, 2000 over the Indian Head site (Figure 1b). The Probe-1 is a "whiskbroom style" instrument that collects data in cross-track direction by mechanical scanning and in along-track direction by movement of the airborne platform. This sensor acquires up-welling radiance in 128 bands in the 400 to 2500 nm wavelengths region. The at-sensor radiance is dispersed by four spectrographs onto four linear detector arrays with 32 bands each. This sensor covers the wavelength range continuously with small gaps in the strong 1380 nm and 1870 nm atmospheric water vapor absorption regions. The bandwidth is between 11 and 18 nm at FWHM. Probe-1 was mounted on a three-axis gyrostabilizer to minimize geometric distortion from the aircraft movement. The flying altitude was 2500 m above ground for a swath width of 3 km and a spatial resolution of 5 m at nadir. A non-differential GPS was recording the location of the aircraft during the flight. Figure 1b shows the 3D cube of the used Probe-1 airborne hyperspectral data.



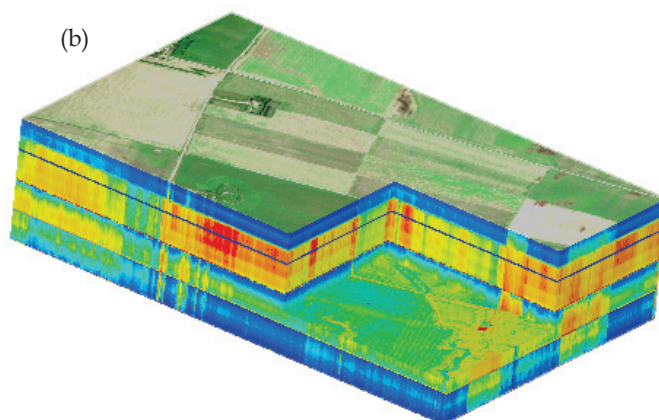


Fig. 1. Hyperspectral Hyperion EO-1 (a) and Probe-1 (b) 3D cubes.

3.4. Field spectroradiometric measurements

In order to achieve the goal of this investigation, spectroradiometric measurements were acquired above 90 bare soil plots selected from different agricultural lands with various optical and physicochemical properties (colour, brightness, roughness, moisture, mineralogical composition, etc.) using an ASD (*Analytical Spectral Devices*) spectroradiometer (ASD Inc., 1999). This instrument is equipped with two detectors operating in the VNIR and SWIR, between 350 and 2500 nm. It acquires a continuous spectrum with a 1.4 nm sampling interval from 350 to 1000 nm and a 2 nm one from 1000 to 2500 nm. The ASD resamples the measurements in 1-nm intervals, which allows the acquisition of 2151 contiguous bands per spectrum. The sensor is characterized by the programming capacity of the integration time, which allows a satisfactory signal-to-noise ratio as well as a great stability (ASD Inc., 1999).

Measurements were taken in the laboratory using two halogen lamps of 500 W each, equipped with an electrical current regulator. The data were acquired at nadir with a FOV of 25° and a solar zenith angle of approximately 5° by averaging twenty-five measurements. The spectroradiometer was installed on a tripod with a height of approximately 30 cm over the target, which makes it possible to observe a surface of approximately 177 cm². A laser beam was used to locate the center of the ASD FOV. The reflectance factor of each soil sample was calculated by rationing target radiance to the radiance obtained from a calibrated "Spectralon panel" in accordance with the method described in Jackson *et al.* (1980). Corrections were made for the wavelength dependence and non-lambertien behavior of the panel. Only the VNIR wavelengths from 430 to 850 nm were required to calculate the chlorophyll indices used in this study (Table 1). The specific and requested wavelength ranges for each chlorophyll index were used from these measurements. Figure 2 shows the reflectance of the soil samples.

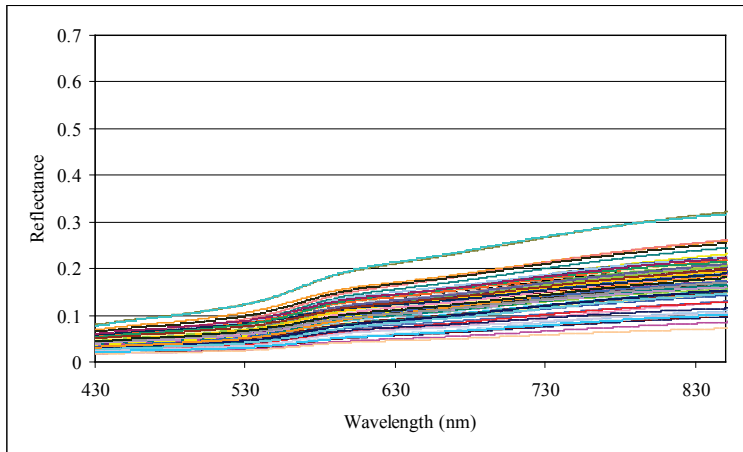


Fig. 2. Spectral signature of soils sampled at the field and measured in the lab using ASD.

3.5. Hyperion and Probe-1 data pre-processing

3.5.1. Hyperion radiometric and spectral calibration

Hyperion EO-1 hyperspectral data were pre-processed with an aim to correct for sensor artifacts and atmospheric and geometric effects (Khurshid *et al.*, 2006). The *Imaging Spectrometer Data Analysis System* (ISDAS) developed at the *Canada Centre for Remote Sensing* (Staez *et al.*, 1998) was used to perform all the pre-processing steps. The procedure begins with geometric corrections (shift and rotation) of the SWIR data to spatially register them to the VNIR data. In fact, the SWIR data were corrected for a single pixel offset from pixels 129 to 256 across-track and rotated by 0.22° . Furthermore, the striping problem due to systematic noise caused by factors such as detector non-linearity, movement of the slit with respect to the focal plane and temperature effects (Kruse *et al.*, 2003) was corrected. In addition, the column dropouts caused by dead pixels (a dead pixel is a functional failure of a single detector element during acquisition) were then removed from the whole image cube (Sun *et al.*, 2008). This was followed by noise reduction using recently developed automated software tools (Khurshid *et al.*, 2006). To achieve this step, the raw imagery (digital numbers) recorded by the sensor was converted to radiance using the radiometric calibration coefficients (gains and offsets) derived in the laboratory and provided by NASA.

The data cube was subsequently analyzed to characterize the distortions of keystone and spectral smile (Neville *et al.*, 2004 and 2008). At this step, the data were cropped to exclude noisy bands resulting in a final data set that spans the spectral range from 426.82 to 2355.20 nm with a total of 192 bands, excluding the overlap bands between the VNIR and SWIR spectrographs. Keystone is a term used in hyperspectral remote sensing to refer to the inter-band spatial mis-registration in imaging spectrometers (Neville *et al.*, 2004). These distortions may be caused by geometric distortions or by chromatic aberration, or a combination of both. Due to these distortions, a particular spatial pixel, corresponding to a specific detector element in the across-track dimension, in one specific band, will not be registered on the ground with the corresponding pixel in the other spectral bands. Neville *et*

al. (2004) reported that the Hyperion sensor has keystone distortions, ranging from - 0.05 to 0.49 pixels for the VNIR spectrometer and - 0.06 to 0.07 pixels for the SWIR spectrometer. For our data, the keystone distortion varied from a minimum of - 0.075 pixels to a maximum of 0.3 pixels for the VNIR and - 0.075 to 0.1 pixels for the SWIR (Khurshid *et al.*, 2006). No keystone corrections were performed due to the lack of an appropriate resampling procedure providing sufficient geometric accuracy without minimal increase in noise.

Furthermore, the spectral smile/frown is a wavelength shift, which is a function of the across-track pixel (column) in the swath (Neville *et al.*, 2008). In an ideal case, all pixels in the across-track dimension correspond to the same wavelength. This wavelength shift is due to many sources, such as spatial distortions caused by the dispersion element, prism or grating, or by aberrations in the collimator and imaging optics (Neville *et al.*, 2008). To achieve spectral calibration, the radiance spectra were analyzed to evaluate the bandwidth and band's center position using five known atmospheric absorption features: 760 nm (oxygen), 940 nm and 1130 nm (water vapor), and 2005 nm and 2055 nm (carbon dioxide). The correct band center wavelengths and bandwidths are determined by correlating the at-sensor Hyperion radiance with a modeled at-sensor radiance calculated with the radiative transfer (RT) code MODTRAN 4.2 (Berk *et al.*, 1999). Wavelength shifts of 1 to 3 nm in the VNIR and SWIR were detected and applied after the atmospheric correction process.

3.5.2. Probe-1 radiometric and spectral calibration

A laboratory calibration was completed for the Probe-1 sensor in April 2000 to obtain the dark current signal, radiometric coefficients, and to ascertain the center position of the spectral bands. However, a vicarious calibration of the sensor was required to correct for errors in gains and band centers, which resulted from the stresses experienced during transportation, installation and operation between the laboratory calibration and the over-flight (Secker *et al.*, 2001). This is an absolute calibration method, which produces a new set of gains that can be used to replace those derived in the laboratory.

To achieve spectral calibration of the Probe-1 data, the raw spectrum (digital numbers) recorded by the sensor was converted to radiance using the radiometric gains and offsets derived in the laboratory. As for the Hyperion data, the derived reflectance spectra were then analyzed to evaluate the band's center positions using five known atmospheric absorption features located at 760 nm (oxygen), 940 nm and 1130 nm (water vapor) and 2005 nm and 2055 nm (carbon dioxide). Wavelength shifts were then calculated, which best corrected the surface reflectance to obtain a smooth spectrum in the regions of these absorption features. These shifts were then applied to the Probe-1 data. The reflectance-based vicarious calibration was then applied in a next step to the data using the reflectance of an asphalt reference target acquired on the ground with an ASD spectroradiometer (Secker *et al.*, 2001). This site was then visually located in the Probe-1 imagery and an average spectrum was extracted for the calibration target. The spectrum from this target was then matched to the averaged ASD spectra of the same site. The differences between the two spectra were calculated and the Probe-1 radiometric coefficients were then adjusted to minimize the absolute reflectance difference until an error threshold was reached, which was set to 0.02 %. This process was carried out in ISDAS (Staez *et al.*, 1998) with an iterative

numerical technique which provided a new set of optimal gains. These new gains were then applied to the raw digital numbers to calculate at-sensor radiance for the dataset.

3.5.3. Hyperion and Probe-1 surface reflectance retrieval

The calibrated at-sensor radiance data were converted to surface reflectance using a look-up table (LUT) approach to correct for the atmospheric effects (Staez and Williams, 1997). Two five-dimensional raw LUTs, each one for a 5% and 60% spectrally flat reflectance, with tunable breakpoints were generated with the MODTRAN 4.2 RT code to provide additive and multiplicative coefficients for the removal of atmospheric scattering and absorption effects. For this purpose, the midlatitude-summer atmosphere model and a continental-rural aerosol model was used. The input parameters for the RT code for Hyperion and Probe-1 data are presented in Table 2. The initial LUTs were then convolved with the Hyperion and Probe-1 spectral sensor characteristics and used in combination with a curve-fitting technique in the 940 and 1130 nm water vapor absorption regions to estimate the atmospheric water vapour content from the datasets themselves on a pixel-by-pixel basis (Green *et al.*, 1991; Gao and Goetz, 1990). The column atmospheric water vapour estimates are then used to interpolate the LUTs to retrieve surface reflectance. Subsequently, the reflectance data were corrected for smile effects.

Input Parameters/Dataset	Probe-1 data	Hyperion data
Date of over flight	June 28, 2000	June 30, 2002
Time of over flight (GMT)	17:10:00	17:36:00
Aircraft heading	110°	N/A
Sensor altitude (above sea level)	3.079 km	705 km
Terrain elevation (above sea level)	579 m	579 m
Solar zenith angle	34.35°	31.72°
Solar azimuth angle	132.55°	142.17°
Atmospheric model	Mid-latitude Summer	Mid-latitude Summer
Aerosol model	Continental (rural)	Continental (rural)
Water vapour content	1.5 g/cm ²	1.5-2.5 g/cm ²
Ozone column (as per model)	0.319 cm-atm	0.319 cm-atm
CO ₂ mixing ratio (as per model)	357.5 ppm	365.00 ppm
Horizontal visibility	50 km	23 km

GMT= Greenwich Mean Time; ASL= Above Sea Level; ppm=parts per million

Table 2. Input parameters for the MODTRAN 4.2 radiative transfer code for Probe-1 data and Hyperion data.

Finally, the post-processing concluded the corrections by removing residuals that still remained after the correction of sensor artifacts and atmospheric effects. This step involves the calculation of correction gains and offsets using a spectrally flat target pixel approach (Staez *et al.*, 1999). The technique assumes that there are a number of pixels whose reflectance spectra are flat or nearly flat (feature-less), and their brightness range covers a major portion of the full range for all the pixels in the scene. A second-order polynomial fit to the reflectance spectra using χ -squared as a goodness of fit measure is calculated on a pixel-by-pixel basis. The pixels with the smallest χ -squared values are selected as "spectrally

flat target pixels". Finally, linear fits are performed on a band-by-band basis providing slopes and offsets, which are used as gain and offset for the correction of residual errors in the reflectance data.

After all these pre-processing steps, 60 spectral signatures representing different bare soils with various optical properties were extracted from bare soil of different agricultural fields from the Hyperion and Probe-1 imagery for the analysis (Figures 3 and 4).

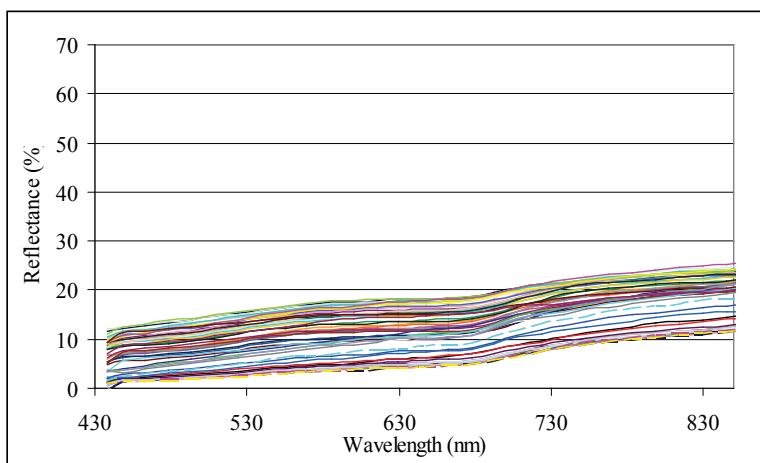


Fig. 3. Spectral signature of soils extracted from airborne Probe-1 imagery.

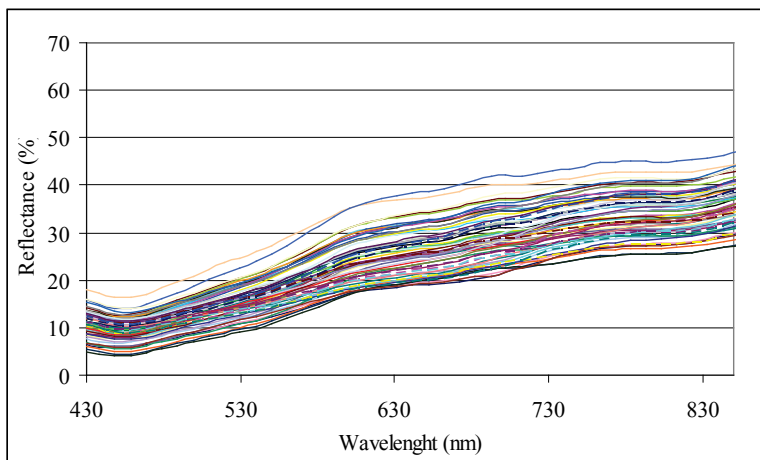


Fig. 4. Spectral signature of soils extracted from Hyperion EO-1 satellite imagery.

4. Results and discussion

For an accurate interpretation of hyperspectral chlorophyll indices, a “true” chlorophyll index value, attributed only to the green vegetation signal and free from any contribution of non-photosynthetic elements, is needed. Theoretically, if only bare soils were considered, with no vegetation, the value of the “ideal” chlorophyll index should be zero regardless of any changes in soil optical properties. Graphically, if a chlorophyll index is computed over bare soils and plotted against the reflectance of the shortest wavelength position used in the index, the clusters of sampling points should then be perfectly superimposed to the theoretical soil line (Huete *et al.*, 1985; Bannari *et al.*, 1996). Unfortunately, chlorophyll indices used to estimate vegetation pigments still show various levels of sensitivity to the effect of soil optical properties.

Index	Required Wavelength Position	Wavelength Position Available From the Hyperion Sensor	Wavelength Position Available From the Probe-1 Sensor
PRI	531 and 550 nm	529.66 and 550.01 nm	537.5 and 552.8 nm
SRPI	430 and 680 nm	428.0 and 682.29 nm	435.7 and 675.7 nm
NDPI	430 and 680 nm	448.47 and 682.50 nm	435.7 and 675.7 nm
CARI	550, 670 and 700 nm	550.22, 672.32 and 702.01 nm	552.8, 675.7 and 705.2 nm
NPCI	430 and 680 nm	448.47 and 682.50 nm	435.7 and 675.7 nm
SIPI	445, 680 and 800 nm	448.47, 682.50 and 803.54 nm	446.2, 675.7 and 797.0 nm
GNDVI	550 and 801 nm	550.22 and 803.54 nm	552.8 and 797.0 nm
PSND _a	680 and 800 nm	682.54 and 803.54 nm	675.7 and 797.0 nm
PSSR _a	680 and 800 nm	682.50 and 803.54 nm	675.7 and 797.0 nm
MCARI	550, 670, 700 and 800 nm	550.22, 672.32, 702.01 and 803.54 nm	552.2, 675.7, 705.2 and 797.0 nm
HNDVI	668 and 827 nm	671.02 and 823.65 nm	675.7 and 827.6 nm
CAI	600 and 735 nm	599.79 and 732.07 nm	599.0 and 735.8 nm
TCARI	550, 670 and 700 nm	550.22, 672.32 and 702.01 nm	552.8, 675.7, 705.2 and 797.0 nm
MTCI	681.25, 708.75 and 753.75 nm	682.29, 712.50 and 753.17 nm	675.7, 705.2 and 751.0 nm

Table 3. Available wavelength positions from the Hyperion and Probe-1 sensors for spectral chlorophyll indices calculation.

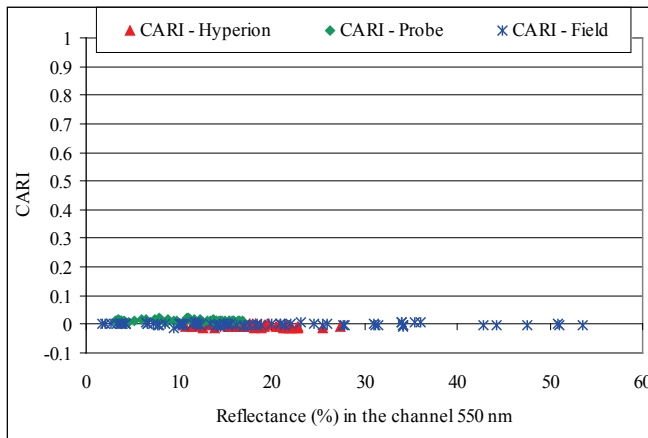
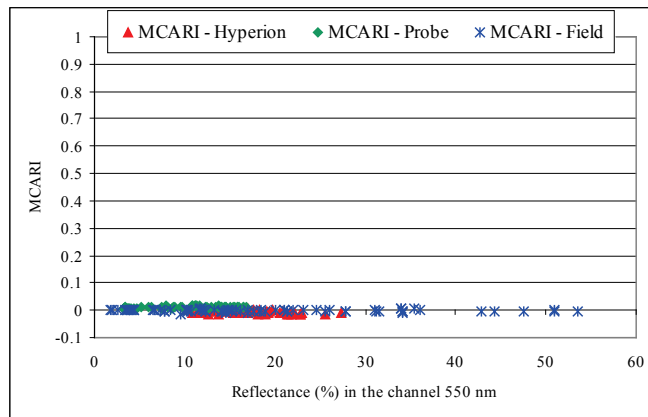
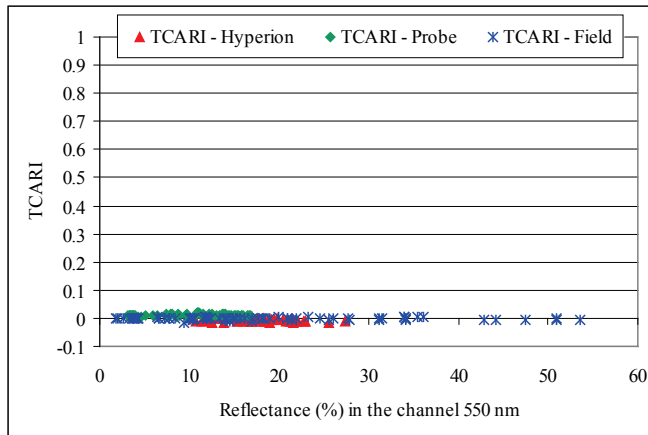
To evaluate the sensitivity of chlorophyll indices, included in this study, to changes of soil optical properties, the relationship between each of these indices and the shortest wavelength position involved in their formula was analyzed for different bare soils observed at three different scales using a field spectroradiometer, an airborne sensor (Probe-1), and a space-borne sensor (Hyperion EO-1). Hence, while field spectroradiometric measurements provide the wavelength positions needed for each index, spectra extracted from Probe-1 and Hyperion images do not contain all the exact wavelength positions

required for these indices. Thus, wavelengths were selected which most closely match the wavelengths position proposed for the indices investigated (Table 3).

The relationships between chlorophyll indices and their shortest wavelengths, over bare soils, are not unique; they show a considerable scatter caused by changes in soil optical properties. In fact, these indices were designed from leaf or canopy spectra to measure vegetation pigments. To understand this influence, chlorophyll indices selected for this study were plotted against their shortest wavelength position as illustrated in Figure 5. It shows that in addition to having different levels of sensitivity to soil properties variation, indices studied exhibit different behaviour and trends expressed in terms of the distance between index values and the theoretical soil line and the scatter magnitude of their clusters within the scatter-plot.

The first group, characterized by a horizontal trend showing clouds of points generally parallel to the X-axis of the scatter-plot (theoretical soil line), include CARI, MCARI, TCARI, PRI and MTCI indices (Figure 5-A). They have in common the use of wavelengths from the red-edge, red, and green spectral regions except for PRI which uses only wavelengths from the green portion of the solar spectrum. Such wavelengths are known to be the most sensitive to leaf chlorophyll variations and relatively influenced by changes in soil optical properties. This may explain their constant behavior as a function of their shortest wavelength. It can be seen in Figure 5-A that these indices behavior was not affected by the source of data: observed trends are similar, with index values falling basically within the same cloud of points for ground, airborne, and space-borne data. Regarding the sensitivity magnitude to the soil optical properties, indices of this group exhibit the best overall performance in terms of resistance to soil background effects with exception of MTCI which is very sensitive.

The second group consists of indices showing higher sensitivity to soil background at low reflectance levels of the shortest wavelength, which corresponds to the conditions of dark or developed soils. Figure 5-B shows that the indices HNDVI, PSNDa, and PSSRa follow a horizontal trend for the shortest wavelength reflectance values exceeding 20%. They appear to be more responsive to soil optical properties when this reflectance tends to decrease, causing a sudden change of the trend with a steep negative slope for reflectance values of less than 20%. Index values of up to 0.5 and 2.5 occur for PSNRa and HNDVI, and PSSRa, respectively. As shown in Table 1, indices of this group either have the particularity of being a simple ratio (PSSRa) or normalized differences (HNDVI and PSNDa) of red and near-infrared wavelengths. These are examples of traditional vegetation indices which are not adjusted for soil optical effects. As expected in this group, PSSRa is the most sensitive to soil influence because as a simple ratio it does not attenuate the background contribution to the observed signal.



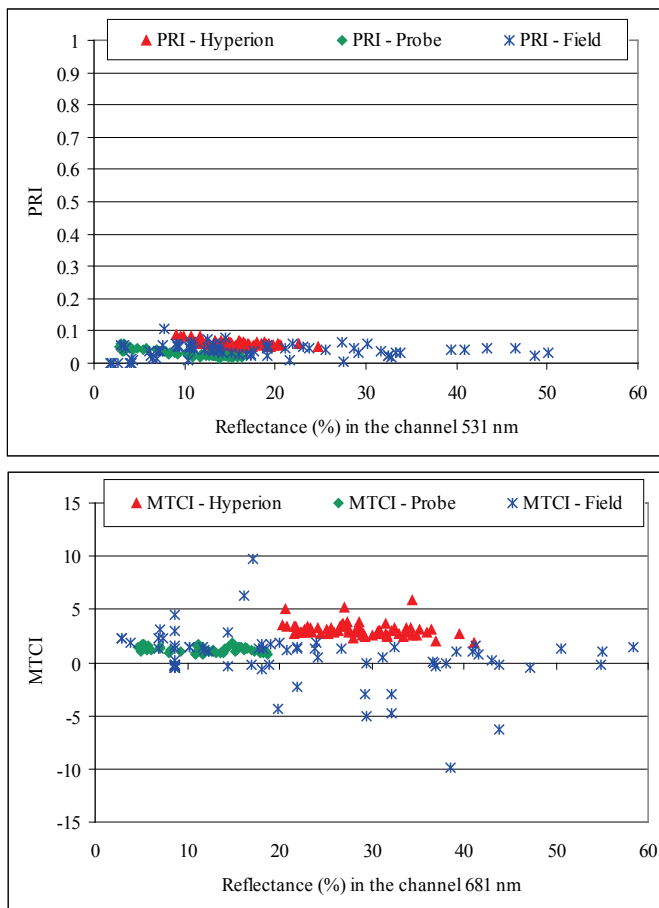


Fig. 5-A. Sensitivity of chlorophyll indices to soil optical properties using field, airborne and satellite hyperspectral data for the indices TCARI, MCARI, CARI, PRI, and MTCI.

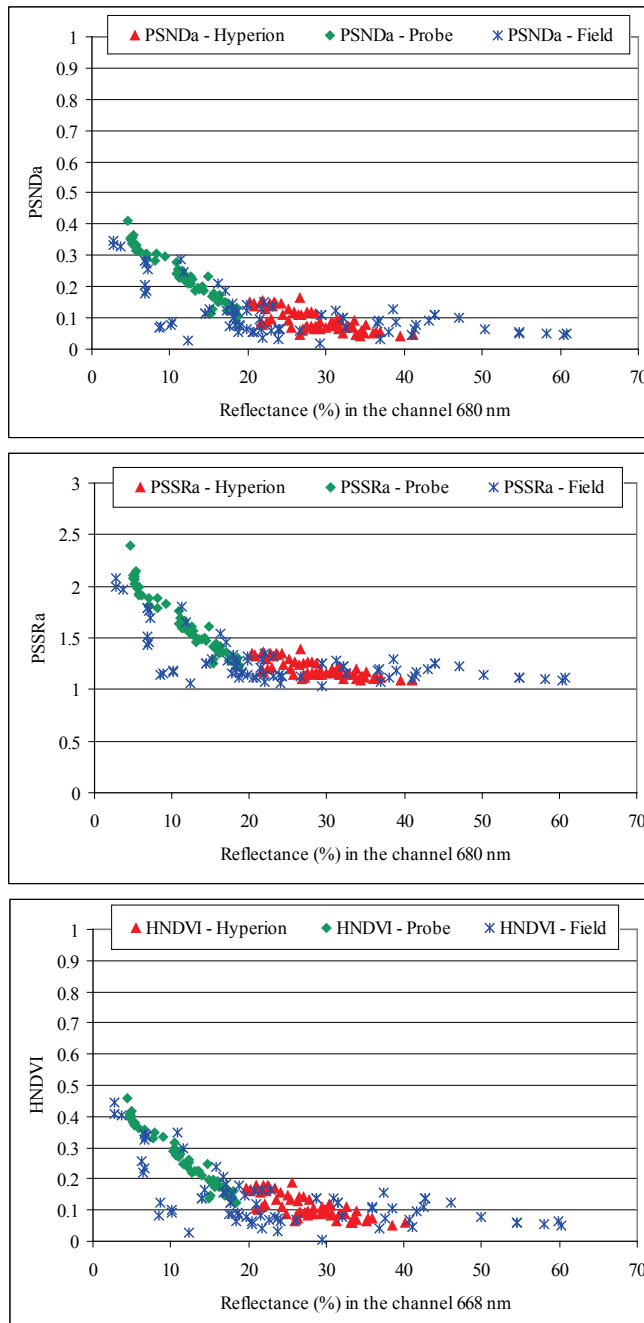


Fig. 5-B. Sensitivity of chlorophyll indices to soil optical properties using field, airborne and satellite hyperspectral data for the indices PSNDa, PSSRa and HPDVI.

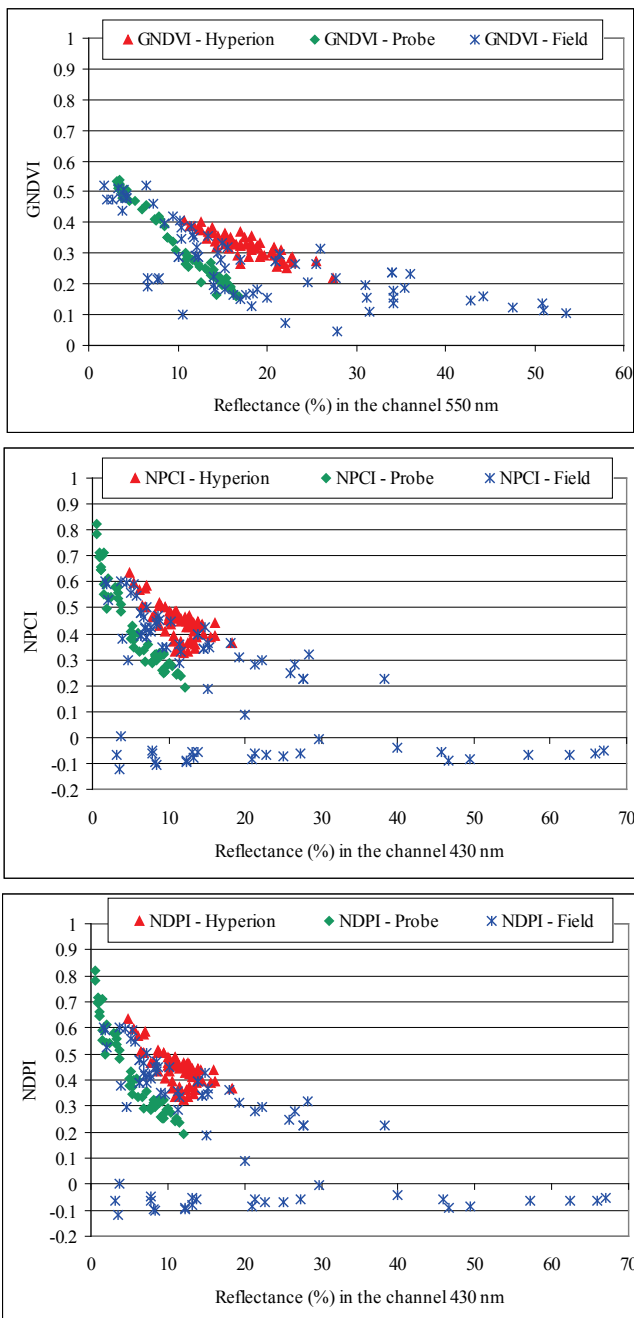


Fig. 5-C. Sensitivity of chlorophyll indices to soil optical properties using field, airborne and satellite hyperspectral data for the indices GNDVI, NPCI and NDPI.

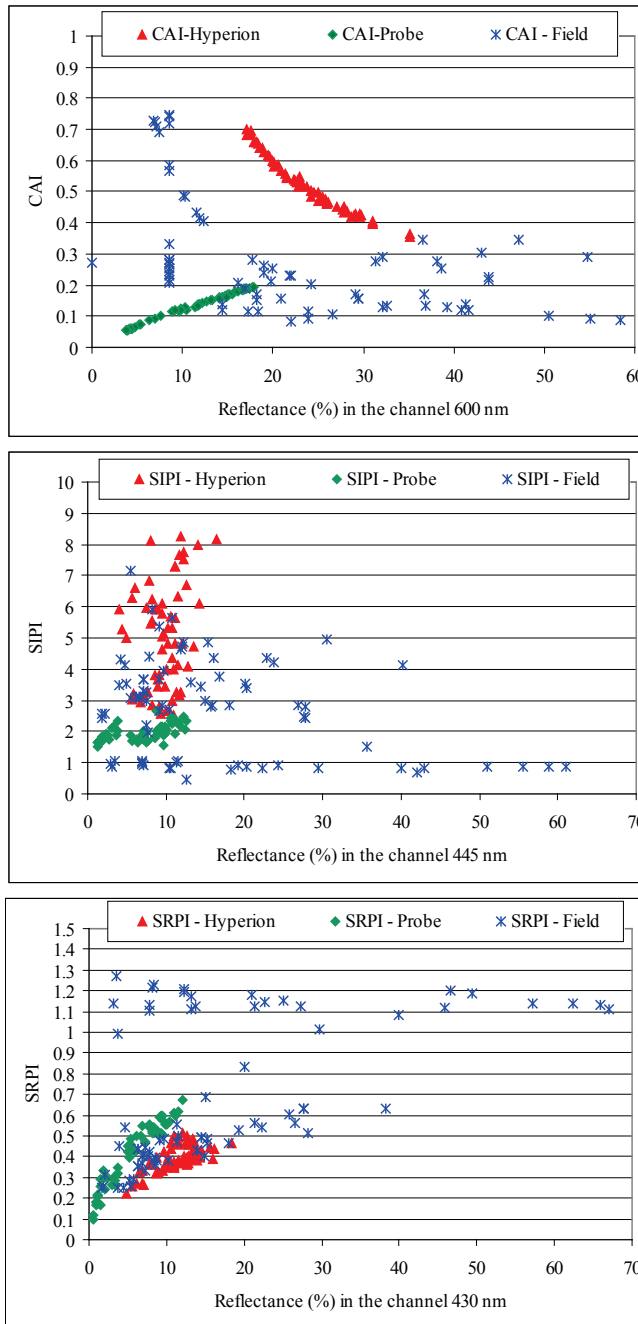


Fig. 5-D. Sensitivity of chlorophyll indices to soil optical properties using field, airborne and satellite hyperspectral data for the indices CAI, SIPI and SRPI.

The third group includes indices formed as normalized differences of reflectances in the blue and red regions (NDPI and NPCI) or the green and near-infrared regions (GNDVI). The distribution of their values within the "index *versus* shortest wavelength" space in Figure 5-C does not follow a clear and well common trend, and they show a considerable spread of the points representing soil spectra regardless of their data source. These indices are significantly inconsistent when the reflectance of the shortest wavelength is lower than 40%. Index values vary between 0.0 and 0.6 for GNDVI and between - 0.2 and 0.9 for NDPI and NPCI, with a negative slope, as the reflectance on the X-axis increases.

The last group has no distinctive common trend as shown in Figure 5-D; each of its indices (CAI, SRPI and SIPI) behaves in a completely different way than all the other indices selected for this study. They have in common the fact that each of them is sensitive up to various degrees to soil background variability. Trend dissimilarities could be explained by the rationale behind the design of each of these indices: SRPI is a simple ratio, SIPI is a ratio of reflectance differences, and CAI is an integral of the chlorophyll absorption feature. Nevertheless, it seems that SIPI is the worst chlorophyll index (up to 8.5) with respect to the resistance against the influence of the changes in soil optical properties; it shows the highest sensitivity amongst all the indices used in the present study (Figure 5-D).

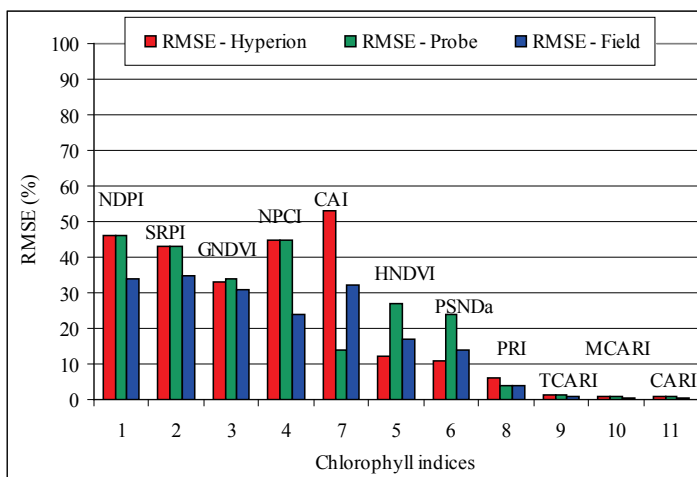


Fig. 6. RMSE related to the sensitivity of the chlorophyll indices to the optical proprieties of bare soils using field, airborne and satellite hyperspectral data (SIPI, MTCI, and PSSRa are not included in this figure, because they show a very high RMSE).

To complete the analyses discussed above, the RMSE was determined for each of the indices selected for this study, using it as a measure to evaluate the performance of the indices in terms of their resistance to soil background effects. Obtained results are summarized in Table 4 and illustrated in Figure 6. As it can be seen in Table 4, the relative magnitude of the RMSE allows the discrimination of three groups of sensitivity to soil effects. The first group includes the SIPI, PSSRa and MTCI indices, which have very high RMSEs related to the optical properties of bare soils. Accordingly, chlorophyll results

would be impossible to interpret properly in sparse canopies. The second group consists of SIPI, SRPI, NDPI, NPCI, GNDVI, CAI and HNDVI which have non-negligible RMSEs related to the optical properties of bare soils, and would be very difficult to interpret at low LAI. The group also includes indices PSND_a and HNDVI which yielded an RMSE lower than 20%, but still significant. The most soil resistant index of this group is the PRI with a RMSE of less than 6%. Finally, in the third group, and regardless of the data source and of the soil background, the indices CARI, MCARI and TCARI with an RMSE of less than 1.2% are basically not sensitive to changes in the soil optical properties. Therefore, they would permit a better estimation of chlorophyll content in sparse crop cover environment in the context of precision agriculture.

Chlorophyll Indices	RMSE (%) From Field	RMSE (%) From Airborne (Probe-1)	RMSE (%) From Satellite (Hyperion EO-1)
SIPI	201.0	199.0	526.0
MTCI	252.0	125.0	314.0
PSSR _a	130.0	165.0	120.0
NDPI	34.0	46.0	46.0
SRPI	35.0	43.0	43.0
CAI	32.0	14.0	53.0
GNDVI	31.0	36.0	33.0
NPCI	24.0	45.0	45.0
HNDVI	17.0	27.0	12.0
PSND _a	14.0	14.0	11.0
PRI	4.0	4.0	6.0
TCARI	0.8	1.2	1.1
MCARI	0.4	0.9	0.8
CARI	0.4	0.9	0.8

Table 4. RMSE related to the sensitivity of the considered chlorophyll indices to the optical properties of bare soils.

5. Conclusion

Hyperspectral indices developed for chlorophyll content estimation using crop canopy reflected radiation are sensitive to other vegetation and environmental parameters like underlying soil reflectance. This chapter focuses on evaluating and comparing the sensitivity of several chlorophyll indices (PRI, NDPI, GNDVI, HNDVI, SIPI, SRPI, NPCI, PSSR_a, PSND_a, CARI, MCARI, TCARI, CAI and MTCI) to bare soil optical property variations. In order to achieve the goal of this investigation, field spectroradiometric measurements were used as well as hyperspectral data acquired with the Probe-1 airborne and Hyperion EO-1 satellite sensors. Spectroradiometric measurements were acquired from 90 bare soil plots with various optical properties and selected from different agricultural lands. After the image data pre-processing steps, sixty spectral signatures of different bare soils with various optical properties were extracted from each image data set and used for the analysis. The results show that SIPI, PSSR_a and MTCI indices have a very high RMSE related to optical background variations; their results would be impossible to interpret correctly. The indices

SIPI, SRPI, NDPI, NPCI, GNDVI, CAI and HNDVI have non-negligible RMSEs related to the optical properties of bare soils, and will be very difficult to interpret at low LAIs such as found in sparse vegetation. The PSNDa and HNDVI show an RMSE less than 20%; this error magnitude is still significant. The PRI index sensitivity varies slightly as a function of soil characteristic variation, and shows an RMSE less than 6%. Independently from the data source (ground, airborne, and space-borne) and from the bare soil background, the indices CARI, MCARI and TCARI with an RMSE of less than 1.2% are basically not sensitive to changes in the soil optical properties; their use will permit a better estimation of chlorophyll content in sparse crop cover environments.

6. Acknowledgments

The authors would like to thank the National Science and Engineering Research Council (NSERC) and the University of Ottawa's Faculty of Arts for their financial support. The authors would also like to acknowledge the Canada Centre for Remote Sensing (CCRS) and Canadian Space Agency (CSA), which provided Probe-1 and Hyperion hyperspectral data. We would like to thank numerous people who were involved in this project and who provided their support and expertise: Lixin Sun (CCRS), Jean-Claude Deguise (CCRS), Dr. Robert Neville (CCRS), Rob Hitchcock (Defense Research and Development Canada, Ottawa) and all the field teams at Indian Head, Saskatchewan.

7. References

- Analytical Spectral Devices, ASD Inc. (1999)
<http://www.asdi.com/products-spectroradiometers.asp>.
- Araus, J.L., Casadesus, J. and Bort, J. (2001) Recent Tools for the Screening of Physiological Traits Determining Yield. *In* Application of Physiology in Wheat Breeding, Edited by Reynolds, M.P., Ortiz-Monasterio, J.I. and McNab, A. CIMMYT, International Maize and Wheat Improvement Center Apdo. Postal 6-641, 06600 Mexico, D.F., Mexico, 240 pages.
- Bannari, A., Morin, D., Bonn, F. and Huete, A.R. (1995) A review of vegetation indices. *Remote Sensing Reviews*, **13**:95-120.
- Bannari, A., Huete, A.R., Morin, D. et Zagolski, F. (1996) Effets de la couleur et de la brillance du sol sur les indices de végétation. *International Journal of Remote Sensing*, **17**(10): 1885-1906.
- Bannari, A., Pacheco, A., Staenz, K., McNairn, H. and Omari, K. (2006) Estimating and Mapping Crop Residue Cover in Agricultural Lands Using Hyperspectral and IKONOS data. *Remote Sensing of Environment*, **104**: 447-459.
- Bannari, A., Khurshid, K.S., Staenz, K. and Schwarz, J. (2007a) A Comparison of Hyperspectral Chlorophyll Indices for Wheat Crop Chlorophyll Content Estimation Using Laboratory Reflectance Measurements. *IEEE Transaction on Geosciences and Remote Sensing*, **45**(10): 3063-3073.
- Bannari, A., Staenz, K. and Khurshid, K.S. (2007b) Remote Sensing of Crop Residue Using Hyperspectral Hyperion (EO-1) Imagery. *International Geosciences and Remote Sensing Symposium (IGARSS-2007)*, 23-27 July 2007, Barcelona, Spain, pp. 2795-2799.

- Bannari, A., Khurshid, S.K., Staenz, K. and Schwarz, J. (2008) Potential of Hyperion EO-1 Hyperspectral Data for Wheat Crop Chlorophyll Content Extraction in Precision Agriculture. *Canadian Journal of Remote Sensing, Special Issue on Hyperspectral Remote Sensing*, **34**(1): 139-157.
- Baret F., Andrieu B. and Guyot G. (1988) A simple model for leaf optical properties in visible and near infrared: application to the analysis of spectral shifts determinism, In *Applications of chlorophyll fluorescence* (H.K. Lichtenthaler, Ed), Kluwer Academic publishers, pp. 345-351.
- Beck, R. (2003) EO-1 User Guide, Version 2.3. <http://eo1.usgs.gov> and <http://eo1.gsfc.nasa.gov>. 74 pages.
- Berk, A., L. S. Bernstein and D.C. Robertson (1999) MODTRAN: A moderate resolution model for LOWTRAN 7. *Final report*, GL-TR-0122. AFGL, Hanscom AFB, Maryland, 42 pages.
- Blackmer, T.M., Schepers, J.S., Varvel, G.E. and Walter-Shea, E.A. (1996) Nitrogen deficiency detection using reflected shortwave radiation from irrigated corn canopies. *Agronomy Journal*, **88**: 1-5.
- Blackmer, T.M., Schepers, J. S., and Varel, G. E. (1994) Light reflectance compared with other nitrogen stress measurements in corn leaves. *Agronomy Journal*, **86**: 934-938.
- Blackburn, G. A. (1998a) Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. *International Journal of Remote Sensing*, **19**: 657-675.
- Blackburn, G. A. (1998b) Quantifying chlorophylls and carotenoids from leaf to canopy scales: an evaluation of some hyperspectral approaches. *Remote Sensing of Environment*, **66**: 273-285.
- Blackburn, G. A. (1999) Relationships between spectral reflectance and pigment concentrations in stacks of deciduous broadleaves. *Remote Sensing of Environment*, **70**: 224-237.
- Blackburn, G. A. and Steele, C. M. (1999) Towards the remote sensing of Matorral vegetation physiology: Relationships between spectral reflectance, pigment, and biophysical characteristics of semiarid bush land canopies. *Remote Sensing of Environment*, **70**: 278-292.
- Carter, G.A. (1994) Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. *International Journal of Remote Sensing*, **15**(3): 697-703.
- Champagne, C., Staenz, K., Bannari, A., Deguise, J.-C. and McNairn, H. (2003) Validation of a hyperspectral curve-fitting model for the estimation of plant water content of agricultural canopies. *Remote Sensing of Environment*, **87**: 148-160.
- Chappelle, E. W., Kim, M. S. and McMurtrey III, J. E. (1992) Ratio analysis of reflectance spectra (RARS): an algorithm for the remote estimation of the concentrations of Chlorophyll A, chlorophyll B and the carotenoids in soybean leaves. *Remote Sensing of Environment*, **39**: 239-247.
- Curran, P. J., Dungan, J. L. and Gholz, H.L. (1990) Exploring the relationship between reflectance, red edge and chlorophyll content in slash pine. *Tree Physiology*, **7**:33-48.
- Curran, P.J., Dungan, J.L., Macler, B.A. and Plummer, S.E. (1991) The effect of a red leaf pigment on the relationship between red edge and chlorophyll concentration. *Remote Sensing of Environment*, **35**: 69-76.

- Dash, J. and Curran, P.J. (2004) The MERIS terrestrial chlorophyll index. *International Journal of Remote Sensing*, **25**(23): 5403-5413.
- Daughtry, C. S. T., Walthall, C. L., Kim, M.S., Brown de Colstoun, E. and McMurtrey III, J. E. (2000) Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, **74**: 229-239.
- Earth Search Sciences Inc. (2001) About Probe-1, Kalispell, Montana, www.earthsearch.com/technology.
- Elvidge, C.D. and Chen, Z.K. (1995) Comparison of broad-band and narrow-band red and near-infrared vegetation indexes. *Remote Sensing of Environ*, **54**: 38-48.
- Filella, I. and Penuelas, J. (1994) The red edge position and shape as indicators of plant chlorophyll concentration, biomass and hydric status. *International Journal of Remote Sensing*, **15**: 1459-1470
- Filella, I., Serrano, L., Serra, J. and Peñuelas, J. (1995) Evaluating wheat nitrogen status with canopy reflectance indices and discriminant analysis. *Crop Science*, **35**: 1400-1405.
- Filella, I., Amaro, T., Araus, J. L. and Peñuelas, J. (1996) Relationship between photosynthetic radiation-use efficiency of barley canopies and photochemical reflectance index (PRI). *Physiologia Plantarum*, **96**: 211-216.
- Gamon J.A., Peñuelas J. and Field, C.B. (1992) A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment*, **41**: 35-44.
- Gao, B.C. and Goetz, A.F.H. (1990) Column atmospheric water vapor and vegetation liquid water retrieval from airborne imaging spectrometer data. *Journal of Geophysical Research*, **95**: 3549-3564.
- Gao, J. (2006) Canopy Chlorophyll Estimation with Hyperspectral Remote Sensing. Ph.D. Thesis, Department of Geography, College of Arts and Sciences, Kansas State University, Manhattan, Kansas, 206 pages.
- Gitelson, A., Merzyak, M. N. and Lichtenthaler, H. K. (1996) Detection of red-edge position and chlorophyll content by reflectance measurements near 700 nm. *Journal of Plant Physiology*, **148**: 501-508.
- Green, R.O., Conel, J.E., Margolis, J.S., Brugge, C.J. and Hoover, G.L. (1991) An inversion algorithm for the retrieval of atmospheric and leaf water absorption from AVIRIS radiance with compensation for atmospheric scattering. *Proceedings of the 3rd Annual Airborne Visible Infrared Imaging Spectrometer (AVIRIS) Workshop*. Pasadena, California, JPL Publication 91-28, pp. 51-61.
- Guyot, G. and Baret, F., 1988. Utilisation de la haute résolution spectral pour suivre l'état des couverts végétaux. In: *Proceedings of the Fourth International colloquium on Physical Measurements and Signatures in Remote Sensing*, European Space Agency, Noordwijk, pp. 279-286.
- Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., and Dextraze, L. (2002) Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, **81**: 416-426.
- Haboudane, D., Tremblay, N., Miller, J.R. and Vigneault, P. (2008) Remote Estimation of Crop Chlorophyll Content Using Spectral Indices Derived From Hyperspectral Data. *IEEE Transaction on Geosciences and Remote Sensing*, **46**(2): 423-437.

- Horler, D. N. H., Dockray, M. and Barber, J. (1983) The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, **4**: 273-288.
- Huete, A.R. (1989) Soil influences in remotely sensed vegetation-canopy spectra. In: Asrar, G. Editor (1989). *Theory and Applications of Optical Remote Sensing*, John Wiley & Sons, Inc., New York, pp. 107-141.
- Huete, A.R., Jackson, R.D. and Post, D.F. (1985) Spectral response of a plant canopy with different soil backgrounds. *Remote Sensing of Environment*, **17**: 37-53.
- Irons, J.R., Weismiller, R.A. and Petersen, G.W. (1989) Soil reflectance. In: Asrar, G. Editor (1989). *Theory and Applications of Optical Remote Sensing*, John Wiley & Sons, Inc., New York, pp. 66-106.
- Jackson, R.D., Pinter, P.J., Paul, J., Reginato, R.J., Robert, J. and Idso, S.B. (1980) Hand-held Radiometry. U.S. Department of Agriculture Science and Education Administration, *Agricultural Reviews and Manuals, ARM-W-19*, Phoenix, Arizona, U.S.A., 66 pages.
- Jacquemoud, S., Ustin, S.L., Verdebout, G., Schmuck, G., Andreoli, G., and Hosgood, B. (1996) Estimating leaf biochemistry using the PROSPECT leaf optical properties model. *Remote Sensing of Environment*, **56**, 194-202.
- Jago, R. A., Cutler, M. E. J., & Curran, P. J. (1999) Estimating canopy chlorophyll concentration from field and airborne spectra. *Remote Sensing of Environment*, **68**: 217- 224.
- Karimi, Y., Prasher, S.O., McNairn, H., Bonnell, R.B., Dutilleul, P. and Goel, P.K. (2005a) Discriminant analysis of hyperspectral data for assessing water and nitrogen stresses in corn. *Trans. ASAE.*, **48**: 805-813.
- Karimi, Y., Prasher, S.O., McNairn, H., Bonnell, R.B., Dutilleul, P. and Goel, P.K. (2005b) Classification accuracy of discriminant analysis, artificial neural networks and decision trees for weed and nitrogen stress detection in corn. *Trans. ASAE.*, **48**: 1261-1268.
- Kim, M. S., Daughtry, C. S. T., Chappelle, E. W., McMurtrey III, J. E. and Walthall, C. L. (1994) The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation (APAR). *Proceedings of the 6th Symposium on Physical Measurements and Signatures in Remote Sensing*, January 17-21, 1994, Val D'Isere, France, pp. 299-306.
- Khurshid, K.S. (2004) Estimation and mapping of wheat crop chlorophyll content using Hyperion hyperspectral data. *Master thesis*, Department of Geography, University of Ottawa, Ottawa (Ontario), Canada, 158 pages.
- Khurshid, K.S., Staenz, K., Sun, L., Neville, R., White, H.P., Bannari, A., Champagne, C.M. and Hitchcock, R. (2006) Preprocessing of EO-1 Hyperion Data. *Canadian Journal of Remote Sensing*, **32**(2): 84-97.
- Kneubühler, M. (2002) Spectral assessment of crop phenology based on spring wheat and winter barley. *Ph.D. Thesis*. Remote Sensing Laboratory. Department of Geography, University of Zurich, Zurich, Switzerland, 148 pages.
- Kruse, F. A., Boardman, J. W. and Huntington, J. F. (2003) Comparison of Airborne Hyperspectral Data EO-1 Hyperion for Mineral Mapping. *IEEE Transactions on Geoscience and Remote Sensing*, **41**: 1388-1400.
- Mariotti, M., Ercoli, L., & Masoni, A. (1996) Spectral properties of iron-deficient corn and sunflower leaves. *Remote Sensing of Environment*, **58**: 282- 288.

- McNairn, H., Deguise, J.C., Pacheco, A., Shang, J. and Rabe, N. (2001) Estimation of crop cover and chlorophyll from hyperspectral remote sensing. Proceedings of the 23rd Canadian Symposium on Remote Sensing. Ste. Foy, Quebec.
- Meggio, F., Zarco-Tejada, P.J., Miller, J.R., Martin, P., Gonzalez, M.R. and Berjon, A. (2008) Row Orientation and Viewing Geometry Effects on Row-Structured Vine Crops for Chlorophyll Content Estimation. *Canadian Journal of Remote Sensing*, **34**(3): 220-234.
- Munden, R., Curran, P. J. and Catt, J. A. (1994) The relationship between red edge and chlorophyll concentration in the broadbalk winter wheat experiment at Rothamsted. *International Journal of Remote Sensing*, **15**(3): 705-709.
- Neville R. A., Sun L. and Staenz K. (2004) Detection of Keystone in Imaging Spectrometer Data. Proceedings of SPIE on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX, Vol. 5425, pp. 208-217.
- Neville R. A., Sun L. and Staenz K. (2008) Spectral Calibration of Imaging Spectrometers by Atmospheric Absorption Feature Matching. *Canadian Journal of Remote Sensing, Special Issue on Hyperspectral Remote Sensing*, **34**(1): 29-42.
- Oppelt, N. and Mauser, W. (2001) The chlorophyll content of maize (*Zea mays*) derived with the airborne imaging spectrometer AVIS. Proceedings of the 8th International Symposium on Physical Measurements & Signatures in Remote Sensing, Aussois, France, pp. 407-412.
- Oppelt, N. and Mauser, W. (2004) Hyperspectral monitoring of physiological parameters of wheat during a vegetation period using AVIS data. *International Journal of Remote Sensing*, **25**: 145-159.
- Oppelt, N. and Mauser, W. (2007) Characterization of Sun and Shade Chlorophyll in Wheat Using Angular CHRIS / PROBA Data. Proceedings of 'Envisat Symposium 2007, Montreux, ESA SP-636, 6 pages
- Pacheco, A., Bannari, A. and McNairn, H. (2001) LAI Measurements in Beans and Corn Canopies with Two Optical Instruments. Proceedings of the 8^e Symposium international des mesures physiques et signatures en télédétection, Aussois (France), 8-12 Janvier 2001, CNES, Toulouse, (France), pp. 374-379758.
- Pacheco, A., Bannari, A., Staenz, K., Deguise, J.-C. and McNairn, H. (2002) Validating LAI Using Hyperspectral Imagery Over Agricultural Canopies. P. 210-215, in J.A. Sobrino (ed.) *International Symposium on Recent Advances in Quantitative Remote Sensing*. Valencia (Spain), 16-20 September 2002, Published by Publications de la Universitat de València, València, Spain, pp. 210-215.
- Pacheco, A., Bannari, A., Staenz, K., Deguise, J.-C. and McNairn, H. (2008) Deriving Percent Crop Cover Over Agriculture Canopies Using Hyperspectral Remote Sensing. *Canadian Journal of Remote Sensing, Special Issue on Hyperspectral Remote Sensing*, **34**(1): 110-123.
- Penuelas, J. and Filella, I. (1998) Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends Plant Science*, **3**: 151-156.
- Penuelas, J., Filella, I., Biel, C., Serrano, L. and Save, R., (1993). The reflectance at the 950-970 nm region as an indicator of plant water status. *International Journal of Remote Sensing*, **14**: 1887-1905.

- Penuelas, J., Gamon, J., Freeden, A., Merino, J. and Field, C. (1994) Reflectance indices associated with physiological changes in nitrogen and water limited sunflower leaves. *Remote Sensing of Environment*, **48**: 135-146.
- Penuelas, J., Baret, F. and Filella, I., (1995) Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica*, **31**: 221-230.
- Penuelas, J., Pinol, J., Ogaya, R. and Filella, I. (1997) Estimation of plant water concentration by the reflectance water index WI (R900/R970). *International Journal of Remote Sensing*, **18**: 2869-2875.
- Pinar, A. and Curran, P. J. (1996) Grass chlorophyll and the reflectance red edge. *International Journal of Remote Sensing*, **17**(2): 351-357.
- Riedell, W.E. and Blackmer, T.M. (1999) Leaf Reflectance Spectra of Cereal Aphid-Damaged Wheat. *Crop Science*, **39**:1835-1840.
- Rollin, E.M. and Milton, E.J. (1998) Processing of High Spectral Resolution Reflectance Data for the Retrieval of Canopy Water Content Information. *Remote Sensing of Environment*, **65**: 86-92.
- Rondeaux, G., Steven, M. and Baret, F. (1996) Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, **55**: 95-107.
- Rouse, J.W., Hass, R.W., Schell, J.A., Deering, D.W. and Harlan, J.C. (1974) Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation. NASA / GSFCT Type III Final report. Greenbelt, Maryland, USA, 164 pages.
- Secker, J., Staenz, K., Gauthier, R.P. and Budkewitsch, P. (2001) Vicarious Calibration of Hyperspectral Sensors in Operational Environments. *Remote Sensing of Environment*, **76**: 81-92.
- Staenz, K. and Williams, D.J. (1997) Retrieval of surface reflectance from hyperspectral data using look-up-table approach. *Canadian Journal of Remote Sensing*, **23**: 354-368.
- Staenz, K., Szeredi, T. and Schwarz J. (1998) ISDAS - A system for processing / analyzing hyperspectral data. *Canadian Journal of Remote of Sensing*, **24**: 99-113.
- Staenz, K., Neville, R.A., Levesque, J., Szeredi, T., Singhroy, V., Borstad, G.A. and Hauff, P. (1999) Evaluation of CASI and SFSI hyperspectral data for environmental and geological applications - two case studies. *Canadian Journal of Remote Sensing*, **25**: 311-322.
- Sun L., Neville R. A., K. Staenz and White H.P. (2008) Automatic Destriping of Hyperion Imagery Based on Spectral Moment Matching. *Canadian Journal of Remote Sensing, Special Issue on Hyperspectral Remote Sensing*, **34**(1): 68-81.
- Thie, J. (2006) Cli-Land Capability of the Regina Map Sheet Area, 72 I. Eco-informatics International Inc., http://www.geostrategis.com/c_cli-regina.htm#
- Verstraete, M.M., Pinty, B. and Curran, P.J. (1999) MERIS potential for land application. *International Journal of Remote Sensing*, **20**(9): 1747-1756.
- Wood, C.W., Reeves, D.W. and Himelrick, D.G. (1993) Relationship between chlorophyll meter reading and leaf chlorophyll concentration, N status, and crop yield: a review. *Proceedings of the Agronomy Society of New Zealand*, Vol. 23, pp. 1-9.
- Wu, C., Niu, Z., Tang, Q. and Huang, W. (2008) Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and Forest Meteorology*, **148**: 1230-1242.

- Zarco-Tejada, P.J. (2000) Hyperspectral remote sensing of closed forest canopies: Estimation of chlorophyll fluorescence and pigment content. Ph.D. Thesis, York University, Toronto, Ontario, Canada, 233 pages.
- Zarco-Tejada, P. J. and Miller, J. R. (1999) Land Cover Mapping at BOREAS Using Red-Edge Spectral Parameters from CASI Imagery. *Journal of Geophysical Research*, **104**(D22): 27-921-933.
- Zarco-Tejada, P.J., Berjón, A., López-Lozano, R., Miller, J.R., Martín, P., Cachorro, V., González, M.R. and de-Frutos, A. (2005) Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sensing of Environment*, **99**(3): 271-287.
- Zhang, Y., Chen, J.M., Miller, J.R. and Noland, T.L. (2008) Leaf chlorophyll content retrieval from airborne hyperspectral remote sensing imagery. *Remote Sensing of Environment*, **112**: 3234-3247.

Simultaneous Estimation of Optical Properties of Asian Dust and Ground Reflectance by Polarization Measurements

Takashi Kusaka and Ryuichi Taniguchi
Kanazawa Institute of Technology

1. Introduction

The Asian dust transported from desert areas in the northern China often covers over East Asia in the late winter and spring seasons. It is regarded as a main source of atmospheric aerosols over East Asia and has significant effects on the climate change. Moreover, fine dust particles in the air have harmful influence on our health on the local and global scales. However, it is as yet very difficult to extract optical properties of the widely spread hazy dust from satellite data over land surfaces because the radiance received by a satellite sensor strongly depends on the surface reflectance. It will be, therefore, necessary to estimate optical properties of the Asian dust and the ground reflectance simultaneously from satellite data. In particular, the polarimetric information will provide the improvement of estimating optical properties of atmospheric aerosols.

Many authors have investigated dust properties over ocean (Nakajima et al.,1998, Tanre et al.,1997) and over dark vegetated areas (Kaufman et al.,1997a, Kaufman et al.,1997b) using satellite measurements. In these cases, the surface contribution to the radiance received by the satellite sensor is small. However, the linear polarization is less commonly measured than the radiance in the remote sensing at optical and near infrared wavelengths.

The ADEOSII/POLDER and PARASOL/POLDER observe the reflectance and polarization of a target quasi-simultaneously in multi-viewing angles at wavelengths of 443nm (ADEOSII/POLDER), 490nm (PARASOL/POLDER), 670nm and 865nm, and so POLDER data provide enough information to determine optical characteristics of atmospheric aerosols and the ground reflectance. We have developed the algorithm for estimating optical properties of dust particles and the surface reflectance simultaneously from POLDER data (Kusaka et al., 2001, 2002b, 2004). However, absorption of light ray by dust particles is not taken into account in the estimation algorithm.

In order to investigate the usefulness of the polarization for aerosol particles, we have made ground-based polarization measurements of the sky radiation by the PSR-1000 (Masuda,1997), which is the multi-spectral polarimeter produced by Opt Research Corporation, Japan and has the same wavelength regions (443nm, 490nm, 565nm, 670nm, 765nm and 865nm) as the POLDER sensor. In the following sections, we will describe the results of ground-based polarization measurements and a new algorithm for estimating

aerosol properties over land using the polarized radiance received by the POLDER. In this new algorithm, absorption of light by aerosols will be taken into account.

2. Ground-based polarization measurements

The vertical profiles and optical properties of aerosols including Asian dust have been investigated by means of Lidar observation (NIES, Japan) and sun photometer measurements (AERONET, NASA) from ground stations. However, ground based polarization measurements are as yet very sparse.

We have made polarimetric measurements of the sky radiation at the ground station in the Kanazawa city, Japan, which is located at the side of the Sea of Japan, in the spring season from 2000. Asian dust clouds often appear together with normal clouds, and so we have only measured the degree of polarization of the sky light when the Kanazawa city was covered with a thin dust cloud.

2.1 Description of PSR-1000

The PSR-1000 measures the intensity of the sky radiation. A Glan-Thompson prism is implemented between the hood (2 degrees field of view) and the interference filters in the PSR-1000 instrument. The prism rotates automatically by a pulse motor and the range of the rotation is 0 to 360 degrees. The measurement by the PSR-1000 is controlled by a personal computer (PC). The optical instrument attached on the tripod is pointed manually to a desired direction to measure the intensity of the sky radiation. Since the received signals change sinusoidally against the angle of rotation of the polarizer (Hansen, et al., 1974), the maximum signal value, I_{\max} and the minimum value, I_{\min} can be easily obtained from the sinusoidal curve. The degree of linear polarization, L , is therefore given by

$$L = \frac{(I_{\max} - I_{dk}) - (I_{\min} - I_{dk})}{((I_{\max} - I_{dk}) + (I_{\min} - I_{dk}))} \quad (1)$$

where I_{dk} is the dark signal that is the signal received in no incident solar radiation. As seen from Eq.(1), no absolute calibration of the optical instrument is needed in the polarization measurement because the degree of polarization is the relative value.

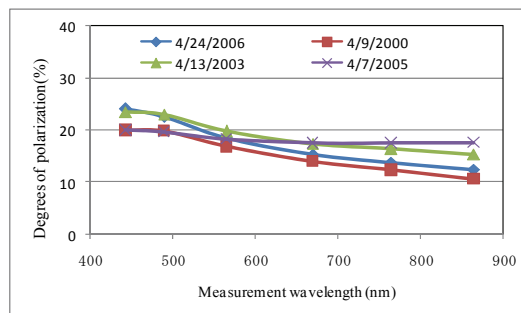


Fig. 1. Degrees of polarization measured at the angle of 90 degrees from the solar direction when Kanazawa city was covered with Asian dust

2.2 Polarization measurements of the sky radiation

Polarizations were measured at angles of 90 and 120 degrees from the solar direction in the principal plane in 2000 to 2002. From 2003, polarization measurements were carried out at angles of 75, 90, 105 and 120 degrees from the solar direction in the principal plane.

Figure 1 shows degrees of polarization measured at the angle of 90° from the solar direction when the hazy dust was recognized at the meteorological observatory in the Kanazawa city. Figure 2 shows degrees of polarization measured at the angle of 90° from the solar direction in the clear sky. Figure 3 shows angular dependencies of polarization at 490nm measured on April 13 and 27, 2003. As seen from Figure 1 to Figure 3, degrees of polarization measured in the sky covered with the hazy dust decrease uniformly as the wavelength increases and are lower than those in the clear sky. The wavelength dependencies are slightly different in the measurement date. In particular, most of degrees of polarization measured in the clear sky show the peak value around the 490nm channel. In polarization measurements in the clear sky, we have often similar polarization patterns as measured on April 27, 2003. In this case, we pointed out that this will be due to light scattering by small aerosols included in the atmosphere (Kusaka, et al., 2002a).

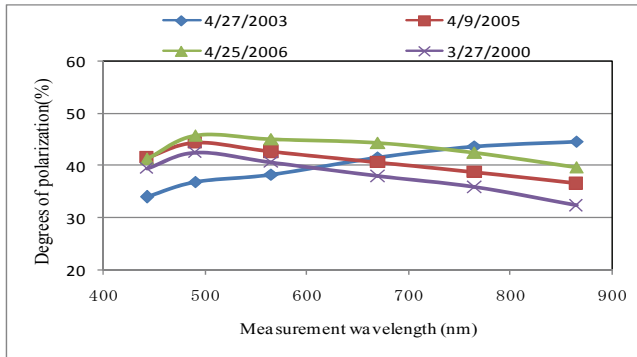


Fig. 2. Degrees of polarization measured at the angle of 90 degrees from the solar direction in the clear sky

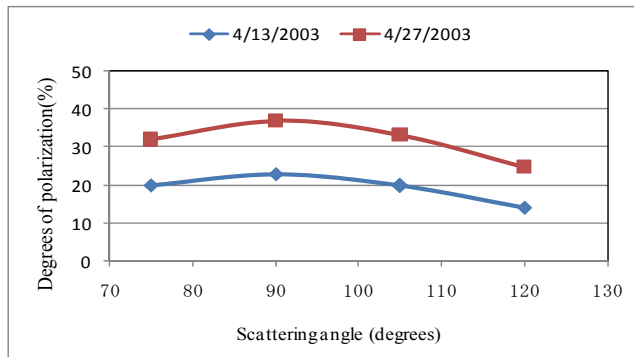


Fig. 3. Degrees of polarization at 490nm measured at angles of 75, 90, 115 and 120 degrees from the solar direction on April 13 and April 27, 2003

3. Estimation of dust properties and surface reflectance

A basic idea for estimating optical properties of aerosols and the surface reflectance is very simple, and aerosol properties and surface reflectance are estimated, by comparing the radiance and polarization obtained from satellite and ground-based measurements with those obtained from the computation of multiple scattered light in the atmosphere-ground system.

The radiance and polarization of scattering light are completely described by the Stokes parameters (I , Q , U , V), where I is the radiance and the other parameters have same dimension. We have $V=0$ in the linear polarization. The linearly polarized radiance I_p and the polarization direction χ can be derived from Q and U as follows (Hansen, et al., 1974):

$$I_p = \sqrt{Q^2 + U^2} \quad (2)$$

$$\tan(2\chi) = U/Q \quad (3)$$

The degree of polarization is defined as the ratio I_p/I .

3.1 Estimation of aerosol properties by ground-based polarization measurements

We estimate optical properties of aerosols using degrees of polarization of sky light measured at four scattering angles by the PSR-1000. To do that, we computed degrees of polarization at the bottom of the atmosphere in the plane parallel uniform atmosphere bounded by the uniform background surface by means of the Monte Carlo integration (O'Brien, 1998, Ishimoto, et al., 2002). In the radiative transfer simulation, it was assumed that the number size distribution of aerosols is represented by the Junge power-law (radius $r < 0.1\mu\text{m}$ $dN/dr = \text{const.}$, $r > 0.1\mu\text{m}$ $dN/dr = cr^{-a}$, minimum radius: $0.05\mu\text{m}$, maximum radius: $15\mu\text{m}$) and the dust particle is spherical, non-absorption matter. Moreover, we assumed that the land surface is the uniform Lambertian reflector.

Therefore, parameters to be estimated from the measured polarizations are the optical thickness of aerosols, t , exponent of Junge power-law, a , refractive index of aerosols, N_r , and the background reflectance, A . We determined the values of t , a , N_r and A , using the following algorithm:

- (1) Degrees of polarizations at the bottom of the atmosphere were computed for typical values of t , N_r , a , and A and were saved in the Lookup table (LUT) for the solar zenith angle at the measurement time. In this case, only degrees of polarization at angles of 75, 90, 105 and 120 degrees from the solar direction in the principal plane were computed.
- (2) We used LUT to determine the values of t , N_r , a and A such that the sum of the square errors, Q , between the measured polarizations and the computed ones in four directions is minimum. In this case, the interpolation scheme by the 3rd order polynomials was adopted to obtain the minimum value of Q .

We used only degrees of polarization at the 490nm channel to estimate values of t , N_r , a , and A because the radiance received at the bottom of the atmosphere not so much strongly depends on the ground reflectance of the suburban area at 490nm.

The method described above was applied to degrees of polarization measured on April 13 and 27, 2003 as shown in Figure 3 (kusaka et al., 2007). As a result, we had optical thickness of aerosols $t=0.8$, refractive index $N_r=1.62$, exponent of Junge power-law $a=4.69$ and

background reflectance $A=0.0$ on April 13, 2003, and $t=0.5$, $Nr=1.48$, $a=4.59$ and $A=0.024$ on April 27, 2003. We used the estimated values of 4 parameters to compute degrees of polarization at the bottom of the atmosphere. The computed degrees of polarization and the measured ones are shown in Figure 4.

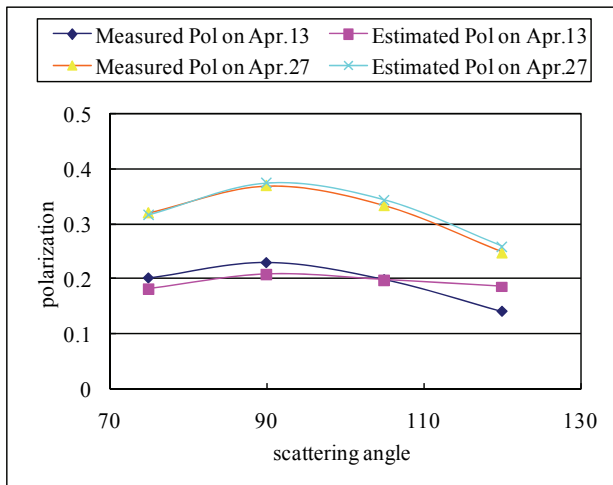


Fig. 4. The estimated degrees of polarization and the measured ones at 490nm are shown in cases of April 13 and 27, 2003

As seen from Figure 4, the computed polarizations are very close to the measured ones on April 27 when the optical thickness of aerosols is thin, but on April 13 when the optical thickness of aerosols is thick, the computed polarizations are significantly different from the measured ones. On April 13 and 27, 2003, the aerosol measurement by the sun photometer and the polarization measurement by the PSR-1000 were carried out at the almost same time. The optical thickness of aerosols at the 500nm channel by the sun photometer was 0.85 on April 13, and 0.45 on April 27. The estimated optical thickness of aerosols at the 490nm channel was 0.8 on April 13 and 0.5 on April 27. The difference between the estimated value and the measured one is small on April 13 and 27. However, on April 13, the estimated degrees of polarization are significantly different from the measured ones at 490nm. Therefore, we may need the improvement of the estimation process in the case of the thick Asian dust.

3.2 Estimation of aerosol properties and surface reflectance by POLDER data

Kusaka, et al. (Kusaka, et al. 2001, 2002b, 2004) proposed a method for estimating optical properties of Asian dust and the surface reflectance simultaneously from the radiance and polarization observed at each of 443P and 670P channels of the ADEOS&ADEOSII /POLDER sensor. It was shown that the method developed by Kusaka, et al. provides reasonable surface reflectance and dust properties such as the optical thickness, the refractive index, the number size distribution of dust particles at the 670nm channel. However, at the 443nm channel, absorption of light ray by dust particles is not taken into account in the algorithm of Kusaka, et al. Therefore, we will need to have a new method for estimating dust properties more accurately at shorter wavelengths.

In this section, we describe a new method for the estimation of aerosol properties using the polarized radiance effectively and apply this method to PARASOL/POLDER data.

Since the ground resolution of a POLDER-measured pixel is $6 \times 7 \text{ km}^2$ at nadir, the radiance and polarization at the top of the atmosphere were obtained from the numerical computation of the radiative transfer equation in a uniform plane parallel atmosphere bounded by the uniform Lambertian reflector. We used the 6SV-1.0B code developed by Vermote et al. (Vermote et al. 2006) to compute values of Stokes parameters at the top of atmosphere. In this case, it was assumed that the number size distribution of aerosols is represented by the Junge power-law and aerosol particles are spherical.

By comparing the radiance and polarization received by the POLDER with those obtained from the radiative transfer simulation, we estimate the optical thickness of aerosols, t , the exponent of Junge power-law, a , the complex refractive index, N_r (real part) and N_i (imaginary part), and the surface reflectance, A .

The optical thickness of aerosols in the near infrared wavelength is generally thinner than that in optical wavelengths and the contribution of light scattering by aerosols to the radiance in the near infrared wavelength received by the satellite sensor is relatively small. In the present study, the retrieval of aerosol properties and the surface reflectance from POLDER data acquired at optical channels will be taken into account.

It is also shown that the contribution of light reflected by the uniform Lambertian reflector to the polarized radiance is very small because the reflected light by the Lambertian surface is depolarized. We investigated the dependency of polarized radiance on the surface reflectance. Figure 5 shows the polarized radiance normalized by $\cos(\theta_s)E/\pi$ against the surface reflectance at 490nm and 670nm channels, where θ_s is the solar zenith angle and E the solar flux at the top of atmosphere. In Figure 5, polarized radiances were computed under the following conditions:

optical properties of aerosols: $t(550)=0.4$, $a=4.0$, $N_r=1.5$ and $N_i=0.0$.

geometric conditions: solar zenith angle 35 degrees, viewing zenith angle 30 degrees,
relative azimuth angle 180 degrees.

atmospheric model: midlatitude winter model.

We can see from Figure 5 that the polarized radiance received by the satellite sensor is independent of the surface reflectance. This indicates that aerosol properties are extracted from polarized radiances received by the satellite sensor.

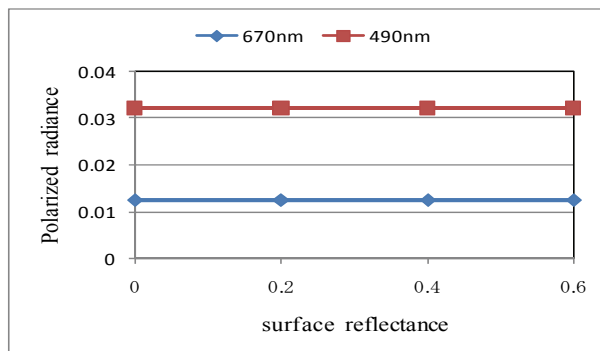


Fig. 5. The dependency of polarized radiances on the surface reflectance at 490nm and 670nm channels. It is assumed that the ground surface is the Lambertian reflector.

3.2.1 Estimation Algorithm at the 670nm channel

(1) Estimation of aerosol properties

Absorption of optical light by dust particles decreases as the wavelength is longer. In this study, we assume that there is very little absorption for dust particles at 670nm, because we have not so much experimental results for it. Therefore, it is assumed that the imaginary part of complex refractive index, N_i , is zero at 670nm.

The polarization (Q and U of Stokes parameters) at a target pixel measured by the POLDER change at geometric conditions of the satellite sensor. We create the look-up table, LUT, for the polarization to retrieve the necessary information easily from it. Using the 6SV-1.0b code, we computed in advance the radiance RAD and the polarized radiance POL and the polarization direction PD at the top of atmosphere as the function of the optical thickness of aerosols at 550nm, $t(550)$, refractive index of aerosols, N_r , and index of Junge power-law, a under a given geometric condition such as solar zenith angle, SZA, viewing zenith angle, VZA, and relative azimuth angle, RAZ. In this computation, the midlatitude winter model given in the 6SV-1.0B code was adopted as the atmospheric model and the surface reflectance A was taken as 0. The values of RAD, POL and PD computed for typical values of $t(550)$, N_r , a , SZA, VZA and RAZ were saved in the look-up table, LUT.

The Stokes parameters I, Q, U in different viewing conditions at each pixel of POLDER data are computed. Consider that the observed Stokes parameters in different N geometrical conditions, each of which consists of a 3-tuple (SZA, VZA, RAZ), were extracted from POLDER data at a target pixel, and polarized radiances POL were computed by Eq. (2). The values of POL against all combinations of N_r , a and $t(550)$ given in LUT were interpolated in all of viewing conditions, $(SZA, VZA, RAZ)_i$, ($i=1, \dots, N$). In this case, we adopted the Lagrange 2nd-order polynomials as the interpolation function. The values of POL_i in the i -th viewing condition, i.e., $(SZA, VZA, RAZ)_i$ were restored in the new file, nLUT.

We use the least square method to estimate optimum values of N_r , a and $t(550)$. Let the polarized radiance observed by the POLDER in a viewing condition, (SZA, VZA, RAZ) , be OPL. The sum of square errors between the observed values and the computed ones, Q , is defined as

$$Q = \sum_{i=1}^n (OPL_i - POL_i)^2 \quad (4)$$

where i represents the i -th geometric condition.

It is necessary to obtain the unique solution that minimizes Q in the 3-dimensional parameter space. In other words, the problem is to find the values of N_r , a and t that correspond to the minimum of Q . In general, the hill climbing algorithm starts by making the initial guess and needs the partial derivatives of a function Q to compute the better solution from the initial solution. In our case, it is difficult to derive the partial derivatives of Q given by Eq.(4). We also used the modified hill climbing algorithm adopted by Kusaka, et al. (Kusaka, et al., 2004) in which the partial derivatives of Q are not used, to get the minimum of Q . The modified hill climbing algorithm provides a local minimum rather than a global one. Therefore, we derived the optimum solution of N_r , a and t in the following two steps.

Step 1: Using the values of POL stored in the file, nLUT, we compute the values of Q in all combinations of variables (N_r, a, t) . Then, we can obtain n values of N_{r_k} , a_k and t_k that

correspond to Q_k ($Q_1 < Q_2 < \dots < Q_n$, $k=1, 2, \dots, n$). The values of Nr_k , a_k and t_k are used as initial values. In practical applications, we chose $n=3$.

Step 2: The values of Nr_k , a_k and t_k ($k=1$) are first applied to the modified hill climbing algorithm and the new solution ($Nr(1)$, $a(1)$, $t(1)$) corresponding to the new local minimum $Q(1)$ is obtained. This procedure is repeated for n initial values. As a result of it, we have n values of 3 parameters corresponding to local minimum values $Q(1)$, \dots , $Q(n)$. As the optimum solution, we choose $Nr(j)$, $a(j)$ and $t(j)$, corresponding to the minimum value, $Q(j)$, among $Q(1)$, \dots , $Q(n)$, where j is one of $1, 2, \dots, n$.

(2) Estimation of surface reflectance

Since optical properties of aerosols such as Nr , a , and t were determined in the previous section, we can easily estimate the surface reflectance using the radiance received by the POLDER. First of all, by using the estimated Nr , a and t , the radiance, CI , at the top of atmosphere is computed for a surface reflectance under a given geometric condition received by the POLDER and is compared with the radiance, OI , received by the POLDER.

After computing the values of CI in typical four surface reflectances, $A1$, $A2$, $A3$ and $A4$ for all viewing conditions of the POLDER, we determine the surface reflectance in the range of $A1$ to $A4$ that minimizes the sum, QR , of square errors between the observed radiances and the computed ones. QR is defined as

$$QR(A) = \sum_{i=1}^n (OI_i - CI_i)^2 \quad (5)$$

We also used the modified hill climbing algorithm to obtain the surface reflectance, A .

3.2.2 Estimation algorithm at 490nm

In general, the refractive index slightly increases as the wavelength is shorter, but the measurement for the wavelength dependency of dust particles is as yet very sparse. In the present study, we assume that the refractive index, Nr , does not depend on the wavelength and so Nr at 490nm is the same as that at 670nm. However, absorption of light by aerosol particles will be taken into account. As a matter of course, number size distribution of aerosols at this wavelength is taken as that estimated at 670nm. Therefore, optical parameters to be estimated at 490nm are the imaginary part of complex refractive index and optical thickness of aerosols.

We generated the look-up table, LUT4, of RAD, POL and PD for typical values of Nr , Ni , a , $t(550)$, SZA, VZA and RAZ in the same way as the case at 670nm. Then, the values of POL and PD for SZA, VZA, RAZ and Nr , a estimated at 670nm were interpolated from LUT4 and were saved in a new file, nLUT4.

We use the nLUT4 file to estimate aerosol properties, Ni and $t(550)$, and the surface reflectance, A in the same way as described at (1) and (2) in the section 3.2.1.

3.2.3. Results

The method described in the previous section was applied to PARASOL/POLDER data (P3L1TBG1032192JD) taken on April 28, 2006. Figure 6 shows the POLDER image over Japan (B: 490nm, G: 865nm, R: 670nm). We estimated the values of t , N_r , N_i , a and A at two target pixels including the symbol + shown in Figure 6.

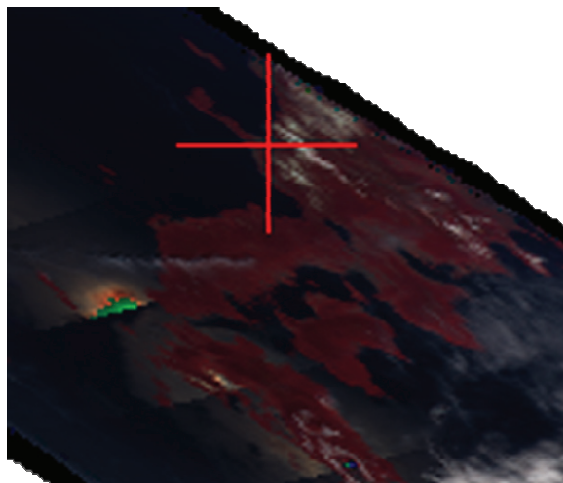


Fig. 6. Parabol/POLDER image over Japan taken on April 28,2006

Point	670nm					490nm				
	N_r	N_i	a	$t(550)$	A	N_r	N_i	a	$t(550)$	A
1	1.35	0	3.92	0.591	0.037	1.35	0.0066	3.92	0.607	0.1
2	1.35	0	3.87	0.6	0.041	1.35	0.0057	3.87	0.6	0.1

Table 1. Estimation results of aerosol properties and surface reflectances

In the estimation process of 3 parameters at 670nm, we chose POLDER data observed in viewing conditions in which the polarized radiances normalized by $\cos(\theta_s)E/\pi$ are larger than 0.02. The results are shown in Table 1. The ground surface at two pixel points selected for the estimation of parameters represents urban and suburban areas. We can see from Table 1 that we have reasonable estimation values for the optical thickness of aerosols and surface reflectance. The polarized radiance and the apparent radiance computed by using parameters at 670 nm estimated at the point 1 in Table 1 and those received by the POLDER are shown in Figure 7, and Figure 8 shows the estimated radiance at the top of atmosphere and the observed ones at 670nm. The estimated polarized radiance and observed one at 490nm are shown in Figure 9 and the computed apparent radiance and observed one at 490nm are shown in Figure 10. The radiance and polarized radiance shown in Figures 7 to 10 are normalized by $\cos(\theta_s)E/\pi$. As seen from Figures 7 to 10, we have a good correspondence between the estimated values and observed ones.

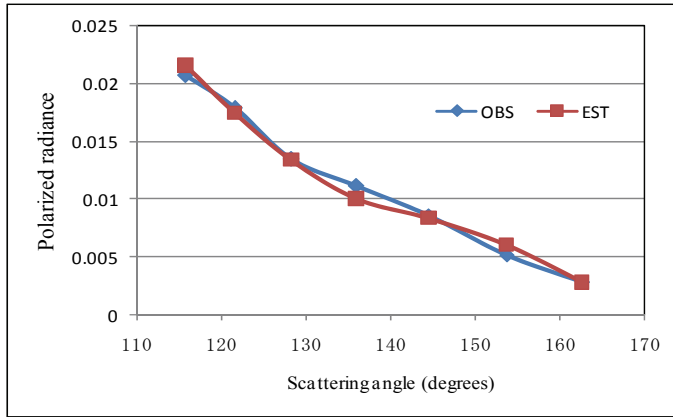


Fig. 7. The estimated polarized radiances (EST) and observed ones (OBS) by the POLDER at 670nm

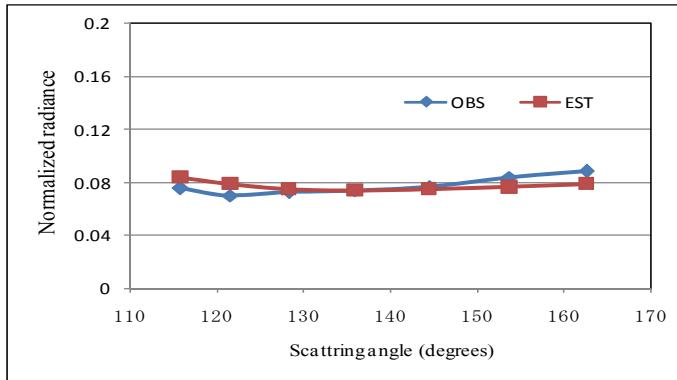


Fig. 8. The estimated radiances (EST) and observed ones (OBS) by the POLDER at 670nm.

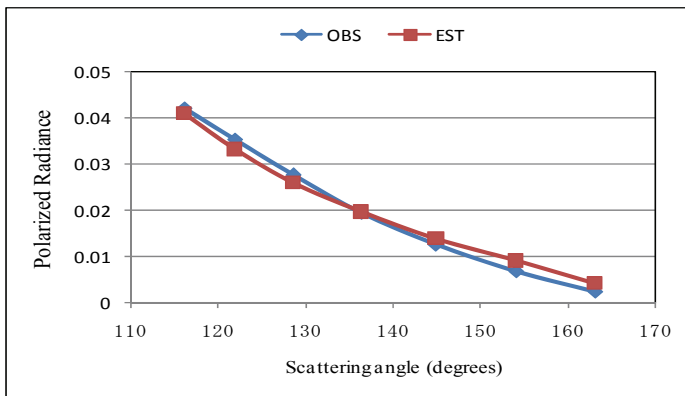


Fig. 9. The estimated polarized radiances (EST) and observed ones (OBS) at 490nm

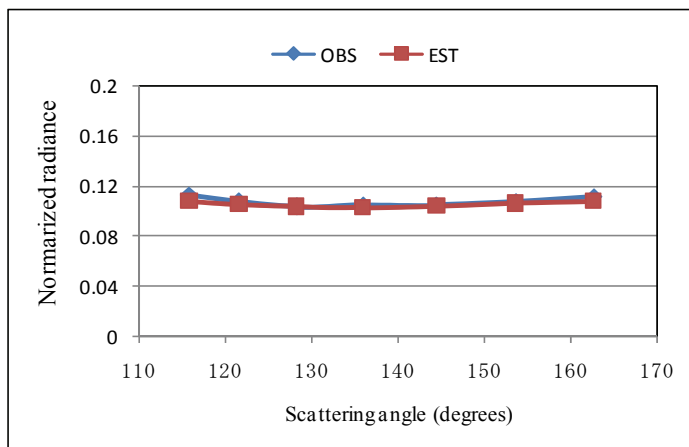


Fig. 10. The estimated radiances (EST) and observed ones by the POLDER at 490nm

4. Conclusions

Ground-based polarization measurements of the sky radiation were made at the Kanazawa city, Japan. The following results were obtained.

(1) Degrees of polarization measured in the sky covered with Asian dust are lower than those measured in the clear sky and decrease uniformly as the wavelength increases.

(2) The wavelength dependency of polarization is slightly different in the measurement date. In particular, most of degrees of polarization measured in the clear sky show the peak value around the 490nm channel.

We also described the method for estimating optical properties of atmospheric aerosols including Asian dust and surface reflectances simultaneously from satellite and ground-based polarization data measured at multi-viewing angles. In addition to it, a new method for the estimation of aerosol properties over land using the polarized radiance measured by the POLDER effectively was proposed. As a result, it was shown that the method described in the present study provides reasonable values for aerosol properties and surface reflectances if it is assumed that the ground surface is the Lambertian diffuse reflector.

5. References

- Hansen, J.E. and L.D. Travis (1974). Light scattering in planetary atmospheres, *Space Science Reviews*, vol.16, pp.527-610.
- Hsu, N.C, S-C. Tsay, M.D.King and J.R.Herman, (2004). Aerosol properties over bright-reflecting source regions, *IEEE trans. Geosci. Remote Sensing*, vol.42, pp.557-569
- Ishimoto, H. and K. Masuda (2002). A Monte Carlo approach for the calculation of polarized light: application to an incident narrow beam, *J. Quant. Spectrosc. Radiat. Transfer*, Vol.72, pp.467-483
- Kaufman, Y.J, D.Tanre, L.A.Remer, E.F.Vermote, D.A.Chu and B.A.Holben, (1997a). Remote sensing of tropospheric aerosol from EOS-MODIS over the land using dark targets and dynamic aerosol models, *J. Geophys. Res.*, vol.17051-17067

- Kaufman, Y.J., A. Walt, L.A. Remer, B.C. Gao, R.R. Li and L. Flynn, (1997b). Remote sensing of aerosol over the continents with the aid of a 2.2 μ m channel, IEEE trans. Geosci. Remote Sensing, vol.35, pp.1286-1298
- Kusaka, T, S. Mori, T. Suzuki and H. Shibata (2001). Estimation of optical parameters of yellow sand dust clouds over desert areas from Satellite_level data, IEEE Proc. IGARSS2001, CD-ROM
- Kusaka, T, F. Satou and Y. Hayato (2002a). Optical properties of Kosa aerosols estimated from the multi-spectral polarization, Proc of Remote Sensing Asia'02 Symposium, SPIE, CD-ROM.
- Kusaka, T, S.Mori and T.Yobuko, (2002b). Estimation of optical parameters of sand dust Clouds and ground Reflectance from satellite data, IEEE proc. of IGARSS2002, CD-ROM
- Kusaka, T and T.Nishisaka, (2004). Simultaneous estimation of optical parameters of the Asian dust and ground reflectance from POLDER data, Proc. of SPIE Vol.5652
- Kusaka, T and H. Kitaguchi, (2007), Evaluation of optical properties of atmospheric aerosols estimated from ground-based polarization measurements, IEEE Proc.of IGARSS2007, Barcelona, Spain, CD-ROM
- Masuda, K and M. Sasaki, (1997). A multi-spectral polarimeter for measurements of direct solar and diffuse sky radiation: Calibration and measurements, Optical Review, Vol.4, pp.496-501
- Nakajima, T and A. Higurashi, (1998). A use of two-channel radiance for an aerosol characterization from space, Geophys. Res. Lett. vol.25, pp.3815-3818, 1998
- NASA, AERONET (Aerosol RObotic NETwork), <http://aeronet.gsfc.nasa.gov/>
- NIES, National Institute for Environmental Studies, Japan, <http://www-lidar.nies.go.jp/>
- O'Brien, D.M. (1998). Monte Carlo integration of the radiative transfer equation in a scattering medium with stochastic reflecting boundary, J. Quant. Spectrosc. Radiat. Transfer, vol.60, pp.573-583
- Tanre, D and M. Legrand, (1991). On the satellite retrieval of Saharan dust optical thickness over land: Two different approaches, J. Geophys. Res., vol.96, pp.5221-5227
- Tanre, D, Y.J. Kaufman, M. Herman and S. Mattoo, (1997). Remote sensing of aerosol over oceans from EOS-MODIS, J. Geophys. Res., vol.102, pp.16971-16988
- Vermote E., D. Tanre, J. L. Deuze, M. Herman, J. J. Morcrette, and S. Y. Kotchenova, (2006). 6S user guide ver.3 in Second simulation of a satellite signal in the solar spectrum - vector (6SV)

Moving Target Detection and Velocity Estimation in Multi-Channel AT-InSAR Systems from Amplitude and Phase Data

Alessandra Budillon

*Dipartimento per le Tecnologie - Università di Napoli "Parthenope"
Naples, Italy*

1. Introduction

Recently, Along Track Interferometric Synthetic Aperture Radar systems (AT-InSAR) have been applied for traffic monitoring of ground vehicles (Meyer et al. 2006, Chapin & Chen, 2006, Hinz et al. 2007).

AT-InSAR systems are composed by more than one SAR antennas (typically two), mounted on the same platform and displaced along the platform moving direction. The separation distance between the antennas is denoted as baseline.

From the acquisitions of two or more image signals these systems are able to recover additional information about the observed scene: they allow the detection of moving targets on the ground and the estimation of their radial velocity (Raney, 1971). This is possible because the interferometric phase, i. e. the $(-\pi, \pi]$ wrapped phase of the signal obtained from the point to point correlation between the complex images acquired from the two interferometric antennas, is related to the radial velocity through a known mapping. Then, after the so-called Phase Unwrapping (PhU) operation, a map of the target range velocity can be retrieved.

Detection and radial velocity estimation of a ground moving target are challenging problems, due to the difficulty of separating the moving target signal from the stationary background (clutter) (Chiu, 2003). Several methods, based on very different approaches, have been proposed in literature, such as Displaced Phase Centre Antennas (DPCA systems) techniques (Gierull & Livingstone, 2004, Chiu & Livingstone 2005), and Space-Time Adaptive Processing (STAP) (Ender, 1999, Klemm, 2002, Gierull & Livingstone, 2004). Interest in investigating AT-InSAR processor is motivated since such alternative techniques attempt to reject or cancel the stationary clutter but have the drawback that can attenuate slowly moving targets (Chiu & Livingstone 2005).

AT-InSAR systems usually use only interferometric phase information in order to estimate radial velocity (Chen, 2004, Budillon et al. 2005, Budillon et al. 2008a) while complex data in place of phase-only data have already been used in AT-InSAR systems detection applications (Gierull, 2004, Zhang et al. 2005, Budillon et al., 2008b) showing that detection performance improve when complex data are used.

In this paper it is proposed to consider both amplitude and phase of the interferometric SAR image since in the case of target velocity estimation of not extended targets, the exploitation of the image amplitude together with the image phase can add more information, even if the amplitude is influenced in a less sensitive way with respect to the phase. AT-InSAR approach can be considered clutter-limited since when the signal to clutter ratio decreases, the velocity estimation becomes gradually more and more critical, till it fails completely. Moreover there are solution ambiguities than can keep the velocity estimation from working correctly due to the wrapped phase measurements.

In this paper both above mentioned problems are solved by using statistical estimation methods, and exploiting multi-channel interferograms. The statistical estimation methods allow taking into account the correct statistics of the involved noise (likelihood model). The use of a multi-channel interferogram, that in this case can be obtained exploiting frequency diversity and/or baseline diversity, has a twofold effect: multi-channel interferograms can help to reduce the variance of the estimation, and, if properly chosen, can allow avoiding solution ambiguities (Budillon et al. 2005, Budillon et al. 2008c).

It is shown that combining the real and imaginary part of more than two acquired images (multi-channel approach) produce significative improvements in the velocity estimation accuracy and a sensitive reduction in the false alarm rate compared with AT-InSAR conventional systems using phase-only data.

In section 2 the AT-InSAR statistical model has been presented and the joint interferogram amplitude and phase distribution has been derived. Based on this distribution, in section 3 a radial velocity estimation maximum likelihood approach using more interferogram channels has been reported. Cramer Lower Bounds and Root Mean Squared Error show the method performance on simulated data using Terra SAR-X parameters and are evaluated in the case of phase-only data and amplitude and phase data. In section 4 a likelihood ratio test is adopted to detect the moving target and performance detection in terms of Probability of detection and false alarm have been examined comparing results obtained on phase-only data and on amplitude and phase data. Moreover a multi-channel detection strategy is proposed and compared with the one based on a single interferogram. Finally follow conclusions in section 5.

2. AT-InSAR statistical model

In this section is presented the statistical model of the AT-InSAR signal. Consider an AT-InSAR system constituted by two antennas moving along the direction x (azimuth) (see Figure 1), and suppose that the two antennas are separated by a baseline b along the azimuth direction x , such that $b \ll H$, where H is the platform quota. Assume a target on the ground moving with a constant velocity $v_T = v_{Tx} x + v_{Tr} r$, where v_{Tx} and v_{Tr} are the velocity components along the azimuth and the line of sight direction (range) r , respectively. The azimuth velocity component v_{Tx} produce a Doppler slope change causing a defocusing in the moving target image. The radial velocity component v_{Tr} produce a Doppler history different from that of the stationary background, and an azimuth displacement of the target. Suppose that $|v_{Tx}|, |v_{Tr}| \ll |v_P|$, where $v_P = v_P x$ is the velocity of the flying platform and $H \gg X$ and $H \gg W$, where X and W are the antenna footprint dimensions.

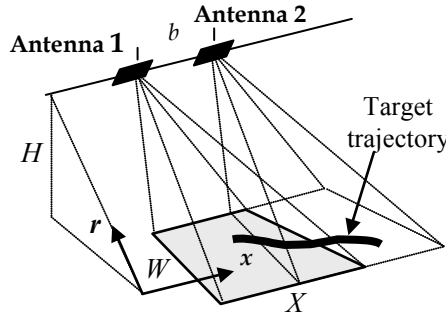


Fig. 1. Along-Track Interferometry system single baseline geometry

The SAR image signal formed by each antenna can be modeled as the superposition of the contributions of the moving target, of the stationary clutter, and of the additive noise. Then in a fixed image pixel we have:

$$\begin{aligned}
 Z_1 &= \begin{cases} Z_{c1} + N_1 + Z_{T1} & \text{in presence of moving target } H_1 \\ Z_{c1} + N_1 & \text{in absence of moving target } H_0 \end{cases} \\
 Z_2 &= \begin{cases} Z_{c2} + N_2 + Z_{T2} & \text{in presence of moving target } H_1 \\ Z_{c2} + N_2 & \text{in absence of moving target } H_0 \end{cases}
 \end{aligned} \quad (1)$$

Z_1 and Z_2 are the computed image signals in the considered pixel, Z_{c1} and Z_{c2} are the clutter signals acquired by the two antennas, N_1 and N_2 represent the receiver thermal noise, and Z_{T1} and Z_{T2} denote the SAR images of the moving target produced by the two interferometric antennas, which will exhibit a phase factor related to the radial velocity:

$$Z_{T1} = A_1, \quad Z_{T2} = A_2 e^{-j\phi_v}, \quad (2)$$

where A_1 and A_2 are the target images obtained for zero velocity, and ϕ_v is given by (Raney, 1971):

$$\phi_v = \left\langle \frac{4\pi b}{\lambda} \frac{v_r}{|v_p|} \right\rangle_{2\pi} = \left\langle \frac{4\pi b}{\lambda} u_r \right\rangle_{2\pi}, \quad (3)$$

where λ is the wavelength corresponding to the working frequency $f=c/\lambda$ of the SAR system, and $\langle \cdot \rangle_{2\pi}$ represents the “modulo- 2π ” operation. In (3) the normalized radial velocity $u_r=v_r/|v_p|$ has been also introduced. Where the moving target is present, $v_r \neq 0$ and consequently $\phi_v \neq 0$, otherwise the along-track interferometric phase (3) is null. From (3) it is easy to derive that the ambiguity velocity value, such that the interferometric phase is equal

to $\pm \pi$, is given by $u_{r,amb} = \pm \lambda / (4b)$. For $|u_r| > \lambda / (4b)$ the interferometric phase wraps, as evidenced also by the “modulo- 2π ” operation. Also disturbing effects have to be taken into account, they are related to different parameters such as the signal to clutter ratio (SCR), the clutter to thermal noise ratio (CNR), and the clutter coherence γ_c . Since the time elapsing between the two interferometric acquisitions is very small (typically of the order of a millisecond) the clutter coherence can be considered equal to one. Then only the effect of SCR and CNR has to be considered.

To analyze the effect that the clutter and noise signals have on the velocity estimation accuracy, a statistical model for the involved signals has to be introduced. It is well known that the clutter signals Z_{c1} and Z_{c2} can be assumed random processes, whose real and imaginary parts are mutually uncorrelated Gaussian signals, with zero mean and same variance σ_c^2 , since they are resulting from the superposition of the signals backscattered from many scattering centres lying in the resolution cell. N_1 and N_2 can be modelled as two additive (to the clutter) zero mean Gaussian complex processes independent of each other, independent on the clutter and with same variance $2\sigma_N^2$.

When the moving target is present, a deterministic model is applicable to the case of a target whose Radar Cross Section (RCS) can be expressed by a deterministic function of the incidence angle (Budillon et al., 2008a). This model applies to canonical scattering objects (such as corner reflectors, spheres, etc.), and to complex or extended targets whose RCS does not rapidly change between the interferometric acquisitions. An accurate knowledge of the average RCS values can be available only for accurately characterized targets (Palubinskas et al. 2004).

A Gaussian model for the target allows to take into account the lack of knowledge of the target RCS values (that can be described in terms of variance σ_T^2), and then of the SCR, and applies to complex or extended targets which can be considered to consist of a large number of isotropic scattering elements, randomly distributed in a region whose dimensions are large compared to the wavelength of the illuminating radiation, and all contributing to the overall signal with the same weight. In the following the target signals Z_{T1} and Z_{T2} , have been modelled as zero mean (complex) Gaussian processes.

Then, when the moving target is absent ($Z_{T1} = Z_{T2} = 0$), the two processes Z_1 and Z_2 are Gaussian with zero-mean and correlation coefficient γ_{H0} given by:

$$\gamma_{H0} = \frac{E[(Z_{c1} + N_1)(Z_{c2} + N_2)^*]}{\sqrt{E[|Z_{c1} + N_1|^2]E[|Z_{c2} + N_2|^2]}} = \frac{\gamma_c}{\left(1 + \frac{1}{\text{CNR}}\right)}, \quad (4)$$

where $E[\cdot]$ denotes the expectation operation, $*$ denotes the conjugate, γ_c is the clutter coherence (real valued, in the ATI application can be considered equal to one), representing the correlation between images Z_{c1} and Z_{c2} , and $\text{CNR} = \sigma_c^2 / \sigma_N^2$, where $2\sigma_c^2$ and $2\sigma_N^2$ are the clutter and thermal noise powers respectively (the factor two is due to the sum of the powers of the real and imaginary parts).

Instead, when we are in presence of the moving target ($Z_{T1} \neq 0$, $Z_{T2} \neq 0$), the expression of correlation coefficient change with respect to (4) and γ_{H1} is given by:

$$\begin{aligned} \gamma_{H_1} &= \frac{E\left[(Z_{c1} + N_1 + Z_{T1})(Z_{c2} + N_2 + Z_{T2})^*\right]}{\sqrt{E\left[|Z_{c1} + N_1 + Z_{T1}|^2\right]} \sqrt{E\left[|Z_{c2} + N_2 + Z_{T2}|^2\right]}} = \\ &= \frac{\gamma_c \sigma_c^2 + \sigma_n^2 \gamma_T}{\sigma_c^2 + \sigma_n^2 + \sigma_T^2} \sqrt{\frac{\gamma_c + \gamma_T \text{SCR}}{1 + \frac{1}{\text{CNR}} + \text{SCR}}} \end{aligned} \quad (5)$$

where $\text{SCR} = \sigma_T^2 / \sigma_c^2$, where $2\sigma_T^2$ is the target power, and γ_T is the target (complex) coherence that depends on the target velocity through the nominal phase (3):

$$\gamma_T = \frac{E\left[Z_{T1} Z_{T2}^*\right]}{\sqrt{E\left[|Z_{T1}|^2\right]} \sqrt{E\left[|Z_{T2}|^2\right]}} = \frac{E\left[A_1 A_2^*\right]}{\sqrt{E\left[|A_1|^2\right]} \sqrt{E\left[|A_2|^2\right]}} e^{j\varphi_0} = \gamma_A e^{j\varphi_0}, \quad (6)$$

where γ_A is the target coherence for zero radial velocity that is usually assumed equal to one.

The two processes $Z_1 = Z_{1r} + jZ_{1i}$ and $Z_2 = Z_{2r} + jZ_{2i}$ are Gaussian, then the joint probability density function of $\mathbf{Z} = [Z_{1r} \ Z_{2r} \ Z_{1i} \ Z_{2i}]^T$, is Gaussian with zero mean and covariance matrix

$$\mathbf{C} = \begin{cases} \mathbf{C}_c + \mathbf{C}_N & H_0 \\ \mathbf{C}_c + \mathbf{C}_N + \mathbf{C}_T & H_1 \end{cases} \quad (7)$$

where the matrices \mathbf{C}_c , \mathbf{C}_N and \mathbf{C}_T are respectively the clutter, noise and target covariance matrix. It can be easily shown (Davenport & Root, 1958) that:

$$\mathbf{C} = \sigma^2 \begin{bmatrix} 1 & \text{Re}(\gamma) & 0 & -\text{Im}(\gamma) \\ \text{Re}(\gamma) & 1 & \text{Im}(\gamma) & 0 \\ 0 & \text{Im}(\gamma) & 1 & \text{Re}(\gamma) \\ -\text{Im}(\gamma) & 0 & \text{Re}(\gamma) & 1 \end{bmatrix}, \quad (8)$$

where, in the hypothesis H_0 , has to be taken $\sigma^2 = \sigma_c^2 + \sigma_N^2$ and $\gamma = \gamma_{H_0}$, in the hypothesis H_1 , $\sigma^2 = \sigma_c^2 + \sigma_N^2 + \sigma_T^2$ and $\gamma = \gamma_{H_1}$.

Then the joint probability density function of $\mathbf{Z} = [Z_{1r} \ Z_{2r} \ Z_{1i} \ Z_{2i}]^T$, is Gaussian, i.e.

$$f_{\mathbf{Z}}(\mathbf{z}) = f_{Z_{1r}, Z_{2r}, Z_{1i}, Z_{2i}}(z_{1r}, z_{2r}, z_{1i}, z_{2i}) = \frac{1}{(2\pi)^2 |\mathbf{C}|^{1/2}} \exp\left\{-\frac{1}{2} \mathbf{z}^T \mathbf{C}^{-1} \mathbf{z}\right\}. \quad (9)$$

The SAR interferometric amplitude and phase distribution can be derived from (9) introducing the interferometric signal I:

$$I = Z_1 Z_2^* = W \exp(j\Phi). \quad (10)$$

The joint W and Φ pdfs, in the hypothesis H_1 and H_0 , derived from (9) via variables transformations (Davenport & Root, 1958), are respectively:

$$f_{W\Phi}(w, \phi; H_1) = \frac{w}{2\pi\sigma_c^2 \left(1 + \frac{1}{\text{CNR}} + \text{SCR}\right)^2 (1 - |\gamma_{H1}|^2)} \cdot K_0 \left\{ \frac{w}{\sigma_c^2 \left(1 + \frac{1}{\text{CNR}} + \text{SCR}\right) (1 - |\gamma_{H1}|^2)} \right\} \exp \left\{ \frac{|\gamma_{H1}| w \cos(\phi - \phi_o)}{\sigma_c^2 \left(1 + \frac{1}{\text{CNR}} + \text{SCR}\right) (1 - |\gamma_{H1}|^2)} \right\}, \tag{11}$$

$$f_{W\Phi}(w, \phi; H_0) = \frac{w}{2\pi\sigma_c^2 \left(1 + \frac{1}{\text{CNR}}\right)^2 (1 - \gamma_{H0}^2)} \cdot K_0 \left\{ \frac{w}{\sigma_c^2 \left(1 + \frac{1}{\text{CNR}}\right) (1 - \gamma_{H0}^2)} \right\} \exp \left\{ -\frac{\gamma_{H0} w \cos(\phi)}{\sigma_c^2 \left(1 + \frac{1}{\text{CNR}}\right) (1 - \gamma_{H0}^2)} \right\}.$$

where K_0 denote the modified Bessel function of order zero, $w \geq 0, 0 \leq \phi \leq 2\pi$ and ϕ is γ_{H1} phase. The analytical expressions of the pdfs (9) and (11) and of the coherences (4) and (5), reveal their dependence on CNR, SCR, radial velocity (through the phase ϕ_v), clutter coherence γ_c and target coherence γ_A . As far as the coherence values are concerned, they are assumed equal to 1 in ATI applications.

In Figure 1 and Figure 2 the joint pdf of W and Φ , respectively in the hypothesis H_0 and H_1 , with CNR= 10 dB, SCR= 10 dB), $u_r = u_{r,amb}/2$ corresponding to phase 1.5 rad, are shown.

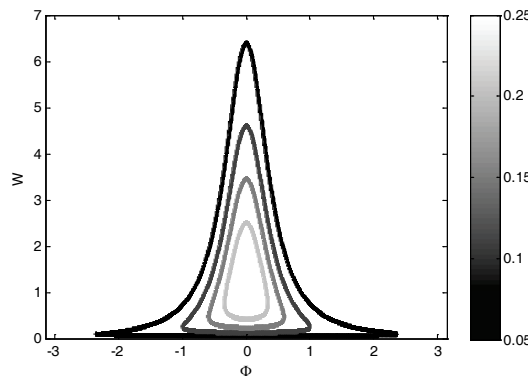


Fig. 2. Interferogram joint amplitude phase pdf in the hypothesis H_0 for CNR=10 dB.

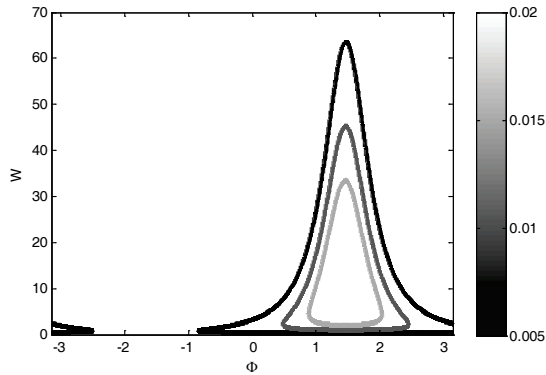


Fig. 3. Interferogram joint amplitude phase pdf in the hypothesis H_1 for $CNR= 10$ dB, $SCR= 10$ dB, $u_r=u_{r,amb}/2$, corresponding to phase 1.5 rad.

3. Multi-Channel AT-InSAR moving target velocity estimation

3.1 Joint estimation of velocity and SCR via Maximum likelihood approach

Since interferometric phase is measured in the interval $(-\pi,\pi]$, then a Phase Unwrapping (PhU) operation is required to retrieve the target radial velocity. The PhU operation presents solution ambiguities when only one phase interferogram (single-channel) is used. It has already been shown in (Budillon et al. 2005, Budillon et al. 2008c) that the joint use of multi-channel configurations (deriving from the use of more than two interferometric images acquired with different baselines or at different working frequencies) and of classical statistical estimation techniques allows to obtain very accurate solutions and to overcome the limitations due to the presence of ambiguous solutions, intrinsic in the single-channel configurations.

Different baseline data sets (at least two) can be generated when the AT-InSAR system is constituted by more than two antennas (at least three). Different frequency data sets can be generated in two ways. In the first, we can suppose that the SAR sensors can operate at different working frequencies, for instance in X and C bands simultaneously. In the second, the multi-frequency interferograms can be obtained by sub-band filtering the interferometric images splitting the overall bandwidth (Budillon et al. 2008c). Note that while the use of a different working frequency or baseline does not affect the SCR values, the generation of adding frequencies by partitioning the available band reduces the SCR value. This SCR reduction is in inverse relation to the number of looks, as the spatial resolution worsens increasing the number of looks.

Likelihood function is easily derived from either pdf (9) or (11) in the H_1 hypothesis. As discussed in the previous section, it depends on CNR , SCR and radial velocity since clutter and target coherence are assumed equal to 1 in ATI applications. CNR value can be easily computed from the data isolating an area where the target is absent. Then, the final estimation can be casted as a joint maximum likelihood estimation (Kay, 1993) of velocity and SCR :

$$\begin{aligned} [\hat{u}_r, \hat{SCR}] &= \arg \max_{u_r, SCR} L(u_r, SCR) \\ L(u_r, SCR) &= \prod_{k=1, \dots, N} \underbrace{f_{\mathbf{Z}}(\mathbf{Z}(k) | u_r, SCR, H_1)}_{\text{single channel likelihood function}} \end{aligned} \quad (12)$$

where the samples $\{\mathbf{Z}(k)\}_{k=1, N}$ represent the SAR signals acquired in N channels. The factorization in (12) comes from the assumed statistical independence of the multi-channel interferograms.

3.2 Performance assessment

Estimation performance evaluation has been carried out using Terra SAR-X parameters in Table 1, but in order to consider a multi-channel system a second baseline $b_2=1.8b_1$ [m] ($b_1=1.2$ [m]), has been adjoined. By sub-band filtering the interferometric images also 4 azimuth looks have been considered, obtaining in total N=8 channels. The maximum normalized radial velocity value that can be unambiguously detected results $|u_{r, \max}| = \lambda / (4b) = 6.5 \times 10^{-3}$, corresponding to a not normalized velocity $|v_{r, \text{amb}}|$ of about 49.4 m/sec (178 Km/h).

TerraSAR-X	
Quota	514.8 Km
Platform velocity	7.6 Km/s
Along track antenna dimension	4.8 m
Across track antenna dimension	0.8 m
Along track baseline	1.2 m
Working frequency	X band- $f_X=9.65$ GHz
Wavelength	3.12 cm
Range bandwidth	150 MHz

Table 1. Main parameters of Terra SAR X system

To evaluate the performance of the ML estimator (12), the Cramer Rao Lower Bounds (CRLBs) (Kay, 1993) and the Root Mean Square Errors (RMSEs) for the unknown parameters (v_r, SCR) have been computed. CRLBs depend on the data model and represent the maximum accuracy attainable with given data. In order to point out the advantages of taking into account amplitude and phase information they have been compared with the ones obtained using a phase-only approach, i.e. a maximum likelihood estimation based on the phase-only distribution (Budillon et al. 2008a).

The $CRLB^{1/2}$ relative to the not normalized radial velocity v_r and SCR are reported respectively in Figures 4 and 5. In Figure 4(a) it is shown the $CRLB^{1/2}$ relative to the estimation of the not normalized radial velocity v_r . Vs. the radial velocity and relevant to the model based on phase-only data. It is evident the significant improvements in the maximum accuracy attainable using the amplitude and phase model reported in Figure 4(b). The $CRLB^{1/2}$ have been evaluated numerically and for different values of v_r in the range (0,

$v_{r,amb}$), for CNR=10 dB, and varying SCR (0,5,10,15,20 dB). It can be appreciated that as expected the $CRLB^{1/2}$ are lower for higher SCRs.

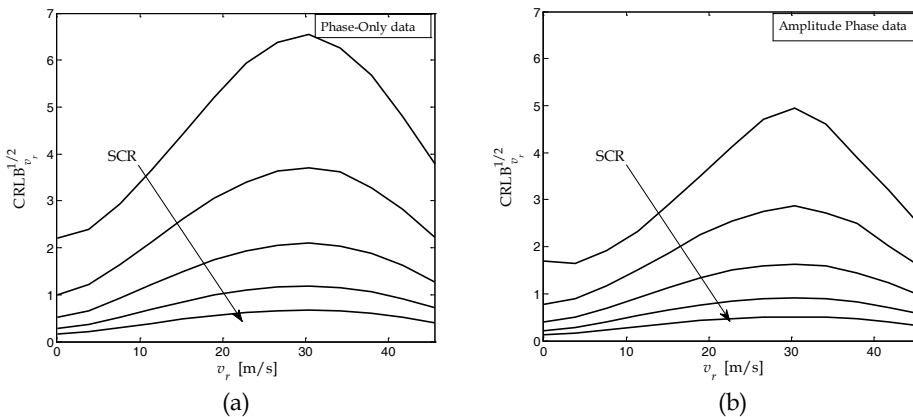


Fig. 4. $CRLB^{1/2}$ relative to the estimation of the radial velocity v_r , Vs. the radial velocity and relevant to the model based on phase-only data (a) and to the model based on amplitude and phase data (b), for CNR=10 dB, and varying SCR (0,5,10,15,20 dB) .

In Figure 5(a) the $CRLB^{1/2}$ relative to the estimation of the SCR, Vs. the SCR and relevant to the model based on phase-only data is shown. Also for this parameter it is evident the significant improvements in the maximum accuracy attainable using the amplitude and phase model reported in Figure 5(b). Moreover it can be seen that using the amplitude and phase model the accuracy is slightly dependent on the SCR values, in both case the $CRLB^{1/2}$ is smaller for higher values of radial velocities, i.e. as expected it is easier to estimate the SCR of a faster target. In Figure 6(a) are shown the $CRLB^{1/2}$ and the RMSE relative to the estimation of the not normalized radial velocity v_r , Vs. the radial velocity and relevant to the model based on phase-only data, for CNR=10 dB, and SCR=10 dB. They can be compared with the correspondent $CRLB^{1/2}$ and the RMSE relevant to the model based on amplitude and phase data in Figure 6(b). It is noticeable that the performances are improved and also the RMSE is closer to the $CRLB^{1/2}$ when the estimation is based on the amplitude and phase model. In Figures. 7(a) and 7(b) the statistical mean values and the RMSEs for different estimated velocities in the range $(0, v_{r,amb})$, for CNR=10 dB and SCR=10 dB, respectively for the model based on phase-only data and to the model based on amplitude and phase data, are reported. Finally in Figures. 8(a) and 8(b) $CRLB^{1/2}$ and the RMSE relative to the estimation of SCR, Vs. the radial velocity are shown respectively for the two models. It is evident again the improvement attainable in case it is assumed the amplitude and phase model.

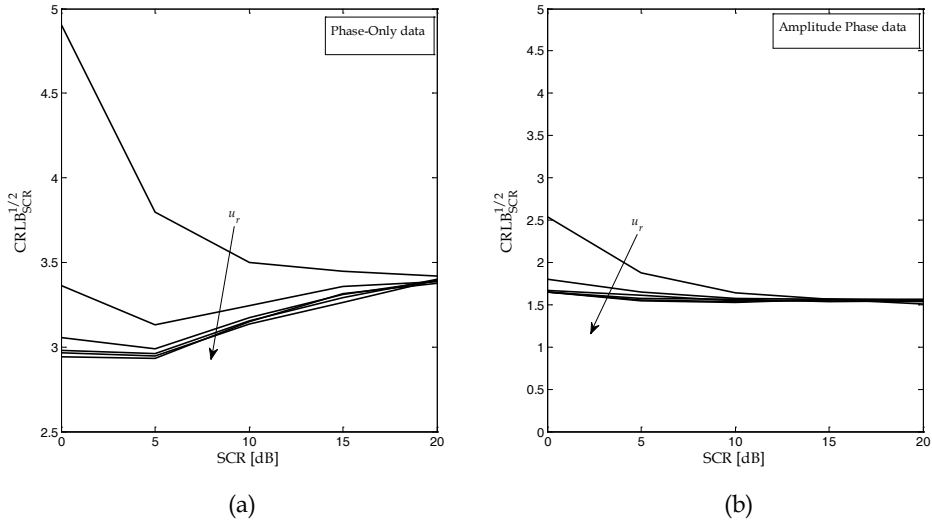


Fig. 5. CRLB^{1/2} relative to the estimation of SCR, Vs. SCR and relevant to the model based on phase-only data (a) and to the model based on amplitude and phase data (b), for CNR=10 dB, and varying v_r in the range $(0, v_{r,amb})$.

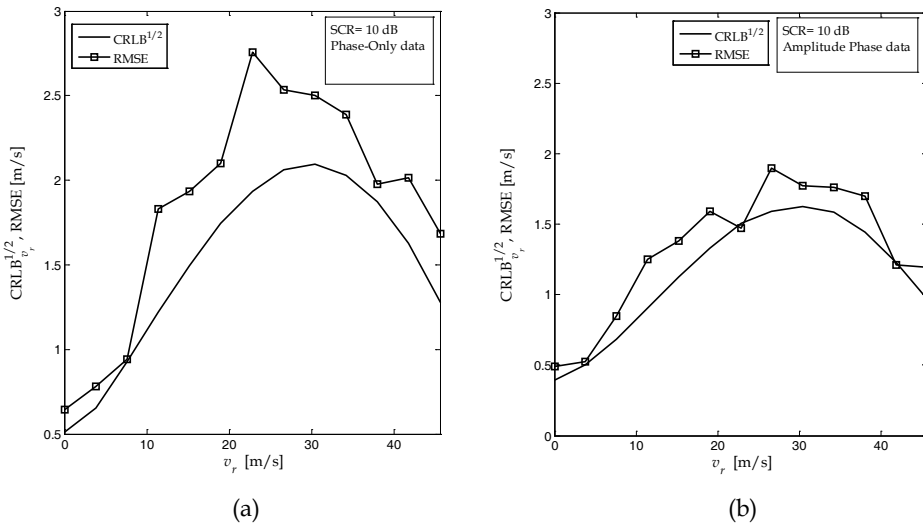


Fig. 6. CRLB^{1/2} and RMSE relative to the estimation of the radial velocity v_r , Vs. the radial velocity and relevant to the model based on phase-only data (a) and to the model based on amplitude and phase data (b), for CNR=10 dB, and SCR=10 dB.

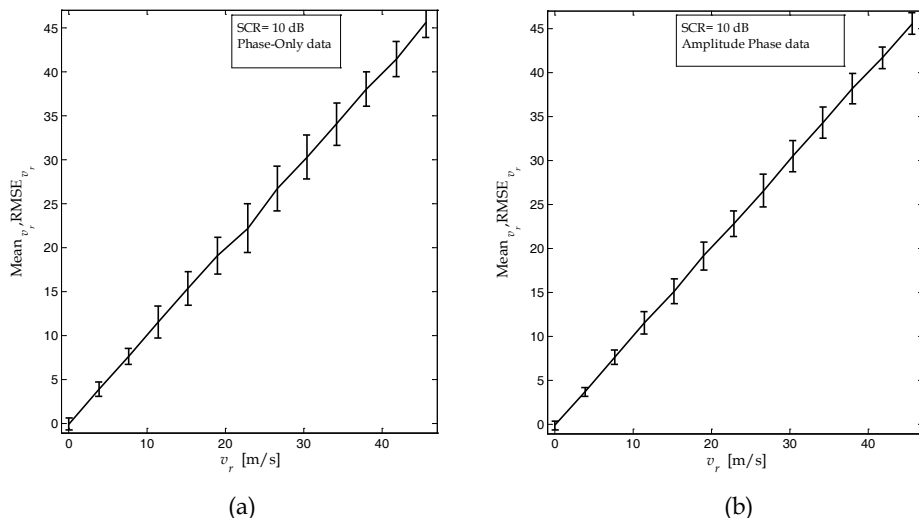


Fig. 7. Mean value and RMSE relative to the estimation of the radial velocity v_r , Vs. the radial velocity and relevant to the model based on phase-only data (a) and to the model based on amplitude and phase data (b), for $CNR=10$ dB, and $SCR=10$ dB.

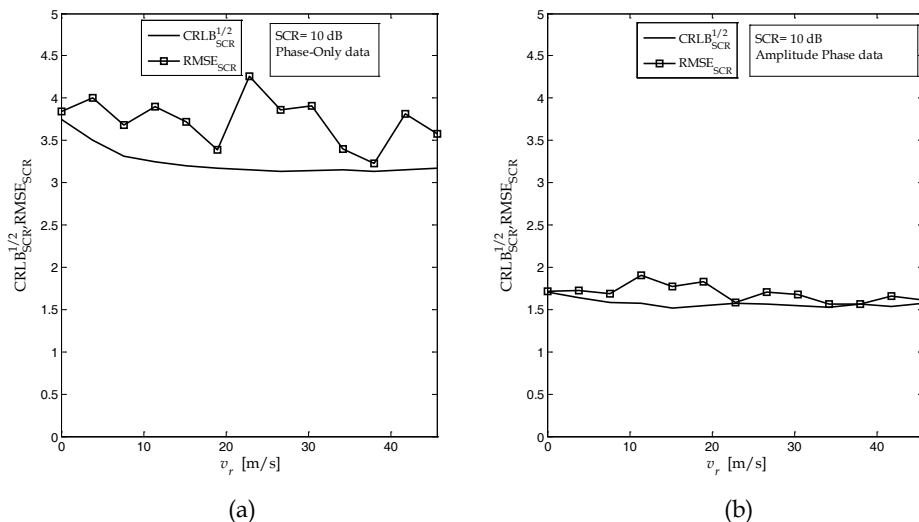


Fig. 8. $CRLB^{1/2}$ and RMSE relative to the estimation of SCR , Vs. the radial velocity and relevant to the model based on phase-only data (a) and to the model based on amplitude and phase data (b), for $CNR=10$ dB, and $SCR=10$ dB.

4. Multi-Channel AT-InSAR moving target detection

4.1 Likelihood ratio test

A moving target can be detected in the conventional way by comparing the interferometric phase ϕ with a threshold η_T in the interval $(-\pi, \pi]$.

The performance of the detection process can be evaluated using the interferometric phase statistics (Budillon et al. 2088a). They are, as expected, better for high values of SCR, i.e. when the moving targets power is significantly larger than the clutter power. For moving targets mingling with the background clutter, the detection capability worsens, so that if one wants low values of P_{FA} , the P_D can decrease to very low values, not consistent with the applications.

As in the case of the velocity estimation (see Section 3) both amplitude and phase of the interferogram are considered instead of taking into account only the interferometric phase. Based on the pdfs (11) a constant false alarm rate (CFAR) detector can be designed.

In order to detect a moving target a likelihood ratio test is proposed, likelihood is derived from (11):

$$\Lambda(z) = \frac{f_{\mathbf{W}\Phi}(w, \phi; \hat{u}_r, \hat{SCR}, H_1)}{f_{\mathbf{W}\Phi}(w, \phi; H_0)} \underset{H_1}{\overset{H_0}{>}} \eta. \quad (13)$$

Probability of false alarm P_{FA} and Probability of detection P_D are derived from (13)

$$\begin{aligned} P_{FA} &= Pr\{A \geq \eta; H_0\} \\ P_D &= Pr\{A \geq \eta; H_1\} \end{aligned} \quad (14)$$

The threshold η depends on a fixed P_{FA} .

4.2 Performance assessment

The detection performance evaluation has been carried out using the same multi-channel system presented in section 3.2.

The proposed approach provides curves of separation between the two classes (see Figure 9-10). In Figure 9 the separation curve for the two hypothesis, presence and absence of a moving target, has been evaluated for CNR=10 dB, SCR=10 dB, $u_r = u_{r,amb}/2 = 3.25 \times 10^{-3}$ (corresponding to the nominal noise-free value $\phi_0 = \pi/2$), a threshold has been chosen such that $P_D=0.91$ and $P_{FA}=0.001$. Figure 10 shows the separation curve for the two hypothesis for CNR=10 dB, SCR= 0 dB, $u_r = u_{r,amb}/2 = 3.25 \times 10^{-3}$, a threshold has been chosen such that $P_D=0.7$ and $P_{FA}=0.05$.

For comparison with the conventional interferometric approach, the two Receiver Operating Characteristics (ROC) have been derived in both case SCR= 10 dB and SCR= 0 dB, for the amplitude and phase approach (solid line) and for the phase-only case (dashed line) (see Figure 11). It is clear the advantage in considering both amplitude and phase, for a fixed P_{FA} a higher P_D can be obtained.

In Figure 12 it can be appreciated the ROC dependence, in the amplitude and phase approach, on the radial velocity (Figure 12(a)), for CNR=10 dB, SCR=10 dB and on SCR

(Figure 12 (b)), for CNR=10 dB and $u_r = u_{r,amb}/2$. As expected it is easier to detect a faster and stronger (in terms of reflectivity) target.

In order to exploit the multi-channel interferograms in the detection process, suppose that the detection probability of one of the channel corresponding to the first baseline is equal to P_{D1} , and that the detection probability of one of the channel corresponding to the second baseline is equal to P_{D2} .

The probability that the target is detected from $(N/2+j)$ channels ($j=1, \dots, N/2$) on a total of N channels results:

$$P_j = \sum_{k=j}^{N/2} \binom{N/2}{k} P_{D1}^k (1 - P_{D1})^{N/2-k} \left[\binom{N/2}{N/2-k+j} P_{D2}^{N/2-k+j} (1 - P_{D2})^{k-j} \right]. \quad (15)$$

The proposed multi-channel detection strategy consists in considering the moving target present when the majority of the interferogram values are above prefixed thresholds, so that the detection probability results:

$$P_{D>N/2} = \sum_{j=1}^{N/2} P_j. \quad (16)$$

For the estimation of the false alarm probability can be used the same reasoning.

In Figure 13 the ROC in the multi-channel amplitude phase approach (solid line) for CNR=10 dB, SCR=10 dB, $N=8$ interferograms compared with the single channel amplitude phase approach is reported (dashed lines). It is evident the advantages in considering a multi-channel approach that allows to keep low P_{FA} and at the same time high P_D .

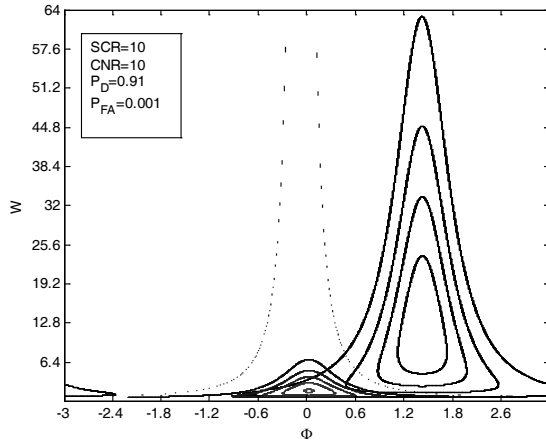


Fig. 9. Interferogram joint amplitude phase pdf in the hypothesis H_0 and H_1 for $u_r = u_{r,amb}/2$, CNR=10 dB, SCR=10 dB, and the separation curve.

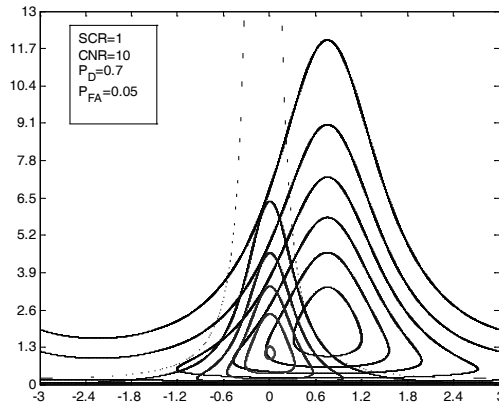


Fig. 10. Interferogram joint amplitude phase pdf in the hypothesis H_0 and H_1 for $u_r = u_{r,amb}/2$, $CNR = 10$ dB, $SCR = 0$ dB, and the separation curve.

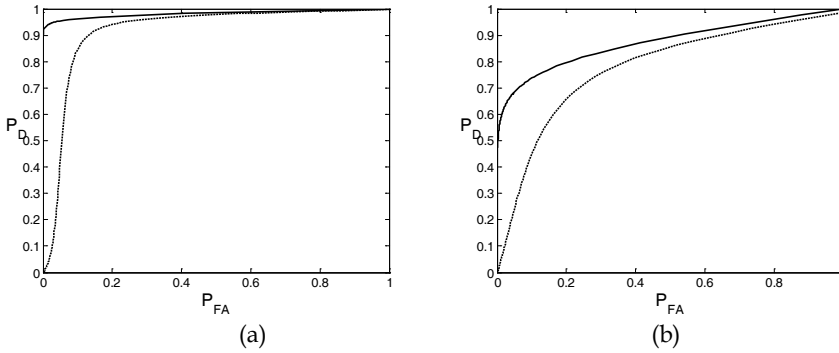


Fig. 11. ROC in the amplitude phase approach (solid line) and in the phase-only approach (dashed line) for $u_r = u_{r,amb}/2$, $CNR = 10$ dB, $SCR = 10$ dB (a) and $SCR = 0$ dB (b).

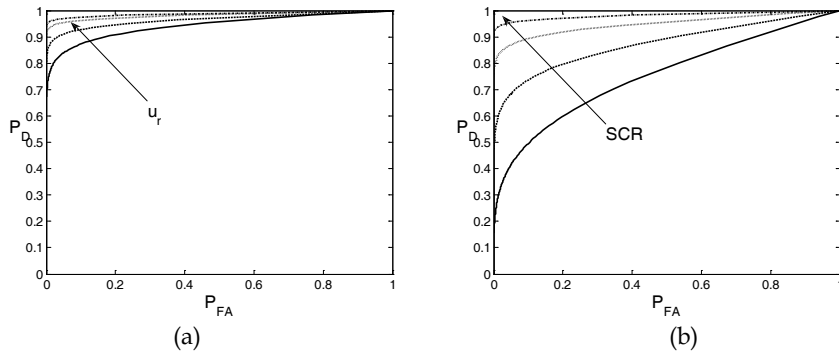


Fig. 12. ROC in the amplitude phase approach for $CNR = 10$ dB, $SCR = 10$ dB, varying u_r in the range $(0, u_{r,amb})$ (a) and for $u_r = u_{r,amb}/2$, varying SCR (-5 dB, 0 dB, 5 dB, 10 dB) (b).

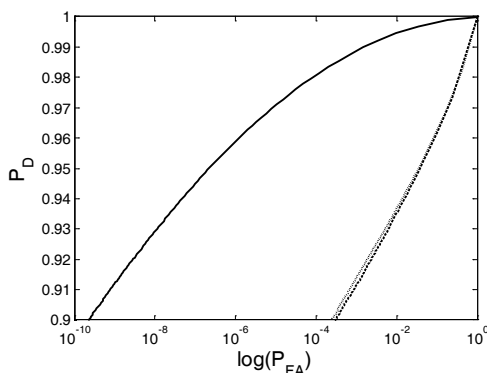


Fig. 13. ROC in the multi-channel amplitude phase approach (solid line) for CNR=10 dB, SCR=10 dB, N=8 interferograms compared with the single channel amplitude phase approach (dashed line baseline b_1 dotted line baseline b_2).

5. Conclusion

In this paper it has been presented the performance evaluation of multi-channel AT-InSAR systems, exploiting both amplitude and phase interferogram information, in terms of target radial velocity estimation accuracy and moving target detection ability.

A Gaussian target response model has been considered and the amplitude and phase joint pdf of the interferogram has been derived. Based on this model a maximum likelihood approach has been used to estimate the target radial velocity. A comparison of the proposed approach with a phase-only system has been reported considering different system parameters, such as radial velocity, SCR, CNR. It reveals the benefits in exploiting both amplitude and phase interferogram information in terms of CRLB and RMSE. The analysis has been performed considering a multi-channel system based on the Terra SAR-X parameters but with a second baseline.

A constant false alarm rate (CFAR) detector has been considered. This approach provides curves of separation between the two classes, hypothesis H_0 , and H_1 , in order to detect the pixels in the SAR images where a moving target is present. The proposed approach outperforms the conventional phase-only approach and in particular is able to detect target with low SCR. ROCs have been presented varying SCR and the normalized radial velocity.

Regarding the detection process, the use of multi-channel interferograms, after the application of a threshold stage to each channel, allows to adopt a binary integration to combine single-channel decisions. Such a strategy, compared with the one based on a single interferogram, provides better results in terms of simultaneous low values of P_{FA} and high values of P_D .

6. References

- Budillon, A.; Ferraiuolo, G.; Pascazio, V. & Schirinzi, G. (2005). "Multi-Channel SAR Interferometry via Classical and Bayesian Estimation Techniques", *J. of Applied Signal Processing*, vol. 20, pp. 3180-3193.

- Budillon, A.; Pascazio, V. & Schirinzi, G. (2008a). "Estimation of Radial Velocity of Moving Targets by Along-Track Interferometric SAR Systems", *IEEE Geosci. Remote Sensing Letters*, vol. 5, pp. 349-353.
- Budillon, A.; Pascazio, V. & Schirinzi, G. (2008b). "Moving Target Detection in Along Track SAR Interferometry from In-Phase and Quadrature Components Data", in *Proc. of IEEE International Geoscience and Remote Sensing Symposium (IGARSS'08)*, pp. III - 1178 - III - 1181.
- Budillon, A.; Pascazio, V. & Schirinzi, G. (2008c). "Multichannel Along-Track Interferometric SAR Systems: Moving targets Detection and Velocity Estimation", *International Journal of Navigation and Observation*, Vol. 2008, 16 pp.
- Chapin, E. & Chen, C.W. (2006). "GMTI Along-Track Interferometry Experiment", *IEEE Aerosp. Electron. Syst. Mag.*, vol. 21, pp. 15-20.
- Chen, C.W. (2004). "Performance Assessment of Along-Track Interferometry for Detecting Ground Moving Targets", in *Proc. of 2004 IEEE Radar Conference*, pp. 99-104.
- Chiu, S. (2003). "Clutter effects on ground moving target velocity estimation with SAR along-track interferometry", in *Proc. of 2003 IEEE International Geoscience and Remote Sensing Symposium (IGARSS '03)*, Toulouse (France), vol. 2, pp. 1314-1319.
- Chiu, S. & Livingstone, C. (2005). "A comparison of displaced phase centre antenna and along-track interferometry techniques for RADARSAT-2 ground moving target indication", *Can. J. Remote Sensing*, vol. 31, No. 1, pp. 37-51.
- Davenport, W.B. & Root, W.L. (1958). *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, Kogakusha,
- Ender, J.H.G. (1999). "Space-Time Processing for Multichannel Synthetic Aperture Radar", *IEE Electronics & Communication Engineering Journal*, vol. 11, pp. 29-38.
- Gierull, C.H. (2004). "Statistical Analysis of Multilook SAR Interferograms for CFAR Detection of Ground Moving Targets," *IEEE Trans. Geosci. Remote Sensing*, vol. 42, pp. 691-701.
- Gierull, C.H. & Livingstone, C. (2004). "SAR-GMTI concept for RADARSAT-2" in *Applications of Space-Time Adaptive Processing*, R. Klemm (Ed.), IEE Publishers, London, UK.
- Hinz, S.; Meyer, F.; Eineider, M. & Bamler, R. (2007). "Traffic Monitoring with Spaceborne SAR—Theory, Simulations, and Experiments", *Computer Vision and Image Understanding*, vol. 106, pp. 231-244.
- Kay, S.M. (1993). *Fundamentals of Statistical Signal Processing: Vol. I, Estimation Theory*, Prentice-Hall.
- Klemm, R.K. (2002). *Principles of space-time adaptive processing*, 2nd edn, London: IEE.
- Meyer, F.; Hinz, S.; Laika, A.; Wehling, D. & Bamler, R. (2006). "Performance Analysis of the TerraSAR-X Traffic Monitoring Concept", *ISPRS J. of Photogr. and Remote Sensing*, vol. 61(3/4), pp 225-242.
- Palubinskas, G.; Runge, H. & Reinartz, P. (2004). "Radar signatures of road vehicles", In *Proc. of 2004 IEEE International Geoscience and Remote Sensing Symposium (IGARSS '04)*, Anchorage (USA), vol. 2, pp. 1498-1501.
- Raney, R. (1971). "Synthetic Aperture Imaging Radar and Moving Targets", *IEEE Trans Aerosp. Electron. Syst.*, vol. 7, pp. 499-505.
- Zhang, Y.; Hajjari, A.; Kim, K. & Himed, B. (2005). "A Dual-Threshold ATI-SAR Approach for Detecting Slow Moving Targets", in *Proc. of 2005 IEEE Radar Conference*, pp. 295-299.

Monitoring tropical peat swamp deforestation and hydrological dynamics by ASAR and PALSAR

Dirk Hoekman

*Wageningen University, Dept. of Environmental Sciences
The Netherlands*

1. Introduction

Thick deposits of peat underneath tropical peat swamp forests are among the world's largest reservoirs of carbon. Although occupying only about 0.3% of the global land surface, they contain as much as 20% of the global peat soil carbon stock, representing 63-148 Giga ton of carbon (Rieley and B. Setiadi, 1997; MacDicken, 2002). A value of approximately 70 Giga ton of carbon is cited by (Sabine *et al.*, 2004). This wide range of values illustrates a large uncertainty. The uncertainty is large as peat depth and carbon densities are poorly described. Tropical peat swamp forests have an uneven global distribution. Most of the areas occur in South-East Asia. The carbon stored in South-East Asian peatlands is estimated to be over 42 Giga ton (Hooijer *et al.*, 2007). The tropical peat swamp forests of South-East Asia account for approximately 26.5 million ha of the total tropical resource of approximately 38 million ha, with Indonesia alone contributing an estimated 17-27 million ha (Waldes and Page, 2002).

Tropical peat swamp forests are threatened by large scale deforestation, canal drainage and forest fire, causing enormous carbon emissions (Goldammer, 1999; IUCN/WWF, 2000; Hooijer *et al.*, 2007; Van der Werf *et al.*, 2008). Large scale conversion of peat swamp forest into, for example, oil palm or Acacia plantations, requires draining. The associated sustained low soil water levels cause oxidation of the peat and, consequently, large emissions of carbon dioxide (e.g. Fargioni *et al.*, 2008). Forest and peat fires are an additional source of carbon dioxide emission. Emissions from peat swamp fires in Indonesia during the strong 1997-1998 El Niño Southern Oscillation (ENSO) event, for example, have been estimated at 0.8-2.5 Giga ton of carbon. This was equivalent to 13-40% of the global annual emission from anthropogenic fossil fuel combustion (Page *et al.*, 2002).

Despite the relevance of this ecosystem for biodiversity and climate, relatively little is known about its functioning and existing maps are often outdated and of poor quality. However, unique observing capabilities of L-band satellite radar may provide a powerful

tool to observe seasonal dynamics of flooding, the impact of drainage by canals and the condition of the peat swamp forest cover.

The use of L-band radar for wetlands monitoring was first demonstrated on large scale with SAR (Synthetic Aperture Radar) data from the Japanese JERS-1 satellite acquired in the period 1992-1998. For all major tropical rain forest areas of the world multi-temporal (2-3 dates) radar mosaics were created, including South-East Asia (Shimada and Isoguchi, 2002), thus providing a benchmark overview for the past decade. Locally, more data were acquired allowing in-depth studies of tropical forest inundation patterns (e.g. Rosenqvist *et al.*, 2002) and tropical coastal vegetation (e.g. Simard *et al.*, 2002). Recently, some first results have been published for tropical peat swamp forests (Hoekman, 2007; Hoekman and Vissers, 2007).

With the launch of the Advanced Land Observing Satellite (ALOS) on January 24, 2006, a new Japanese spaceborne L-band radar system became available. The Phased Array L-band Synthetic Aperture Radar (PALSAR) on-board ALOS has several observations modes. The PALSAR observation strategy has been designed to provide consistent wall-to-wall observations at fine resolution (Fine Beam mode) of all land areas on Earth on a repetitive basis. For the world's major wetlands areas up to eight additional observations per year in ScanSAR mode are made to capture seasonal dynamics (Rosenqvist *et al.*, 2007a; Rosenqvist *et al.*, 2008). The entire island of Borneo is one of these major wetland areas. Of particular interest is the ability of PALSAR to contribute to objectives of the Ramsar (wetlands) UN convention (Davidson and Finlayson, 2007; Rosenqvist *et al.*, 2007b).

In this chapter methodologies are discussed for mapping biophysical parameters, hydrological modelling and monitoring based on historical JERS-1 radar data, and currently available ALOS PALSAR. Also the use of C-band ENVISAT ASAR, which is of special interest for peat swamp deforestation monitoring, is discussed. Unique features of radar for observation of peat swamp forests are briefly outlined in Section 2. A test site located in the Mawas peat swamp conservation area in Central Kalimantan is used for method development and features a 23 km long research bridge, which crosses an entire intact peat dome. This test site is discussed in Section 3. Sections 4 until 7 discuss various methodologies and results.

2. Radar observation

The use of spaceborne radar to map and monitor peat swamp forests has certain unique advantages. In the first place, observation by radar systems is unimpeded by cloud cover, which is an advantage over optical data in the humid tropics. In the second place, radar can penetrate vegetation cover to a certain extent, depending on wavelength. The JERS-1 and ALOS imaging radar (or SAR) systems use a relatively long wavelength (23.5 cm, or 1.275 GHz), also referred to as L-band. It allows observation of flooding under a closed forest canopy. Hence, in principle, seasonal flooding dynamics can be revealed well. The ENVISAT ASAR C-band radar has a shorter wavelength (5.6 cm, or 5.331 GHz) and, compared to L-band, observes higher parts of the vegetation canopy. Though ASAR, for this reason, is less suitable to observe hydrological features of wetlands, it is still of large interest to monitor deforestation for technical reasons to be discussed later.

A certain level of understanding of the physical interaction between the radar wave and the terrain is necessary to allow for an accurate interpretation of L-band SAR images. Biomass and flooding are the two main terrain parameters and polarisation is one of the most important radar wave parameters describing this interaction. The effect of biomass is an increase of the radar echo (or backscatter) intensity with increasing biomass up to a level of around 100 ton/ha. Notably the so-called HV-polarisation is sensitive for biomass variation. Above this biomass level the radar image intensity saturates and the radar wave does not penetrate the vegetation well. Below this biomass level, or in open canopies, the effect of flooding is noticeable. In this case the interaction mechanism is somewhat different. Since radar instruments are side-looking and the water surface acts as a mirror, smooth open water surfaces yield no radar return, i.e. these areas appear black in the image. However, when vegetation is present it causes additional reflection (mainly by tree trunks) in the direction of the radar, or the so-called backscatter direction. This effect is particularly strong for the HH-polarisation. In practice, for forested peat domes, the combined effect of flooding and biomass is a variation in the image intensity for which the range of variation is mainly determined by the biomass level (i.e. low biomass areas show large variations in time; high biomass areas small variations) and for which the relative brightness is mainly determined by the intensity of flooding (i.e. dry terrain shows a relatively low intensity; flooded terrain a relatively high intensity). Examples for a variety of vegetation cover will be shown later.

Both PALSAR and ASAR are useful for detection of deforestation. Though the contrast between forest and recently deforested terrain is highest for the L-band with HV-polarisation, also L-band with HH-polarisation and C-band shows a certain level of sensitivity. The contrast also strongly depends on the elapsed time since deforestation. Depending on the vigour of regeneration the contrast fades away quickly in L-band (within approx. 4-6 months), and even faster in C-band (within approx. 2-3 months). The preference for C-band is related to the fact that L-band HV observations are only made once a year, L-band HH observations are less sensitive and ASAR C-band dual-polarisation data (APP mode) can be observed routinely every 35 days. Moreover, ASAR data can be made available very quickly, within two days of satellite overpass, which allows fast response to supposed illegal logging. Table 1 summarises the main characteristics of the radar systems discussed in this chapter.

	JERS-1	PALSAR Fine beam	PALSAR ScanSAR	ASAR Alternating polarisation
Centre frequency	1275 MHz	1270 MHz	1270 MHz	5331 MHz
Image mode (Polarisation)	(HH)	- FBS (HH) - FBB (HH/HV)	WB (HH)	APP (VV/HV)
Incidence range	angle 36°~42°	36.6°~40.9°	18.1°~43.0°	- IS2: 19.2°~26.7° - IS4: 31.0°~36.3°
Swath width	75 km	70 km	360 km	- IS2: 105 km - IS4: 88 km
Ground resolution	~18 m	~10 m ~20 m	~100 m	~30 m

Table 1. Brief overview of radar system characteristics

3. Field station and hydrological characterisation

To study peat swamp hydrology, ecology and radar wave interaction in a systematic way a dedicated research station has been established in the Mawas peat swamp forest conservation area, which is located some 80 km east of Palangkaraya, in the province Central Kalimantan. The main feature is a research bridge, 23 km in length, crossing an entire peat dome (Figure 1). Instruments placed along this bridge automatically measure rainfall and water level every hour. In December 2004, an airborne radar survey (the ESA INDREX-2 campaign) was carried out along this bridge to test a variety of advanced imaging radar techniques (Hajsek *et al.*, 2005; Hajsek and Hoekman, 2006). The intention is to collect data over an extended period (i.e. 10 years) to develop hydrological modelling, examine relationships between hydrological, soil and vegetation characteristics, study carbon sequestration and to relate biomass and water (flooding) levels to L-band radar observations of the ALOS PALSAR instrument.



Fig. 1. Field photograph of a section of the 23 km long research transect in the Mawas peat swamp conservation area. The transect crosses an entire ombrogenous peat dome. Along the transect ground water level dynamics are recorded.

Peat domes are formed in ombrogenous peat swamp areas, which are purely rain-fed and, consequently, nutrient poor. Vegetation types are located in concentric zones, with the 'poorer' forest types located towards the centre of the dome. Typically, the outer ring consists of relatively dense and high 'mixed' peat swamp forest, which gradually changes in a lower, more open, 'pole' peat swamp type. At the top the open 'padang' shrubland type may be found. To characterize the hydrology of such a dome, where water is flowing from the top in the centre towards the edges, the water level variation along the flow is monitored. An example result for one of the instruments along the bridge is shown in Figure 2 (Hoekman, 2007).

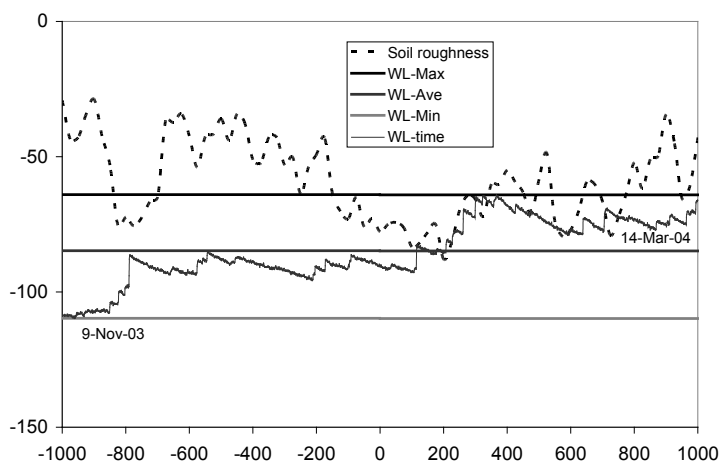


Fig. 2. Water table variation WL-time (solid curve) and peat soil surface roughness (dashed curve). The vertical axis shows water level and soil surface height (both in cm). The horizontal axis shows horizontal distance (in cm) along the soil surface roughness profile (i.e. from -1000 to 1000 cm) as well as time (i.e. from 9-Nov-03 to 14-Mar-04). The position of the water table measurement is at the centre of this profile. These measurements are made every hour. The results for the period 9 Nov2003 until 14 March 2004 are shown (also along the horizontal axis). The three horizontal lines show the maximum (WL-Max), average (WL-Ave) and minimum (WL-Min) water level. The percentage terrain flooding, thus, can be deduced from the combined roughness and water table measurements.

4. Observation of severe peat dome degradation events by JERS-1

Time series of L-band radar data can provide information on hydrology in peat swamps. For many peat swamp areas in Borneo and Sumatra large series of JERS-1 images (i.e. 15-30) collected in the period 1992-1998 exist. Figures 3 and 4 give examples of biophysical characteristics and events observed for the Mawas conservation area and the adjacent Kahiyu area. Figure 3 shows temporal dynamics in flooding, which reveals three large domes. The areas labelled as A are a complex of two flooded domes divided by a river originating from a central depression (B). The feature labelled as C is a relatively flat and wet fringe of a dry dome. Since tropical rainfall can be very localised and surface run-off is fast, the availability of large time series strongly supports proper interpretation.

Another combination of three images, all collected in the dry season, is shown in Figure 4. It shows deforestation caused by excess drainage as the three large blue areas intersected by canals labelled as A and B in the image. These areas appear blue because the radar echo strength in the third image (of this composite time series image) is much higher due to the combined effect of flooding and presence of sparse vegetation, while in the first two images (the red and green channels) the vegetation was still dense. Smaller blue areas labelled as C along rivers relate to fire scars and shifting cultivation in secondary forest.

The large blue area labelled as B in Figure 4 is one of the domes. In this area all trees died of drought and ground fires, which burned the root system causing the remaining trunks to fall down. The dome's destruction is shown in more detail in the time sequence of events in Figure 5. Until 1996 the dome was still hydrologically intact. In 1997 the construction of a very wide canal is visible. A low-altitude aerial photograph of this canal is shown in Figure 6. In Figure 5(c) (September) the canal is filled with water (the canal becomes black) and a small somewhat brighter area appears. This area grows very fast and becomes even brighter as shown in Figure 5(e) until the destruction is completed (January 1998). The physical interpretation of the radar brightness changes in the dome area can be associated with an initial period of excess drought under a dense canopy, i.e. the area becomes somewhat darker, like in Figure 5(b), followed by a period in which trees collapse and the brightness increases due to direct radar reflections from exposed trunks, like in Figures 5(d-e). It is interesting to note that the spatial extent of the destruction halted at the relatively flat and wet fringe well visible in Figure 3. The obvious cause of the destruction is the huge drainage caused by the wide canal. The coinciding strong El Niño Southern Oscillation (ENSO) period may have accelerated the process.

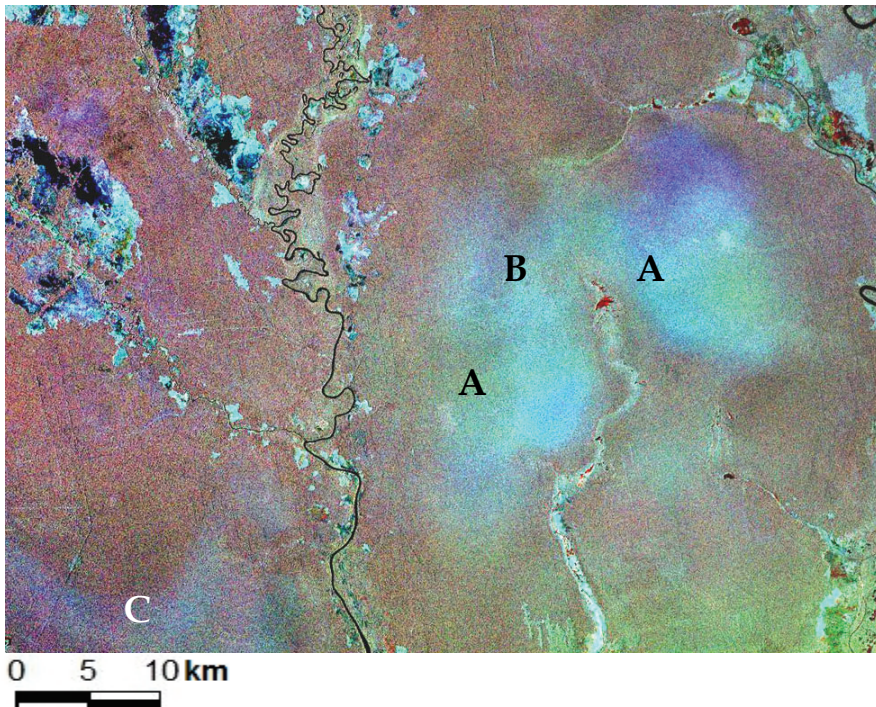


Fig. 3. Temporal dynamics in flooding intensity can be related to the hydrology of ombrogenous peat swamp forests and, indirectly, to peat depth. The blue areas labelled as A are flooded parts of the relatively flat tops of a complex of two peat domes, with a river originating from a central depression (B). The feature labelled as C shows the relatively flat and wet fringe of a dry peat dome. Mawas area, Central Kalimantan; JERS-1 SAR multi-temporal composite image (Red 7 Sep 1994; Green 12 Jul 1995; Blue 4 Jan 1996).

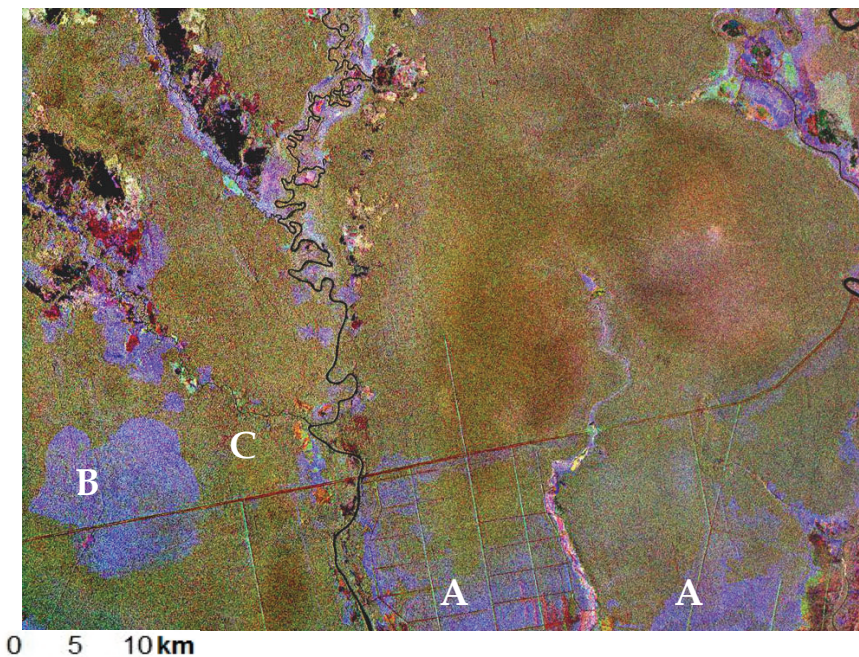


Fig. 4. Deforestation in Central Kalimantan caused by excess peat swamp forest drainage shows up as the three large blue areas intersected by canals labelled as A and B. Smaller blue areas along rivers labelled as C relate to fire scars and shifting cultivation in secondary forest. JERS-1 SAR multi-temporal composite image (Red 25 Jul 1994; Green 24 Jul 1997; Blue 16 Jul 1998).

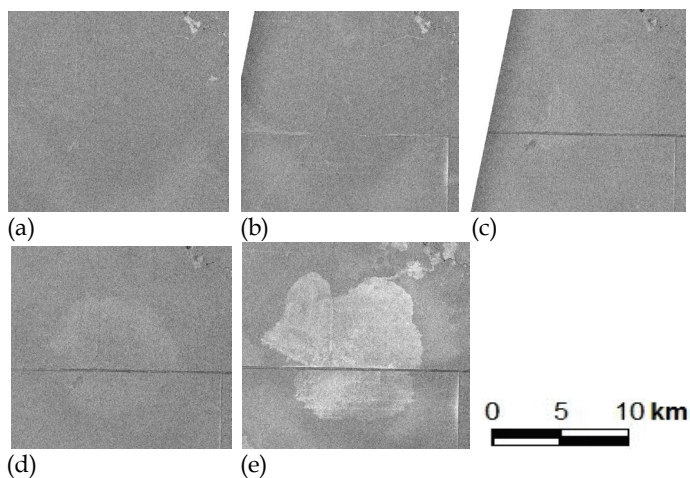


Fig. 5. JERS-1 SAR time series of the collapse of the peat dome in Kahiyyu: (a) 12 Jul 1995; (b) 19 Mar 1997; (c) 11 Sep 1997; (d) 25 Oct 1997; (e) 21 Jan 1998.



Fig. 6. Low altitude aerial photograph of the main East-West oriented double canal system passing South of Mawas. It shows the crossing of the Mentangai river, canal blocking activities and stretches of burnt forest areas along the canal, covered with small bushes and ferns. January 2005. Courtesy: Ruandha Agung Sugardiman, Indonesian Ministry of Forestry.

The possibility to observe peat swamp forest hydrology ceased at the end of the lifetime of the JERS-1 SAR instrument in 1998. Only since the year 2006, with the launch of ALOS, a new window of opportunity has been opened. During this eight year time span drastic changes occurred in many of the South-East Asian peatlands. This is illustrated in Figure 7 where new PALSAR observations are compared with the historical JERS-1 SAR data shown in Figs 3 and 4. The most striking features are the wet *padang* vegetation areas on top of two of the three domes (A, C). Compared to Fig. 3 the second dome (B) is now dryer, which may be an effect of local rainfall. The top of the Kahiya dome (C) is wet. It shows regeneration of *padang* peat swamp bush (bright centre) and denser vegetation (the red fringe around this bright centre, D). South of the main East-West canal the presence of a dry and low biomass peat area (blue area, E) is an indication of further degradation of the vegetation cover.

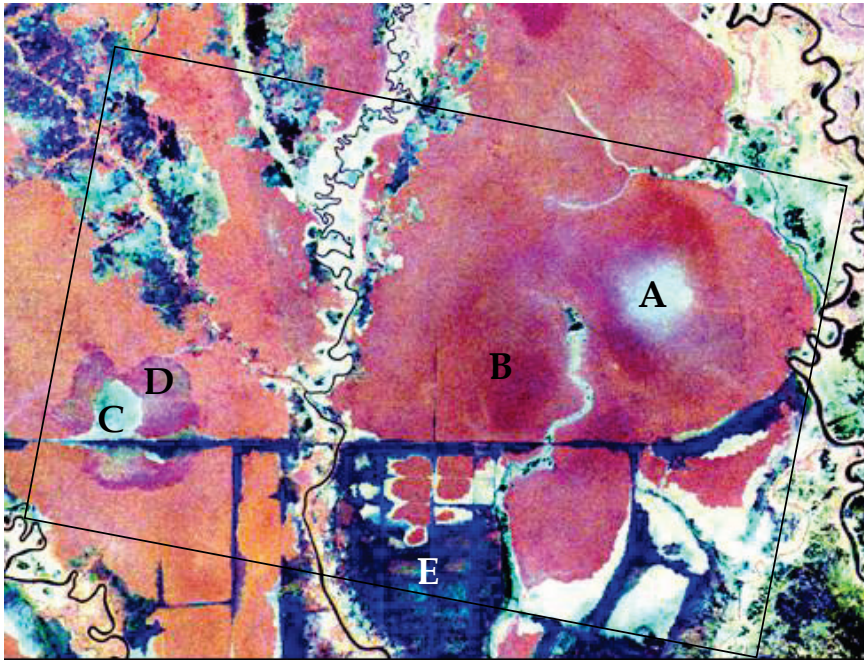


Fig. 7. Decadal change as observed by PALSAR data in 2007. The black frame outlines the area of Figs 4 and 5. The most striking features are the wet *padang* vegetation areas on top of the domes (A, C); a dry dome top (B); indications of re-generation on the top of the Kahiyu dome (D); and the dry low biomass peat area (blue area) indicating further degradation of the vegetation cover (E). PALSAR multi-temporal composite (Red FBD HV; Green FBD HH; Blue WB HH; FBD 7 and 24 Aug 2007 (2 images); WB 29 Mar 2007). Courtesy: ALOS K&C © JAXA/METI.

5. Monitoring of fire damage and deforestation by ENVISAT ASAR

Due to smoke and persistent cloud cover optical satellite sensors fail to detect forest cover area change in a timely manner. To monitor deforestation over large areas in a feasible way, a system using both traditional satellite imagery (i.e. Landsat ETM+) and ASAR APP radar imagery from the European Space Agency's ENVISAT satellite has been proposed, developed and implemented. This was done for a 60,000 km² area of peatland in Central-Kalimantan to support peatland restoration and protection activities carried out in the framework of the Central Kalimantan Peat Programme (CKPP, 2009).

For this area more than 90 ENVISAT ASAR APP radar images were collected between 2005 and 2007 and systematically analysed using semi-automated computer techniques to detect change. The approach works best using two polarisations (HH and HV or VV and HV) and incorporates analysis of changes in both the strength and polarisation of the radar return signal both within a small timeframe (every 35 days, which is the revisiting cycle of the satellite) as well as in a large timeframe (1 year). This is necessary to improve accuracy of the

changes and reduce false alarms. For this system a relatively small incidence angle was used (IS2; see Table 1; see footnote¹) to provide continuity with the predecessor of ENVISAT ASAR, viz. the ESA ERS-2 SAR. Whenever available, Landsat ETM+ was integrated. Output of this change detection process is a consistent series of change maps showing forest, forest cover change (deforestation, fire damage, road building etc.), other land and water. These results have been used to support law enforcement and projects for the generation of voluntary carbon credits. Some results are shown in Figs 8-10.

Figure 8 is a low altitude aerial photograph of sub-surface peat forest fires along a canal in the Sebangau National Park. This photograph clearly illustrates the need for radar. Optical observation fails to detect the ground fires because the forest seems intact, and observation is obscured by smoke, haze, and or clouds. Thermal infrared (hot spot) observation, such as from the MODIS, AVHRR or AATSR instruments, fail because the fire is underground and under the forest. L-band radar (HH-polarisation) works because it detects the excess drought in the soil. ENVISAT APP radar detects damage very fast because it registers falling trees directly (with an update frequency of 35 days). Figure 9a shows the cumulative damage for the year 2006 as recorded by ASAR, which was a dry year because of a moderate El Niño. Figure 9b shows fire hot spots which are detectable as soon as the ground fires have developed in open fires. The correspondence is large. Figure 9c shows the first available cloud free Landsat ETM+ scene of the same area after the fire period. It shows the burned forest area (in cyan) at exactly the same locations where ENVISAT already mapped deforestation 10 months (!) earlier.



Fig. 8. Low altitude aerial photograph of sub-surface peat forest fires along a canal in the Sebangau National Park. 6 September 2004. (Photo: Dirk Hoekman).

¹ Though IS2 images were used here, and demonstrated to be suitable for deforestation monitoring, it is noted that IS4 images are even better suited because of a higher incidence angle, which increases the contrast between forest and non-forest. Even higher incidence angles are possible (from IS5-7 modes) but these provide no full coverage at the equator.

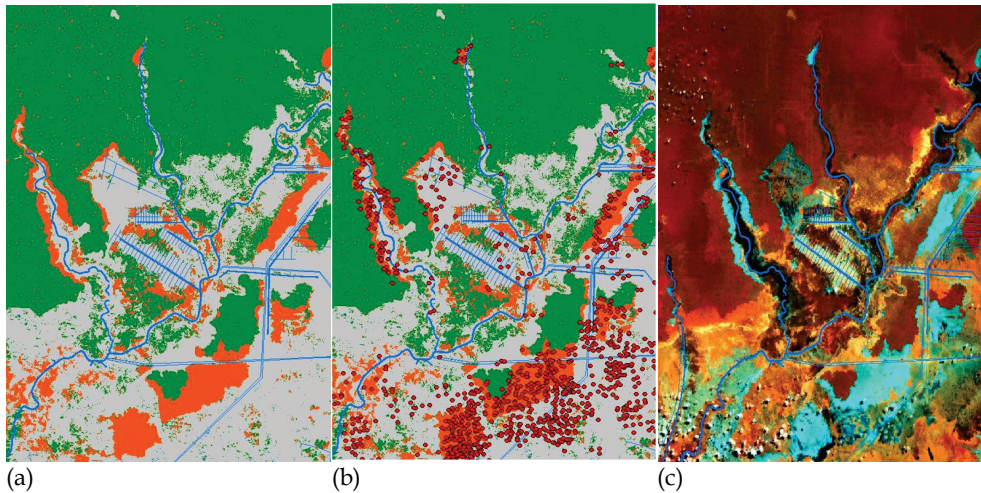


Fig. 9. South-East section of Sebangau National Park, East-Kalimantan. (a) Cumulative deforestation recorded in the year 2006 by ENVISAT ASAR (Green: forest; Orange: forest loss); (b) Idem, with MODIS hot spot fire detections (small red circles) of the 2006 dry season superimposed; (c) Landsat ETM+ image of 4 July 2007 (RGB: bands 4-5-7). Central Kalimantan Peatlands Programme. ASAR APP data courtesy ESA. Image processing and analysis by SarVision & Wageningen University, 2007.

This unique capability on ASAR APP to follow deforestation patterns nearly real time (i.e. within ± 5 weeks) is nicely illustrated in Figure 9. It shows the development of a (probable illegal) road from an already deforested area in the direction of a small rock outcrop in the Sebangau National Park. Already in December 2005 the first section is visible and construction work is proceeding until September 2006. Good trafficability on a road in peat swamp forest requires the construction of canals for drainage on both sides of the road. However, these canals drain a large area of the surrounding peat soil and make it vulnerable to fire. The October-December map sequence shows the damaging effects of forest fire.

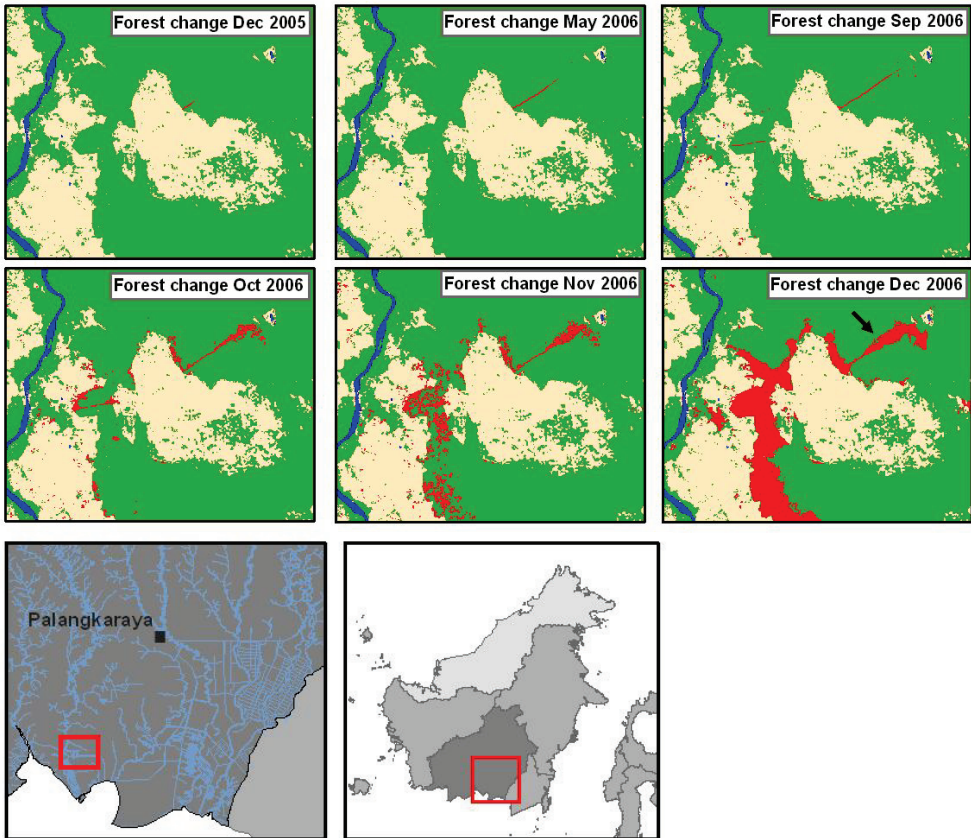


Fig. 10. ASAR Alternating Polarisation radar deforestation time series example showing forest (green), water (blue), non-forest areas (yellow) and recent deforestation (red). The top 3 images show the development of a new road in the forest (period December 2005 until September 2006). The lower 3 images show major deforestation because of forest fires in the dry season along this new road and along the forest boundaries (period October-December 2006). This example covers an area of ~ 30 km by 20 km and is part of a larger area of ~ 300 km \times 200 km where this new radar monitoring technique has been applied pre-operationally. Central Kalimantan Peatlands Programme. ASAR APP data courtesy ESA. Image processing and analysis by SarVision & Wageningen University, 2007.

6. Peat swamp restoration impact assessment using L-band backscatter change

Historical JERS-1 L-band radar data provide insight in the pre-disturbance or early disturbance state of the hydrological functioning of peat domes and may be used as a baseline for restoration planning. In Mawas, in the framework of CKPP, canal blocking was performed. The effect of such activities may be assessed and monitored by PALSAR images,

and auxiliary data. An example is given in Figures 11 and 12. In the JERS-1 image of January 1998 (dry period) shown in Figure 11 the area demarcated by the red line is an area within the Mawas area suffering from excess drought. In the PALSAR image of 9 November 2006 (dry period) this area has decreased above the main East-West canal because of the construction of dams in the canal going North (canal Neraka). In the area south of the main East-West canal a large network of canals is still present and the continued drainage has worsened the situation. Note the very low radar backscatter (intense black) caused by very dry bare peat areas and the bright white area, which is a strongly degraded open forest with fire damage. The areas demarcated in blue are hydrologically intact, allowing forests previously damaged to regenerate. The fire damage is visible in the PALSAR area as a very bright area (B) associated with sparse open vegetation with many dead standing trunks. This area is also visible in the ENVISAT deforestation map created directly after the fire damage (Fig.12a) and Landsat ETM+ 10 months later (Fig.12b).

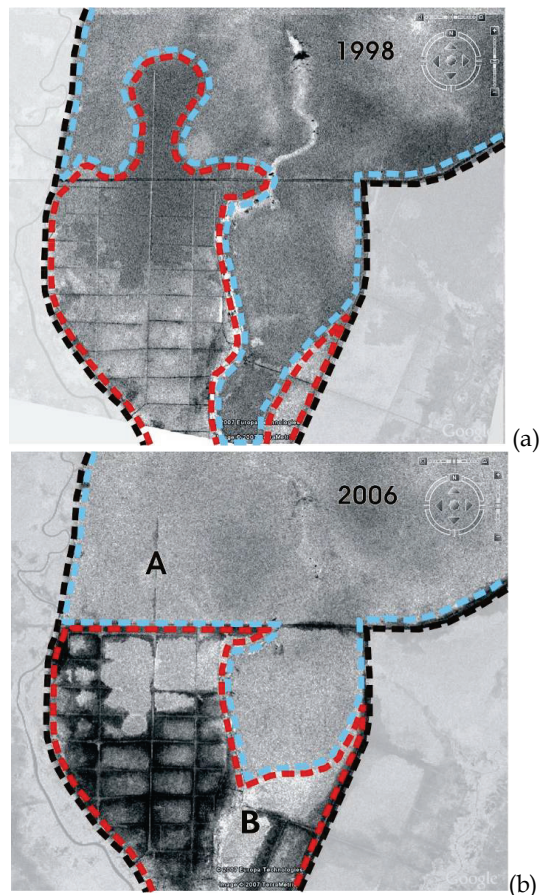


Fig. 11. Peat swamp degradation (B) and restoration (A) in the Mawas area between 1998 (JERS-1) (a) and 2006 (PALSAR) (b). The red area is degraded; the blue area is intact or regenerating. Courtesy: ALOS K&C © JAXA/METI.

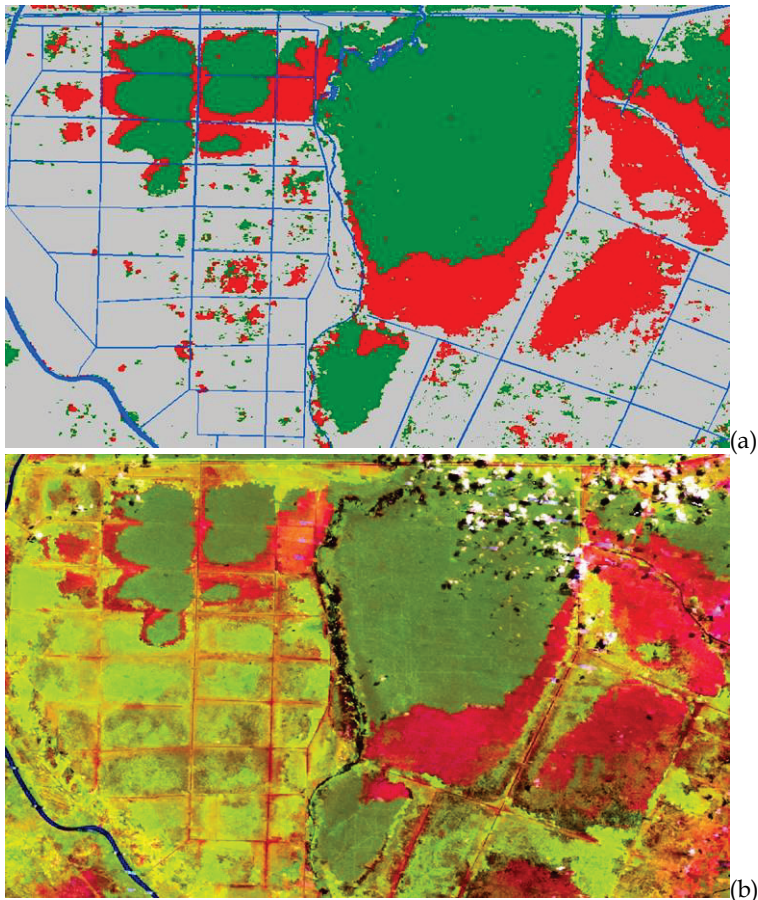


Fig. 12. Forest loss south of Mawas during the 2006 moderate El Niño period. (a) ENVISAT ASAR deforestation map (Forest: green; Burnt forest: red); (b) Landsat ETM+ image of 4 July 2007 (RGB: bands 5-4-3). The correspondence between both images is striking. The burnt forest areas mapped by ASAR directly after the fires (September 2006) are also observed in the first available Landsat image acquired 10 months later.

7. Flood frequency map Central Kalimantan

To support peatland restoration efforts knowledge on hydrological dynamics are imperative. The PALSAR ScanSAR mode provides a unique capability to assess these dynamics. As explained before (Section 2) the effect of flooding on the radar image intensity depends on the amount of vegetation and the height of vegetation. This is exemplified in Figure 13 where the temporal signature of HH-polarisation backscatter is plotted for the nine observations made in the period November 2006 until December 2007, as listed in Table 2. Terrain with moderate but high vegetation cover shows a strong increase in backscatter because of flooding. Terrain with low vegetation shows a decrease in backscatter because of

flooding. Mangrove, mixed peat swamp forest and pole peat swamp forest show a moderate increase during the wet season. Therefore, it is necessary to make a classification of the area first (this can be done with PALSAR) before thresholding the backscatter intensities (per class) to determine the incidences of flooding. Figure 14 shows a mapping result of the flood frequency or flood duration.

11-Nov-2006	29-Mar-2007	29-Sep-2007
27-Dec-2006	14-May-2007	14-Nov-2007
11-Feb-2007	14-Aug-2007	30-Dec-2007

Table 2. ALOS PALSAR ScanSAR HH (WB1) input data of nine consecutive (46 day) cycles used for the production of the Central-Kalimantan flood frequency map.

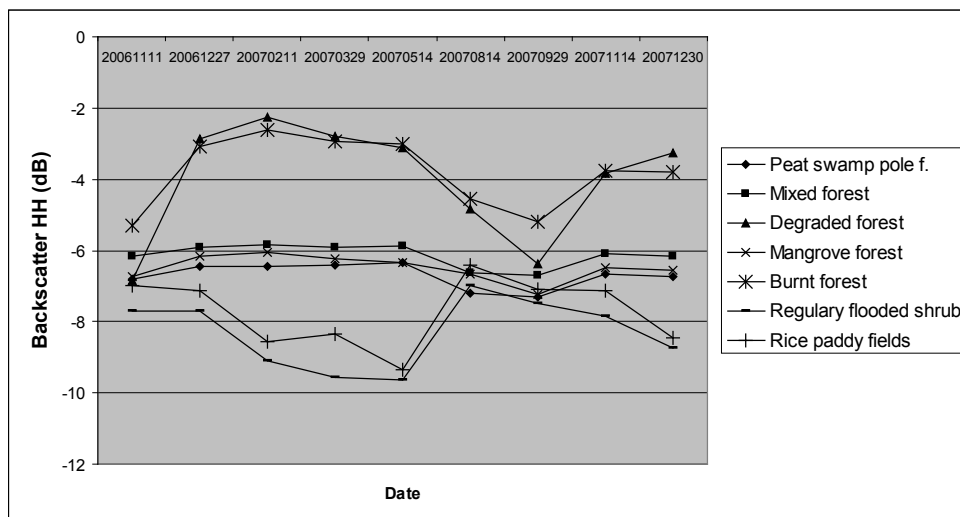


Fig. 13. Temporal signatures of L-band HH-polarisation backscatter for several key vegetation types. The first observation is made at the end of the dry season, at 11 November 2006. During the next wet season terrain with moderate but high vegetation cover shows a strong increase in backscatter because of flooding. Terrain with low vegetation shows a decrease in backscatter because of flooding. Mangrove, mixed peat swamp forest and pole peat swamp forest show a moderate increase during the wet season. Also the return to dry conditions during the 2007 dry season is well visible. PALSAR ScanSAR, period November 2006 until December 2007.

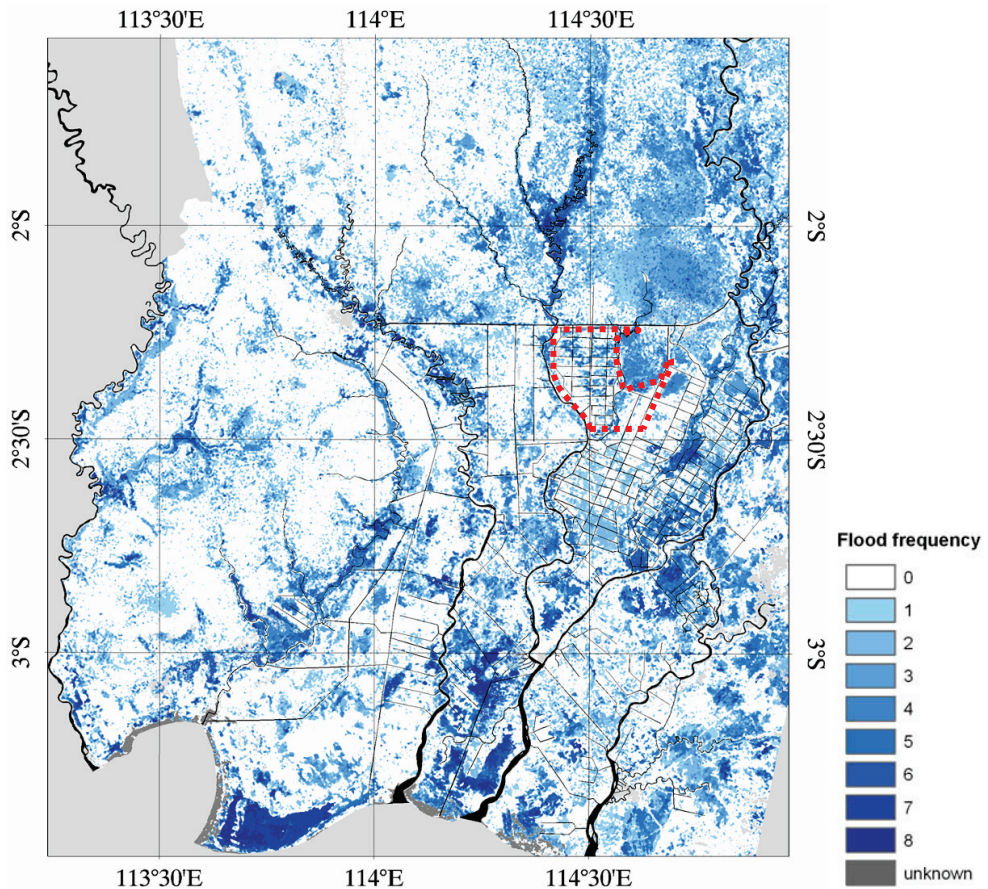


Fig. 14. Map of flooding frequency in 2007 for the Ex-Mega Rice Project (EMRP) area and Sebangau National park in Central Kalimantan based on nine PALSAR WB1 HH images. For ease of reference the degraded area indicated in Figure 11 is demarcated with red dots. PALSAR data courtesy: ALOS K&C © JAXA/METI.

8. Discussion

Many of the tropical peat swamp forests in Borneo and Sumatra are seriously threatened by (illegal and legal) logging and conversion to plantations for the oil palm and pulp and paper industries. In all cases the hydrology is affected by excess drainage, leading to destruction of remaining forests, notably in dry years. Beyond a certain point the hydrological integrity of ombrogenous areas is lost, leading to an irreversible process of total destruction and the combustion and oxidation of the remaining thick peat layers. Unless rigorous measures are taken very soon, this most likely will lead to major negative effects on biodiversity and global climate.

More information is needed to support protection and restoration efforts. The availability of better vegetation and peat depth maps may be very useful. However, the most crucial factors may appear to be the knowledge on the hydrological functioning and the relationships between hydrological and ecological characteristics. These latter points are still poorly understood. Radar, unimpeded by cloud cover, can provide continuous observations which can be related to hydrological characteristics, may become a key instrument in future protection and restoration efforts. Exploitation of PALSAR time series collected by the ALOS satellite may provide important support for peat land management, protection and restoration, such as described in the Ramsar "Guidelines for Global Action on Peatlands (GGAP, 2002)". Moreover, it may significantly support other international treaties, such as the CBD and the Kyoto Protocol, a possible post-Kyoto protocol, and carbon cycle science.

The methodology may eventually be applied on a large scale using systematic observations of PALSAR and ENVISAT, and its successors PALSAR-2 and SENTINEL-1. The latter two instruments may be available from 2013 onwards, providing continuity of L- and C-band radar observation. PALSAR-2 ScanSAR observations will be even powerful because it uses dual polarisation, providing HV-polarisation in addition, which is important to improve assessment of biomass level dynamics and deforestation. SENTINEL-1 is a major improvement over ENVISAT ASAR because it allows a 4 times higher temporal observation, i.e. (illegal) deforestation may be reported every 8 days, instead of the current 35 days.

PALSAR radar proved particularly useful for improving information related to flooded cover types and biomass levels. ASAR deforestation maps provide at least as much accuracy and detail as the best available maps based on visual interpretation of Landsat imagery, however, provide this information near real time. Many of the results shown in this chapter are operationally used by local governmental and non-governmental agencies for spatial planning of sustainable peatland management strategies.

9. Acknowledgments

This work has been undertaken in part within the framework of the JAXA Kyoto & Carbon Initiative. JAXA is acknowledged for providing JERS-1 SAR and ALOS PALSAR data. SarVision is acknowledged for providing image processing and analysis support. BOS is acknowledged for providing research facilities and support in the Mawas area.

10. References

- CKPP, 2009, Final report; <http://www.ckpp.org/Portals/16/CKPP%20products/CKPP%20Provisional%20Report%20ENG%20final.pdf>
- Davidson, N.C., and, C.M. Finlayson , 2007, Earth observation for wetland inventory, assessment and monitoring, Aquatic Conservation: Marine and Freshwater Ecosystems. Special edition title "Satellite-based radar - developing tools for wetlands management", Vol.17, pp.219-228.
- Fargioni, J., J. Hill, D. Tilman, S. Polasky, and P. Hawthorne, 2008, Land clearing and the biofuel carbon debt, Science, Vol.29, pp.1235-1238.
- GGAP, 2002; <http://www.imcg.net/docum/ggap.htm>

- Goldammer, J. G. (1999). Forest on Fire. *Science* 284: 1782-1783.
- Hajnsek, I., F.Kugler, K. Papathanassiou, R. Scheiber, R.Horn, A. Moreira, D.H. Hoekman, M. Davidson, and E.P.W. Attema, 2005, INDREX 2 - Indonesian airborne radar experiment campaign over tropical forest in L- and P-band, POLinSAR 2-nd Int. Workshop on Applications of Polarimetry and Polarimetric Interferometry, ESA-ESRIN, Frascati, Italy, 17-21 January 2005, ESA report SP-586 on CD-ROM.
- Hajnsek, I, and D.H. Hoekman, 2006, Final Report, INDREX II - Indonesian Radar Experiment Campaign over Tropical Forest in L- and P-band, Version 1, 14 June 2006; ESA Contract RFQ/3-11077/04/NL/CB; Report ESA, DLR and Wageningen University; 142 pages
- Hoekman, D.H., 2007, Satellite radar observation of tropical peat swamp forest as a tool for hydrological modelling and environmental protection. *Aquatic Conservation: Marine and Freshwater Ecosystems*. Special edition title "Satellite-based radar - developing tools for wetlands management", Vol.17, pp.265-275.
- Hoekman, D.H., and M.A.M. Vissers, 2007, ALOS PALSAR radar observation of tropical peat swamp forest as a monitoring tool for environmental protection and restoration, *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, 23-27 July 2007, Barcelona, Spain (CD-ROM).
- Hooijer, A., Silvius, M., Wösten, H. and S. Page, 2007: PEAT-CO₂. Assessment of CO₂ emissions from drained peatlands in SE Asia. Delft Hydraulics report Q3944.
- IUCN/WWF (2000). *Global Review of Forest Fires*. The World Conservation Union (IUCN) and Worldwide Fund for Nature (WWF), Gland Switzerland. 64 pp.
- MacDicken, K.G., 2002. Cash for tropical peat: land use change and forestry projects for climate change mitigation. In: Rieley, J.O., and page, S.E. (eds.) with B. Setiadi. *Peatlands for people: natural resource functions and sustainable management*. *Proceedings of the International Symposium on Tropical Peatland*, 22-23 August 2001, Jakarta, Indonesia. BPPT and Indonesian Peat association. 272 pp.
- Page, S.E., F. Siegert, J.O. Rieley, H.V. Boehm, A. Jaya and S. Limin, 2002. The amount of carbon released from peat and forest fires in Indonesia during 1997. *Nature* 20(Nov. 7):61-65.
- Rieley, J.O., and B. Setiadi, 1997. Role of tropical peatlands in the global carbon balance: preliminary findings from the high peats of Central Kalimantan, Indonesia. *Alami* 2 (1): 52-56.
- Rosenqvist, A., Forsberg, B.R., Pimentel, T.P., Rausch, Y.A. and Richey, J.E., 2002, The use of spaceborne radar data to model inundation patterns and trace gas emissions in the Central Amazon floodplain, *Int. Journal of Remote Sensing*, Vol.23, No.7, pp.1303-1328.
- Rosenqvist A., M. Shimada, N. Ito and M. Watanabe, 2007a, ALOS PALSAR: A pathfinder mission for global-scale monitoring of the environment, *IEEE Transactions on Geoscience and Remote Sensing*, Vol.45, No.11, pp.3307-3316.
- Rosenqvist, A. , A., C. M. Finlayson, J. Lowry, D. Taylor, 2007b, The potential of long-wavelength satellite-borne radar to support implementation of the Ramsar wetlands convention, *Aquatic Conservation: Marine and Freshwater Ecosystems*. Special edition title "Satellite-based radar - developing tools for wetlands management", Vol.17, pp.229-244.

- Rosenqvist, A., M. Shimada, R. Lucas, J. Lowry, P. Paillou, B. Chapman [eds.], 2008, The ALOS Kyoto & Carbon Initiative, Science Plan (v.3.1), JAXA EORC, March, 2008. [Online]: http://www.eorc.jaxa.jp/ALOS/kyoto/KC-Science-Plan_v3.1.pdf
- Sabine, C. L., M. Heimann, P. Artaxo, D.C.E. Bakker, C.T.A. Chen, C.B. Field, N. Gruber, C. Le Quéré, R. Prinn, J.E. Richey, P.R. Lankao, J.A. Sathaye and R. Valentini, 2004, Current status and past trends of the global carbon cycle. In: C.B. Field, M.R. Raupach (Eds.), *The global carbon cycle*, Island Press, Washington, pp.17-44.
- Shimada, M., and O. Isoguchi, 2002, JERS-1 SAR mosaics of Southeast Asia using calibrated path images, *Int. Journal of Remote Sensing*, Vol.23, No.7, pp. 1507-1526.
- Simard, M., G. De Grandi, S. Saatchi, and P. Mayaux, 2002, Mapping tropical coastal vegetation using JERS-1 and ERS-1 radar data with a decision tree classifier, *Int. Journal of Remote Sensing*, Vol.23, No.7, pp.1461-1474.
- Van der Werf, G., R.J. Dempewolf, S.N. Trigg, J.T. Randerson, P.S. Kasibhatla, L. Giglio, D. Murdiyarso, W. Peters, D.C. Morton, G.J. Collatz, A.J. Dolmana, and R.S. DeFries, 2008, Climate regulation of fire emissions and deforestation in equatorial Asia, *PNAS*, December 23, 2008, Vol.105, No.51, pp.20350-20355.
- Waldes, J.L., and S.E. Page, 2002. Forest structure and tree diversity of a peat swamp forest in Central Kalimantan, Indonesia. In: Rieley, J.O., and page, S.E. (eds.) with B. Setiadi. *Peatlands for people: natural resource functions and sustainable management. Proceedings of the International Symposium on Tropical Peatland, 22-23 August 2001, Jakarta, Indonesia*. BPPT and Indonesian Peat association. 272 pp.

Multivariate Differencing Techniques for Land Cover Change Detection: the Normalized Difference Reflectance Approach

Paolo Villa*, Giovanmaria Lechi** and Mario A. Gomasasca*

* *Institute for Electromagnetic Sensing of the Environment (CNR-IREA),
National Research Council, Milan, Italy*

** *Bulding Environment Sciences and Technology (BEST) Department,
Polytechnic of Milan, Italy*

1. Introduction

The importance of the dynamic side of natural and man-made phenomena has become an urgent need when trying to mitigate the human impact on environment. Remote Sensing is one of the most effective way to quantify and map the changes of environmental conditions on our planet: the tools used for this purpose are called Change Detection Techniques. Techniques among which an important role is played by those methodologies based on multi-spectral remote sensing data and exploiting multivariate analysis derived methodologies, also demonstrating their capabilities through some test cases, covering flood events and urban growth studies.

Multi-temporal and multi-spectral techniques for Change Detection exist in a wide variety of approaches, often far too sector oriented and not straightforward. Compression and decorrelation techniques, on the other side, tend not to exploit the whole spectral content of remotely sensed data. The Normalized Difference Reflectance (NDR) here introduced is a general approach for bi-temporal land cover change mapping and detection that exploits the whole spectral capabilities of panchromatic, multi-spectral or hyper-spectral images. NDR is a general and simple measure that can be used in the frame of what are called Normalized Difference Change Detection Techniques (NDCD), which starts using as a input the NDR derived results. This Chapter includes a large test case which is a good benchmark for NDR approach, using Minimum Noise Fraction implementation of NDCD for mapping Hurricane Katrina aftermaths over the city of New Orleans, U.S., thus fusing together urban and flood change applications.

The purpose of the chapter is to give an overview of multivariate difference-based techniques for land cover change mapping using multispectral remote sensing data, and to introduce and demonstrate the Normalized Difference Reflectance approach in the frame of Normalized Difference Change Detection techniques. Two examples of NDCD results are given as a complement to theoretical aspects of the methodology, and an application study has been used as benchmark for the technique performances evaluation, in comparison with other established Change Detection techniques.

2. Multivariate Differences in Change Detection

Multi-temporal differencing is an established change detection technique for environmental mapping and monitoring with remotely sensed data (Singh, 1989; Lu et al., 2004; Coppin et al., 2004). Following a difference normalization approach, introduced in remote sensing for vegetation studies with the normalized difference vegetation index (NDVI), a multi-temporal implementation of this standardization technique for forest change analysis was first proposed for univariate vegetation indexes (VIs) (Coppin & Bauer, 1994), and then in comparison with other change detection methodologies (Coppin et al., 2001), always for forest mapping purposes.

During this work it has been introduced a quantitative method to evaluate land cover change through multi-spectral variation in radiometric response of surface features. In order to detect interesting changes, a pair of satellite scenes, geometrically registered and atmospherically corrected, is to be radiometrically normalized. After that, a map of spectral variations is produced using a multi-spectral difference index named Normalized Difference Reflectance (NDR). The NDR is therefore an approach to Change Feature Identification phase in Change Detection.

The phase of Change Mapping is then performed using NDR measures as inputs for Change Detection methodologies and techniques, in the frame of what are called Normalized Difference Change Detection (NDCD) techniques. The NDCD is a technique which, given an image pair, performs calculations on radiometric normalized reflectance data through the definition of the normalized difference reflectance (NDR) and produces a standardized difference of the reflectance values.

The use of NDR and NDCD will be presented through case studies showing change analysis covering flood events and urban environment: one case is over a flood event occurred in Bangladesh and exploiting Landsat-7/ETM+ scenes, another case regards the urban expansion scenario of Washington outskirts, U.S., and exploiting Terra ASTER data, the last and most complete case study is the analysis of the damages to the urban area of the city of New Orleans (Louisiana, USA) resulting from the passage of hurricane Katrina, using both SPOT-4/HRVIR and Landsat-5/TM data.

3. Normalized Difference Reflectance (NDR)

The Normalized Difference Reflectance (NDR) here introduced is a general approach for bi-temporal land cover change mapping and detection that exploits the whole spectral capabilities of panchromatic, multi-spectral or hyper-spectral images. Given an image pair, the NDR produces a standardized difference by analyzing the changes in the reflectance properties of each spectral band, so without losing any spectral richness as when applying indexes, feature reduction or compression techniques (Villa & Lechi, 2007).

The image reflectance differences were modified to a normalized version on the sum of spectral values, in order to minimize the confusion among difference values which are numerically equal, but come from different land cover change events.

Hence, for every spectral band the NDR is defined as follows:

$$NDR_j = \frac{R_j^{norm}(post) - R_j^{norm}(pre)}{R_j^{norm}(post) + R_j^{norm}(pre)} \quad (1)$$

where:

NDR = normalized difference reflectance

$R_j^{norm}(post)$ = normalized reflectance for the post flood scene

$R_j^{norm}(pre)$ = normalized reflectance for the pre flood scene

j = spectral band number

The NDR is a multi-spectral quantity which spans over the range of values from -1 to +1 and shows the amount of change in surface reflectance for every band in the original data, in terms of the relative difference in spectral signature of ground objects (-1.00 = maximum reflectance decrease, 0.00 = no change, +1.00= maximum reflectance increase). This approach is a quantitative base for building the successive phase of change detection, through the use of multi-spectral normalized reflectance values. A first and simple visual inspection of NDR band compositions permits a prompt and clear preliminary assessment of changes, thus supporting the choice of an apt change detection algorithm or technique for the phase of change mapping. Figure 1 shows an example of NDR calculated for an urban scenario, located in Maryland, U.S, using multi-temporal Terra ASTER data.

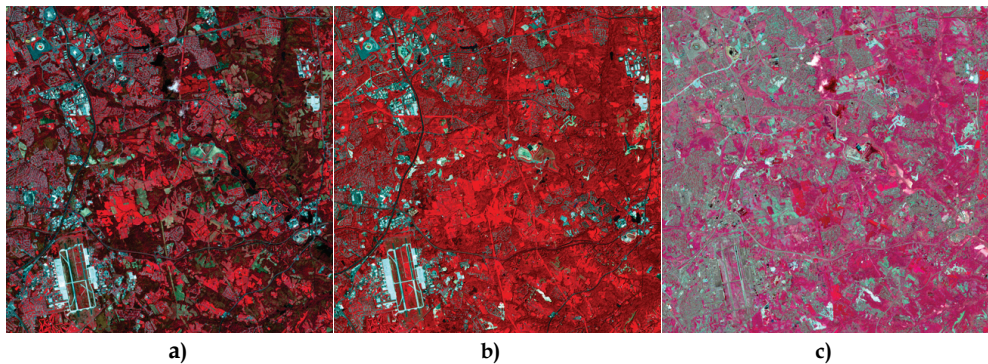


Fig. 1. NDR calculated for an urban scenario, located in Maryland, U.S.: CIR visualization (RGB=3N,2,1), ASTER scene of April 9th, 2000 (a); CIR visualization (RGB=3N,2,1), ASTER scene of August 24th, 2003 (b); CIR visualization (RGB=3N,2,1), NDR values derived (c). Different colours in (c) are inked to different kind of variations in surface reflectance between (a) and (b) images: grey areas represent not changed features, cyan areas represent a decreasing response in near infrared (linked to newly exposed areas, construction sites and new impervious surfaces), red areas represent increasing near infrared response (linked to phenological conditions of vegetation, going to an April scene to an August one), white areas represent increased response in all the visualized bands.

This approach allow to promptly visualize in RGB channels different triplets of bands at a time, thus bringing the user to a straightforward inspection of multi-spectral change features of surface objects; beginning with this visualization of multi-date information every end-user has the possibility to decide which may be the best Change Detection Technique to retrieve a land cover change map. NDR not only permits an easy and straightforward multi-

spectral comparison and evaluation of land cover changes, but permits enhanced individualization of radiometric response change, in comparison with simple Reflectance Differencing (RD). In fact, the same amount of reflectance difference between two surface features can be due to different land cover changes depending on the reference amount of spectral response. The NDR approach takes into account this issue and outputs different values of NDR for the same RD situation, when corresponding to different changes, as illustrated in the example of Figure 2 and Table 1.

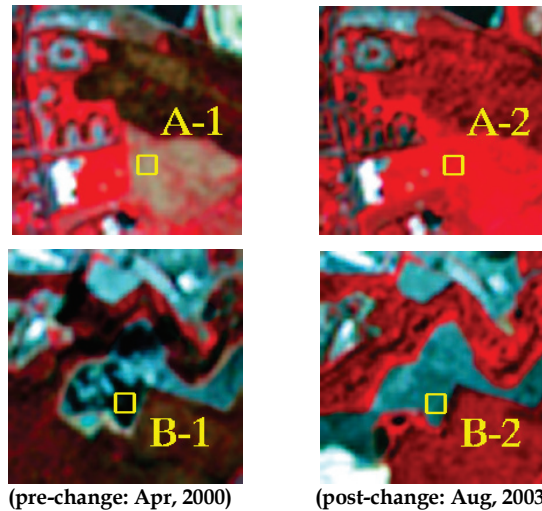


Fig. 2. Particular of two changed areas in an urban environment, located in Maryland, U.S.: CIR visualization (RGB=3N,2,1), ASTER scene of April 9th, 2000 (pre-change: A-1; B-1); CIR visualization (RGB=3N,2,1), ASTER scene of August 24th, 2003 (post-change: A-2; B-2): in the A area (above) a change in land cover from Bare Soil to Vegetation is highlighted in the yellow square, whereas in the B area a change in land cover from dark construction asphalt to paving concrete is highlighted in the yellow square.

Area	Land Cover		NIR Reflectance Response [760-900 nm]	Reflectance Difference (RD)	Normalized Difference Reflectance (NDR)	
Fig. 2	Pre (A-1; B-1)	Post (A-2; B-2)	pre	post		
A	Bare Soil	Grass Vegetation	0.311	0.412	0.101	0.140
B	Construction Asphalt	Construction Concrete	0.103	0.207	0.104	0.335

Table 1. Normalized Reflectance Difference (NDR) results compared with common Reflectance Difference (RD) results, calculated for particular spots in Figure 2, to show the enhanced discrimination capabilities of NDR.

4. Normalized Difference Change Detyection (NDCD)

The further step to exploit the NDR approach defined and described in the previous section is its implementation in the frame of the so called Normalized Difference Change Detection (NDCD) techniques.

The NDCD technique uses the NDR defined in Equation (4.1) as input variables for deriving a land cover/land use change map through the use of one particular change detection method, thus leading to a specific implementation of the NDCD. Out of a range of techniques, such as multi-spectral transforms (e.g. Principal Components Analysis and Minimum Noise Fraction), image classification techniques (both supervised and unsupervised), image segmentation algorithms, Neural Networks or Support Vector Machines, one could be used (Lu et al., 2004).

The possible applications and purposes of this approach are manifold and diverse. In the following we will show the effectiveness of the NDCD for flood mapping. Nevertheless, this approach is a general one and might be applied not only for such mapping purposes, but also for urban growth, burnt areas mapping and other land cover change analyses.

During this work we particularly focused our research on the Minimum Noise Fraction (MNF) (Green et al., 1988; Gianinetto & Villa, 2007) implementation of the normalized difference change detection technique (NDCD-MNF), where the MNF transform is applied to the NDR data to obtain the final change detection map, for the case study analysis of the flood event due to Hurricane Katrina aftermath over the city of New Orleans, in Louisiana, U.S.. The case study and its results will be presented in the next section.

In order to give a demonstration of how the NDR and NDCD approaches work, in the following paragraphs a couple of implemented examples are show, covering a flood hazard mapping case for the monsoonal flood occurred in autumn 2000 and an urban sprawl assessment case for the suburban areas of Washington, U.S..

4.1 Flood Hazard, an Example

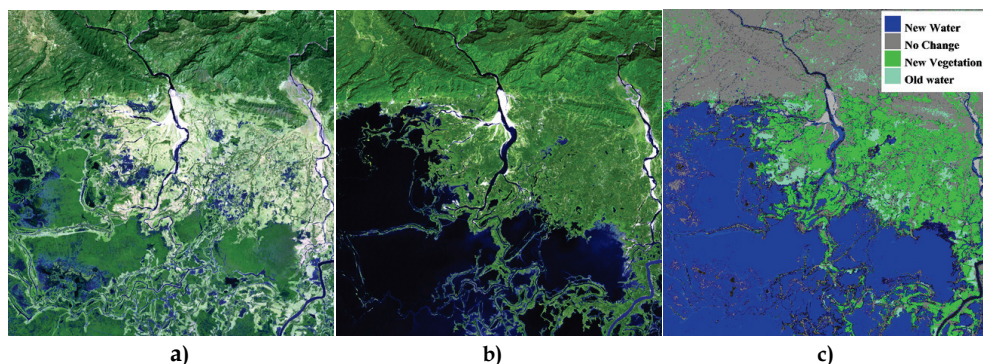


Fig. 3. Change Detection for an monsoon flood event, which took place in the Haor region, North-East of Bangladesh: band composition visualization (RGB=7,5,3), ETM+ scene of February 28th, 2000 (a); band composition visualization (RGB=7,5,3), ETM+ scene of October 25th, 2000 (b); Change map derived with Max. Likelihood classification of NDR values.

The first example deal with a change detection application for post-flood analysis, a topic already taken into consideration by previous works of the authors (Gianinetto & Villa, 2006). The inundation event is a monsoon flooding which drawn the North of Bangladesh and North-eastern part of India in autumn 2000. A pair of Landsat ETM+ scenes covering the Haor region in north-eastern Bangladesh (the pre-event image of February 28th was normalized using post-event image of October 25th as reference) was processed and radiometrically normalized with Pseudo Invariant features (PIFs) selection and linear regression, to produce NDR values as using equation 1.

In order to map surface features changes the couple of images was inspected and regions of change were chosen as ground truth for producing a Maximum Likelihood classification and therefore a map of changed areas, shown in Figure 3. This way, not only the area covered by flooding water could be identified, but also the different vegetation phenological features due to seasonal variations was mapped (Rogan et al., 2002).

4.2 Urban Area, an Example

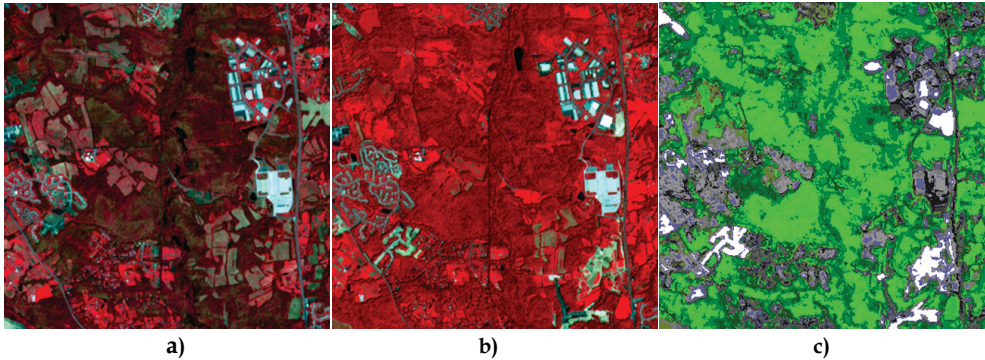


Fig. 4. Change Detection for an urban scenario, located in Maryland, U.S., particular of an area of residential and commercial growth in 2000-2003 period: CIR visualization (RGB=3N,2,1), ASTER scene of April 9th, 2000 (a); CIR visualization (RGB=3N,2,1), ASTER scene of August 24th, 2003 (b); Change map derived with ISODATA classification of NDR values. Gray tones represent not changed features, green hues represent increasing vegetation vigour, bright areas represent changed surface cover: mainly newly exposed areas, construction sites and new impervious surfaces.

Another example focuses on a change detection analysis of an urbanized area for urban sprawl and its impact on environment description (Chou et al., 2005). The area covered by Terra ASTER satellite data (VIS, NIR and SWIR subset bands) is located in Maryland, U.S., in the outskirts of Washington, around 15 kilometres East of the capital's centre: the pre-change image dates back to April 9th, 2000 and was radiometrically normalized using post-change image of August 24th, 2003 as reference, using a linear regression model and PIFs.

After pre-processing and radiometric normalization, the dataset was converted to NDR values, using equation 1, and an unsupervised approach was chosen to classify changes occurred between the two dates (Bruzzone & Prieto, 2000). ISODATA classification was then performed over NDR bands and post classification labelling was utilized to assign to

retrieved classes a land cover change significance. The results are displayed in Figure 4 for a small area of detail and class colour code illustrated in caption.

The two examples presented exploit supervised or unsupervised classification of NDR values to produce change maps of a flood event (see Figure 3) or an urban growth situation (see Figure 4); the case studies tests showed a good performance in change areas delineation and identification, as a visual inspection of resulting maps witnesses. It should be pointed out that those example are only representative of a first assessment of NDR approach as an aid to Change Detection; in fact, a thoroughly assessment of the NDCD approach capabilities , together with a comparison with other Change Detection techniques results, will be done in the next section over the complete case study covering Hurricane Katrina struck New Orleans city.

5. Application Study – Flood damage assessment with NDCD

5.1 Introductory section

Recent years have seen a tremendous increase in economic and human losses from weather hazards all over the world. Major global climatic alterations are projected to occur during the 21st century and there is great concern about expected negative economic and social consequences resulting from such changes (United Nations, 2007).

Hurricane Katrina was the costliest and one of the deadliest hurricanes in the history of the USA. It was the sixth-strongest Atlantic hurricane ever recorded and the third-strongest land falling U.S. hurricane on record. At its highest intensity, Katrina was a category 5 storm on the Saffir-Simpson scale (Simpson, 1974) with wind speeds of 280 km/h.

The storm made initial landfall at Plaquemines Parish in south-eastern Louisiana on the morning of August 29 2005, and the cities of New Orleans (Louisiana), Mobile (Alabama) and Gulfport (Mississippi) bore the brunt of Katrina's force as it moved inland.

Thanks to the increasing number and observation capabilities of operational remote sensing satellites, remote sensing technology is becoming more and more used for natural hazards monitoring and management, with the great advantage of providing a synoptic vision over a wide area in a short time and in a very cost effective manner (Wang et al., 2002; Brivio et al., 2002; Sanyal & Lu, 2004; Villa & Gianinetto, 2006). In particular, remotely sensed data collected both by radar and optical satellites have been largely used for flood extent evaluation during the last 20 years and now the processing techniques are mature for an operational use (Imhoff et al., 1987; Hess et al., 1995; Frazier et al., 2003; Wang, 2004; Villa & Gianinetto, 2006; Gianinetto & Villa, 2007).

This case study exploits a new method for change detection based on the normalized difference change detection technique (NDCD). The NDCD is a technique which, given an image pair, performs calculations on radiometric normalized reflectance data through the definition of the normalized difference reflectance (NDR) and produces a standardized difference of the reflectance values.

The NDCD was used to detect the damages to the urban area of the city of New Orleans (Louisiana, USA) resulting from the passage of hurricane Katrina. Flood maps were both obtained from the image processing of SPOT-4/HRVIR and Landsat-5/TM imagery, with a suitable spatial resolution for supporting political institutions with a rapid response, effective and prompt decision maker tool.

The maps' accuracy were verified with respect to the inundation maps produced at the Dartmouth Flood Observatory, Dartmouth College (USA). A comparison was also performed between the results of the NDCD technique and that of other standard change detection methods as NIR normalized difference and spectral-temporal minimum noise fraction technique (ST-MNF).

5.2 Dataset

Remotely Sensed Dataset

The flooding caused by Hurricane Katrina over the city of New Orleans (29° 57' 33" latitude north, 90° 03' 36" longitude west) was studied using SPOT-4/HRVIR images supplied by SpotImage and the Centre National d'Etudes Spatiales (CNES) under the Optimising Access to Spot Infrastructure for Science (OASIS) Programme and Landsat-5/TM images made available from the United States Geological Survey's Earth Resources Observation and Science (USGS EROS) through the Hurricane Katrina disaster response project.

The SPOT-4/HRVIR data set was composed of:

- One 20-meters SPOT-4/HRVIR image collected on January 17, 2005 (scene ID 4 601-290 05-01-17 17:03:19 2 I) with orientation angle of 11.5 degree and incidence angle of 19.9 degree left and geocoded in UTM-WGS84 F16N projection. This image was used as pre flood image;
- One 20-meters SPOT-4/HRVIR image collected on September 19, 2005 (Scene ID 4 601-290 05-09-19 16:50:34 1 I) with orientation angle 10.0 degree and incidence angle 0.3 degree right and geocoded in UTM-WGS84 F16N projection. This image was used as post flood image.

The Landsat-5/TM data set was composed of:

- One 30-meters Landsat-5/TM image collected on June 19, 2005 (scene ID 5022039000517010), WRS-2 path 022 row 039, used as pre flood image;
- One 30-meters Landsat-5/TM image collected on September 7, 2005 (scene ID 5022039000525010), WRS-2 path 022 row 039, used as post flood image.

Additional Dataset

For the urban analysis some additional vector maps were used. The 30-meters National Land Cover Database Imperviousness Layer (NLCDIL) raster file representing urbanized and infrastructural features (impervious areas) of the city and surroundings of New Orleans (Yang et al., 2003), made available by USGS through its website (U.S. Geological Survey, 2006) was used for deriving separate mapping for the urban areas only and for the non-urban areas only.

5.3 Methodological Approach

Pre-processing

As typical in change detection applications and as envisaged in the earlier part of this work, about pre-processing of data for change analysis, geocoding and atmospheric correction are always needed. For this purposes the satellite data were first georeferenced in the UTM-WGS84 projection, using reference data. Original at-sensor radiance data were atmospherically corrected using a low resolution Radiative Transfer Code, combined with aerosol retrieval based on band reflectance ratios and with adjacency correction of path radiance (Berk et al., 1999; Vermote et al., 1997).

A further step is the radiometric normalization of multispectral data, carried out using a parabolic parametric model:

$$R_j^{norm} = a_j (R_j^{raw})^{b_j} \quad (2)$$

where:

R^{norm} = normalized reflectance

R^{raw} = input reflectance of the slave image

a = multiplicative coefficient of the parametric model

b = exponential coefficient of the parametric model

j = spectral band number

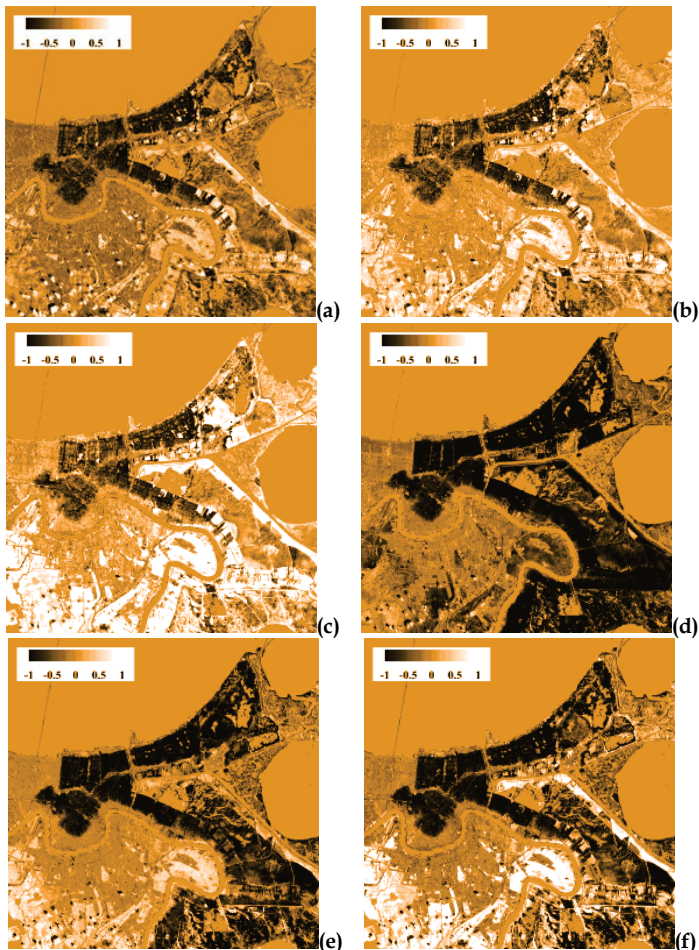


Fig. 6. Example of normalized difference reflectance (NDR) calculated for the Landsat-5/TM dataset. (a) Spectral band nr.1; (b) Spectral band nr.2; (c) Spectral band nr.3; (d) Spectral band nr.4; (e) Spectral band nr.5; (f) Spectral band nr.7.

The radiometric normalization of reflectance data was performed using a parametric parabolic model based on equation 2 through standard linearized least square matching based on a parametric model and an iteration approach to solution of the linearized basic observation equation. The transformation coefficients were computed using standard linearized least squares matching, through an iteration approach to solution of the linearized basic observation equation.

Radiometrically normalized data were used for calculating NDR values for both Landsat-5/TM and SPOT-4/HRVIR data, using the approach described in the previous sections and calculated with equation (4.1). The NDR values were finally used as inputs for Minimum Noise Fraction (MNF) transform, thus structuring the implementation of the NDCD-MNF technique for change mapping and flooded area delineation.

Mapping Hurricane Katrina's aftermaths in New Orleans

The widespread destruction in New Orleans was mapped using the NDCD-MNF technique. The SPOT-4/HRVIR and Landsat-5/TM images were first radiometrically normalized using the parametric model of equation (2.2) and the NDR were computed using equation (4.1). Following, to the multi-spectral NDR values it was applied the MNF transform, generating the normalized difference reflectance-Minimum Noise Fraction (NDR-MNF) components. From all the NDR-MNF components generated, only the first and the second were retained, being the most representative of a good identification of water related land cover. By visual interpretation of the post flood images, the final selection of the best representative NDR-MNF component (component nr.1 or component nr.2) was carried out and the final mapping was realized by using an adaptive threshold. Figure 7 shows the NDR-MNF component nr.2 for SPOT-4/HRVIR (Figure 7a) and Landsat-5/TM (Figure 7b), subsequently used for the mapping.

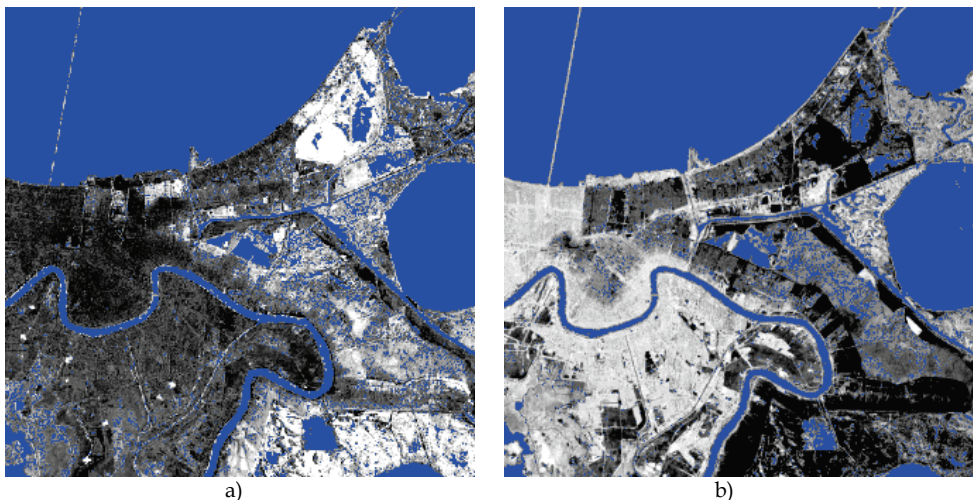


Fig. 7. Normalized Difference Reflectance-Minimum Noise Fraction (NDR-MNF) component nr.2 used for the mapping. (left) SPOT-4/HRVIR; (right) Landsat-5/TM.

The criteria used for the threshold selection was based on the detection of the maximum separability interval between flooded and non-flooded areas. For the SPOT-4/HRVIR and Landsat-5/TM image data some samples of the selected NDCD-MNF component were independently extracted, belonging to the urban and non-urban land cover classes, both for the flooded and non-flooded areas.

For the Landsat-5/TM data set, 900 pixels were selected for the urbanized areas (400 pixels in flooded area and 500 in non-flooded area), covering nearly 0.2% of the total urbanized areas, while 1,200 pixels were selected for the non-urbanized areas (700 pixels in flooded area and 500 in non-flooded area), covering nearly 0.2% of the total non-urbanized areas.

For the SPOT-4/HRVIR data set, 1,500 pixels were selected for the urbanized areas (600 pixels in flooded area and 900 in non-flooded area), covering nearly 0.15% of the total urbanized areas, while 2,000 pixels were selected for the non-urbanized areas (1,200 pixels in flooded area and 800 in non-flooded area), covering nearly 0.15% of the total non-urbanized areas.

For all these samples, the first and second-order statistics were computed and the maximum separability interval between the flooded and non-flooded areas was identified by testing different threshold values belonging to the interval; finally the global flood maps were produced.

Next, using the USGS's NLCDIL as supplementary input data, three other products were generated from the SPOT-4/HRVIR and the Landsat-5/TM data sets: i) a flood map for the 'urban areas only'; ii) a flood map for the 'non-urban areas only'; and iii) a 'fused' flood map:

- i) The flood map for the **urban areas** only was built using the non-impervious surface layer of the NLCDIL as mask for excluding from the processing all the image pixels collected on non-urban areas;
- ii) The flood map for the **non-urban areas** only was built using the impervious surface layer of the NLCDIL as mask for excluding from the processing all the image pixels collected on urban areas;
- iii) The **fused** flood map was built fusing together the results previously obtained for the urban areas only and the non-urban areas only. This processing returned a product comparable to the global flood map above described, but it has proven more accurate.

To boost the spatial coherency and homogeneity of the final mapping, all the flood maps were refined with classical segmentation and clumping techniques.

5.4 Performance Evaluation and Comparison to other techniques

The accuracies of all the maps produced with the NDCD-MNF technique were verified using as ground truth the flood extension map of the city of New Orleans (Figure 8) produced at the Dartmouth Flood Observatory (Dartmouth College, USA) and provided by courtesy of Prof. G.R. Brakenridge and Dr. E. Anderson (Dartmouth College, USA).

The potentialities and performances of the NDCD technique for flood mapping were also compared to following standard change detection methods characterized by different complexity:

- Change detection based on the near-infrared normalized difference (Hayes and Sader, 2001);
- Spectral-Temporal Minimum Noise Fraction (ST-MNF) technique previously developed by authors for flood mapping (Gianinetto & Villa, 2007).

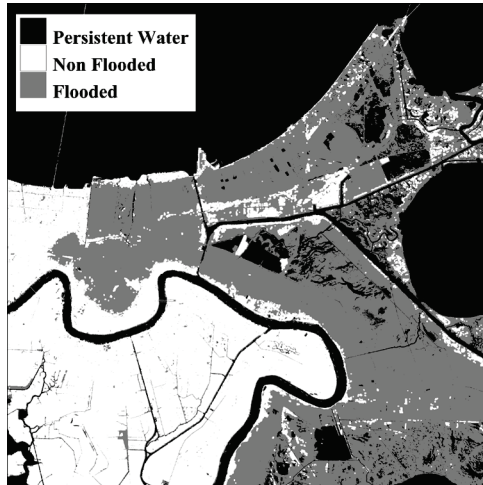


Fig. 8. 20-meters resolution raster image derived from the vector flood extension map produced at the Dartmouth Flood Observatory (Dartmouth College, USA).

NIR normalized difference change detection

The simplest change detection technique used to compare the results obtained using the NDCD-MNF was based on the NIR normalized difference (Hayes and Sader, 2001). Using only the infrared band (TM4 for Landsat-5/TM and XS3 for SPOT-4/HRVIR) it was produced a flood map by thresholding the normalized difference between the post-flood and pre-flood images.

Similarly to the processing carried on with the NDCD-MNF, also using the NIR normalized difference it were separately calculated: i) a flood map for the urban areas only, ii) a the flood map for the non-urban areas only, and iii) a global flood map, both for the SPOT-4/HRVIR and the Landsat-5/TM data set.

Spectral-Temporal Minimum Noise Fraction technique

Another term of comparison for the NDCD-MNF method was the ST-MNF technique previously developed by the authors (Gianinetto & Villa, 2006; Gianinetto & Villa, 2007).

In this case only the global flood maps were generated for both the Landsat-5/TM and SPOT-4/HRVIR data set by processing together both the impervious and non-impervious land cover features. Starting from the pre-processed normalized images, a synthetic n-band file (with n=8 for SPOT-4/HRVIR and n=12 for Landsat-5/TM) was created including first the reflective bands of the pre flood scene followed by the homologous bands of the post flood scene, stacked together. To this Spectral-Temporal merging it was applied the MNF transform and a thresholding to derive the flood extension map, whereas a complete description of the ST-MNF technique can be found in (Gianinetto & Villa, 2007).

5.5 Results and Discussion

Sampling for accuracy assessment

The testing samples used for the accuracy assessment of the flood maps were selected following a stratified random sampling approach over the datasets. In detail, accuracy test samples were collected as:

- a) For the SPOT-4/HRVIR data set:
 - i) 9,921 samples on reference ground truth data (1.0% of the total) for the urbanized areas: 4,017 samples in flooded areas (40.5%) and 5,904 samples in non-flooded areas (59.5%).
 - ii) 13,568 samples on reference ground truth data (1.0% of the total) for the non-urbanized areas: 8,158 samples in flooded areas (60.1%) and 5,410 samples in non-flooded (39.9%).
 - iii) 11,812 samples on reference ground truth data (0.5% of the total) for the whole area, 6,449 samples in flooded areas (54.6%) and 5,363 samples in non-flooded areas (45.4%).
- b) For the Landsat-5/TM data set:
 - i) 8,726 samples on reference ground truth data (2.0% of the total) for the urbanized areas: 3,401 samples in flooded (39.0%) and 5,325 samples in non-flooded areas (61.0%).
 - ii) 11,608 samples on reference ground truth data (2.0% of the total) for the non-urbanized areas: 6,580 samples in flooded areas (56.7%) and 5,028 samples in non-flooded areas (43.3%).
 - iii) 15,596 samples on reference ground truth data (1.5% of the total) for the whole area: 7,878 samples in flooded areas (50.5%) and 7,718 samples in non-flooded areas (49.5%).

Mapping accuracy using the NDCD-MNF technique

Flood maps for the 'urban areas only' and for the 'non-urban areas only', along with a 'fused' flood map were obtained by thresholding the NDR-MNF component nr.1.

Regarding the flood mapping in the urban areas only, the data processing performed on the Landsat-5/TM imagery led to higher accuracy than those performed on the SPOT-4/HRVIR imagery (Table 2, Figure 9). For the former it was obtained a best Overall Accuracy (OA) of 92.05% and a kappa coefficient (K) of 0.83, while for the latter the results gave an OA of 86.37% and a K of 0.72.

The threshold selection was not a critical issue for the Landsat-5/TM data, while for the SPOT-4/HRVIR data, approaching to the upper (positive) limit of the separability interval the accuracy became worse (OA=72.30, K=0.48). In any case, regardless the threshold value selection, the mapping based on the Landsat-5/TM images was always superior to those based on the SPOT-4/HRVIR images.

On the contrary, with respect to the flood mapping in the 'non-urban areas only', the data processing performed on the Landsat-5/TM imagery led to lower accuracy than those performed on the SPOT-4/HRVIR imagery (Table 3, Figure 10). For the former it was obtained a best OA of 75.70% and a K of 0.49, while for the latter the results gave an OA of 86.31% and a K of 0.71.

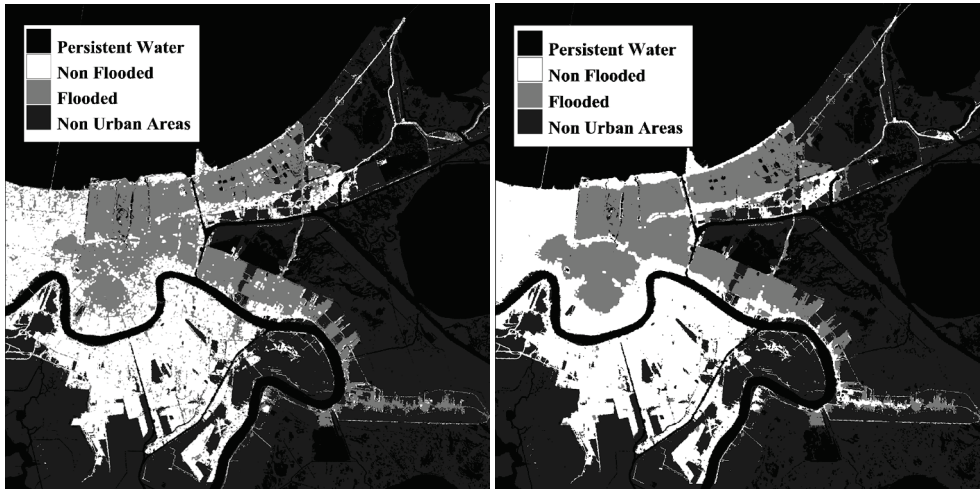


Fig. 9. New Orleans (Louisiana). Flood mapping in the urban areas only using the NDCD-MNF technique. (left) Derived from SPOT-4/HRVIR data; (right) Derived from Landsat-5/TM data.

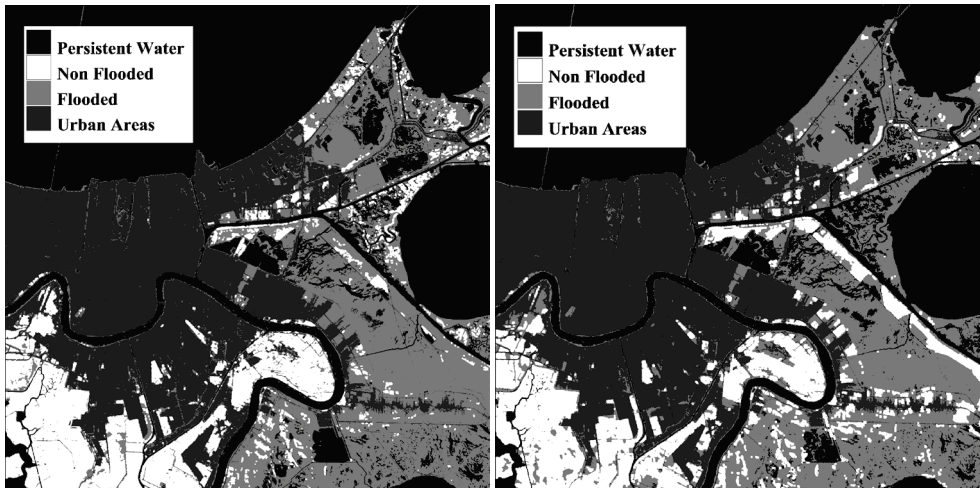


Fig. 10. New Orleans (Louisiana). Flood mapping in the non-urban areas only using the NDCD-MNF technique. (left) Derived from SPOT-4/HRVIR data; (right) Derived from Landsat-5/TM data.

The threshold selection was not a critical issue both for the Landsat-5/TM and for the SPOT-4/HRVIR data. In any case, regardless of the threshold value selection, this time the mapping based on the SPOT-4/HRVIR images was always superior to those based on the Landsat-5/TM images.

	Threshold value	Overall Accuracy * (%)	K coefficient
<i>SPOT-4/HRVIR data set</i>	-2.0	85.72	0.70
	-1.5	86.37	0.72
	-1.0	84.96	0.69
	0.0	72.30	0.48
<i>Landsat-5/TM data set</i>	-1.0	89.69	0.79
	0.0	90.52	0.80
	1.0	92.05	0.83
	2.0	91.16	0.81

* Values in bold indicate the best accuracy.

Table 2. Flood mapping in the ‘urban areas only’ using the NDCD-MNF technique. Threshold selection and mapping accuracy.

	Threshold value	Overall Accuracy * (%)	K coefficient
<i>SPOT-4/HRVIR data set</i>	0.0	85.56	0.69
	0.5	86.31	0.71
	1.0	86.17	0.70
	2.0	84.0	0.65
<i>Landsat-5/TM data set</i>	-1.5	75.59	0.49
	-1.0	75.70	0.49
	-0.5	75.11	0.48
	0.0	74.34	0.47

* Values in bold indicate the best accuracy.

Table 3. Flood mapping in the ‘non-urban areas only’ using the NDCD-MNF technique. Threshold selection and mapping accuracy.

The reason of the poor mapping in non-urban areas using the Landsat-5/TM dataset may be found in the closeness of the post-flood image (September 7, 2005) to Katrina landfall (August 29, 2005). In the Landsat-5/TM post-flood image many wet areas (rain-washed), mainly located in non urbanized areas (impervious surfaces), were incorrectly detected as flooded. This phenomena was not observed in the SPOT-4/HRVIR post-flood image because of the longer time elapsed from the passage of Katrina (September 19, 2005).

A global ‘fused’ flood map was obtained by fusing together of the urbanized and non-urbanized flood maps separately computed with the NDCD-MNF technique (Figure 11). In this case, the comparison of global results from the Landsat-5/TM (OA=84.03% and K=0.68) and SPOT-4/HRVIR (OA=86.36% and K=0.73) data processing are similar, with a little advantage for the SPOT-based mapping. This result is justified by the non homogeneous accuracy of the Landsat-based mapping in urban and non urban areas, as previous described (Table 2 and Table 3).

A simpler and less computational expensive global flood map was derived by processing together both the urban and non-urban areas in a single step (Figure 12). Differently to the previous cases, the NDR-MNF component nr.2 was used this time for the thresholding of both the dataset.

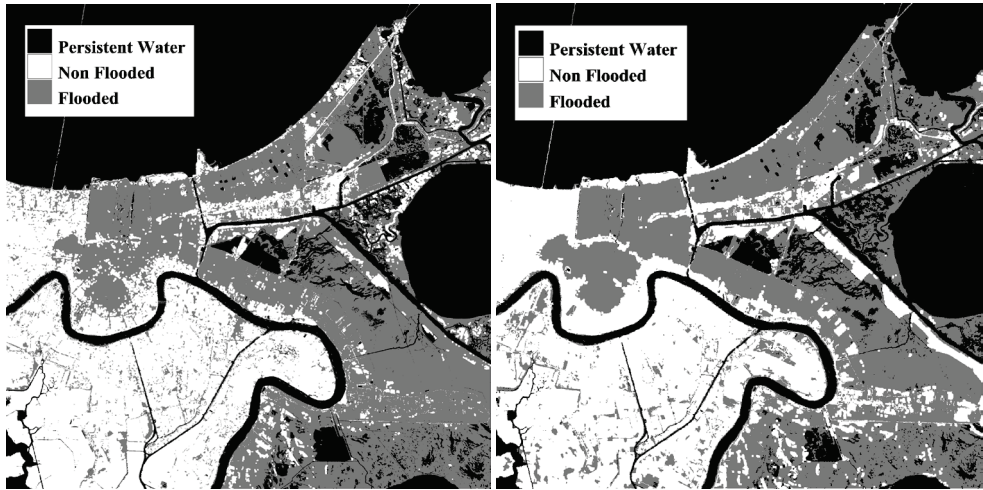


Fig. 11. New Orleans (Louisiana). Global 'fused' mapping using the NDCD-MNF technique. (left) Derived from SPOT-4/HRVIR data; (right) Derived from Landsat-5/TM data.

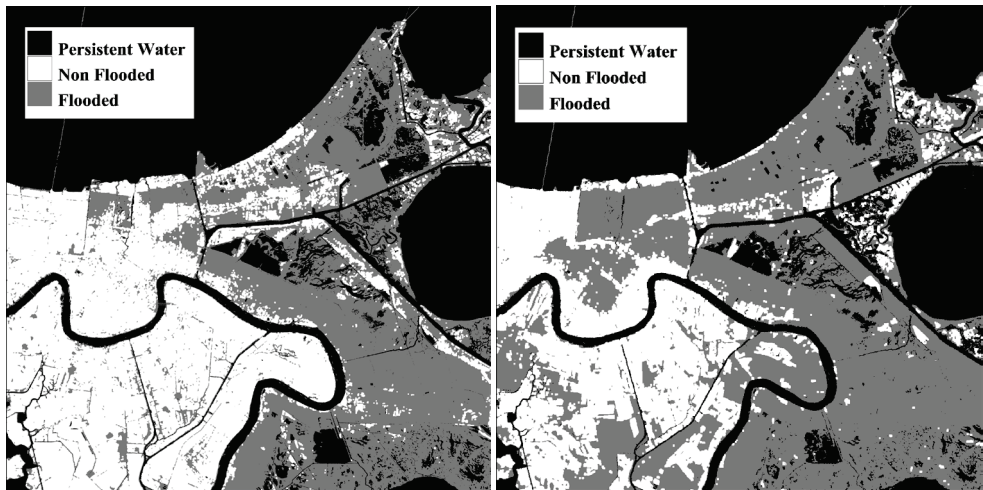


Fig. 12. New Orleans (Louisiana). Global flood map using the NDCD-MNF technique. (left) Derived from SPOT-4/HRVIR data; (right) Derived from Landsat-5/TM data.

Again the Landsat-5/TM data processing led to lower accuracy (OA=77.32% and K=0.54) than the SPOT-4/HRVIR data processing (OA=83.20% and K=0.67), when compared to ground truth data, mainly due to its lower accuracy in the non-urban areas (Table 4).

	Threshold value	Overall Accuracy * (%)	K coefficient
<i>SPOT-4/HRVIR data set</i>	-2.0	66.11	0.28
	-1.0	82.02	0.63
	-0.5	83.20	0.67
	0.0	82.24	0.65
	1.0	75.92	0.53
<i>Landsat-5/TM data set</i>	0.0	76.60	0.53
	1.0	77.32	0.54
	2.0	75.81	0.51
	3.0	72.50	0.45

* Values in bold indicate the best accuracy.

Table 4. Global flood mapping using the NDCD-MNF technique when processing together both the impervious and non-impervious surfaces. Threshold selection and mapping accuracy.

This time, for the SPOT-4/HRVIR dataset both the NDR-MNF component selection and the threshold selection are critical and the accuracy largely depends upon their correct identification. Regarding the NDR-MNF component selection, when using component 1 instead of component 2, as in previous cases, for the Landsat-5/TM data processing we had a little decrease in the mapping accuracy (from 77.32% to 75.50% for the OA), while for the SPOT-4/HRVIR data processing a greater decrease in the mapping accuracy was observed (from 83.20% to 67.74% for the OA).

The foremost advantage of this single step global mapping is that no urban mask is required, so no a priori information is needed to separate impervious from non impervious surfaces. On the other hand, the mapping accuracy is always worse when compared to the global ‘fused’ map obtained by fusing together of the urbanized and non-urbanized flood maps separately computed (Figure 11). For the SPOT-4/HRVIR data it was observed a decrease in the OA from 86.36% to 83.20% and a decrease in the K from 0.73 to 0.67, while for the Landsat-5/TM data it was observed a larger decrease both in the OA from 84.36% to 77.32% and in the K from 0.68 to 0.54 (Table 4).

Comparing the mapping accuracy of the NDCD-MNF to the NIR-normalized difference change detection

Using the NIR-normalized difference change detection it was generated a global flood map by processing together both the urban and the non-urban areas (Figure 13) and a global ‘fused’ flood map by processing separately the urban and the non-urban areas (Figure 14). This technique seems to be insensitive to both the data set used (SPOT-4/HRVIR or Landsat-5/TM) and to the data processing adopted (global or ‘fused’ map), leading to an OA between 81.66% and 82.75% and a K between 0.63 and 0.66. Table 5 shows a summary of results.

When comparing the NDCD-MNF to the NIR-normalized difference change detection it emerged the superiority of the former in all the ‘fused’ products (OA=86.36 and K=0.73 for the SPOT-4/HRVIR data set and OA=84.36 and K=0.68 for the Landsat-5/TM data set) and a better accuracy for the latter with respect to only the Landsat-5/TM global flood map (OA=82.03% and K=0.64).

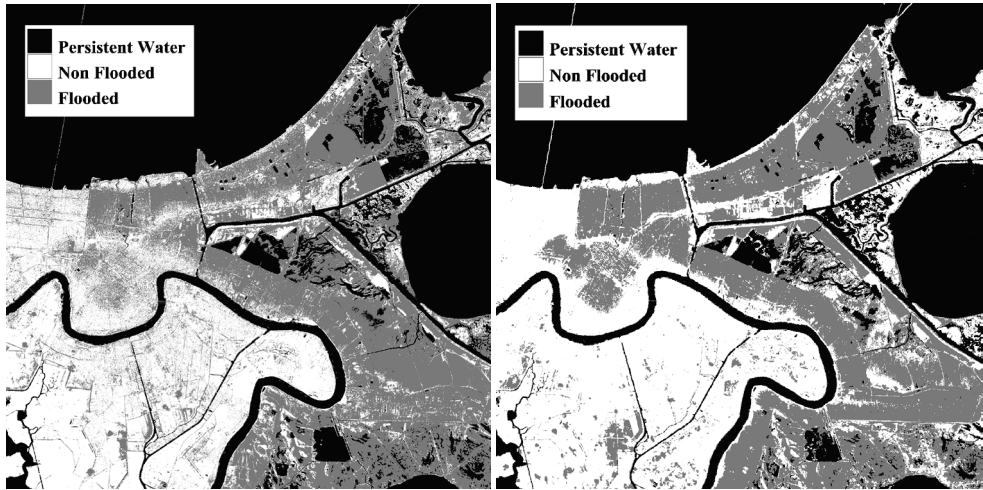


Fig. 13. New Orleans (Louisiana). Global flood mapping using the NIR-normalized difference technique. (left) Derived from SPOT-4/HRVIR; (right) Derived from Landsat-5/TM.

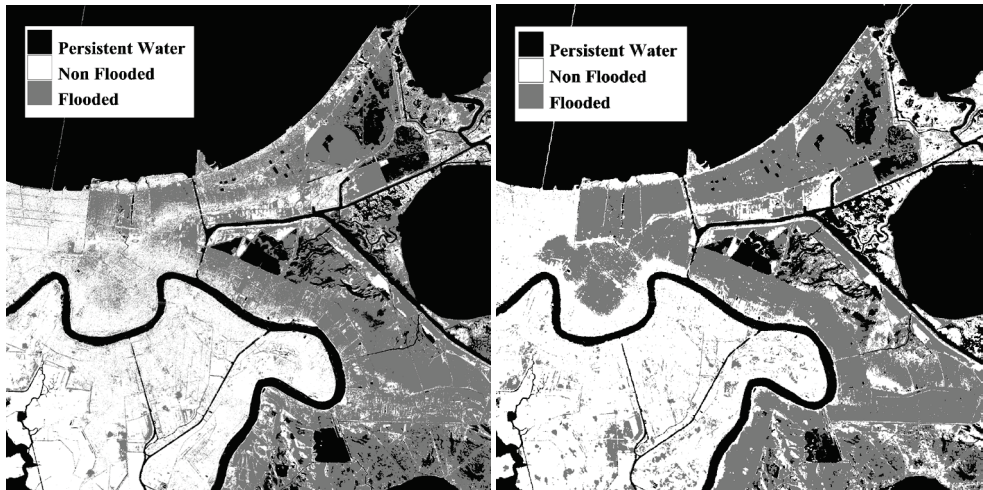


Fig. 14. New Orleans (Louisiana). Global 'fused' map using the NIR-normalized difference technique. (left) Derived from SPOT-4/HRVIR; (right) Derived from Landsat-5/TM.

	Threshold Value	Overall Accuracy * (%)	K coefficient
<i>SPOT-4/HRVIR data set</i>	-0.100	79.13	0.59
	-0.075	81.07	0.62
	-0.050	81.73	0.63
	-0.025	80.83	0.61
<i>Landsat-5/TM data set</i>	-0.35	79.92	0.60
	-0.30	82.03	0.64
	-0.25	81.95	0.64
	-0.20	80.16	0.60

* Values in bold indicate the best accuracy.

Table 5. Global flood mapping using the NIR-normalized difference change detection technique. Threshold selection and mapping accuracy.

Comparing the mapping accuracy of the NDCD-MNF to the Spectral-Temporal Minimum Noise Fraction technique

Using the ST-MNF technique it was generated a global flood map by processing together both the urban and the non-urban areas on the basis of the MNF component nr. 1 for both the data set (Figure 15).

The threshold selection here revealed to be not a critical issue for the mapping accuracy. By using the ST-MNF technique, the Landsat-5/TM data processing led to higher accuracy (OA=90.17% and K=0.80) than the SPOT-4/HRVIR data processing (OA=81.87% and K=0.63) when compared to ground truth data. Table 6 shows a summary of results.

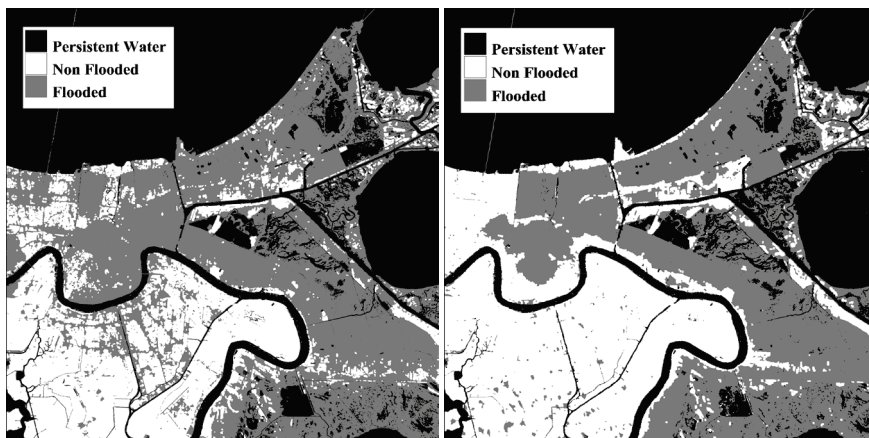


Fig. 15. New Orleans (Louisiana). Global damages mapping using the ST-MNF technique. (left) Derived from SPOT-4/HRVIR data; (right) Derived from Landsat-5/TM data.

A comparison between the NDCD-MNF and the ST-MNF shows that the former always performed better on the SPOT-4/HRVIR data set, regardless the data processing used for the global mapping used (with or without urban areas masking). While with respect to the Landsat-5/TM data set, results are more difficult to analyse. Looking at the OA only it seems that the ST-MNF led to higher accuracy (OA=90.17 for ST-MNF and OA=84.36 for NDCD-MNF), but the K score of the NDCD-MNF is higher for the ‘fused’ map (K=0.68 for

NDCD-MNF and $K=0.63$ for ST-MNF). So it is difficult to say which performed better with respect to the overall situation.

	Threshold value	Overall Accuracy * (%)	K coefficient
<i>SPOT-4/HRVIR data set</i>	-1.0	81.76	0.63
	-0.5	81.87	0.63
	0.0	81.43	0.62
	1.0	79.33	0.58
<i>Landsat-5/TM data set</i>	-1.0	89.37	0.78
	0.0	89.79	0.79
	0.5	90.12	0.80
	1.0	90.17	0.80
	2.0	89.71	0.79

* Values in bold indicate the best accuracy.

Table 6. Global flood mapping using the ST-MNF technique. Threshold selection and mapping accuracy.

5.6 Summary and Conclusions

This case study tested the normalized difference change detection technique effectiveness for change mapping, also in comparison with other literature methodologies, starting from the processing of the normalized difference reflectance data.

The radiometric normalization of data influenced the accuracy of the mapping. A parametric normalization with coefficients calculated with standard linearized least squares adjustment and iterative solution was found a better solution than a standard linear normalization, and thus adopted. However, the general definition of the NDCD leaves the possibility to develop processing techniques based on different radiometric normalization schemes.

Using its MNF implementation, the NDCD technique was used for mapping and evaluating the havoc on the city of New Orleans (Louisiana, USA) wreaked by Hurricane Katrina landfall in August 2005, using both a SPOT-4/HRVIR and a Landsat-5/TM data set.

As a term of comparison for evaluating the potentialities and performances of the NDCD-MNF technique, several other standard change detection methods have been tested: from the simple NIR normalized difference to the more complex Spectral-Temporal Minimum Noise Fraction technique.

Comparing the global mapping accuracy when using the SPOT-4/HRVIR data, the NDCD-MNF technique always led to better results than all the others methods here taken into consideration. Moreover, results were better when processing separately the urban and the non-urban areas in the so called 'fused' product.

With regards to the Landsat-5/TM data, the NDCD-MNF technique poorly performed in the non-urban areas (probably due to the closeness of the post-flood image to Katrina landfall), thus affecting the final global mapping. However, with respect to the only urban areas, which may be of major interest in most cases, the NDCD-MNF always performed better.

Finally, regarding the threshold selection, a number of studies [Yuan et al., 1998; Chen et al., 2003] have pointed out that a major weakness of all spectral change detection approaches is that the selection of a minimum threshold to signify change is often arbitrary [Warner,

2005]. For example, a threshold value of two standard deviations above the mean is sometimes selected (Sohl, 1999). To address this problem, some studies used a noise model to select the threshold (Dwyer et al., 1996). As an alternative, other studies have developed a systematic method using training data (Chen et al., 2003). In their approach, areas of change are digitized, as well as a surrounding window of no change. These training areas are then classified into 'change' and 'no change' classes, using a small number of arbitrarily chosen thresholds spread over a wide range of possible values. Based on the accuracies of these classifications, the range of thresholds is narrowed successively to focus on the region where the accuracy is highest. In this iterative fashion, an optimal threshold is selected.

The accuracy gained through NDR approach derived changes mapping have been proven very satisfying most of the times, with Overall Accuracies percentage figures ranging from 80% to over 90% of correctness, that is to say error percentage in change mapping around 10% which is to be considered as a really good result for analysis performed and based only on remote sensing satellite data.

Nevertheless, more testing and a more fine tuning of the processing chain can be implemented and done in the future, including not yet explored land cover change application fields, and the good results achieved until now are a great encouragement to continue on the path already traced with the works described in this chapter.

6. References

- Berk, A., Anderson, G. P., Bernstein, L. S., Acharya, P. K., Dothe, H., Matthew, M. W., Adler-Golden, S. M., Chetwynd, J. H., Richtsmeier, S. C., Pukall, B., Allred, C. L., Jeong, L. S., and Hoke, M. L., (1999); MODTRAN4 Radiative Transfer Modeling for Atmospheric Correction, *SPIE Proceedings, Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research III*, vol. 3756, pp.348-353
- Brivio, P. A., Colombo, R., Maggi, M., and Tomasoni, R., (2002); Integration of remote sensing data and GIS for accurate mapping of flooded areas, *International Journal of Remote Sensing*, 23(3), pp. 429-441.
- Bruzzone, L., and Fernández Prieto, D., (2000); Automatic Analysis of the Difference Image for Unsupervised Change Detection, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp.1171-1182.
- Chen, J., Gong, P. and Shi, P., (2003); Land-use/land-cover change detection using improved change-vector analysis, *Photogrammetric Engineering and Remote Sensing*, 69 , pp. 369-379.
- Chou, T. Y., Lei, T. C., Wan, S., and Yang, L. S., (2005); Spatial knowledge databases as applied to the detection of changes in urban land use, *International Journal of Remote Sensing*, vol. 26, no. 14, pp. 3047-3068.
- Coppin, P., and Bauer, M.E., (1994); Digital Processing of Multitemporal Landsat TM Imagery to Optimize Extraction of Forest Cover Change Features, *IEEE Transactions on Geoscience and Remote Sensing*, 32(4), pp. 918-927.
- Coppin, P., Nackaerts, K., Queen, L., and Brewer, K., (2001); Operational Monitoring of Green Biomass Change for Forest Management, *Photogrammetric Engineering & Remote Sensing*, 67(5), pp. 603-611.

- Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., and Lambin, E., (2004); Digital change detection methods in ecosystem monitoring: a review, *International Journal of Remote Sensing*, 25(9), pp. 1565-1596.
- Dwyer, J.L., Saylor, K.L., and Zylstra, G.J., (1996); Landsat pathfinder data sets for landscape change analysis. *Proceedings of the 1996 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 1996*, Lincoln, Nebraska, pp. 547-550.
- Frazier, P., Page, K., Louis, J., Briggs, S., and Robertson, A. I., (2003); Relating wetland inundation to river flow using Landsat TM data, *International Journal of Remote Sensing*, 24(19), pp. 3755-3770.
- Gianinetto, M., and Villa, P., (2006); Monsoon Flooding Response: a Multi-scale Approach to Water-extent Change Detection, *The International Archive of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVI(7), Enschede, the Netherlands, pp. 128-133.
- Gianinetto, M., and Villa, P., (2007); Rapid Response Flood Assessment Using Minimum Noise Fraction And Composed Spline Interpolation, *IEEE Transactions on Geoscience and Remote*, 45(10), pp. 3204-3211.
- Green, A.A., Berman, M., Switzer, P., and Craig, M.D., (1988); A transformation for ordering multispectral data in terms of image quality with implications for noise removal, *IEEE Transactions on Geoscience and Remote Sensing*, 26(1), pp. 65-74.
- Hayes, D.J., and Sader, S.A., (2001); Comparison of Change-Detection Techniques for Monitoring Tropical Forest Clearing and Vegetation Regrowth in a Time Series, *Photogrammetric Engineering and Remote Sensing*, 67(9), pp. 1067-1075.
- Hess, L.L., Melack, J.M., Filoso, S., and Wang, Y., (1995); Realtime mapping of inundation on the Amazon floodplain with the SIR-C/X-SAR synthetic aperture radar, *IEEE Transactions on Geoscience and Remote Sensing*, 33, pp. 896-904.
- Imhoff, M. L., Vermillion, C., Story, M. H., Choudhury, A. M., and Gafoor, A., (1987); Monsoon flood boundary delineation and damage assessment using spaceborne imaging radar and Landsat data, *Photogrammetric Engineering and Remote Sensing*, 4, pp. 405-413.
- Lu, D., Mausel, P., Brondizio, E., and Moran, E., (2004); Change detection techniques, *International Journal of Remote Sensing*, 25, pp. 2365-2407.
- Rogan, J., Franklin, J., Roberts, D. A., (2002); A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery, *Remote Sensing of Environment*, vol. 80, pp. 143-156.
- Sanyal, J., and Lu X., X., (2004); Application of the Remote Sensing in Flood Management with Special Reference to Monsoon Asia: a Review, *Natural Hazards*, 33, pp. 283-301.
- Simpson, R.H., (1974); The hurricane disaster-potential scale, *Weatherwise*, 27, pp. 169.
- Singh, A., (1989); Digital change detection techniques using remotely-sensed data, *International Journal of Remote Sensing*, 10(6), pp. 989-1003.
- Sohl, T.L., (1999); Change analysis in the United Arab Emirates: an investigation of techniques, *Photogrammetric Engineering and Remote Sensing*, 65, pp. 475-484.
- U.S. Geological Survey, (2001); *National Land Cover Database 2001 (NLCD 2001)*, digital resource available online at: www.mrlc.gov/mrlc2k_nlcd.asp.
- United Nations, (2007); *Evidence is now 'unequivocal' that humans are causing global warming*, UN report.

- Vermote E., Tanré D., Deuzé J. L., Herman M., Morcrette J. J., (1997); Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: An Overview, *IEEE Transactions on Geoscience and Remote Sensing*, 35(3), pp. 675-686.
- Villa, P., and Gianinetto, M., (2006); Multispectral transform and Spline Interpolation for Mapping Flood Damages, *Proceedings of the 2006 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2006*, Denver, U.S., pp. 275-279.
- Villa, P., and Lechi, G., (2007); Normalized Difference Reflectance: An Approach to Quantitative Change Detection, *Proceedings of the 2007 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2007*, Barcelona, Spain, pp. 2366-2369.
- Wang, Y., Colby, J.D., and Mulcahy, K.A., (2002); An efficient method for mapping flood extent in a coastal flood plain using Landsat TM and DEM data, *International Journal of Remote Sensing*, 23(18), pp. 3681-3696.
- Wang, Y., (2004); Using Landsat 7 TM data acquired days after a flood event to delineate the maximum flood extent on a coastal floodplain, *International Journal of Remote Sensing*, 25(5), pp. 959-974.
- Yang, L., Huang, C., Homer, C.G., Wylie, B.K., and Coan, M.J., (2003); An approach for mapping large-area impervious surfaces: synergistic use of Landsat-7 ETM+ and high spatial resolution imagery, *Canadian Journal of Remote Sensing*, 29(2), pp. 230-240.

Using Kernel Methods in a Learning Machine Approach for Multispectral Data Classification. An Application in Agriculture

Adrián González*, José Moreno†, Graham Russell‡ and Astrid Márquez*

**Romulo Gallegos University, Venezuela*

†Central University of Venezuela

*‡University of Edinburgh
Scotland, UK*

Abstract

Most pattern recognition applications within the Geoscience field involve the clustering and classification of remote sensed multispectral data, which basically aims to allocate the right class of ground category to a reflectance or radiance signal. Generally, the complexity of this problem is related to the incorporation of spatial characteristics that are complementary to the nonlinearities of land surface heterogeneity, remote sensing effects and multispectral features. The present chapter describes recent developments in the performance of a kernel method applied to the representation and classification of agricultural land use systems described by multispectral responses. In particular, we focus on the practical applicability of learning machine methods to the task of inducing a relationship between the spectral response of farms land cover to their informational typology from a representative set of instances. Such methodologies are not traditionally used in agricultural studies. Nevertheless, the list of references reviewed here show that its applications have emerged very fast and are leading to simple and theoretically robust classification models. This chapter will cover the following phases: a) learning from instances in agriculture; b) feature extraction of both multispectral and attributive data and; c) kernel supervised classification. The first provides the conceptual foundations and a historical perspective of the field. The second belongs to the unsupervised learning field, which mainly involves the appropriate description of input data in a lower dimensional space. The last is a method based on statistical learning theory, which has been successfully applied to supervised classification problems and to generate models described by implicit functions.

1. Introduction

A farming type or modality is a representation of a population of farms that share the same n dimensional traits. Typically, farming system studies seek to define separate groups of farms by looking for a natural structure among the observations. The objective is to maximize homogeneity within clusters and heterogeneity between them (Dixon et al., 2001; Hair et al., 1998). Information about properties of farming systems such as censuses and surveys have long been the most widely used instruments to gather data on agrarian activities; indeed, historically they have proved to be a useful means of gaining knowledge of such diverse agrarian

features as: dominant patterns of farm activities and household livelihoods, including field crops, livestock, trees, aquaculture, grazing and forest areas, crop-livestock integration, technology, farm size and land tenure, to mention but a few. Nevertheless, the high requirements in terms of human and monetary resources of censuses and surveys prevent their application with the frequency and extent required to tackle the complexity of many agricultural issues.

The rapid development shown by land observation satellites over the last three decades has made a great deal of information about land surfaces available. This has widely been used to study land cover changes by the general model of pattern recognition process; which can be divided into a sequence of three main elements: a) generation of input random vectors with the information to be classified (sensor); b) translation of data into a statistically independent representation code that preserves their most relevant characteristics (feature extraction); and c) a system that, based on extracted features, can develop a function space where an operator might be built to serve as an answer predictor to any input generated by the sensor (classification). In this sense, within the field of pattern recognition, one of the most studied subjects is the idea of approximating relationships from the within-farm land surface processes and their emerging spectral response; using methods that can fit the complexity of these processes. This is vitally important for the study of crop-livestock production systems, given that these are critical to the livelihood of an important portion of the rural population at a worldwide level (Bouwman et al., 2005; Seré & Steinfeld, 1996). In addition, projections indicate that the demand for livestock food products is increasing globally (Delgado et al., 1999; Wint et al., 2000), and concern about the potential response of these systems is generally justified.

On this issue, a problem that remains open is the spatial monitoring of crop-livestock systems especially for those involving open range feeding, from which sometimes only time- and site-specific data can be approximated through field methods. These are usually not cost effective and suffer from poor spatial resolution. It is also true, in a broader context, that public availability of space-based remote sensing has helped with the monitoring of land surface biophysical properties. Some approaches have been concerned with the correction of observational data to create valued-added time series (Gleason et al., 2002; Green & Hay, 2002). Others in turn stress the use of optical, thermal and microwave data to model atmospheric and soil moisture (Dubayah, 1992; McVicar & Jupp, 2002); exploiting radiative transfer theory to estimate biophysical properties of vegetation (Goel, 1987; Myneni et al., 1992; Wylie et al., 2002); and macroscale modelling (Asrar & Dozier, 1994; Kimes et al., 2002). In summary, most methodologies monitor and map land surface processes by classification or detecting change (Song et al., 2001). Nevertheless, there is no evidence of using the gathered spectral data in recognising patterns associated with agricultural land management where an optimal discrimination of pixel mixture might be inferred beyond a training set. It has been in this context that the general aim of this chapter was defined to provide a unified framework and examples of using learning machines to accomplish the task of pattern recognition for complex mosaics of within-farm land use in crop-livestock systems from multi-spectral data.

These methodologies are based on feature induction from a training set by establishing a separating hyperplane between any two classes whose margin is maximum. Additionally, they include the inherent advantage of kernel functions, through which solutions are not built in the input space but into one with a higher dimensionality. In this feature space, it is possible that linear functions are enough to separate classes; given that input data are taken to this space by a nonlinear transformation whose diversity adds richness to the process of finding - if it exists - a solution. This flexibility is considered critical within the field of learning machines,

to deal with complex task of using multispectral data for pattern recognition in crop-livestock systems.

2. Historical elements of statistical learning from instances in agriculture

2.1 General problem

The process of estimating an unknown input-output dependence and generalising it beyond a limited training set of observations is acknowledged as learning from instances, which had its origin in the pioneering work of Rosenblatt (1958). During the 1960's the application of this paradigm was seriously hampered as a result of the work of Minsky & Papert (1969). By this time it was thought that complex applications in the real world would require representational hypotheses much more expressive than linear functions, given that the target concept could not normally be represented as a simple linear combination of data attributes. As a result, some fields of study such as learning machine and pattern recognition were negatively affected, preventing their use on applied research including farming systems. It was subsequently demonstrated that the theories of Minsky & Papert were wrong.

Creating typologies of farming systems has been one of the major approaches within the field of agricultural systems in which research has been conducted. This paradigm mainly refers to those methods characterised by inductive non-supervised clustering of farms within a taxonomy; where farm likeness is represented according to a finite set of m -dimensional variables (Berdegue & Escobar, 1990; Köbrich et al., 2003; Kostrowicki, 1977). During the 70's most of the learning techniques used in the agricultural system field were influenced by the wave of learning linear decision surfaces (Capillon, 1985; Hart, 1990; Kostrowicki, 1977). That kind of representation was preferred given that its theoretical properties were well understood. After the 80s, researchers trying to move away from the limitations of linear models started using non-linear models in decision trees decision trees and artificial neural networks. These techniques were rapidly employed within the agriculture domain. However, the main problems of these approaches were their theoretical weakness and that their solution space had many local minima.

The consolidation and application of statistical learning theory during the mid-90's allowed the development of efficient algorithms to learn non-linear functions. These ideas completely recast the pioneering work of Rosenblatt (1958); and were theoretically supported in statistical learning theory (Vapnik, 1995, 1998; Vapnik & Chervonenkis, 1974). Vapnik and Chervonenkis formalised the learning problem as a function estimation; where given an empirical data set generated by a regular stochastic distribution, the algorithm pursues the extraction of regularities in the data by a general model of learning that might be summarised in a sequence of components: a) an input vector generator; b) a system that produces an output value and c) a linear machine.

Contrasting with the statistical learning theory, which appeared on the scene quite recently, another current solution implementation is based on kernel functions (Aronszajn, 1950; Mercer, 1909), which were first studied about a century ago, and which have been playing an important role in increasing the representational capacity of the solutions especially in agricultural applications involving remote sensing. Their use within the learning task relates closely to data pre-processing; and along with the learning machine, constitutes a compact body.

2.2 Particular cases

Supervised and unsupervised learning are among the most investigated applications in agriculture. The former approach pursues building relations between input vectors and target

outputs. The outputs may be expressed at different scales: categorically or numerically, corresponding to classification and regression problems respectively. The unsupervised approach, rather than approximating input data to a target label, seeks to approximate data by similarity expressions, generally distance functions, from which groups of data that resemble each other can be built. This paradigm is usually referred to as clustering (Bishop, 2006).

The remote sensing works of Hermes et al. (1999) and Huang et al. (2002) are precursors of the classification approach in agriculture, where, given a spatially dispersed set of pixels, different forms of land cover (closed forest, open forest and woodland) are classified according to their spectral response. Other research of this kind includes the work of Keuchel et al. (2003) which progressively compares land cover classification using three methods (support vector machines, maximum likelihood and iterated conditional models); and the work of Su et al. (2007) which uses the multi-angle approach and its corresponding spectro-radiometer image to accurately map grassland types by support vector machines. A good application of learning machines on the regression problem is the work of Yang et al. (2007) within the forestry field. In that research, the target vector used was eddy covariance-based gross primary production (GPP) and three remotely sensed variables (land surface temperature, enhanced vegetation index and land cover) in order to predict flux-based GPP at a continental scale.

Regarding the clustering problem in the unsupervised ground, Diez et al. (2006) combined a kernel-based similarity function and a support vector machine to permit the identification of public beef product preferences stratified by market segment. In addition, within the unsupervised family can be found density estimators, which mainly project data from a high onto a lower dimensional space to determine its distribution in the input space in order to add visual richness to the solutions represented (Bishop, 2006).

In summary, these methodologies are based on feature induction from a representative set of instances, where it may be possible to produce a model able to generalise beyond the training instances. In this way a description of relationships present in the original data is possible, and their representation is simplified at the same time that their main features are preserved. Today, there is still a wide usage of linear paradigms in farming systems studies (Dobremez & Bousset, 1995; Köbrich et al., 2003; Milá et al., 2006) while extensive applications of linear machine techniques in agriculture are still scarce. The forerunners have shown that models generated are flexible, theoretically robust and provide expressive solutions. Some of the preliminary results of the present topic may be found in González et al. (2007). For those seeking a deep understanding in the machine learning field the following publications are suggested: Bishop (2006); Cristianini & Shawe-Taylor (2000); Shawe-Taylor & Cristianini (2006) and Vapnik (1995, 1998).

3. Feature extraction of both multispectral and attributive data

Feature extraction constitutes an important task within multidimensional crop-livestock pattern classification. The idea behind it is, among others, to isolate those statistical characteristics of the data that portray essential elements of them, and to provide a better understanding about the underlying processes that generate the data (Guyon & Elisseeff, 2003). Feature extraction is also very effective for avoiding the redundancy that characterises crop-livestock systems (crop production, land use, livestock production, management, etc) by finding meaningful projections, of even low dimensional input data, into a feature space. Principal components analysis (PCA) is one of the standard techniques to obtain features from input data (Jolliffe, 2002). This is achieved by maximising the projected variance onto mutually orthogonal eigenvectors along the directions of higher eigenvalues through iterative algorithms that

minimise information losses. PCA basically performs a linear decomposition of input vectors, into a space whose coordinate system is hierarchically organised by data variability (Bishop, 2006).

Feature extraction through principal component analysis (also referred to as the *Karhunen-Loève* transform) can be traced back to the pioneering work of Pearson (1901) and Hotelling (1933a,b). Today PCA is one of the feature extraction methods most used in farming systems (Berdegue & Escobar, 1990; Köbrich et al., 2003), and there has been considerable research surrounding the application of this technique in different topics of pattern recognition (Bishop, 2006; Duda et al., 2001; Jolliffe, 2002). Basically, the method involves the finding of a lower dimensional space by the orthogonal transformation of the coordinate system where a given data set is described, with the aim of identifying directions of maximum variability. Let us consider a set of observations such that:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & & & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \quad (1)$$

where X is the original data set $m \times n$ matrix, n is the number of samples, which conform m -dimensional vectors ($\alpha = x_1 \dots x_m \in \mathbb{R}^N$) of random variables in an arbitrary space. These vectors are linearly decomposed into another coordinate system whose first axis is a projection of each observation and respond to the linear function $\alpha_1^T x$. This new $m = 1$ -dimensional subspace is oriented to the direction where the elements of X show their highest variability. The subsequent axes are orthogonally aligned in X to the next highest direction through recursive linear decompositions until m vectors have been aligned $\alpha_m^T x$. The axes of this new coordinate system are organized hierarchically according to data variability, and are normally referred to as principal components. It might happen that those components in directions of very low variability are practically constant for all vectors (Jolliffe, 2002), and can be eliminated since they do not contribute new information. Therefore, a substantial dimensionality reduction ($\ll m$) of the problem is usually achieved, given that typically a few axes are enough to retain most of the data structure, if this exists.

Generally the feature extraction and dimensionality reduction proceeds as described above. However, it is worth pointing out the following observations: to obtain the new coordinate system data must be projected to the direction aligned with the maximum variance; this best fit axis passes through the mean of the data cloud which is given by:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x \quad (2)$$

In order to establish this direction, data is projected onto the $d = 1$ -dimensional vector whose scalar value projection is defined by $\alpha_1^T x$ with a projected data variability such that:

$$\frac{1}{n} \sum_{i=1}^n \{\alpha_1^T x - \alpha_1^T \bar{x}\}^2 = \alpha_1^T S \alpha_1 \quad (3)$$

Variability maximisation is pursued in such a way that the sum of squares of element on α_1 equals 1 ($\alpha_1^T S \alpha_1 = 1$), where S is defined by:

$$S = \frac{1}{n} \sum_{i=1}^n (x_n - \bar{x})(x_n - \bar{x})^T \quad (4)$$

At this stage, the main task is the minimisation of redundancy present in the covariance and maximisation of useful information provided by the variance. Diagonal elements of the covariance matrix summarise the data dynamic of interest as long as they are high. Otherwise, they are associated with noise. Maximisation of $\alpha_1^T S \alpha_1$ is performed incorporating a Lagrange multiplier λ :

$$\alpha_1^T S \alpha_1 + \lambda_1 (1 - \alpha_1^T \alpha_1) \quad (5)$$

whose derivative with respect to α_1 yields:

$$S \alpha_1 = \lambda_1 \alpha_1 \quad (6)$$

Considering that the eigenvalues are ordered in a decreasing sequence ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$) being $\lambda^1 = \lambda_{max}$ and proceeding by mathematical induction, it is assumed that principal components from 1 to $m - 1$ can be found along the first $m - 1$ directions of eigenvectors. The principal component m^{th} is constrained to be orthogonal to such directions. In the variance expression in this direction $\alpha_1 \cdot \dots \cdot \alpha_{m-1} = 0$. So maximising S subject to this condition and being a unitary vector $|\alpha| = 1$, or $S \alpha = \lambda$

$$\alpha_1^T S \alpha_1 = \lambda_1 \quad (7)$$

Hence, the principal component m^{th} can be found along with the eigenvalue m^{th} and it can be established that the variance equals the eigenvalue m^{th} when α_1 is aligned to the direction of the m^{th} principal component (Bishop, 2006; Jolliffe, 2002).

In the literature correlation and covariance matrices can be presented as alternatives. To be completely accurate, the covariance matrix is the mean scalar product of patterns minus the mean, while the correlation matrix is a standardized version of the covariance matrix, given that the correlation originates from the mean scalar products of the patterns divided by the product of the standard deviation of patterns (Field, 2005). When this kind of analysis is performed from centred data ($\sum_{i=1}^m x_i = 0$) both matrices are equivalent.

Principal component analysis has been shown to be a very powerful technique for finding orthogonal derived variables that in succession maximise the variance of a given data set (Jolliffe, 2002; Mardia et al., 1979). However, sources of nonlinearities and complexities in real-world problems might require to be hypothesised in sub-spaces much richer than a linear combination of features (Cristianini & Shawe-Taylor, 2000). Therefore, nonlinear generalisations of principal components analysis play an important role in pattern analysis through the inclusion of kernel functions.

PCA has performed well in previous studies related to farming systems, especially for dimensionality reduction and for interpreting multiple crop-livestock signals (Köbrich et al., 2003). However, crop-livestock system variables interact in a non-linear dynamic, which in turn usually produces complex outcomes of landscape heterogeneity, livestock activity, and vegetation interactions. In consequence, most of these crop-livestock systems traits are subject to limited description within the second order correlation approach of linear PCA. One

solution to this problem is the generalisation of linear PCA setting to an application of kernel principal component analysis (KPCA) (Schölkopf et al., 1998). This algorithm combines the simplicity of linear PCA with the capability of integral operators, known as kernel functions; to express data from input space as dot products in the feature space. This method enables the construction of nonlinear versions of the original variables in a high dimensional context (Shawe-Taylor & Cristianini, 2006).

3.1 Coping with non linearities

The kernel “trick” permits the generalising of any algorithm that uniquely depends on inner products (Aizerman et al., 1964). This approach has proven to be particularly helpful for those statistical problems that involve feature extraction (Schölkopf et al., 1998); classification (Boser et al., 1992); regression (Williams, 1998) and clustering (Crammer & Singer, 2002; Graepel & Obermayer, 1998). Generally it can be said that kernel methods serve to induct non-linear functions in feature spaces usually of high dimensionality, and also may be incorporated into the dual form of most algorithms in such a way that it is not necessary to calculate explicitly the transformation to the feature space (Shawe-Taylor & Cristianini, 2006).

A result of the inclusion of the kernel idea within the dual representation, is that the computation task is not affected by the feature space dimensionality (Cristianini & Shawe-Taylor, 2000), and given that the gram-matrix is the unique information used in the feature space, the amount of work required to calculate the inner product is not necessarily proportional to the feature number. Thus the use of kernels can be seen as a means to establish an implicit correspondence between the original data and the feature space, without the limitations associated with the computation of such correspondence.

Within a broad context, the study of statistical aspects of pattern analysis has been approached from two main paradigms: the Bayesian approach (Duda et al., 2001) and empirical processes (Vapnik, 1995). Boser et al. (1992) pioneered the merging of kernel methods and statistical learning theory (empirical processes approach) through large margin classifiers. However, most of the theoretical development on kernel methods has its origin in the research of Mercer (1909) and Aronszajn (1950) where fundamental issues of Mercer’s theory and Hilbert’s spaces were treated respectively. After the crisis of the main linear approaches commonly used in the learning machine field (Fisher, 1936; Rosenblatt, 1958) as a result of the publication of Minsky & Papert (1969), one of the alternatives proposed was the threshold multilayer structures, which led to the development of neural networks (generalised perceptron) with associated algorithm as back propagation (Hertz et al., 1991).

The other approach was data pre-processing: in other words, the projection of data into a higher dimensional space to increase the computational power by including redundancies in their representation and assuring an effective feature extraction process from very complex data. An interesting alternative method to accomplish the above task, was the use of kernel methods, whose functions and corresponding feature spaces theory derive from integral operators studies (Aronszajn, 1950; Berg et al., 1984; Sahitoh, 1988). The inclusion of these constructs into a nonlinear generalisation of principal components analysis was led by Schölkopf et al. (1998). One of the main achievements of the study was to express the feature extraction based on eigen-decomposition, as a process that pursues the finding of orthonormalized directions in a kernel-defined feature space by dual representation, along which data variability is maximised.

Nonlinear PCA might be expressed as an eigenvalue problem. Consider a feature space \mathcal{H} associated to the input space \mathbb{R}^m by a non-linear transformation:

$$\Phi : X \Rightarrow \mathcal{H}, \quad x \Rightarrow \Phi(x) \quad (8)$$

The feature space \mathcal{H} can show an arbitrarily large dimensionality ($m \times m$), that is potentially infinite. Assuming that in this space data are centred according to $\sum_{i=1}^m x_i = 0$, the covariance matrix can be written in \mathcal{H} as following:

$$Cov = \frac{1}{p} \sum_{i=1}^p \Phi(x^i) \Phi(x^i)^T \quad (9)$$

Having a feature space that possesses infinite dimensions, $\Phi(x^j) \Phi(x^j)^T$ can be considered the linear operator in \mathcal{H} that performs the transformation $x \Rightarrow \langle \Phi(x^j) \Phi(x^j)^T \cdot x \rangle$. The main objective then consists of finding the solution to an eigenvalue problem that satisfies $\lambda v = Cov v$, without working explicitly in the feature space. By analogy to the input space analysis, all solutions v with $\lambda \neq 0$ are encountered in the sub-space generated by $\Phi(x^1), \dots, \Phi(x^p)$. This includes two helpful implications:

1. The following equation can be used:

$$\lambda \langle \Phi(x^n) \cdot v \rangle = \langle \Phi(x^n) \cdot Cov v \rangle \quad \forall n = 1, \dots, p \quad (10)$$

2. Provided that $\lambda \geq 0$ are found subject to the existence of non null eigenvectors $v \in \mathcal{H} \setminus \{0\}$; and given that coefficients belonging to $\alpha_i (i = 1, \dots, p)$ are determined by linear combinations of $\Phi(x^n)$, v can be written as:

$$v = \sum_{i=1}^p \alpha_i \Phi(x^i) \quad (11)$$

These expressions can be merged by substituting both into $\lambda v = Cov v$ and multiplying both sides by $\Phi(x)^T$ in order to express them as kernel terms $K(x^i, x^j) = \Phi(x^i)^T \Phi(x^j)$:

$$\lambda \sum_{i=1}^p \alpha_i \langle \Phi(x^n) \cdot \Phi(x^i) \rangle = \frac{1}{p} \sum_{i=1}^p \alpha_i \left\langle \Phi(x^n) \cdot \sum_{i=1}^p \Phi(x^j) \langle \Phi(x^j) \cdot \Phi(x^i) \rangle \right\rangle \quad (12)$$

$$\forall n = 1, \dots, p$$

which in terms of the matrix (Gram $p \times p$) notation, integrated by the elements $K_{ij} = \langle \Phi(x^i) \cdot \Phi(x^j) \rangle$, the equation for all n are consolidated in:

$$p\lambda K\alpha = K^2\alpha \quad (13)$$

where α represents the column vector integrated by elements $\alpha_1, \dots, \alpha_p$. Finding solutions to the previous equation requires an eigenvalue problem to be solved:

$$p\lambda\alpha = K\alpha \quad \forall \lambda \neq 0 \quad (14)$$

It can be demonstrated that this simplification (removing K from both sides) leads to (14) without those K that showed zero eigenvalues, not affecting the projection of principal components and bringing all useful solutions from (13). So if $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ are the eigenvalues of K ($p\lambda$ solutions) and $\alpha^1, \dots, \alpha^p$ the whole corresponding eigenvectors set, being λ_q the last non-zero eigenvalue (assuming that Φ is not identically 0). The condition of unitary norm ($\langle v^n \cdot v^n \rangle = 1$) for corresponding vectors in the feature space leads to the following solution of normalisation over $\alpha^1, \dots, \alpha^q$ when (11) and (14) are used:

$$\begin{aligned}
 1 &= \sum_{i,j=1}^p \alpha_i^n \alpha_j^n \langle \Phi(x^i) \cdot \Phi(x^j) \rangle = \sum_{i,j=1}^p \alpha_i^n \alpha_j^n K_{ij} \\
 1 &= \langle \alpha^n \cdot K \alpha^n \rangle = \lambda_n \langle \alpha^n \cdot \alpha^n \rangle
 \end{aligned}
 \tag{15}$$

The principal components projections can be calculated by projecting an x test point with an image $\Phi(x)$ onto eigenvectors v in the feature space with $n = 1, \dots, q$; and expressing them in kernel notation using (11); that way principal components can be extracted:

$$\langle v^n \cdot \Phi(x) \rangle = \sum_{i=1}^p \alpha_i^n \langle \Phi(x^i) \cdot \Phi(x) \rangle
 \tag{16}$$

These are the non-linear principal components or features corresponding to Φ (Bishop, 2006; Schölkopf et al., 1998).

To illustrate the above descriptions, differences in performance between linear (LPCA) (Hotelling, 1933a,b; Pearson, 1901) and kernel principal component analysis (KPCA) (Schölkopf et al., 1998) will be depicted in the following lines, based on the effectiveness of extracted features to yield meaningful and compact farm groups (dependent variable) within unsupervised classification by hierarchical clustering procedures (Johnson, 1967; Ward, 1963), using as few principal components as possible. For the purpose of this illustration, meaningful groups were defined as those clusters whose means were significantly different from each other, showing strong similarities within groups and possessing high variability between groups. Such estimations were based on a discriminant analysis approach (Fisher, 1936) using the statistics of Wilks' lambda ($W\lambda$), Hotelling's test (T^2), Pillai's trace test (P); Roy's maximum root (RM); and average squared canonical correlation (r^2) using data from farming systems located in the central plains of Venezuela.

An example of a comparison between the best performing configuration of kernel methods and the linear approach whose feature extraction required six principal directions is presented in Table 1. The profiles of clustering performance after discriminant analysis for the Gaussian kernel show that means of farm classes of the selected variables were different in the population given the closeness of Wilks' lambda statistic to zero and comparatively higher values of the Pillai, Hotelling and Roy tests with respect to the linear and polynomial approaches. Also, classification based on Gaussian feature extraction, showed higher average squared canonical correlations (r^2) supporting the idea of well separated groups accounting for a high percentage (69%) of the total variance explained.

The percentage of farms classified correctly was slightly higher when feature extraction was performed using polynomial kernels compared to inserting linear and Gaussian kernels. However, this feature extraction method did not provide enough information to find directions in the feature space along which farm groups were as well separated as with the Gaussian kernel. Even so, its performance was much better than classification based on linearly extracted feature vectors.

Kernel	%C	$W\lambda$	PT	T^2	RM	r^2
Linear	88.3	0.15	1.21	3.36	2.30	0.60
Gaussian	90.3	0.09	1.38	3.50	2.43	0.69
Polynomic	91.5	0.11	1.31	3.83	1.94	0.65

%C: percentage classified correct; $W\lambda$: Wilks' lambda; PT : Pillai's trace; T^2 : Hotelling's test; RM : Roy's minimum root; r^2 : squared average canonical correlation

Table 1. Impact of kernel function on clustering performance using linear, Gaussian and polynomic approaches of feature extraction, after stepwise discriminant analysis for a group of farms in Venezuela.

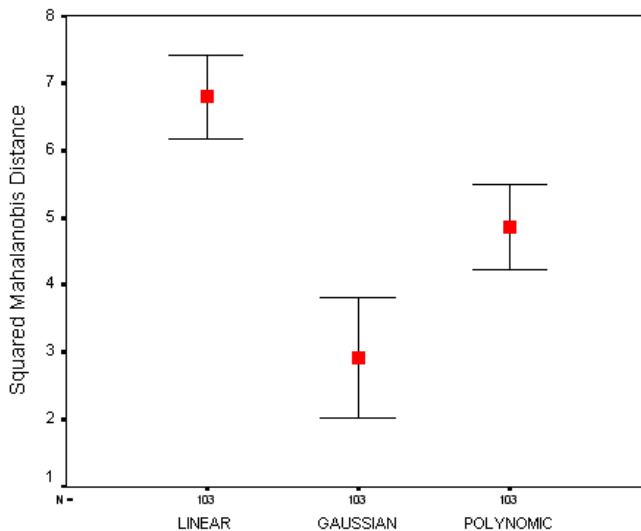


Fig. 1. Adjusted means and confidence intervals (95%) of squared Mahalanobis distance by selected feature extraction methods (linear, gaussian and polynomic) for a group of farms in Venezuela, after stepwise discriminant analysis.

Within canonical discriminant analysis, if a farm belongs to a particular class, it must fulfill some distance constraints with respect to the centroid of its class and projections of these groups onto some discriminant direction are expected to be compact and to show minimum overlaps. Hence, an easy way to assess the compactness of a given class is to look at the proximity of an observation set to its class-centroid. A visual approximation of these differences can be seen in Fig.1, where squared adjusted means of the Mahalanobis distance and their respective confidence intervals (95%) are shown for each feature extraction method. As can be observed, clusters segmented from feature vectors extracted by the linear approach and the Gaussian kernel were shown to be comparatively more scattered with respect to the clusters achieved from the polynomic feature extraction method, which showed a higher proximity (minimum distance) between a within-class object and its cluster centroid.

This effect is illustrated in Fig. 2, where farm objects are projected onto their first three principal directions with different levels of class overlap for the three feature extraction methods used. Only one of the three algorithms (Gaussian kernel) leads to a classification model that describes in a suitable way (without overlapping) the groups suggested by the instances cloud. The linear and polynomic-kernel methods were completely ineffective for cluster separation. This is mainly due to the topology of the sample covariance matrix as a result of the effect that the feature extraction method had on class-object component coordinates.

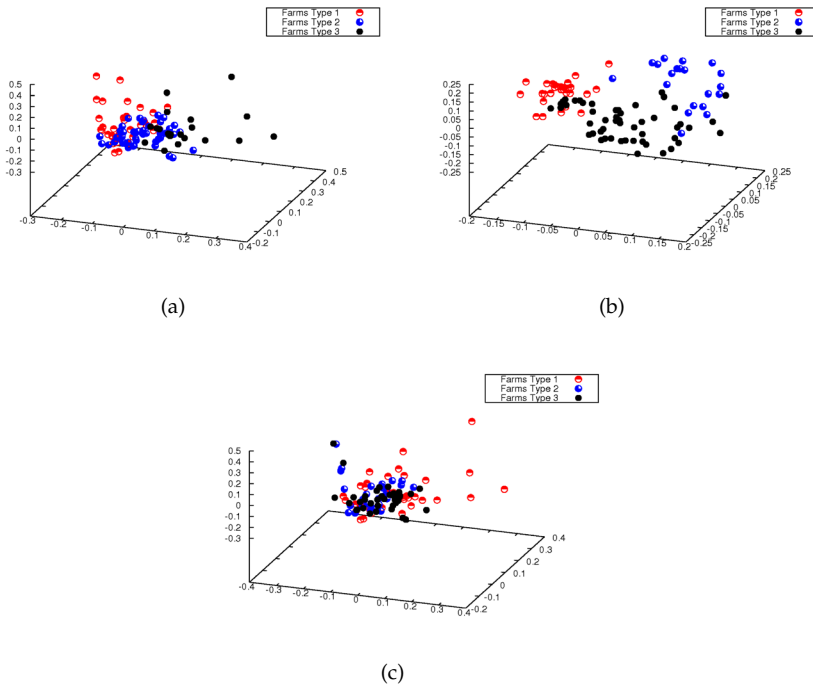


Fig. 2. Projection onto the three first principal components by farm class for linear (a), Gaussian (b), and polynomic kernel (c), feature extraction approaches.

4. Kernel supervised classification of multispectral data

Once farm labels (informational classes) have been generated as illustrated in previous section, multispectral data can be used to perform supervised classification of farms' spectral responses. Traditionally, multispectral data, such as those from the Landsat series of satellites, have been used for mapping geology, geobotany, forestry, agriculture, soil and land cover. They have rarely been used to identify continuous pixel groups integrated in a class such as a farm, which is a mosaic of land covers. However, kernel methods coupled to a maxim-margin classifier can achieve the difficult task of discriminating farm types using their land cover spectral response as recorded in a satellite image as indicators. The resultant representation is flexible, uniform over the pattern presented, and preserves the topology of the input space.

4.1 Classification

Spatial land cover classification has been mainly approached through the following paradigms: maximum likelihood classifier (MLC) (Strahler, 1980); fuzzy clustering (Kosko & Isaka, 1993); and artificial neural networks (ANN) (Miller et al., 1995). However, farm classes are abstractions which are sometimes difficult to observe directly, and this leads to a number of limitations of these methods. For instance, MLC methods are not free from distribution assumptions, given their parametric premises. Fuzzy clustering represents the solutions in terms of probabilities, where both fuzzy rules and membership functions are subjected to the bias of the interpreter. The ANN method has theoretical weaknesses because of its black box character, preventing the proper repeatability of the results. The presence of local minima and of the time-consuming training process (referred to as lack of convergence) are also significant limitations.

There are two main practical approaches to induce linear classifier parameters; on the one hand are those methods based on modelling conditional density functions (generative models) such as: linear discriminant analysis (Fisher, 1936; Lachenbruch, 1975) and Naive Bayes Classifier (Domingos & Pazzani, 1997); on the other hand, there are those that pursue the maximization of the outputs quality over a training set (discriminative models). These devices include: logistic regression (Hosmer & Lemeshow, 2000), the perceptron (Rosenblatt, 1958) and support vector machine (Vapnik, 1995; Vapnik & Chernovenkis, 1974). The main characteristic of support vector machines is that they seek to find a maximal margin hyperplane (Fig. 3). This is achieved using optimization procedures that can place severe computational demands. These problems were central to developing the kernel-adatron method which takes advantage of the adatron simplicity (Anlauf & Biehl, 1989), generalizing it to operate in a high dimensional feature space by the introduction of kernel functions. It solves the optimization problem of the Lagrangian formalism performing the margin-maximization through the application of a gradient ascent algorithm, resulting in an enhanced capability to learn nonlinear boundaries with a rate of convergence that is exponentially fast.

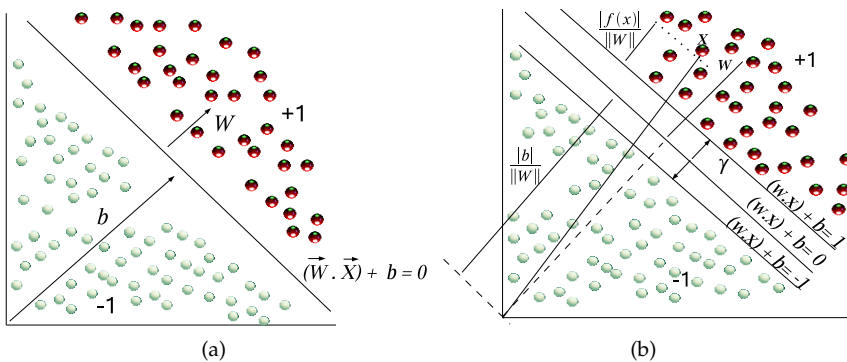
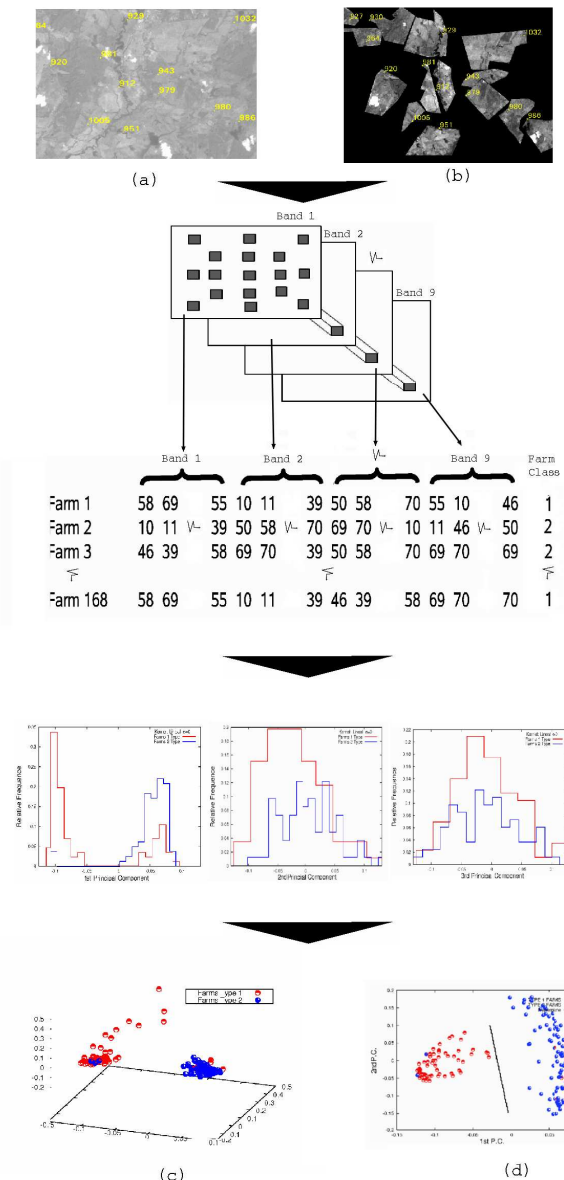


Fig. 3. Toy example of decision boundaries between classes; modified from Cristianini & Shawe-Taylor (2000).

Practical applications of these approaches can be addressed through the “hybrid” algorithm known as the kernel-adatron, first proposed by Friest et al. (1998). This uses a classifier that is based on a linear decision function whose estimated output is given by $y = f(\vec{x} \cdot \vec{w}) =$



From a Landsat (ETM+) image (a) subset are created (b), following as clipping criterion the farms' perimeters or boundaries.

The Sampling is performed within each farm subset. The nine multi-band dataset is sampled producing 20 values over each spectral channel, resulting in 180 pixel values per farm instance.

After values are collected, a vector of 180 columns per farm is organised by the concatenation of the 20 pixel values coming from the nine spectral channels

The matrix of 180 columns is centered, and a KPCA performed in order to extract features and reduce dimensionality. Then two or three principal directions with potential for class separation are selected.

If data plotted in the subspace spanned by the selected principal directions (c), show that samples can be linearly separated with minimum errors; a linear machine is then trained with this data in order to find a separating hyperplane (d)

Fig. 4. Landsat image segmentation procedure.

$f(\sum_i x_i w_i)$, where \vec{x} represents the input feature vector to the classifier; \vec{w} is the vector of weights defining the separating boundary and f is a function that projects input values x on w . In this way input patterns are linearly separated by dividing the input space with a hyperplane (Fig. 3).

Fig. 4 depicts a Landsat image segmentation procedure used in a learning machine classification context. As can be seen, the nine multi-band raster dataset is sampled producing a collection of pixel values over each band, following an amplified von Neumann vicinity in a pre-selected area of interest within the farm's perimeter. This training data set was used as input to a dimension reduction procedure, using principal component analysis with kernel (KPCA).

Using kernels to learn potential nonlinear representation hypotheses based on the function of the form $f(x) = \sum_i^n \alpha_i y_i K(x_i, x) + b$, essentially involves the simulation of the nonlinear projection of the input data in a higher dimensional space (Schaback & Wendland, 2006):

$$\begin{aligned} \Phi : S \in \mathbb{R}^d &\rightarrow \mathcal{F} \in \mathcal{H} \\ x &\mapsto \Phi(x) \end{aligned} \tag{17}$$

where \mathcal{F} denotes a feature space; and, \mathcal{H} represents a dot product space, within which, a learning relationship could be induced between a pattern $\Phi(x)$ and a label y . In this way, having as theoretical context Mercer's theorem (Aizerman et al., 1964; Mercer, 1909); (18) represents the kernel matrix, where each entry is a measure of similarity between two objects. Thus, a symmetric function $K(x_i, x)$ was a kernel if it fulfilled Mercer's condition, i.e. the function K is (semi) positive definite. When this is the case there exists a mapping ϕ such that it is possible to write $K(\mathbf{x}, \mathbf{y}) = \langle \phi(x_i) \cdot \phi(x) \rangle$.

$$K(x_i, x) \triangleq \langle \phi(x_i) \cdot \phi(x) \rangle \Rightarrow \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & \dots \\ K(x_2, x_1) & \ddots & \\ \vdots & & \end{bmatrix} \tag{18}$$

The kernel represents a dot product on a feature space \mathcal{F} into which the original vectors were mapped (Fig. 5). In this way a kernel function defines an embedding of memory patterns into (high or infinite dimensional) feature vectors and allows the algorithm to be carried out in this space without the need to represent it explicitly (Cristianini & Shawe-Taylor, 2000; Schölkopf & Smola, 2002). Further details on the way this procedure was implemented are outside the scope of this paper. Nevertheless, for those seeking deeper understanding of the ideas behind kernel-based learning theory there are fuller descriptions in Aizerman et al. (1964); Aronszajn (1950); Mercer (1909) and Schölkopf & Smola (2002). Applications of kernel methods and learning machines may also be reviewed in García & Moreno (2004a,b,c)

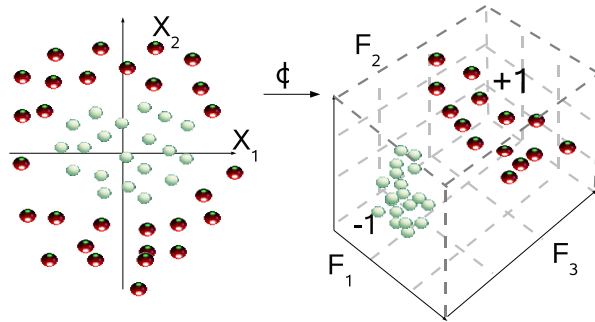


Fig. 5. Toy example illustration of the effect of mapping a simple binary problem to a higher dimensional feature space on the ability to separate complex relations.

The basic KA algorithm is a binary classifier that makes use of an optimization procedure based on the descent gradient to find the maxim-margin hyperplane that separates two groups. For the classification of farms from a multi-class problem (existence of three or more informational categories), a one against the rest strategy might be adopted. Basically, three machines (one per each class) can be trained organized in such an assembly that the class of interest is compared against the other two (Fig. 6).

Table 2 presents the performance accuracy of the three KA machines trained for an experimental group. As can be seen the KA appears to be more sensitive for class 1, given the highest accuracy reached, and its degree of overlap seems to be with class 3. This may be explained by the levels of farming intensification observed in farm class 1, with an important degree of fragmentation of the land cover mosaic, which probably facilitated its differentiation from those instances that resemble the more natural scenes typical of the less intensive farm classes 2 and 3 (Drury, 2001). The tendency to wrongly allocate farm type 3 as class 1 might be because these groups of farms share similar attributes in their proportions of pasture, forage and forest cover. Misclassification between classes 2 and 3 can be explained by the lack of anthropogenic changes leading to the occupation of less discrete areas of the feature space as a function of the natural environment context (Richards and Jia, 2006; Landgrebe, 2007). In this kind of study, farms are seen as bags of pixels representing different land covers in a space where each dimension is associated with a spectral channel. Because this vector space was sensibly transformed by non linear feature extraction to improve representation, and with this to ensure equivalent land covers mapped to similar feature vectors, it is possible to reach an acceptable level of accuracy with this approach.

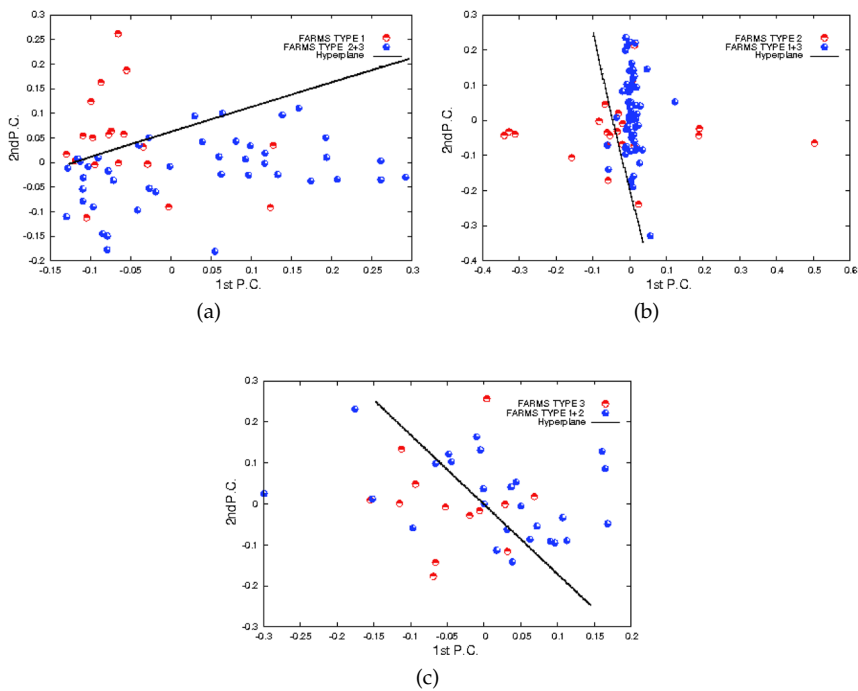


Fig. 6. Separating hyperplanes for farms class 1 (a) and 3 (c) using a Gaussian kernel ($\sigma = 200$), and class 2 (b) using a polynomial kernel (order= 3; $\sigma = 4$).

KA		Predicted				Σ	Accuracy (%)
		Class 1	Class 2	Class 3			
Actual	Class 1	35	0	4	39	89.8	
	Class 2	2	34	6	42	80.95	
	Class 3	3	2	17	22	77.27	
	Σ	40	36	27	103		
Accuracy (%)		87.5	94.44	62.96		Overall Accuracy (%) 83.49	

KA: Kernel Adatron

Table 2. Confusion matrix for the segmentation of three farm categories trained on 14, 20, and 16 cases for class 1, 2, and 3 respectively using the KA machine.

Finding these separating decision functions on the segmentation of farm classes is particularly significant given the non-stationary spatial behaviour of the spectral response of this kind of object; and because of the small training set size with respect to the dimensionality of the input space. Another important consideration is that this approach only focuses on extreme samples for its training, making possible the derivation of comparable levels of performance at a lower cost. This fact would confirm the argued advantages of previous applications of kernel methods in the land use domain, in which decision functions have been induced without any other *a priori* knowledge about the land cover than labels (Huang et al., 2002; Zhu & Blumberg, 2002). This implies a considerable resource saving in practical application to livestock systems monitoring.

4.2 The multispectral data

The use of multispectral data to distinguish one type of land cover from another, has been an effective way of linking anthropomorphic intervention to a physical environment, particularly within the agricultural sector (Campbell, 2002). For instance, Wylie et al. (2002) combined optical and thermal data to estimate biophysical properties of vegetation. Other approaches use the land cover mosaic, to induct farm typologies based on their relative spectral similarities, as in the case of Duvernoy (2000). The popularity of using visible and near infrared (VIR) imagery on the classification of areas covered by agricultural activities, is because plant cell structures, morphology, chlorophyll and other pigments have a marked effect on this wavelength range (Drury, 2001), and on the temperature brightness of thermal infrared (TIR) radiation incident on living plants (Rees, 2007).

The configuration of multispectral sensors, such as Landsat 7 Enhanced Thematic Mapper Plus (ETM+), is particularly well suited to perceive the energy field, in the form of VIR and TIR radiation emanating from vegetation covers (Richards & Jia, 2006). This feature makes many multispectral data sensible to spatial patterns tied to crop calendars, and vegetative growth-lessening as a result of phenophases (Campbell, 2002; Richards & Jia, 2006). For instance, the spectral bands per pixel in Landsat 7 are delineated by six VIR bands, where band 6 is split into two channels defined by filters that control the radiance that reach the sensor; and a panchromatic band (Barsi et al., 2003; Heckenlaible et al., 2007). These radiometric features make Landsat 7 a good choice within the context of farming system research at household resolution level. The precision to which this sensor registers the radiation power, for a particular pixel in a given wavelength is 8 bits (256 levels) (Richards & Jia, 2006). This feature enhances the ability of the sensor to distinguish the spectral responses from different materials, when human-scale factors such as agriculture need to be addressed (Campbell, 2002; Landgrebe, 2007).

As with radiometric resolution, the spatial resolution of Landsat 7, which ranges from 15 to 60 meters per pixel across all the spectral bands, is rich (small or fine) compared to farms, which are the usual objects of study in farming systems research and where a pixel smaller than the agricultural field to be studied is usually preferred (Landgrebe, 2007). To these spatial characteristics of Landsat, should be added its scanning features, whose cover swath is 185 km², which means that each scene sample observes an area of 34.225 km². Such an overlay represents an advantage for farming system research given the scale of the typical study area (10.000 km²), and because the whole can be extracted from one image. However depending on the size of the farms under study misclassification risk might occur, from the impact that spatial resolution has on the separability of informational classes (Landgrebe, 2007). Spatial resolution has been shown to have a significant influence on spectral class separability because

of the hierarchy that generally characterizes informational categories (Campbell, 2002; Rees, 2007); and there is reason to believe that similar effects occur with collections of land cover such as farms (Landgrebe, 2007). For studies of farming systems, the spatial resolution of Landsat 7 (ETM+) data, might be too fine for the purposes of classification, in the sense that sometimes it is desirable to have pixel sizes smaller, but not excessively smaller, than the field under study, because too fine a resolution may lead to pixels that spectrally do not represent the field of interest but only part of it. In farm classification, most of the time interest is focussed on pixels that integrate across what is desired to be called a field, which in this study would be a farm, rather than a small part of a particular cover of crop, grassland or forest.

From that viewpoint, an alternative possibility is to use a source of data with a coarse spatial resolution, such as the Moderate Resolution Image Spectrometer (MODIS) (NASA, 2008). This sensor is one of the principal instruments aboard EOS¹ AM-1 (TERRA); and its spatial resolution ranges from 500m to 1 km, with a viewing swath width of 2.330 km. The possibility that the use of this sensor would lead to an improvement in farm classification accuracy, is in line with the reviews of Landgrebe (2007) and Drury (2001) in the sense that compared with Landsat 7, each pixel in MODIS would be made up of a mixture of "Landsat-size" pixels on categories such as grass and crops that may lead to an improved representation of a farm as a field of interest. The advantages of MODIS are not restricted to its spatial resolution; its spectral resolution, 36 channels covering from visible to thermal infrared spectral regions, also presents some benefit compared to the 7 bands of Landsat. This spectral richness should increase the accuracy of discrimination of complex classes, because of the high volume space. Evidence for the significance of spectral resolution on discrimination accuracy comes from Bazi & Melgani (2006); Foody & Mathur (2004); Melgani & Bruzzone (2004) and Muñoz-Marí et al. (2007). They exploit high spectral resolution sources, spreading the data out as much as possible in the feature space to make the most of the spectral richness, that generally results in small classification errors.

5. References

- Aizerman, M., Braverman, E. & Rozonoer, L. (1964). Theoretical foundations of the potential function method in pattern recognition learning, *Automations and Remote Control* **25**: 821–837.
- Anlauf, J. & Biehl, M. (1989). The adatron: an adaptive perceptron algorithm, *Europhysics Letters* **10**: 687–692.
- Aronszajn, N. (1950). Theory of reproducing kernels, *Trans. Amer. Math. Soc.* **68**: 337–404.
- Asrar, G. & Dozier, J. (1994). *EOS-science strategy for the Earth observing system*, AIP Press, Woodbury, NY.
- Barsi, J., Schott, J., Palluconi, F., Helder, D., Hook, S., Markham, B., Chander, G. & O'Donnell, E. (2003). Landsat tm and etm+ thermalband calibration, *Canadian Journal of Remote Sensing* **29**(2): 141–153.
- Bazi, Y. & Melgani, F. (2006). Toward an optimal SVM classification system for hyperspectral remote sensing images, *IEEE Transaction on Geoscience and Remote Sensing* **44**(11): 3374–3385.
- Berdegue, J. & Escobar, G. (1990). Conceptos y metodologías para la tipificación de sistemas de finca, *Tipificación de Sistemas de Producción Agrícola*, RIMISP, Santiago de Chile, pp. 13–43.

¹ Earth Observation System

- Berg, C., Christensen, J. & Ressel, P. (1984). *Harmonic analysis on semigroups*, Springer-Verlag, New York, USA.
- Bishop, C. (2006). *Pattern recognition and machine learning*, Springer, Singapore.
- Boser, B., Guyon, M. & Vapnik, V. (1992). A training algorithm for optimal margin classifiers, *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory* pp. 144–152.
- Bouwman, A., der Hoek, K. V., Eickhout, B. & Soenario, I. (2005). Exploring changes in world ruminant production systems, *Agricultural Systems* **84**(2): 121–153.
- Campbell, J. (2002). *Introduction to remote sensing*, Taylor & Francis, London, UK.
- Capillon, A. (1985). Connaitre la diversité des exploitations: un préalable à la recherche de références techniques regionales, *Agriscopie* **6**: 31–40.
- Crammer, K. & Singer, Y. (2002). Pranking with ranking, in T. Dietterich, S. Becker & Z. Ghahramani (eds), *Advances in Neural Information Processing Systems*, Vol. 14, MIT Press, pp. 641–647.
- Cristianini, N. & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*, Cambridge University Press, Cambridge, UK.
- Delgado, C., Rosengrant, M., Steinfeld, H., Ehui, S. & Courbois, V. (1999). Livestock to 2020. the next food revolution, *Food, Agriculture and the Environment Discussion Paper 28*, International Food Policy Research Institute, Food and Agriculture Organization of the United Nations, International Livestock Research Institute, Washington, DC.
- Diez, J., Coz, J., Bahamonde, A., nudo, C. S., Olleta, J., Macie, S., Campo, M., Panea, B. & Alberti, P. (2006). Identifying market segments in beef: breed, slaughter weight and ageing time implications, *Meat Science* **74**(4): 667–675.
- Dixon, J., Gulliver, A. & Gibbon, D. (2001). *Farming systems and poverty. Improving farmer's livelihoods in a changing world*, FAO and World Bank, Malcolm Hall, chapter 1st.
- Dobremez, L. & Bousset, J. (1995). *Rendre compte de la diversité des exploitations agricoles. Une démarche d'analyse par exploration conjointe de sources statistiques, comptables et technico-économiques*, Cemagref, France.
- Domingos, P. & Pazzani, M. (1997). On the optimality of the simple bayesian classifier under zero-one loss, *Machine Learning* **29**: 103–137.
- Drury, S. (2001). *Image Interpretation in Geology*, 3rd edn, Nelson Thornes Ltd, Cheltenham, UK.
- Dubayah, R. (1992). Estimating net solar radiation using landsat thematic mapper and digital elevation data, *Water Resources Research* **28**: 2469–2484.
- Duda, R., Hart, P. & Stork, D. (2001). *Pattern classification*, 2nd. edn, John Wiley & Sons, INC, New York, USA.
- Duvernoy, I. (2000). Use of land cover model to identify farm types in the Misiones agrarian frontier (Argentina), *Agricultural Systems* (64): 137–149.
- Field, A. (2005). *Discovering statistics using SPSS*, 2nd edn, SAGE Publications, London, UK.
- Fisher, R. (1936). The use of multiple measurements in taxonomic problems, *Annals of Eugenics* **7**: 179–188.
- Foody, M. & Mathur, A. (2004). Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification, *Remote Sensing of Environment* **93**: 107–117.
- Friest, T., Campbell, C. & Cristianini, N. (1998). *The kernel-adatron: A fast and simple learning procedure for support vector machines*, In: Proceedings of the Fifteenth International Conference on Machine Learning, Morgan-Kaufmann, San Francisco, USA.

- García, C. & Moreno, J. (2004a). *The hopfield associative memory network: Improving performance with the kernel "trick"*, In: C. Lemaître; C.A. Reyes and J.A. Gonzalez (Eds). IBERAMIA 2004, LNAI 3315, Springer-Verlag, Berlin, Germany, pp. 871–880.
- García, C. & Moreno, J. (2004b). *Kernel based method for segmentation and modeling of magnetic resonance images*, In: C. Lemaître; C.A. Reyes and J.A. Gonzalez (Eds). IBERAMIA 2004, LNAI 3315, Springer-Verlag, Berlin, Germany, pp. 636–645.
- García, C. & Moreno, J. (2004c). *The kernel hopfield memory network*, In: P.M.A. Sloot; B. Chopard and A.G. Hoekstra (Eds). ACRI 2004, LNCS 3305, Springer-Verlag, Berlin, Germany, pp. 755–764.
- Gleason, A., Prince, S., Goetz, S. & Small, J. (2002). Effects of orbital drift on land surface temperature measured by avhrr thermal sensors, *Remote Sensing of Environment* **79**: 147–165.
- Goel, N. (1987). Models of vegetation canopy reflectance and their use in estimation of biophysical parameters from reflectance data, *Remote Sensing Review* **3**: 1–212.
- González, A., Russell, G., Márquez, A., Moreno, J., García, C., Domínguez, C., Colmenares, O. & Machado, J. (2007). Supervised farm classification from remote sensing images based on the kernel adatron algorithm, *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS07)*, Barcelona, Spain.
- Graepel, T. & Obermayer, K. (1998). Fuzzy topographic kernek clustering, in W. Bauer (ed.), *Proceeding of the 5th GI Workshop Fuzzy Neuro Systems '98*, pp. 90–97.
- Green, R. & Hay, S. (2002). The potential of pathfinder avhrr data for providing surrogate climatic variables across africa and europe for epidemiological applications, *Remote Sensing of Environment* **79**: 166–175.
- Guyon, I. & Elisseeff, A. (2003). An introduction to variable and feature selection, *Journal of Machine Learning Research* **3**: 1157–1182.
- Hair, J., Anderson, R., Tatham, R. & Black, W. (1998). *Multivariate data analysis*, 5th edn, Prentice Hall, Upper Saddle River, New Jersey.
- Hart, R. (1990). Componentes, subsistemas y propiedades del sistema finca como base para un método de clasificación, *Tipificación de Sistemas de Producción Agrícola*, RIMISP, Santiago de Chile, pp. 45–61.
- Heckenlaible, D., Meyerink, A., Torbert, C. & Lacasse, J. (2007). Landsat 7 (L7) enhanced thematic mapper plus(ETM+ level zero-r distribution product (LORP) data format control book (DFCB), *Technical report*, Department of the Interior U.S. Geological Survey, Sioux Falls, South Dakota.
- Hermes, L., Friauff, J., Puzicha, J. & Buhmann, J. (1999). Support vector machines for land usage classification in landsat tm imagery, *Geoscience and Remote Sensing Symposium*, Vol. 1 of IGARSS 99, IEEE international, pp. 348–350.
- Hertz, J., Krogh, A. & Palmer, R. (1991). *Introduction to the theory of neural computation*, Addison-Wesley.
- Hosmer, D. & Lemeshow, S. (2000). *Applied Logistic Regression*, 2nd edn, Wiley, New York.
- Hotelling, H. (1933a). Analysis of a complex of statistical variables into principal components, *The Journal of Educational Psychology* **24**(6): 417–441.
- Hotelling, H. (1933b). Analysis of a complex of statistical variables into principal components, *The Journal of Educational Psychology* **24**(7): 498–520.
- Huang, C., Davis, L. & Townshend, J. (2002). An assesment of support vector machines for land cover classification, *International Journal of Remote Sensing* **23**: 725–749.
- Johnson, S. (1967). Hierarchical clustering schemes, *Psychometrika* **32**(3): 241–254.

- Jolliffe, I. (2002). *Principal components analysis*, Springer, New York, USA.
- Keuchel, J., Naumann, S., Heiler, M. & Siegmund, A. (2003). automatic land cover analysis for tenerife by supervised classification using remotely sensed data, *Remote Sensing of Environment* **86**(4): 530–541.
- Kimes, D., Gastellu-Etchegorry, J. & Esteve, P. (2002). Recovery of forest canopy characteristics through inversion of a complex 3d model, *Remote Sensing of Environment* **79**: 320–328.
- Köbrich, C., Rehman, T. & Khan, M. (2003). Typification of farming systems for constructing representative farm models: two illustrations of the application of multi-variate analyses in chile and pakistan, *Agricultural Systems* **76**(1): 141–157.
- Kosko, B. & Isaka, S. (1993). Fuzzy logic, *Scientific American* **271**: 76–81.
- Kostrowicki, J. (1977). Agricultural typology concept and method, *Agricultural System* **2**: 33–45.
- Lachenbruch, P. (1975). *Discriminant analysis*, Hafner Press, New York, USA.
- Landgrebe, D. (2007). Multispectral thematic mapping of land areas, some fundamentals, *IEEE Geosciences and Remote Sensing Society Newsletter* (145): 11–15.
- Mardia, K., Kent, J. & Bibby, J. (1979). *Multivariate analysis*, Academic Press, London, UK.
- McVicar, T. & Jupp, D. (2002). Using covariates to spatially interpolate moisture availability in the murray-darling basing. a novel use of remotely sensed data, *Remote Sensing of Environment* **79**: 199–212.
- Melgani, F. & Bruzzone, L. (2004). Classification of hyperpectral remote sensing images with support vector machines, *IEEE Transactions on Geosciences and Remote Sensing* **42**(8): 1778–1790.
- Mercer, J. (1909). *Functions of positive and negative type and their connection with the theory of integral equations*, Philosophical Transactions of the Royal Society of London, London.
- Milá, M., Bartolomé, J., Quintanilla, R., García-Cachán, M., Espejo, M., Herráiz, P., Sánchez-Recio, J. & Piedrafita, J. (2006). Structural characterisation and typology of beef cattle farms of spanish wooded rangelands (dehesas), *Livestock Science* **99**: 197–209.
- Miller, D., Kaminsky, E. & Rana, S. (1995). Neural network classification of remote-sensing data, *Computers and Geosciences* **21**: 377–386.
- Minsky, M. & Papert, S. (1969). *Perceptrons: an introduction to computational geometry*, MIT Press.
- Muñoz-Marí, J., Bruzzone, L. & Camps-Valls, G. (2007). A support vector domain description approach to supervised classification of remote sensing images, *IEEE Transaction on Geosciences and Remote Sensing* **45**(8): 2683–2692.
- Myneni, R., Asrar, G. & Hall, F. (1992). A three-dimensional radiative transfer method for optical remote sensing of vegetated land surface, *Remote Sensing of Environment* **41**: 105–121.
- NASA (2008). Moderate Resolution Image Spectrometer (modis), *Online*. Retrieved April 25, 2008 from <http://modis.gsfc.nasa.gov/>.
- Pearson, K. (1901). On lines and planes of closets fit to points in space, *Phylosophical Magazine* **2**: 559–572.
- Rees, W. (2007). *Physical principles of remote sensing*, Cambridge University Press, Cambridge, UK.
- Richards, J. & Jia, X. (2006). *Remote sensing digital image analysis*, Springer-Verlag, Berlin, Germany.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Review* **65**(6): 386–408.

- Sahitoh, S. (1988). *Theory of reproducing kernels and its applications*, Logman Scientific & Technical, Harlow, UK.
- Schaback, R. & Wendland, H. (2006). *Kernel techniques: from machine learning to meshless methods*, Acta numerica (2006), Cambridge University Press.
- Schölkopf, B. & Smola, A. (2002). *Learning with kernels. Support vector machines, regularization, optimization, and beyond*, The MIT Press, Cambridge, Massachusetts, USA.
- Schölkopf, B., Smola, A. & Müller, K. (1998). Nonlinear component analysis as a kernel eigenvalue problem, *Neural Computation*. **10**: 1299–1319.
- Seré, C. & Steinfeld, H. (1996). *World livestock production systems: Current status, issues and trends*, FAO Animal Production And Health Paper, Rome, Italy.
- Shawe-Taylor, J. & Cristianini, N. (2006). *Kernel methods for pattern analysis*, Cambridge Univ. Press, Cambridge, UK.
- Song, C., Woodcock, C., Seto, K., Lenney, M. & Macomber, S. (2001). Classification and change detection using landsat tm data: When and how to correct atmospheric effects?, *Remote Sensing of Environment* **75**: 230–244.
- Strahler, A. (1980). The use of prior probabilities in maximum likelihood classification of remotely sensed data, *Remote Sensing of Environment* **10**: 135–163.
- Su, L., Chopping, M., Rango, A., Martonchik, J. & Peters, D. (2007). Support vector machines for recognition of semi-arid vegetation types using misr multi-angle imagery, *Remote Sensing of Environment* **107**(1-2): 299–311.
- Vapnik, V. (1995). *The nature of statistical learning theory*, Springer-Verlag, New York, USA.
- Vapnik, V. (1998). *Statistical learning theory*, John Wiley & Sons.
- Vapnik, V. & Chernovenkis, A. (1974). *Theory of pattern recognition*, Nauka, Moscow.
- Ward, J. (1963). Hierarchical grouping to optimize an objective function, *Journal of the American Statistical Association* **58**(301): 236–244.
- Williams, C. (1998). Prediction with Gaussian processes: From linear regression to linear prediction and beyond, in M. Jordan (ed.), *Learning and inference in graphical models*, Kluwer.
- Wint, W., Slingenbergh, J. & Rogers, D. (2000). Livestock distribution, production and diseases, *Towards a global livestock atlas*, Food and Agriculture Organisation of the United Nations, Rome. www.fao.org/ag/againfo/resources/en/glipha/default.html.
- Wylie, B., Meyer, D., Tieszen, L. & Mannel, S. (2002). satellite mapping of surface biophysical parameters at the biome scale over the north american grasslands, *Remote Sensing of Environment* **79**: 266–278.
- Yang, F., Ichii, K., White, M., Hashimoto, H., Michaelis, A., Votava, P., Zhu, A., Huete, A., Running, S. & Nemani, R. (2007). Developing a continental-scale measure of gross primary production by combining modis and americaflux data through support vector machine approach, *Remote Sensing of Environment* **110**(1): 109–122.
- Zhu, G. & Blumberg, D. (2002). Classification using ASTER data and SVM algorithms; the case study of Beer Sheva, Israel, *Remote Sensing of Environment* **80**: 233–240.

Multivariate Time Series Support Vector Machine for Multispectral Remote Sensing Image Classifications

Pei-Gee Peter Ho
*Naval Undersea Warfare Center, Newport Rhode Island
USA*

1. Introduction

Satellite and airborne Remote Sensing for observing the earth surface, land monitoring and geographical information systems are the big issues in world's daily life as well as country defense projects. The source of information was primarily acquired by imaging sensors and spectroradiometer in remote sensing multispectral image stack format. The traditional image processing either by single picture image processing or compressing pictures stack via Principle Component Analysis (PCA) or Independent Component Analysis (ICA) into a single image component for further pixel classification or region segmentation is not enough to describe the true information extracted from multispectral satellite sensors. In an effort to significantly improve the existing classification and segmentation performance in this research, the contextual information between pixels or pixel vectors is characterized by a time series model for the remote sensing image processing.

Time Series statistical models such as Autoregressive Moving Average (ARMA) were considered useful in describing the texture and contextual information of an remote sensing image. To simplify the computation, a two-dimensional (2-D) Autoregressive (AR) model was used instead. In the first phase, the 2-D univariate time series based imaging model was derived mathematically (Ho, 2008) to extract the features for further terrain segmentations. The effectiveness of the model was demonstrated in region segmentation of a multispectral image of the Lake Mulargias region in Italy. Due to the nature of remote sensing images such as SAR (Synthetic Aperture Radar) and TM (Thermal Mapper) which are mostly in multispectral image stack format, a 2-D Multivariate Vector AR (ARV) time series model with pixel vectors of multiple elements (i.e. 15 elements in the case of TM+SAR remote sensing) are examined. The 2-D system parameter matrix and white noise error covariance matrix are estimated for further classifications in the 2nd phase of algorithm development. To compute the time series ARV system parameter matrix and estimate the error covariance matrix efficiently, a new method based on modern numerical analysis is developed by introducing the Schur complement matrix, the QR (orthogonal, upper triangular) matrix and the Cholesky factorizations in the ARV model formulation. As for pixel classification, the powerful Support Vector Machine (SVM) kernel based learning machine is applied in

conjunction with the 2-D time series ARV model. The SVM is particularly suitable for the high dimensional vector measurement as the “curse of dimensionality” problem is avoided. The performance improvement over the popular Markov random field is demonstrated. The 2-D multivariate time series model is particularly suitable to capture the rich contextual information in single and multiple images at the same time. A single composite image is constructed from the vector pixels through ARV based Support Vector Machine classifications.

2. Remote Sensing Image Data Set

The remote sensing image data of the earth surface is from either satellite or aircraft in digital multispectral or hyperspectral format. Both multispectral and hyperspectral imaging techniques are the process of capturing the same scene at different wavelengths that yield a 2-D spatial dimensions and one spectral dimension hypercube. The main properties of a remote sensing image are the wavelength bands it interprets. As far as remote sensing physical phenomena is concerned, some are the measurements of the spatial information reflected by the solar radiation in terms of visible and ultraviolet frequency range of wave (Schowengerdt, 1997). This type of remote sensing is the passive type. Some are the spatial distribution of energy emitted by the earth in the thermal infrared. Others are in the microwave band of wavelengths, measuring the energy returns from the earth that was transmitted from the vehicle which is the active type of remote sensing (Richards, 1999). The remote sensing image data is based on the concept of the “spectral signature” of an object (Prieto, 2000). Therefore, different land covers have different spectral signatures. The system produces multi-spectral images stack that each pixel is represented by the mathematical pixel vector which contains a set of brightness values for the pixels arranged in column form:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_N \end{bmatrix}$$

Table 1. (Bruzzone, 1998) below lists examples of the spectral regions on earth remote sensing.

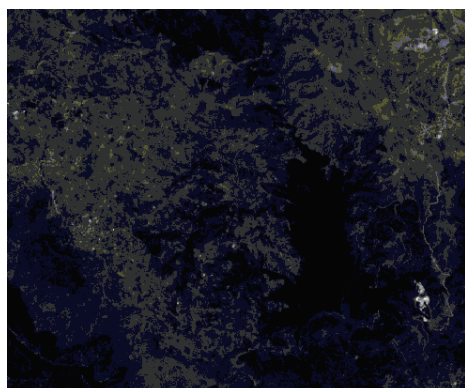
Spectral Signature	Wavelength	Radiation source	Surface
Visible	0.4-0.7 mm	Solar	Reflectance
Infrared	0.7-1.1 mm	Solar	Reflectance
Short Infrared	1.1-2.5 mm	Solar	Reflectance
Mid Infrared	3-5 mm	Solar and Thermal	Ref. and Temp.
Thermal Infrared	0.95 mm	Thermal	Temperature
Radar band Ka	0.8-1.1 cm	Thermal	Temperature

Radar band K	1.1-1.7 cm	Thermal	Temperature
Radar band Ku	1.7-2.4 cm	Thermal	Temperature
Radar band X	2.4-3.8 cm	Thermal	Temperature
Radar band C	3.8-7.5 cm	Thermal	Temperature
Radar band S	7.5-15 cm	Thermal	Temperature
Radar band L	15-30 cm	Thermal	Temperature
Radar band P	30-100 cm	Thermal	Temperature

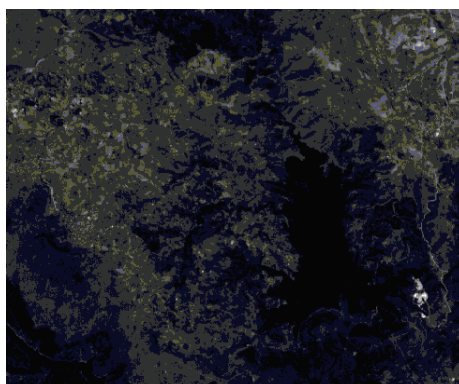
Table 1. The spectral signature of different bands used in remote sensing

Thermal Mapper (TM) Remote Sensing Images Data:

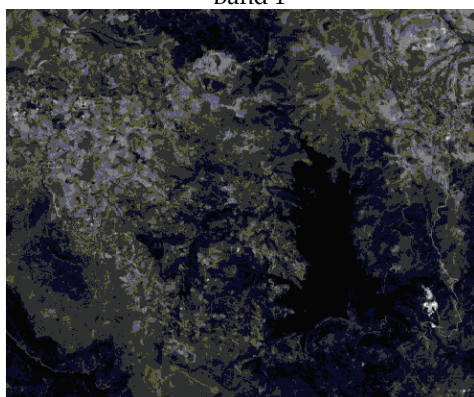
The Landsat earth resources satellite system was the first designed to provide global coverage of the earth surface via remote sensing techniques. Three imaging instruments have been used with this satellite. These are the Return Beam Vidicon (RBV), the Multi-spectral Scanner (MSS) and the Thermal Mapper (TM). Landsat uses multispectral instruments to record the remote sensing images stack. Figure 1. shows one example of 6 different bands of Thermal Mapper acquired by the Landsat 5 satellite in the area of the Mulargias lake in July 1996.



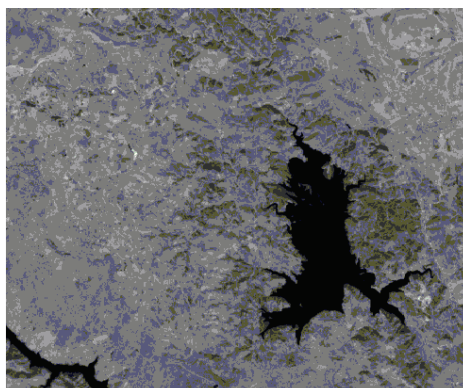
Band 1



Band 2



Band 3



Band 4

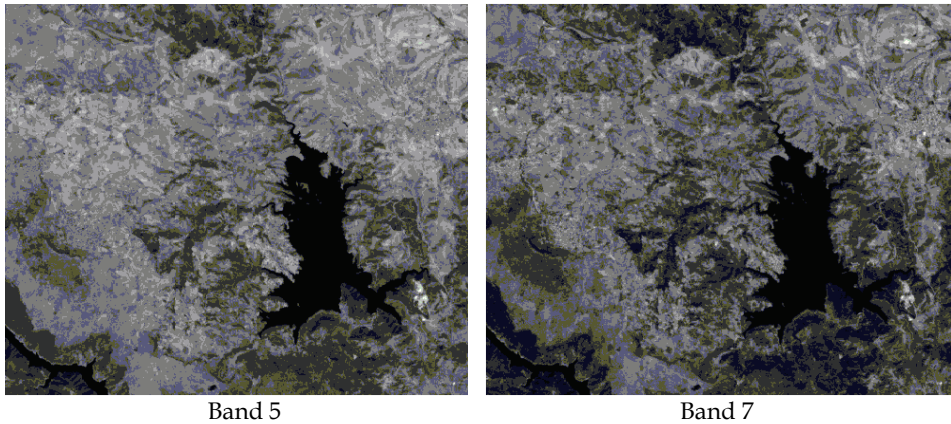


Fig. 1. Examples of Thermal Mapper

Synthetic Aperture Radar (SAR) Remote Sensing Images Data:

Active remote sensing techniques employ an artificial source of radiation. The resulting signal scatters back to the sensor which reflect atmosphere or earth characteristics. Synthetic Aperture Radar is an imaging technology that the radiation is emitted in a beam from a moving sensor. The backscattered components returned from the ground are measured. An image of the backscatter spatial distribution is reconstructed by digital processing of the amplitude and phase of the returned signal. Samples of SAR polarization images are shown in figure 2.

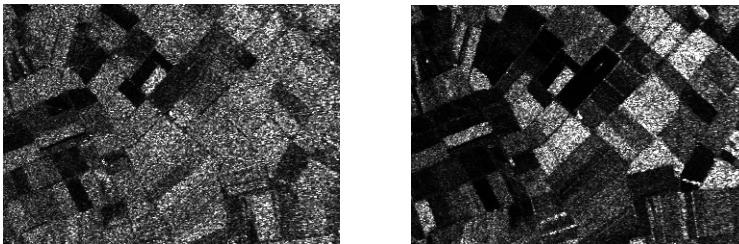


Fig. 2. Samples of Synthetic Aperture Radar Images (C-HV and L-HV band)

Though SAR images are popular due to its accessibility, the speckle noise is all over the places which degrades in the image quality.

Hyperspectral Remote Sensing Images:

The other type of multiple spectral images data are produced by spectrometers which is different from multippectral instruments. One example is the NASA's Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) optical sensor that delivers calibrated images of the upwelling spectral radiance in 224 contiguous spectral bands. The AVIRIS detector operates with a wavelength of 10 nanometers to cover the entire range. Figure 3 shows the AVIRIS hyperspectral sample images through different spectral bands.

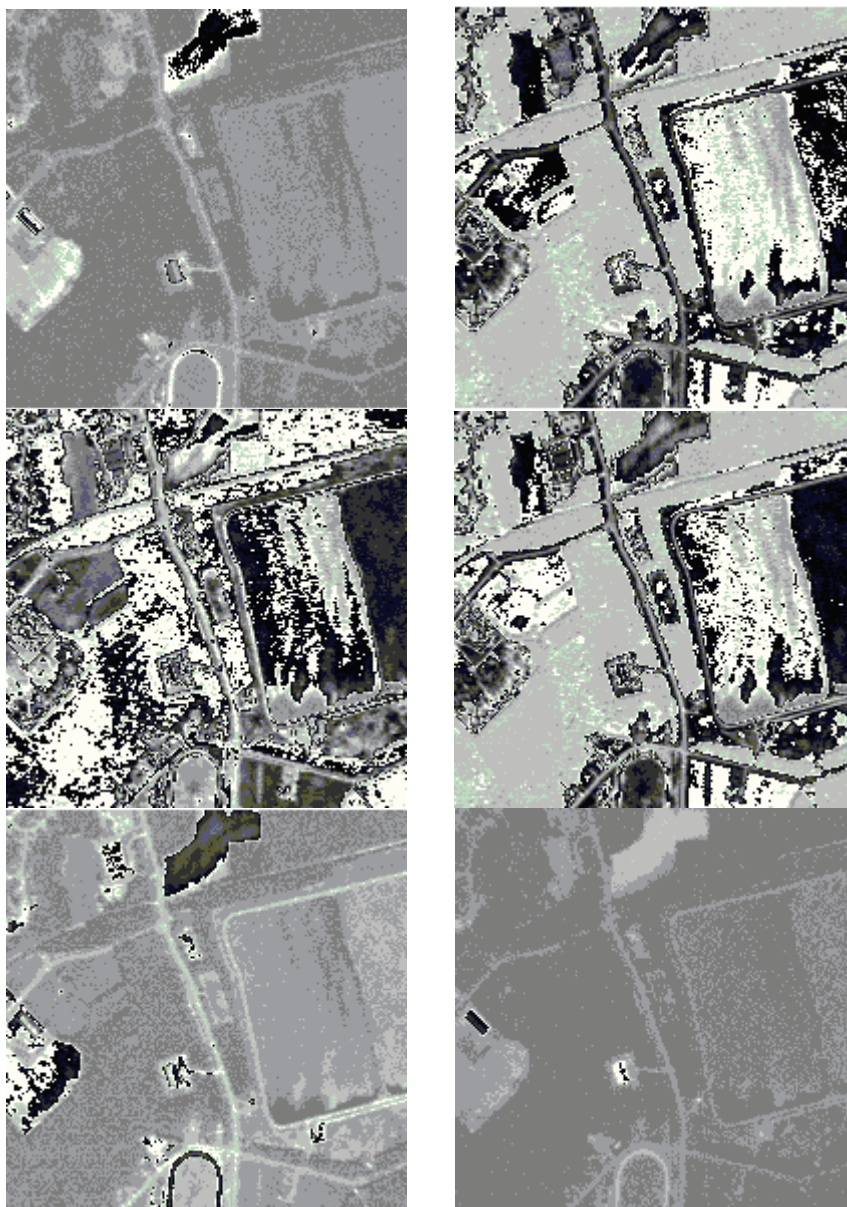


Fig. 3. AVIRIS Hyperspectral Image Examples

Either multi-spectral image stack (TM or SAR) or hyperspectral remote sensing stack are suitable for demonstrating the classification capability of the newly developed Multivariate Autoregressive image model based SVM method. Unfortunately, the MOFFET data set of AVIRIS remote sensing on NASA's website is no longer available. The remote sensing image

data for testing algorithm in this chapter is limited to the multispectral (TM and SAR) images stack for the purpose of comparisons to other classification algorithms.

3. Methodologies on Time Series Remote Sensing Image Analysis

A general scheme of Time Series Remote Sensing Image Processing as described in this chapter is shown in figure 4 and 5. below:

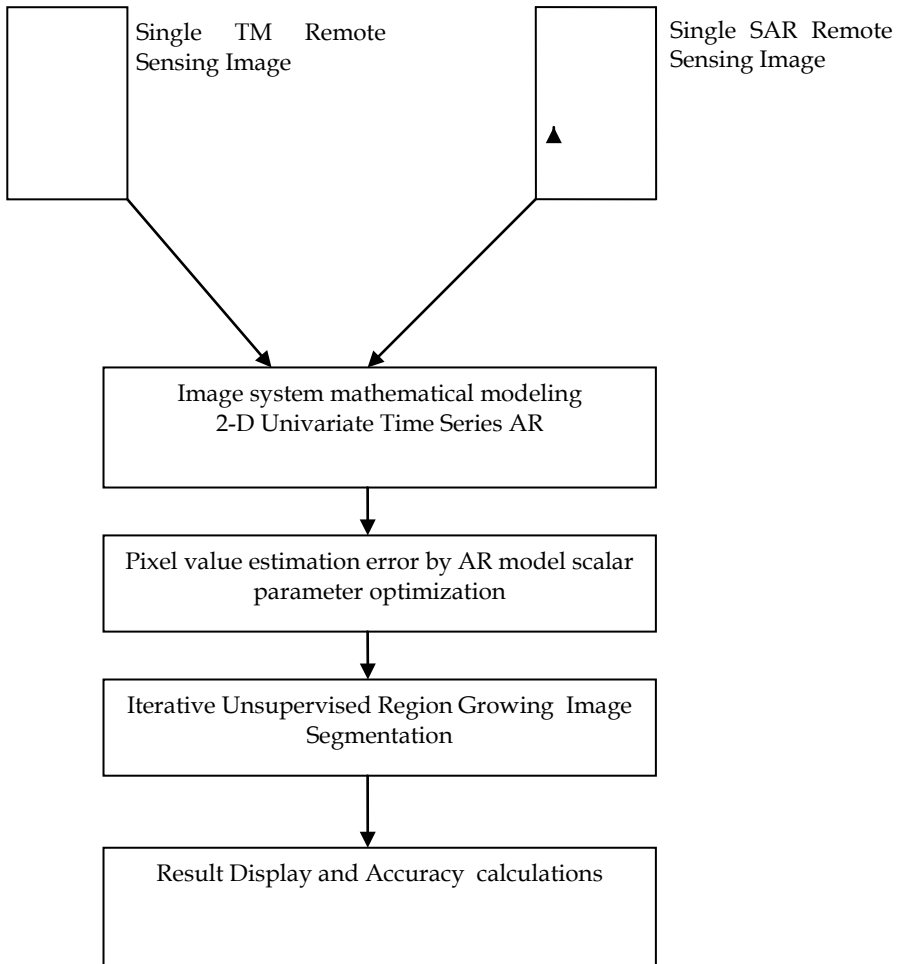


Fig. 4. Univariate Time Series Region Growing image processing system scheme

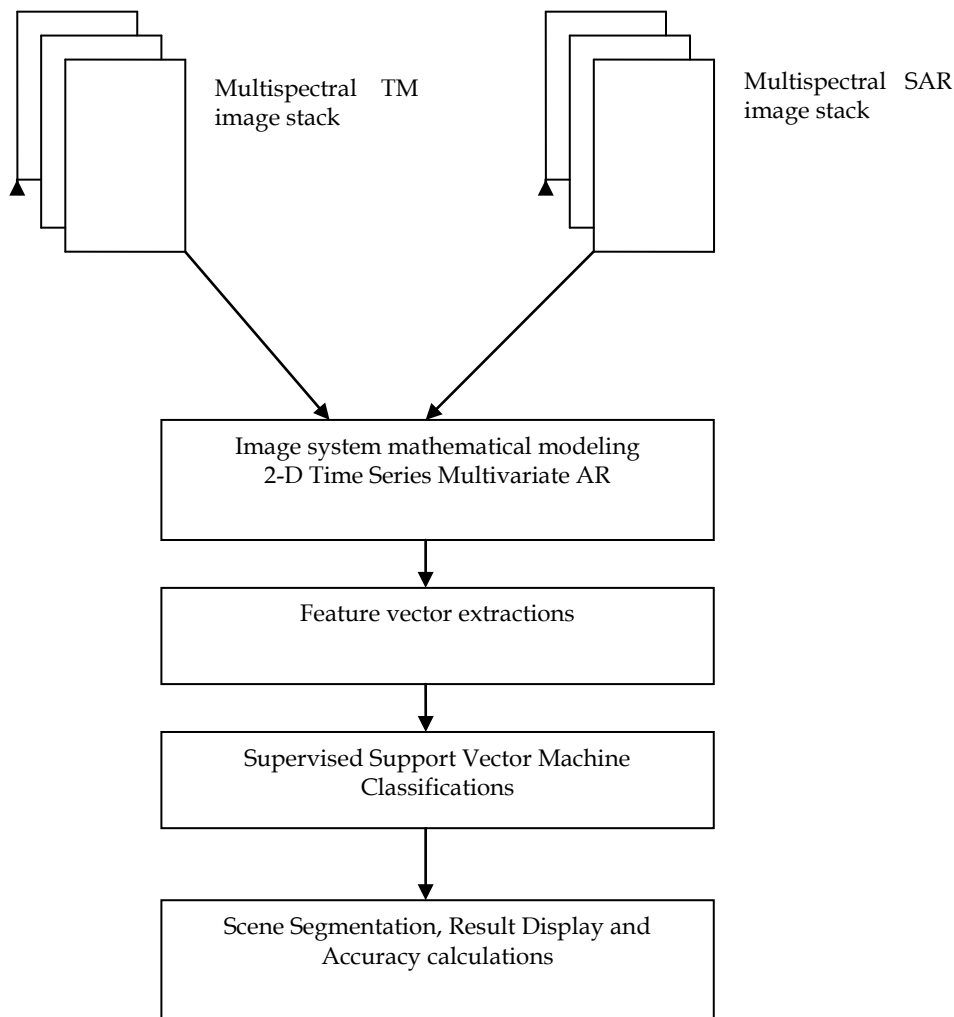


Fig. 5. Multivariate Time Series AR parameter matrix estimation SVM system scheme

4. Univariate Time Series Region Growing

We assume airborne and satellite's remote sensing can be defined as a collection of random variables indexed according to the order they are obtained in time-space. For example, we consider a sequence of random variables x_1, x_2, x_3, \dots , in general, $\{x_t\}$ indexed by t is the stochastic process. The adjacent points in time are correlated. Therefore, the value of series x_t at time t depends in some fashion on the past values x_{t-1}, x_{t-2}, \dots . Suppose that we let the value of the time series at some point of time t to be denoted by x_t . A stationary time series is the one for which the probabilistic behavior of $x_{t_1}, x_{t_2}, \dots, x_{t_k}$ is

identical to that of the shifted set $x_{t_1+h}, x_{t_2+h}, \dots, x_{t_k+h}$. In our remote sensing application, the 2-D image was scanned from left upper corner to right bottom as a sequence of time series pixel values. Further, to simplify the numerical calculations, we model each class of surface textures by 1st order and 2nd order Autoregressive stationary time series models. In another way of thinking, the two-dimensional Markov model is a similar mathematical model to describe an image area per remote sensing texture class. By using time series model, when the within-object interpixel correlation varies significantly from object to object, we can build effective classifiers. The unsupervised Region Growing is a powerful image segmentation method for use in shape classification and analysis. The LANDSAT 5 database in the area of Italy's Lake Mulargias remote sensing image data acquired in July 1996 to be used for the computing experiments with satisfactory results. The advanced statistical techniques, such as Gaussian distributed white noise error confidence interval calculations, sampling statistics based on mean and variance properties are adopted for automatic threshold finding during Region Growing iterations. The linear regression analysis with least mean squares error estimation is implemented as a time series system optimization scheme (Chen, 2003). The classification and segmentation results are shown in Figure 6,7,8.

Remote Sensing Original Image



Fig. 6. Original Lake Region Remote Sensing band 5 image

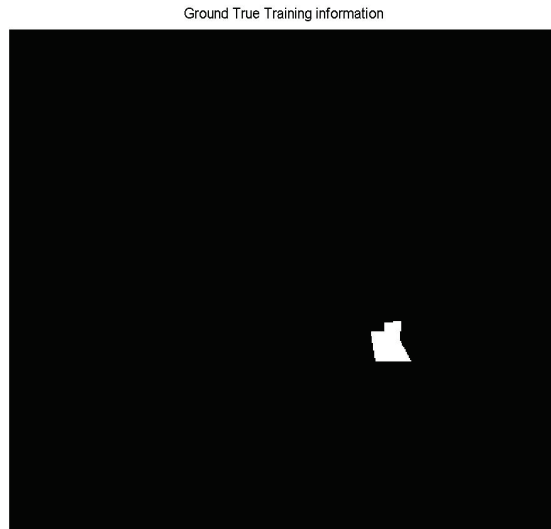


Fig. 7. Ground Truth Information

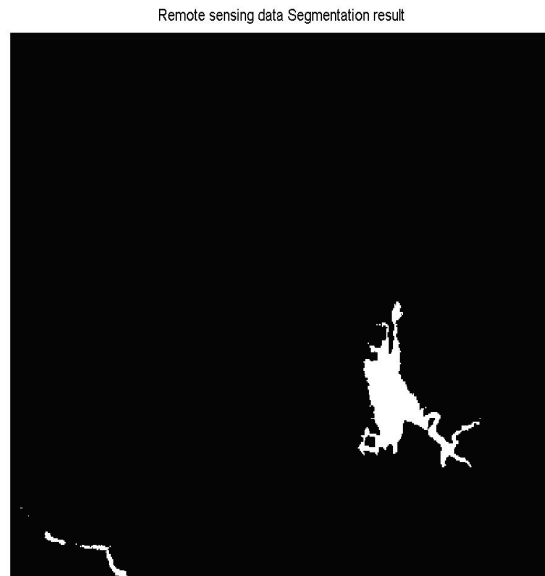


Fig. 8. Segmentation Result After Region Growing Based On univariate AR Model

5. Multivariate AR Model and Error Covariance Matrix Estimation

The remote sensing image data widely used in military, geographic survey and scientific research such as SAR (Synthetic Aperture Radar), TM (Thermal Mapper) are in multi-spectral format. As shown in figure 9, it consists of a stack of images. There are correlations between pixels in a single image as well as correlations among image slices. The Autoregressive (AR) model described in univariate time series model was based on Box-Jenkins system (Box, 1970) which is not enough to describe information extracted from multi-spectral satellite sensors. The innovative 2-D Multivariate Vector AR form (ARV) time series model described in this section is aimed to solve the problem.

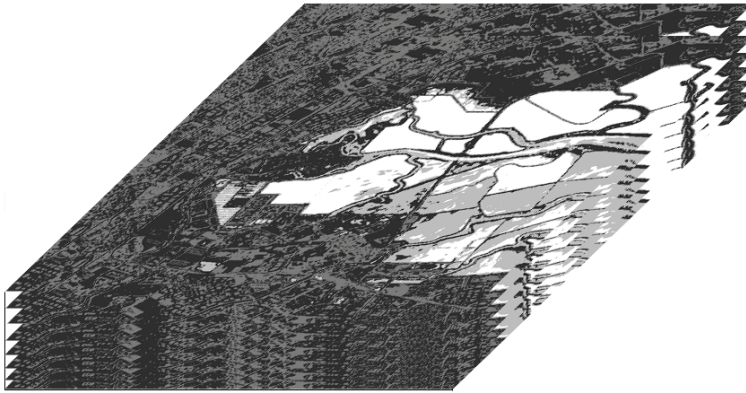


Fig. 9. Multi-spectral remote sensing image stack

In order to preserve most of the stacking remote sensing information for further 2-D image processing, there are methods such as Classification using Markov Random Fields and PCA method (Chen, 2002), Independent Component Analysis (Chen, 2002), ...etc which require data compression before using 2-D single image classification technique. Both PCA and ICA methods are really time consuming. Lengthy computation time plus undetermined components to select make them unfeasible. On the contrary, the computation involved with the multivariate ARV method is the simplest matrix operations which are faster (roughly 2 times faster in computations). Besides, the classification accuracy results have shown that this new Multivariate Vector AR method is feasible and superior. The Multivariate Time Series data analysis model is a generalized vector form of the Box-Jenkins' univariate time series model. It is also called the ARMAV (AutoRegressive Moving Average Vector) model. The fact is that each time series observation is a vector containing multiple factors.

Let $X_i = [x_{1i}, x_{2i}, x_{3i}, \dots, x_{mi}]^T \quad -\infty \leq i \leq \infty$

$$X_i = W + \phi_1 X_{i-1} + \phi_2 X_{i-2} + \dots + \phi_p X_{i-p} + \varepsilon_i - \theta_1 \varepsilon_{i-1} - \theta_2 \varepsilon_{i-2} - \dots - \theta_q \varepsilon_{i-q} \quad (1)$$

X_i : m-by-1 column vector, time series state variable

ε_i : m-by-1 column vector, multivariate white noise

$\phi_k : k = 1,2,3,\dots,p$ m-by-m autoregressive parameter matrix

$\theta_k : k = 1,2,3,\dots,q$ m-by-m moving average parameter matrix

W : m-by-1 Constant Vector (deterministic DC term)

AutoRegressive Vector (ARV) is reduced to:

$$X_i = W + \phi_1 X_{i-1} + \phi_2 X_{i-2} + \dots + \phi_p X_{i-p} + \varepsilon_i \tag{2}$$

Example: m=2, p=2, W=0 of ARV model

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \phi_{1,11} & \phi_{1,12} \\ \phi_{1,21} & \phi_{1,22} \end{bmatrix} \begin{bmatrix} x_{i-1} \\ y_{i-1} \end{bmatrix} + \begin{bmatrix} \phi_{2,11} & \phi_{2,12} \\ \phi_{2,21} & \phi_{2,22} \end{bmatrix} \begin{bmatrix} x_{i-2} \\ y_{i-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} \tag{3}$$

Estimation of the Multivariate AR model Parameter Matrix:

Optimization Method: Least Squares

Let's assume the system is a m-dimensional time series

An AR(p) time series model can be expressed as the following regression model:

$$X_v = BU_v + \varepsilon_v \tag{4}$$

where ε_v = noise vector with covariance matrix C $v = 1,2,\dots,n$ and n is the total number of samples.

Therefore, $C = E(\varepsilon_v \varepsilon_v^T)$

The Parameter Matrix $B = [W \ \phi_1 \ \phi_2 \ \dots \ \phi_p]$

$$\text{Define } U_v = \begin{bmatrix} 1 \\ X_{v-1} \\ X_{v-2} \\ \vdots \\ X_{v-p} \end{bmatrix} \qquad V = \begin{bmatrix} X_1^T \\ X_2^T \\ \cdot \\ \cdot \\ \cdot \\ X_n^T \end{bmatrix} \tag{5}$$

Let's also define:

$$T = \sum_{v=1}^n U_v U_v^T \quad (6)$$

$$X = \sum_{v=1}^n X_v X_v^T \quad (7)$$

$$S = \sum_{v=1}^n X_v U_v^T \quad (8)$$

The least squares estimation of the B matrix can be found as

$$X_v = B U_v + \varepsilon_v \quad (9)$$

$$X_v U_v^T = B U_v U_v^T + \varepsilon_v U_v^T \quad (10)$$

$$\sum_{v=1}^n X_v U_v^T = \sum_{v=1}^n \hat{B} U_v U_v^T \quad (11)$$

$$S = \hat{B} T \quad (12)$$

Therefore, the estimated parameter matrix is

$$\hat{B} = S T^{-1} \quad (13)$$

The error covariance matrix is

$$\hat{C} = \frac{1}{n - nf} \sum_{v=1}^n \hat{\varepsilon}_v \hat{\varepsilon}_v^T \quad (14)$$

Where n is the total number of samples and nf is the degree of freedom (Glantz,2002). In normal cases, this parameter can be ignored.

We have

$$\hat{\varepsilon}_v = X_v - \hat{B} U_v \quad (15)$$

$$\hat{\varepsilon}_v^T = X_v^T - U_v^T \hat{B}^T \quad (16)$$

$$\hat{\varepsilon}_v \hat{\varepsilon}_v^T = X_v X_v^T - \hat{B} U_v X_v^T - X_v U_v^T \hat{B}^T + \hat{B} U_v U_v^T \hat{B}^T \quad (17)$$

$$\hat{\varepsilon}_v \hat{\varepsilon}_v^T = X_v X_v^T - ST^{-1} U_v X_v^T - X_v U_v^T T^{-1} S^T + ST^{-1} U_v U_v^T T^{-1} S^T \quad (18)$$

$$\sum_{v=1}^n \hat{\varepsilon}_v \hat{\varepsilon}_v^T = \sum_{v=1}^n X_v X_v^T - ST^{-1} \sum_{v=1}^n U_v X_v^T - \sum_{v=1}^n X_v U_v^T T^{-1} S^T + ST^{-1} \sum_{v=1}^n U_v U_v^T T^{-1} S^T \quad (19)$$

$$\sum_{v=1}^n \hat{\varepsilon}_v \hat{\varepsilon}_v^T = \sum_{v=1}^n X_v X_v^T - ST^{-1} S^T - ST^{-1} S^T + ST^{-1} T T^{-1} S^T \quad (20)$$

$$\sum_{v=1}^n \hat{\varepsilon}_v \hat{\varepsilon}_v^T = X - ST^{-1} S^T \quad (21)$$

Therefore,

$$\hat{C} = \frac{1}{n - nf} (X - ST^{-1} S^T) \quad (22)$$

The error covariance matrix \hat{C} is similar to a Schur complement matrix
 Let's define the Schur complement matrix as

$$M = \begin{bmatrix} T & S^T \\ S & X \end{bmatrix} = \sum_{v=1}^n \begin{bmatrix} U_v \\ X_v \end{bmatrix} \begin{bmatrix} U_v^T & X_v^T \end{bmatrix} \quad (23)$$

which is the moment matrix

$$M = K^T K \quad (24)$$

where

$$K = \begin{bmatrix} U_1^T & X_1^T \\ U_2^T & X_2^T \\ \cdot & \cdot \\ U_n^T & X_n^T \end{bmatrix} \quad (25)$$

The least squares estimate can be computed from a QR factorization of the data matrix K . According to QR decomposition theorem (Anton, 2000), (Moler, 2004), (Cheney, 1999) and (Chapra 2002), if K is n by $(1 + (p + 1) * m)$ matrix with linearly independent column vectors, then K can be factored as

$$K = QR \quad (26)$$

where Q is an $n \times n$ matrix with orthonormal column vectors, and R is an n by $(1 + (p + 1) * m)$ upper triangular matrix.

$$R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \\ 0 & 0 \end{bmatrix} \quad (27)$$

$$R^T = \begin{bmatrix} R_{11}^T & 0 & 0 \\ R_{12}^T & R_{22}^T & 0 \end{bmatrix} \quad (28)$$

The submatrices R_{11} and R_{22} are upper triangular square matrices of order $p \times m + 1$ and m , respectively. The submatrix R_{12} has a dimension of $p \times m + 1$ by m .

According to matrix numerical analysis, the QR factorization of the data matrix K leads to the Cholesky factorization

$$M = K^T K = R^T Q^T Q R = R^T I R = R^T R \quad (29)$$

of the moment matrix.

Therefore,

$$M = \begin{bmatrix} T & S^T \\ S & X \end{bmatrix} = \sum_{v=1}^n \begin{bmatrix} U_v \\ X_v \end{bmatrix} \begin{bmatrix} U_v^T & X_v^T \end{bmatrix} = R^T R = \begin{bmatrix} R_{11}^T R_{11} & R_{11}^T R_{12} \\ R_{12}^T R_{11} & R_{12}^T R_{12} + R_{22}^T R_{22} \end{bmatrix} \quad (30)$$

Therefore,

$$S = R_{12}^T R_{11} \quad (31)$$

$$T = R_{11}^T R_{11} \quad (32)$$

From equation (13) $\hat{B} = ST^{-1}$ and equation (22) $\hat{C} = \frac{1}{n - nf} (X - ST^{-1}S^T)$

we can then derive the estimated parameter matrix \hat{B} and error covariance matrix \hat{C} :

$$\hat{B} = (R_{11}^{-1} R_{12})^T \quad (33)$$

$$\hat{C} = \frac{1}{n - nf} (R_{22}^T R_{22}) \quad (34)$$

where again n is the sample size and nf is the degree of freedom based on statistical sampling theory.

The error covariance matrix \hat{C} is further factorized by Cholesky decomposition. The diagonal terms of this factorized Cholesky matrix are used as feature vector inputs for Support Vector Machine (SVM).

$$\hat{C} = E^T E \tag{35}$$

Where E is the upper triangular matrix.

Suppose E is an m -by- m upper triangular, lower triangular or diagonal matrix, the eigenvalues of E are the entries on the main diagonal of E . The eigenvalues of E matrix are the characteristics that E contains. Based on theorem, the E matrix contains the major information on the estimated error covariance matrix \hat{C} .

Note that the theory behind QR decomposition, Cholesky factorization and the Schur Complement form will be detailed in section 5.2, section 5.3 and section 5.4 respectively.

5.1. Numerical Simulation

The following 2 simulated experimental examples will explain and verify the multivariate ARV model algorithm described above.

Example 1:

From Equation (3)

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \phi_{1,11} & \phi_{1,12} \\ \phi_{1,21} & \phi_{1,22} \end{bmatrix} \begin{bmatrix} x_{i-1} \\ y_{i-1} \end{bmatrix} + \begin{bmatrix} \phi_{2,11} & \phi_{2,12} \\ \phi_{2,21} & \phi_{2,22} \end{bmatrix} \begin{bmatrix} x_{i-2} \\ y_{i-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} \tag{36}$$

$$B_1 = \begin{bmatrix} \phi_{1,11} & \phi_{1,12} \\ \phi_{1,21} & \phi_{1,22} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.28 \\ 1.3 & 1 \end{bmatrix} \tag{37}$$

$$B_2 = \begin{bmatrix} \phi_{2,11} & \phi_{2,12} \\ \phi_{2,21} & \phi_{2,22} \end{bmatrix} = \begin{bmatrix} 0.35 & -0.4 \\ -0.3 & -0.5 \end{bmatrix} \tag{38}$$

Let $B = [B_1 \ B_2]$

$$B = \begin{bmatrix} 0.5 & 0.28 & 0.35 & -0.4 \\ 1.3 & 1 & -0.3 & -0.5 \end{bmatrix} \tag{39}$$

$$C = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{bmatrix} \tag{40}$$

The 200 simulated 2 elements vector of Multivariate Time Series random process data set $V_{200 \times 2}$ was generated by defining this 2nd order ARV parameter matrix:

$$V = \begin{bmatrix} 0.8778 & -2.1116 \\ 2.717 & 4.0011 \\ 1.8159 & 6.3244 \\ 1.5357 & 6.9345 \\ 1.2415 & 4.9586 \\ 0.5507 & 2.3339 \\ \cdot & \cdot \\ \cdot & \cdot \\ 3.2515 & 12.5413 \\ 2.8186 & 10.1881 \end{bmatrix} \quad \text{where } V \text{ was defined in (5)} \quad (41)$$

After multivariate ARV time series model estimation process, we get the estimated:

$$\hat{B} = \begin{bmatrix} 0.5985 & 0.2055 & 0.3347 & -0.3394 \\ 1.4657 & 1.0258 & -0.5929 & -0.4753 \end{bmatrix} \quad (42)$$

$$\hat{C} = \begin{bmatrix} 1.1896 & 0.6626 \\ 0.6626 & 1.7277 \end{bmatrix} \quad (43)$$

We can see that both the estimated parameter matrix and error covariance matrix by the above least squares algorithm is accurate by measuring a small percentage of matrix norm of matrices difference:

$$\hat{B} \approx B \quad \hat{C} \approx C$$

Example 2:

$$B = \begin{bmatrix} 0.5 & 1.3 & 0.35 & -0.3 \\ 0.28 & 1 & -0.4 & -0.5 \end{bmatrix} \quad (44)$$

$$C = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{bmatrix} \quad (45)$$

Estimation result:

$$\hat{B} = \begin{bmatrix} 0.35437 & 1.3809 & 0.4864 & -0.27223 \\ 0.25398 & 1.0096 & -0.36946 & -0.47929 \end{bmatrix} \quad (46)$$

The estimated parameter matrix is also accurate in this 2nd example.

5.2 QR Algorithm

To find the eigenvalue decomposition is to find a diagonal matrix Λ and a nonsingular matrix S such that

$$A = SAS^{-1} \tag{47}$$

Two problems might occur. First is that the decomposition may not exist. Secondly, even if the decomposition exists, it might not be robust enough. The way to solve the problems is to get as close to diagonal as possible. Ideally we would like to find the Jordan canonical form of the matrix, however, this is not practical to do in finite precision arithmetic. Instead we compute the Schur decomposition of the matrix $A = TBT^H$, where B is the upper triangular and T is unitary. Every square matrix has a Schur decomposition which can be computed using an iterative algorithm. The algorithm is called the QR algorithm since it performs a QR factorization in each iteration. The eigenvalues of A are on the diagonal of its Schur form of B . Since the unitary transformations are perfectly well conditioned, they do not magnify errors. The diagonal elements of B are the eigenvalues of A if A is symmetric and B is diagonal. In this case, the column vectors of T are orthonormal eigenvectors of A . In general, the large off-diagonal elements of B measure the lack of symmetry in A . In the non-symmetric case, the eigenvectors of A are the column vectors of $G = TX$, where X is a matrix contains the eigenvectors of the upper triangular matrix B . The QR algorithm computes the eigenvalues of real symmetric matrices, real nonsymmetric matrices and complex matrices. The singular values of general matrices are computed using the Golub-Reinsch algorithm which is based on the QR algorithm.

5.3 Cholesky Factorization

In mathematics, the Cholesky factorization (or decomposition) (Kreyszig, 1999) is named after Andre-Louis Cholesky, who found that a symmetric positive-definite matrix can be decomposed into a lower triangular matrix and the transpose of the lower triangular matrix. The lower triangular matrix is the Cholesky triangle of the original, positive-definite matrix. Cholesky's result has since been extended to matrices with complex entries.

Any square matrix A with non-zero pivots can be written as the product of a lower triangular matrix L and an upper triangular matrix U ; this is called the LU decomposition. However, if A is symmetric and positive definite, we can choose the factors such that U is the transpose of L , and this is called the Cholesky decomposition. Both the LU and the Cholesky decomposition are used to solve systems of linear equations. When it is applicable, the Cholesky decomposition is twice as efficient as the LU decomposition.

5.4 The Schur Complement Form

Let's understand what is Schur complement form. It is a block of a matrix within the larger matrix which is defined as follows. Suppose A, B, C, D are respectively $p \times p, p \times q, q \times p$ and $q \times q$ matrices, and D is invertible. Let

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad (48)$$

so that M is a $(p+q) \times (p+q)$ matrix.

6. Support Vector Machine classifier

Support Vector Machine (SVM), a new class of machine learning with superior classification capability over other learning algorithms. It can be analyzed theoretically using concepts from statistical learning theory (Vapnik, 1998) and at the same time to achieve good performance when applied to real problems. We have implemented software on top of SVM-KM matlab toolbox developed by Dr. A. Rakotomamonjy of INSA, France for remote sensing image classification and region segmentations. The results on both Multivariate AR model (ARV) and non-AR pixel-by-pixel based methods are reasonably good. Of course, the ARV's additional contextual information gives a better performance. The main idea of this classification is to construct a hyperplane as a decision surface in a way that the margin of separation between positive and negative classes is maximized. The support vector machine can provide a good generalization performance on pattern classification problems. The SVM constructs models that are complex, it contains a large class of neural networks, radial basis function and polynomial classifiers as the special cases. Nevertheless, it is also simple enough to be analyzed mathematically due to the fact that it can be shown to correspond to a linear method in a high dimensional feature space nonlinearly related to input space. By use of kernels, all needed computations are performed directly in the input space. There are different cases of SVM, the first case is the linear separable data to be trained on linear machine. The 2nd case is the nonlinear SVM trained on non-separable data, this will result in a quadratic programming problem. The SVM can work with high dimensional data as the method is less dependent on the statistical distribution of the data. The SVM avoids the "Curse of Dimensionality" problems typically experienced in statistical pattern classifications (Chen, 1999).

SVM were specifically designed for binary classification. The multiclass applications on SVM is still an on-going popular research topics. A few straight forward methods have been proposed in such a way that multiclass classifier can be constructed by combining several binary classifiers. Some other researchers are proposing to create multi-classifiers that process multi-class data at once but it is most likely less accurate. The multi-class remote sensing SVM classifier we developed that describes in this chapter is based on One-Again-One pairwise with majority votes.

Support Vector Machine Concepts

The SVM in most cases is competitive among the existing classification methods. It is relatively easy to use. We assume the data vectors $\mathbf{X} = [x_1, x_2, \dots, x_n]^T$

$$x_j \quad j = 1, \dots, n \quad \text{where } n \text{ is the number of features in the feature space}$$

Let's also consider simple case of two classes data set :

Define classification output indicator vector $Y = [y_1, y_2, \dots, y_l]^T$

$$y_i = \begin{cases} 1 & \text{if } \mathbf{x}_i \text{ in class 1} \\ -1 & \text{if } \mathbf{x}_i \text{ in class 2} \end{cases}$$

A hyperplane $\mathbf{w}^T \mathbf{x} + b = 0$ is to be found to separate all data.

Where \mathbf{w} is an weight vector and b is a bias from the origin.

A separable hyperplane is shown in figure 10.

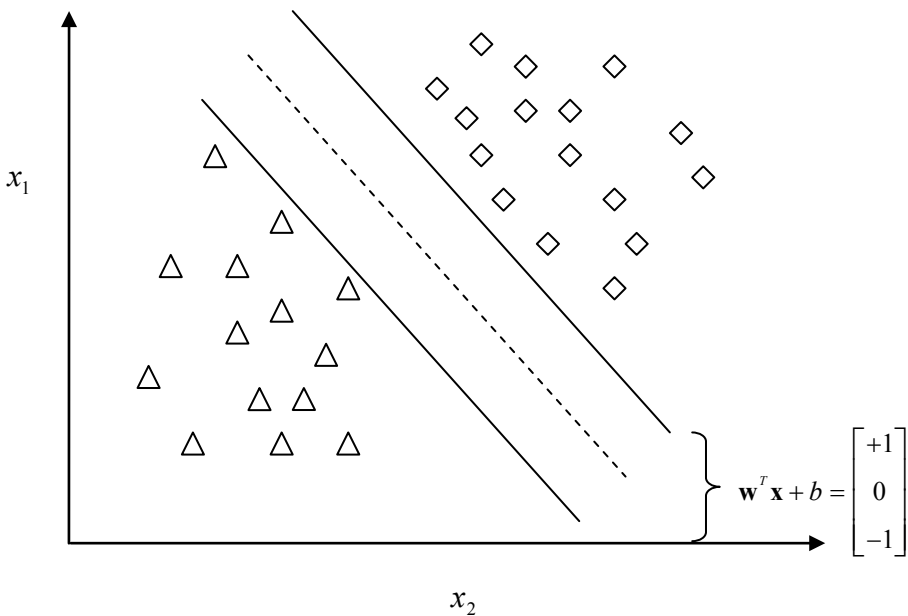


Fig. 10. Optimal hyperplane separator (in the case of 2 dimension feature space)

To illustrate the concept of SVM kernel mapping, we show in figure 11 how to map the input space to a feature space such that the non-separable input space can be separable in the feature space. This is done by nonlinear mapping to higher dimension and constructing a separating hyperplane with a maximum margin.

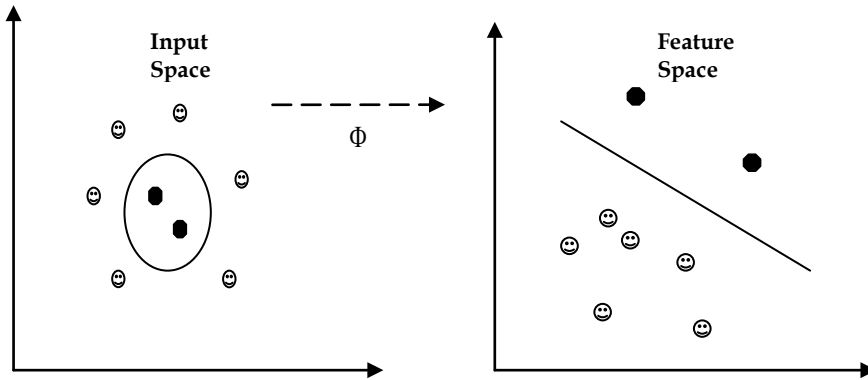


Fig. 11. Map training data nonlinearly into a higher dimensional feature space

Feature space kernel mapping example:

Figure 11 shows the basic idea of SVM which maps data into dot product space (feature space) F by a nonlinear mapping:

$$\Phi : R^N \rightarrow F \quad (49)$$

The dot product is

$$k(\mathbf{x}, \mathbf{y}) := (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{y})) \quad (50)$$

If F is in high dimension, the right hand side of equation will be expensive to compute.

7. Experimental Results

The UK village remote sensing data set were used and the Multivariate Time Series SVM performance is compared with other existing classification techniques. This image data set was kindly offered by Italian Remote Sensing Research group (Serpico 1995, Bruzzone 1999). It consists of a set of 250x350 pixel images acquired by two imaging sensors installed on a 1268 Airborne Thematic Mapper (ATM) scanner and a PLC band, full polarimetric NASA JPL SAR sensor. For performance comparison, Table 2 shows the experimenting results of the classification accuracies by several remote sensing popular classifiers (Ho, 2008):

Method 1: Structured Neural Networks (TLN) by University of Genoa, Italy (Serpico 1995)

Method 2: K-mean + PCA (Ho, 2008)

Method 3: K-mean + ICA (Ho, 2008)

Method 4: K-mean + (PCA+MRF) (Ho, 2008)

Method 5: FCM + (PCA+MRF) (Ho, 2008)

Method 6: Multivariate ARV Support Vector Machine (Ho, 2008)

Method 1: Structured Neural Networks (TLN) method:

This method uses the architecture of structured multilayer feedforward networks. The networks are trained to solve the problem by the error backpropagation algorithm. They are transformed into equivalent networks to obtain a simplified representation. It is considered to be a hierarchical committee that accomplishes the classification task by checking on a set of explicit constraints on input data.

Method 2: K-mean + PCA:

Principal Component Analysis (PCA) is a technique to reduce multidimensional data to a lower dimension for analysis. By constructing the transformation matrix which consists of eigenvectors and ordering the eigenvalues, the statistically uncorrelated components can be extracted. It is an unsupervised approach that finds the best features from data. K-means clustering is a form of stochastic hill climbing in the log-likelihood function. The contours in the feature space represents equal log-likelihood values. It iteratively calculates the mean vectors in the selected classes. As more data are inputted, it dynamically adjusts until there is no change on mean vectors.

Method 3: K-mean + ICA:

Independent Component Analysis (ICA) is a computational method to separate a multivariate signal into additive subcomponents which are statistically independent and at least one of which is a non-Gaussian source signal. Method 3 is the same as Method 2 except that PCA is replaced by ICA.

Method 4: K-mean + (PCA+MRF):

Markov Random Field (MRF) theory provides a basis for modeling contextual constraints in image processing. It is a model of the joint distribution of a set of random variables. The estimated pixels based on MRF model form a new image. PCA is applied to each pixel vector of the new image stack. This is followed by K-mean operation.

Method 5: FCM + (PCA+MRF):

Fuzzy-C-Mean (FCM) is an iterative clustering method. In every iteration of the classical K-Means procedure, each data point is assumed to belong to exactly one cluster. But in FCM, we relax this condition and assume that each sample has some graded or fuzzy membership in a cluster. Method 5 is the same as Method 4 except that K-mean is replaced by FCM operation.

Method 6: Multivariate ARV Support Vector Machine

This method was newly developed as described in the sections earlier.

Figure 12,13,14,15 show the results of region segmentation by this novel multivariate Time Series model based SVM classification method as described in this chapter.

Figure 12. is the original th-c-hh (one example of 15 remote sensing input data). Figure 13. is the Multivariate AR - SVM without smooth post-processing result. Figure 14. is the Multivariate AR - SVM with post-processing result and Figure 15. is the Multivariate AR - SVM with post-processing and tone re-scaling result.

	Method 1	Method 2	Method 3	Method 4	Method 5	Method 6
Accuracy	86.49%	64.93%	65.79%	72.46%	75.27%	87.11%

Table 2. Classification performance comparison

The followings are the color bar and crop identifications in UK village remote sensing image data set:






	Class 1: Sugar Beets
	Class 2: Stubble
	Class 3: Bare Soil
	Class 4: Potatoe
	Class 5: Carrots

Figure 12. Original th-c-hh (one example of 15 remote sensing input data)

Figure 13. Multivariate AR - SVM without smooth post-processing result

Figure 14. Multivariate AR - SVM with post-processing result

Figure 15. Multivariate AR - SVM with post-processing and tone re-scaling result

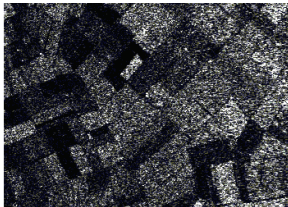


Fig.12

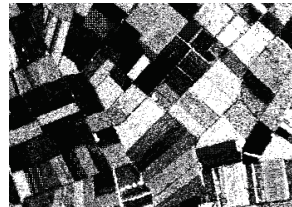


Fig. 13

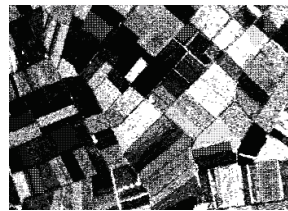


Fig. 14

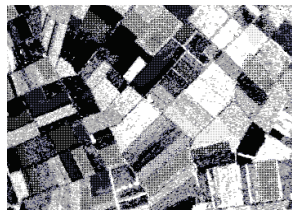


Fig. 15

8. Conclusion

This chapter has focused on contexture information in order to improve remote sensing pixel classification accuracy. The study of time series is primarily concerned with time or spatial correlation structure. The time series model has its world wide applications in the area of finance and oceanography but not much in remote sensing image processing. We took the big challenge to develop this new idea for remote sensing information processing. As we mentioned earlier, the time series models were started from univariate system to move gradually toward complicated multivariate matrix structures in order to solve the multiple image pixel stack problems. The system optimization and image pixel estimation solutions were changed from simple derivative, linear system to complex matrix decompositions by numerical analysis. It does open up a new era away from traditional approaches as far as remote sensing image processing is concerned. Numerical methods are the arithmetic operations to solve the mathematically formulated problems. Although they are involved with large numbers of tedious calculations, the most recent fast and efficient

digital computer can solve them quickly. The role of numerical analysis in solving the engineering problem has increased dramatically. They are capable of handling large systems of equations, nonlinearities and geometries that are often impossible to solve analytically. The remote sensing data mining is huge. The system is initially gained by empirical means through observations and experiments. As new measurements are taken, the generalizations might be modified or newly developed. On the other hand, generalizations can have a strong influence on the observations. Hopefully the conclusions can be drawn eventually. From an electrical and computer engineering problem solving perspective, a system's mathematical model has to be usefully expressed. We are hoping that system mathematical modeling and the powerful numerical methods that take full advantage of fast computing tools can resolve most of the remote sensing issues. Take one example in this chapter: the Multivariate Time Series model with system parameter matrix estimation and error covariance matrix (solved by Cholesky decomposition and the stable QR algorithm) might be able to capture remote sensing research attentions for their complex unsolvable image stacking problems. Though originality of this method exists, the algorithm stability does require years of testing by researchers and interested parties. The novel Estimation of Parameters Matrix of Multivariate Time Series techniques we developed in this chapter can be useful not only for image classification but also good for other research areas such as accurate stock market predictions, video prediction for wireless multimedia applications, adaptive frame prediction for scalable video coding ...etc. The reasons are that multivariate time series can handle multiple data factors at every single time stamp. It can also expand the traditional Digital Signal Processing capability which is mostly in univariate time sequences. Advanced methods for automatic analysis of multisensor and multitemporal remote sensing images is still on-going in the years to come. Image processing and pattern recognition researches have been going on for over half century since the digital computer was invented. Many excellent and useful tools such as Maple, Matlab, IDS, Labview ...etc have been created accordingly. Millions of image processing algorithms and great research results were also published. Remote sensing, image processing and pattern recognition related community have generated many journal papers and held developer conferences each year around the world to exchange ideas. Still, none of the "final" universal optimal algorithm has been done successfully yet. Take an example of the remote sensing texture classifications, it is difficult to obtain a good texture representation and to have the true adaptive function for segmentation for all kinds of remote sensing data.

9. Future Work

Though the multivariate ARV model was developed, to select the optimal order of an ARV model is one of complicated problem to solve. The error covariance matrix \hat{C} has to be computed and the order selection criterion such as famous Akaike's Final Prediction Error (FPE) (Akaike, 1971) criterion, Schwarz's Bayesian Criterion (SBC) (Schwarz, 1978) and Lutkepohl's improved order criterion (Lutkepohl, 1985) has to be determined in order to decide the optimal ARV system order. As far as today's most decent and generalized data classification method -Support Vector Machine, there are still many holes to be filled and explored. To mention a few, improving system model formulation for extracting more useful feature vectors for SVM, new kernel functions developments for better SVM

classification performance, convex and non-convex optimization theory, intelligent numeric computational methods ...etc are the open research area for the near future.

10. References

- Akaike, H. Autoregressive Model Fitting for Control. *Ann. Inst. Statistical Mathematics* 23, 1971, pp. 163-180.
- Anton, H. Rorres, C. *Elementary Linear Algebra, Applications*, John Wiley and Sons 2000.
- Bruzzone, L. and Prieto, D. A Technique for The Selection Of Kernel Function Parameters in RBF neural Networks for Classification of Remote Sensing Images, *IEEE Trans. Geosci. Remote Sensing*, Vol. 37, No.2, pp. 1179-1184, March 1999.
- Bruzzone, L. *Advanced Methods for the Analysis of Multisensor and Multitemporal Remote-Sensing Images*, PhD dissertation 1998, Electronics and Computer Science, Univeristy of Genova, Italy.
- Chapra, S. Canale, R. *Numerical Methods for Engineers with Software and Programming Applications*, 4th edition, McGraw Hill 2002.
- Chen, C. editor, *Information Processing for Remote Sensing*, World Scientific 1999, pp. 12-13
- Chen, C. and Ho, P. On the ARMA model based Region Growing method for extracting lake region in a Remote Sensing image, on *Proceedings of SPIE*, Vol. 5238, Paper No. 5238-13, September 8-12, 2003, Barcelona, Spain.
- Chen, H. Remote Sensing Data Unsupervised Classification Using Gauss-Markov Random Fields and PCA, MS thesis 2002 ECE University of Massachusetts at Dartmouth.
- Cheney, W. and Kincaid, D. *Numerical Mathematics and Computing*, 4th edition, Publisher: Gary W. Ostedt Book 1999.
- Glantz, S. *Primer of Biostatistics*, 5th edition, McGraw-Hill 2002.
- Box, G. and Jenkins, G. *Time Series Analysis: Forecast and Control*, San Francisco, Holden-Day, 1970.
- Ho, P. *Multivariate time series model based support vector machine for multiclass remote sensing image classification and region segmentation*, Ph.D dissertation, Univ. of Massachusetts Dartmouth, Jan. 2008.
- Kreyszig, E. *Advanced Engineering Mathematics*, 8th edition, John Wiley and Sons 1999.
- Lutkepohl, H. Comparison of Criteria for Estimating the Order of a Vector Autoregressive Process. *Journal of Time Series*, Vol 6, 1985, pp. 35-52.
- Moler, C. *Numerical Computating with Matlab*, Chapter 10, SIAM 2004 .
- Prieto, D. *Automatic Analysis of Multisource and Multitemporal Remote-Sensing Data for Land-Cover Monitoring Application*, Ph.D Thesis, University of Genova, 2000.
- Richards, J. and Jia, X. *Remote Sensing Digital Image Analysis*, Springer-Verlag 1999.
- Schowengerdt, R. *Remote Sensing, Models and Methods for Image Processing*. 2nd edition, Academic Press, 1997.
- Serpico S. and Roli, F. Classification of Multisensor Remote Sensing Images by Structured Neural Netorks, *IEEE Transactions on Geoscience and Remote Sensing*, Vol 33, No.3, pp. 562-578, May 1995.
- Schwarz, G. Estimating the Dimension of a Model. *Ann. Statistics* Vol. 6, 1978, pp. 461-464.

Surface approximation from rapidly varying data: Applications to geophysical surfaces and seafloor surfaces

¹Apprato Dominique, ²Gout Christian and ³Le Guyader Carole

¹*Université de Pau et des Pays de l'Adour,*

²*Université de Valenciennes et du Hainaut Cambrésis,*

³*INSA de Rouen,*

France

1. Introduction

We propose an approximation method for surfaces with fault and /or large variations. We use image segmentation tools, meshing constraints, finite element methods and spline theory.

Curve and surface fitting using spline functions from rapidly varying data is a difficult problem (see Salkauskas, 1974, or Franke and Nielson, 1984, or Franke, 1982). In the bivariate case and without information about the location of large variations, usual approximation methods lead to instability phenomena or undesirable oscillations that can locally and even globally hinder the approximation (Gibbs phenomenon).

So, we propose a new method which uses scale transformations (see Apprato and Gout, 2000). The originality of the method consists of a pre-processing and a post-processing of the data. Instead of trying to find directly an approximant, we first apply a scale transformation to the z-values of the function. In the particular case of the approximation of surfaces, the originality of the method consists in removing the variations of the unknown function using a scale transformation in the pre-processing. And so the pre-processed data do not exhibit large variations. So we can use a usual approximant which will not create oscillations.

In case of vertical fault, we also propose an algorithm in order to find the location of large variations: the right approach to get a good approximant consists, in effect, in applying first a segmentation process to precisely define the locations of large variations and faults, and exploiting then a discrete approximation technique. To perform the segmentation step, we propose a quasi-automatic algorithm that uses a level set method to obtain from the given (gridded or scattered) Lagrange data, several patches delimited by large gradients (or faults). Then, with the knowledge of the location of the discontinuities of the surface, we generate a triangular mesh (which takes into account the identified set of discontinuities) on which a D^m -spline approximant (see Manzanilla 1986, Apprato et al., 1987, or Arcangéli et., 1997, or Arcangéli et al. 2004) is constructed.

We apply our method to different datasets (bathymetry, Lagrange data...): Piton de la Fournaise volcano in La Réunion island (see Gout and Komatitsch, 2000), Pyrénées

mountains in France (see Apprato et al., 2000), Marianas trench in the Pacific (see Apprato et al. 2002). We also give an example around the Hawaiian hot spot. The topography and bathymetry of the Hawaiian Islands in the Pacific ocean result from the activity of a huge hot spot combined with the effect of erosion. This hot spot has been more or less active since the Late Cretaceous, and as a result, the Big Island continues to grow, and to the East a new island is being formed.

2. Mathematical Modelling for surface approximation from Lagrange dataset

Unfortunately, when applied to the approximation of surfaces from rapidly varying data, usual methods like splines lead to strong oscillations near steep gradients, as illustrated in Figure 1. When the location of the large variations in the dataset is known, Salkauskas (1974) has proposed methods that use a spline under tension with a nonconstant smoothing parameter, and Hsieh and Chang (1994) have proposed a concept of virtual nodes inserted at the level of the large variations in the case of an approximant in the context of computer-aided geometric design. In the more general context where the location of the large variations in the dataset is not known *a priori*, Franke (1982) and Bouhamidi and Le Méhauté (1999) have proposed splines under tension belonging to more general spaces. These methods give good results in the case of curve fitting, but less accurate results in the case of surface fitting.

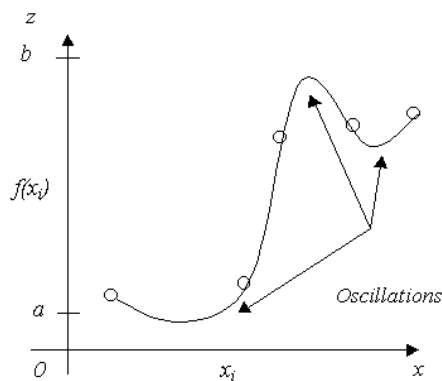


Fig. 1. When classical splines (for instance, here a C^1 spline) are used to interpolate data points $(x_i ; f(x_i))$ with large local variations, strong spurious oscillations are generated near steep gradients.

The new method we introduce here uses scale transformations, and is applied without any particular *a priori* knowledge on the data. The philosophy of the method is similar to interpolation methods based upon anamorphosed data commonly used in geostatistics (see, for instance, Issaks and Srivastava, 1989). In the first part of this article, a construction of the scale transformation families is presented. Results concerning the convergence of the approximation are given. We also show the efficiency of this innovative approach by applying it to the topography of the summit of the Piton de la Fournaise volcano, located in the Réunion Island (Indian Ocean, France). This volcano exhibits large and rapid variations

in steep river valleys in its southwestern part, as well as in a caldera, where the behavior of the method is tested.

The method we propose uses two scale transformations—namely φ_d for the preprocessing and ψ_d for the postprocessing. The first one, φ_d , is used to transform the z values representing the height of the unknown surface f into values (u_i) , regularly distributed in an interval chosen by the user, as illustrated in Figures 2A and 2B. The preprocessing function φ_d is such that the transformed data do not exhibit large local variations, and therefore a usual spline operator T_d can subsequently be applied without generating significant oscillations, as shown in Figure 2C. The second scale transformation ψ_d is then applied to the approximated values to map them back and obtain the approximated values of z (Figure 2D). It is important to underline that the proposed scale transformations do not create spurious oscillations. Moreover, this method is applied without any particular knowledge on the location of the large variations in the dataset. Let us consider a dataset

$(x_i^d, z_i^d)_{i=1, \dots, N(d)}$ indexed with a real d , such that when d tends to 0, the number of data points $N(d)$ tends to infinity. For the purpose of a theoretical study of the approximation convergence, we introduce a function $f : \Omega \rightarrow [a, b]$, such that the data set becomes

$$(x_i^d, z_i^d = f(x_i^d))_{i=1, \dots, N(d)}.$$

The functions introduced above have the following expressions, for $m \in \mathbb{N}$:

$$- \varphi_d : [a, b] \rightarrow [\alpha, \beta] \subset \mathbb{R},$$

$$- T^d : (\varphi_d \circ f) \in H^m(\Omega, [\alpha, \beta]) \rightarrow T^d(\varphi_d \circ f) \in H^m(\Omega, [\alpha, \beta]), \tag{1}$$

$$- \psi_d \circ T^d(\varphi_d \circ f) \in H^m(\Omega, [a, b]),$$

where the preprocessing φ_d and the postprocessing ψ_d are continuous scale transformation families, where T^d is an approximation operator, for instance a spline, and where $H^m(\Omega, \cdot)$ denotes the usual Sobolev space. More precisely, we introduce a nonempty bounded connected set Ω of \mathbb{R}^2 with Lipschitz boundary, and an unknown function $f \in H^{m'}(\Omega, [a, b])$ that we want to approximate, this hypothesis allowing to have $(\varphi_d \circ f)$ bounded in $C^m(\overline{\Omega})$ (with $m' > m + 1$) a property used to establish the convergence of the approximation (Apprato and Gout, 2000). We also consider a subset A^d of $N = N(d)$ distinct points of $\overline{\Omega}$ such that

$$\sup_{x \in \overline{\Omega}} \delta(x, A^d) = d \tag{2}$$

where δ is the Euclidean distance in \mathbb{R}^2 ; the index d represents the radius of the biggest sphere included in Ω that does not intersect with any point of A^d , and thus, when d tends to 0, the number of data points tends to infinity. We also introduce the set Z_1^d of $N = N(d)$ real numbers such that

$$\forall x_i^d \in A^d, f(x_i^d) \in Z_1^d \tag{3}$$

and the sequence Z_2^d of $p(d)$ distinct z values obtained from the ordering of Z_1^d ,

$$\forall \tilde{z}_i^d \in Z_2^d, i = 1, \dots, p(d),$$

$$a = \tilde{z}_1^d < \tilde{z}_2^d < \tilde{z}_3^d < \dots < \tilde{z}_{p(d)-1}^d < \tilde{z}_{p(d)}^d = b \tag{4}$$

where $[a, b] = \text{Im}(f)$. The sequence Z_2^d will be used for the construction of the scale transformation families in the following section. In what follows, for convenience, we also write z_i^d instead of \tilde{z}_i^d .

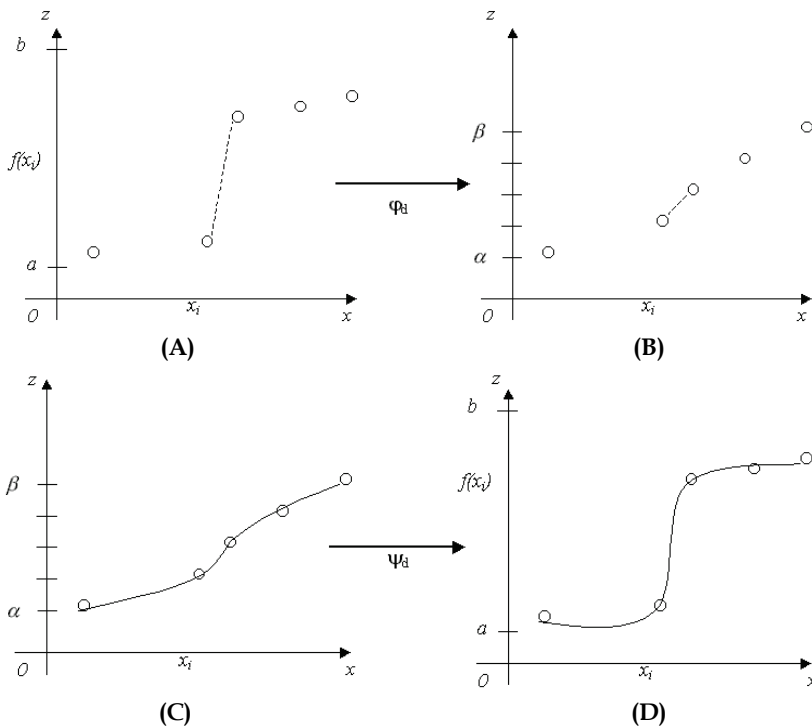


Fig. 2. The preprocessing phase, (A) and (B), transforms the values $f(x_i)$ using a scale transformation ϕ_d . After preprocessing, B, the local variations in the data have been

drastically reduced. Therefore, it is possible to obtain a regular approximant T^d with no significant oscillations using a usual C^1 spline operator (see subsection 2.2), as shown in (C). A second scale transformation ψ_d is subsequently applied to the values of the approximant in a postprocessing phase, (D), to map them back and obtain the final approximant. It is important to mention that the scale transformations used do not create spurious oscillations, as illustrated in (D).

2.1 Scale transformations

In this section, we give a construction of the scale transformation families by generalizing the technique seen in Torrens, 1991. These scale transformations are realistic in the sense that, as classical transformations, they are monotonous.

Preprocessing of the Data: Family φ_d of Scale Transformations

The goal of the scale transformation φ_d is to reduce the variations in the data set. We first construct φ_d , and, in order to study the convergence of the approximation, we then establish the convergence of φ_d to a function φ when the number of data points tends to infinity (i.e., $d \rightarrow 0$): Let $[\alpha, \beta]$ be an interval of \mathbb{R} , and $\{u_i\}_{i=1, \dots, p(d)}$ the following regular subdivision, for $i = 1, \dots, p(d)$:

$$\alpha = u_1 < u_2 < u_3 < \dots < u_{p(d)-1} < u_{p(d)} = \beta \text{ and } u_{i+1} - u_i = \frac{\beta - \alpha}{p(d) - 1} \quad (5)$$

These interval and subdivision are chosen by the user. When dealing with surface approximation from rapidly varying data, we choose the interval to be $[0, 1]$, and an even subdivision of the $\{u_i\}$ that is used to reduce the local variations of the (z_i) . After applying φ_d , we obtain a new data set (x_i, u_i) related to the initial data by $u_i = \varphi_d(z_i)$. When this technique is applied to other problems however, for instance in some applications in imaging when one has an image with homogeneous regions, it can be of interest to increase the variations between pixel values—the (z_i) —; in such a case, Apprato and Gout, 2000 showed that it is possible to choose a nonregular distribution in the interval $[\alpha, \beta]$ to generate variations, and therefore to enhance some features present in the image to facilitate its segmentation.

We introduce $\varphi : [a, b] \rightarrow [\alpha, \beta]$ the C^∞ diffeomorphism that transforms $[a, b]$ into $[\alpha, \beta]$ (such families of transformations are usually called anamorphosis in the geostatistics literature):

$$\varphi(z) = \frac{\beta - \alpha}{b - a}(z - a) + \alpha \quad (6)$$

We also introduce the function φ_d , for $i=1, \dots, p(d)-1$ and for any $z \in [z_i, z_{i+1}]$,

$$\varphi_d(z) = u_i q_{0m}^0 \left(\frac{z - z_i}{z_{i+1} - z_i} \right) + u_{i+1} q_{0m}^1 \left(\frac{z - z_i}{z_{i+1} - z_i} \right) + \alpha_1(z_i)(z_{i+1} - z_i) q_{1m}^0 \left(\frac{z - z_i}{z_{i+1} - z_i} \right) + \alpha_1(z_{i+1})(z_{i+1} - z_i) q_{1m}^1 \left(\frac{z - z_i}{z_{i+1} - z_i} \right) \quad (7)$$

where the q_{lm}^i , for $i = (0, 1)$, and $l = (0, 1)$, are the basis functions of the finite element of class C^m on $[0, 1]$ (see Ciarlet, 1978) and where, for any $i = 1, \dots, p(d)-1$,

$$\alpha_1(z_i) = \frac{u_{i+1} - u_i}{z_{i+1} - z_i} \text{ and } \alpha_1(z_{p(d)}) = \alpha_1(z_{p(d)-1}) \quad (8)$$

Using relations (5)-(8), we obtain the following results: φ_d implements the interpolation of the (u_i) and φ_d belongs to $C^m([a, b])$:

- (i) $\varphi_d(z_i) = u_i$, for $i=1, \dots, p(d)$;
- (ii) $\varphi_d \in C^m([a, b])$.

We now consider a *sufficient* convergence hypothesis, which implies that the distribution of the data (z_i) has an asymptotic regularity in the interval $[a; b]$ when d tends to 0, and which is used to establish the convergence of the approximation. This hypothesis is that there exists $C > 0$ and an integer m'' verifying $m'' \geq m \geq 2$ such that, for d small enough; and for any $i = 1, \dots, p(d)-2$, we have

$$\left| 1 - \frac{z_{i+1} - z_i}{z_{i+2} - z_{i+1}} \right| \leq C \left(\frac{b-a}{p(d)-1} \right)^{m''} \tag{9}$$

We also suppose that the set A^d introduced above is such that there exists $C' > 0$ such that

$$p(d) \leq \frac{C'}{d^2} \tag{10}$$

Inequation (10), introduced by Arcangéli (1989), expresses a property of asymptotic regularity of the distribution of the data set A^d in $\overline{\Omega}$. Using a compactness argument, Gout, 2002 established that hypotheses (9) and (10) imply that there exists $C'' > 0$, such that $\|\varphi_d\|_{C^m([a,b])} \leq C''$ and

$$\lim_{d \rightarrow 0} \varphi_d = \varphi \text{ in } C^0([a, b]), \tag{11}$$

where φ_d is defined by (7), and φ is defined by (6).

One can notice that the construction of the scale transformations φ_d made in (7) uses a finite difference scheme of order 1 to construct, from the u_i , the first derivatives of φ_d at the points \tilde{z}_i , $i = 1, \dots, p(d)$. Moreover, the option retained in (6), which is to cancel the l derivatives of φ_d at the points \tilde{z}_i for any $l=2, \dots, m$, could be substituted by the option consisting in using a finite difference scheme of order l to define these l derivatives. Let us also mention that we have chosen to construct scale transformations on a finite element basis in order to be able to study the convergence of the approximation.

Postprocessing of the Data: Family ψ_d of Scale Transformations

Similarly to the way we constructed the scale transformations φ_d , we now define a scale transformation family ψ_d that implements the postprocessing of the calculation. We recall that after the preprocessing, the large local variations in the dataset have been drastically reduced; therefore it is possible to approximate the data using a usual spline operator T^d without generating significant oscillations. To map these values back and obtain the approximated values of z , we need to use a postprocessing step, and therefore need to

introduce a family ψ_d , which is almost the inverse of φ_d : as φ_d converges to φ , we construct ψ_d such that ψ_d converges to φ^{-1} . To do so, we define the C^∞ diffeomorphism $\varphi^{-1} : [\alpha, \beta] \rightarrow [a, b]$ inverse of φ defined in Equation (6):

$$\varphi^{-1}(u) = \frac{(u - \alpha)(b - a)}{\beta - \alpha} + a. \tag{12}$$

We also define ψ_d the function, for $i = 1, \dots, p(d) - 1$, and for any $u \in [u_i, u_{i+1}]$,

$$\psi_d(z) = z_i q_{0m}^0 \left(\frac{u - u_i}{u_{i+1} - u_i} \right) + z_{i+1} q_{0m}^1 \left(\frac{u - u_i}{u_{i+1} - u_i} \right) + \beta_1(u_i) \chi_{(u_{i+1} - u_i)} q_{1m}^0 \left(\frac{u - u_i}{u_{i+1} - u_i} \right) + \beta_1(u_{i+1}) \chi_{(u_{i+1} - u_i)} q_{1m}^1 \left(\frac{u - u_i}{u_{i+1} - u_i} \right) \tag{13}$$

where the q_{lm}^i , for $i = (0, 1)$, and $l = (0, 1)$, are the basis functions of the finite element of class C^m on $[0; 1]$ (Ciarlet, 1977) and where, for any $i = 1, \dots, p(d) - 1$:

$$\beta_1(u_i) = \frac{z_{i+1} - z_i}{u_{i+1} - u_i} \text{ and } \alpha_1(u_{p(d)}) = \alpha_1(u_{p(d)-1}). \tag{14}$$

Under hypotheses (9) and (10), Apprato and Gout, 2000 established the following relations:

- (i) $\psi_d(u_i) = z_i, i = 1, \dots, p(d)$;
- (ii) $\psi_d \in C^m([\alpha, \beta])$;
- (iii) there exists $C > 0$; such that $\|\psi_d\|_{C^m([\alpha, \beta])} \leq C$;
- (iv) $\lim_{d \rightarrow 0} \psi_d = \varphi^{-1}$ in $C^0([\alpha, \beta])$.

It is important to mention that (15-i) is one of the key points of the algorithm, that (15-ii) enables us to obtain approximants with high regularity, and that (15-iii) and (15-iv) are used to establish the convergence of the approximation.

2.2 The spline operator

Given a Lagrange dataset $(x_i, (\varphi_d \circ f)(x_i) = \varphi_d(z_i))_{i,d}$ we have to solve the classical problem of constructing an approximant T^d of class C^k (with $k = 1$ or 2 in practice). In this work, we use a smoothing D^m spline, as defined in Arcangéli et al., 2004, which has many advantages: it is possible to implement a local refinement, the matrix of the linear system to solve is banded, and it is possible to study the convergence of the approximation. We have chosen to use a smoothing D^m spline and not an interpolation spline because we want to be able to work with large datasets of up to several hundreds of thousands of points, and in that case, a smoothing spline is far less expensive than an interpolation spline. We consider the functional, for any Φ belonging to $H^m(\Omega)$,

$$J_\varepsilon^d(\Phi) = \left\langle \rho^d(\Phi - \varphi_d \circ f) \right\rangle_{p(d)}^2 + \varepsilon |\Phi|_{m,\Omega}^2 \tag{16}$$

where $\rho^d \in L(H^m(\Omega), \mathbb{R}^{p(d)})$ is defined by $\rho^d f = (f(a))_{a \in A^d} \in \mathbb{R}^{p(d)}$, $\|\bullet\|_{m,\Omega}$ is the usual semi-norm on $H^m(\Omega)$, $\langle \bullet \rangle_{p(d)}$ is the Euclidean norm in $\mathbb{R}^{p(d)}$, and ε a smoothing parameter.

We call σ_ε^d the D^m -smoothing spline on Ω relative to $\varphi_d \circ f$ which is the unique solution of the minimization problem: find $\sigma_\varepsilon^d \in H^m(\Omega)$ such that for any Φ belonging to $H^m(\Omega)$:

$$J_\varepsilon^d(\sigma_\varepsilon^d) \leq J_\varepsilon^d(\Phi). \tag{17}$$

The solution σ_ε^d to this problem is also the unique solution of the variational problem: find $\sigma_\varepsilon^d \in H^m(\Omega)$ such that for any Φ belonging to $H^m(\Omega)$:

$$\langle \rho^d \sigma_\varepsilon^d, \rho^d \Phi \rangle_{p(d)} + \varepsilon \langle \sigma_\varepsilon^d, \Phi \rangle_{m,\Omega} = \langle \rho^d(\varphi_d \circ f), \rho^d \Phi \rangle_{p(d)}. \tag{18}$$

Uniqueness of the solution can be proved using the Lax-Milgram lemma and results by Necas, 1967 to establish an equivalence of norms.

In order to compute σ_ε^d , we choose to discretize it on a finite element basis, which enables us to obtain a small sparse linear system. We choose the generic Bogner-Fox-Schmit (BFS) rectangular finite element (see Ciarlet, 1977). In what follows, we use either the BFS of class C^0 or of class C^1 in order to obtain a C^0 or C^1 approximant. In the following, we write σ_ε^d instead of T^d .

2.3 Convergence results

We first give the convergence of the D^m spline operator σ_ε^d related to the transformed data $(\varphi_d \circ f)$ to the function $(\varphi \circ f)$ when d tends to 0. We obtain this result using the convergence of φ_d to φ , using the fact that Apprato et al., 1987, showed that, for any function g , we have $\lim_{d \rightarrow 0} \sigma_\varepsilon^d(g) = g$.

Keeping the notation of the previous sections, and since $(\varphi_d \circ f)$ is bounded in $C^m(\overline{\Omega})$, Apprato and Gout, 2000, proved that

$$\lim_{d \rightarrow 0} \sigma_\varepsilon^d(\varphi_d \circ f) = \varphi \circ f \text{ in } C^0(\overline{\Omega}). \tag{19}$$

From this result, using a compactness argument, Apprato and Gout, 2000 established a theoretical result concerning the convergence of the approximation:

$$\lim_{d \rightarrow 0} (\psi_d \circ \sigma_\varepsilon^d(\varphi_d \circ f)) = \varphi^{-1} \circ \varphi \circ f = f \text{ in } H^{m-\Theta}(\Omega), \tag{20}$$

for any $\Theta > 0$ such that $\Theta < m - 1$ (\Rightarrow continuous embedding of $H^{m-\Theta}(\Omega)$ into $C^0(\overline{\Omega})$). Note that if we take $n = 2$ and $m = 3$, the convergence takes place in $H^{2-\Theta}(\Omega)$ for any $\Theta \in]0, 1[$.

2.4 Numerical examples

The Piton de la Fournaise is a volcano located in the Indian Ocean, in the Réunion Island, France. This volcano exhibits strong topographic variations near its summit, due to the presence of a caldera and of two steep river valleys in its southwestern part, as can be seen on the picture of the volcano presented in Figure 3. The maximum height of the volcano is 2.6 km, and the depth of the valleys reaches more than 1000 m in several places. Being able to describe the topography of such regions exhibiting rapid local variations with at least C^0 regularity, or even C^1 regularity, is important in many fields in geophysics. For example, this description of the topography can be an input to numerical modeling codes that study the propagation of pyroclastic flows or lava flows, and related hazards; other examples are seismic site effects and ground motion amplification due to topographic features. In both cases, to avoid creating numerical artefacts, it is important not to introduce spurious oscillations in the description of the model itself. Otherwise, it is well known that in the context of curvilinear spectral element modeling of elastic wave propagation, artificial diffraction points appear at the edges between elements, which significantly affects the behavior of surface waves. To demonstrate the efficiency of our method, we create C^0 and C^1 approximants from a set of 8208 data points taken from a DEM of the summit. The data points in the DEM have been obtained by digitizing a map of the area. In this DEM, the height is given on an evenly spaced grid of 76×108 points, with a grid spacing of 200 m. Therefore the considered region has a dimension of 15 km in the East-West direction, and 21.4 km in the North-South direction. This DEM is shown in Figure 4 using a top view with isocontours representing the height of the topography every 0.2 km.

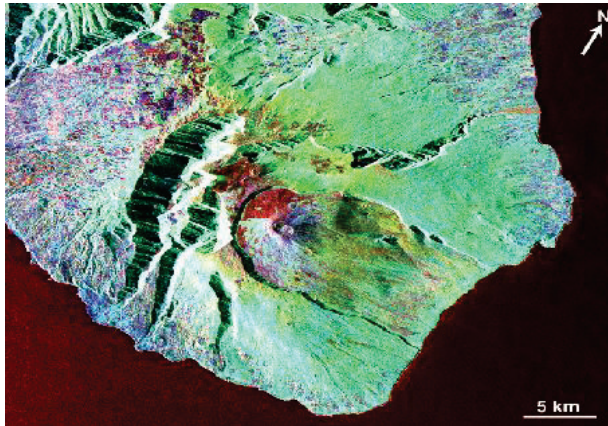


Fig. 3. Image of the Piton de la Fournaise volcano in the Réunion Island, Indian Ocean, France. One can clearly see the summital caldera, and the two steep valleys in the South-West. The size of the region represented is approximately 40×35 km. The height of the volcano is 2.6 km. Image taken as part of the Space Shuttle SIR-C/X-SAR radar missions, courtesy of Pete Mouginiis-Mark, University of Hawaii.

In the preprocessing step, we choose a regular distribution of the u_i in $[\alpha, \beta] = [0, 1]$ in order to reduce the large variations in the data set. The approximants are subsequently obtained by discretizing the D^m spline in a finite-element space. In the case of the C^0 approximant, we use 30×40 rectangular C^0 BFS finite elements, each having four degrees of freedom. In the case of the C^1 approximant, we use 15×20 rectangular C^1 -BFS finite elements, each having sixteen degrees of freedom. In both cases, the smoothing parameter ε is taken to be 10^{-6} .

In Figure 4, we show a three-dimensional representation of the C^1 approximant after postprocessing, evaluated on an evenly spaced grid comprising 200×200 points. The grid spacing in the East-West direction is therefore 107.54 m, and the one in the North-South direction is 75.37 m. From the figure it is clear that the results do not exhibit strong oscillations, even though the use of such a dense grid for the evaluation of the approximant is expected to enhance the artefacts generated by the approximation method. To compare this approximant to the original dataset more precisely, in Figure 5 we present a top view of the approximated values, with isocontours representing the height every 0.2 km, in addition to the same plot for the original dataset. It is clear from these plots that the approximant is very close to the original data, with local variations smoothed as expected. One can notice that the approximant does not exhibit significant oscillations even in the difficult regions of the model, particularly the two valleys. To demonstrate this more quantitatively, we evaluate the quadratic error for the two approximants. In the case of the C^0 approximant, we find that the error is 4.96×10^{-4} ; in the case of the C^1 approximant it is 4.01×10^{-4} . Such values are considered as very good ones in the context of surface approximation, and show the efficiency of the proposed approach for this case with rapidly varying data. In the entire dataset, the maximum error measured is 5.5%, corresponding to an absolute error of 56 m. This maximum error occurs in a region located on the edge of the steep valleys, where the local variations are the strongest, as expected. More detailed studies of the approximation error, and evidence that the rate of convergence is higher in this method than in usual approaches with no preprocessing, such as thin plate spline or splines under tension can be found in Schoenberg, 1960.

We have presented a new method to fit rapidly varying geophysical data. The ability to suppress, or at least significantly reduce, oscillations of the surface near steep gradients has been demonstrated. The scale transformation families introduced provide more control on the behavior of the approximant, without any particular *a priori* knowledge of the location of the large variations in the dataset. The regularity obtained, which can be C^0 , C^1 , or higher, enables us to describe the topography of real geophysical surfaces accurately. We have shown the good properties of this approach by applying it to the real case of the Piton de la Fournaise volcano.

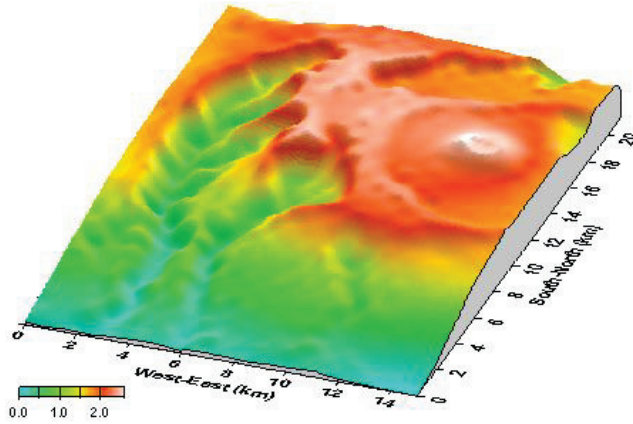


Fig. 4. Three-dimensional view of the C^1 approximant, after post-processing, obtained for the Piton de la Fournaise volcano from the Digital Elevation Model. The scale represents the height of the topography, from 0 to 2.6 km. The image has been generated with no vertical exaggeration. The approximant has been evaluated on an evenly spaced grid comprising 200 x 200 points. No significant oscillations can be observed, even in the difficult regions of the model, which are mainly the two valleys, and also the caldera. In this example, we have discretized the spline using 15 x 20 BFS finite elements, each having sixteen degrees of freedom.

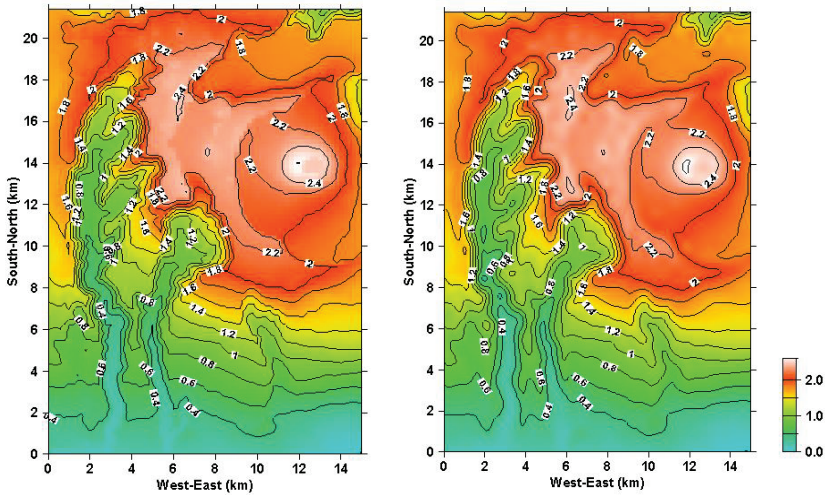


Fig. 5. Comparison between the isocontours obtained from the original dataset of the DEM. Left : Isocontours of the DEM of the Piton de la Fournaise volcano. The DEM is given on a grid of 76 x 108 points, with a uniform grid spacing of 200 m. The height of the summit is 2.6 km. One can clearly observe the slopes of the two steep valleys, and the isocontours of the C^1 approximant after postprocessing (right), as in the three-dimensional view of Figure 4.

The general agreement is excellent, and it is important to notice that no significant oscillations can be observed, even in the two steep valleys. Isocontours represent the height of the topography every 0.2 km. The gray scale also indicates the height of the topography, from 0 to 2.6 km.

3. Seafloor surface approximation from bathymetric dataset

3.1 Modelling

The problem of surface approximation from a given set of curves can be formulated as follows: from a finite set of curves $F_j, j = 1, \dots, N$ (the bathymetry ship track curves in our case) in the closure of a bounded nonempty open set $\Omega \subset \mathbb{R}^2$, and from a function f defined on $F = \bigcup_{j=1, \dots, N} F_j$, construct a regular function Φ on Ω approximating f on F , i.e.:

$$\Phi|_F \cong f|_F. \quad (21)$$

We can assume that Ω is a connected set, with a Lipschitz-continuous boundary (following the definition of Necas, 1967), that for any integer j , with $j=1, \dots, N$, F_j is a nonempty connected subset in F , and that, for simplicity, f is the restriction on F of a function, still denoted by f , that belongs to the usual Sobolev space $H^m(\Omega)$, with the integer $m > 1$. We also assume that the approximant Φ belongs to $H^m(\Omega) \cap C^k(\bar{\Omega})$ with $k = 1$ or 2 , where $\bar{\bullet}$ denotes the closure. The main interest of such a regularity for Φ is that it allows one to obtain a final surface that can later be used directly as an input model in a different application, such as ray tracing, image synthesis, or numerical simulation.

Let us define, for any v belonging to $H^m(\Omega)$, $\rho v = v|_F$ where ρ is a linear operator and let us introduce the convex set $K = \{v \in H^m(\Omega), \rho v = \rho f\}$. Then we consider the minimization problem of finding $\sigma \in K$ such that for any $v \in K$:

$$|\sigma|_{m,\Omega} \leq |v|_{m,\Omega}, \quad (22)$$

where $|\bullet|_{m,\Omega}$ is the usual semi-norm on $H^m(\Omega)$. If $L^2(F)$ is equipped with the usual norm

$$\|v\|_{0,F} = \left(\sum_{j=1}^N \int_{F_j} v^2 ds \right)^{1/2}, \quad (23)$$

and under the hypothesis that for any $p \in P_{m-1}(\bar{F})$, $p|_F = 0 \Rightarrow p \equiv 0$, we know, based upon a compactness argument (Necas, 1967), that the function $\|\bullet\|$ defined by

$$\|u\| = \left(\|\rho u\|_{0,\Omega}^2 + |u|_{m,\Omega}^2 \right)^{1/2} \tag{24}$$

is a norm on $H^m(\Omega)$ which is equivalent to the usual norm $\|\bullet\|_{m,\Omega}$ on $H^m(\Omega)$. Then the solution σ of the interpolation problem (22) is the unique element of minimal norm $\|\bullet\|$ in K that is convex, nonempty, and closed in $H^m(\Omega)$. Hence we could take the solution $\Phi = \sigma$ when $m > k + 1$. Unfortunately, it is often impossible to compute σ using a discretization of problem (22), because in a finite dimensional space, it is generally not possible to satisfy an infinity of interpolation conditions. Therefore, to take into account the continuous aspect of the data on F , we instead choose to define the approximant Φ as a fitting surface on the set:

$$\left\{ (x_1, x_2, x_3) \in \mathbb{R}^3, x_3 = f(x_1, x_2), (x_1, x_2) \in F_j, j = 1, \dots, N \right\}. \tag{25}$$

In this work, we propose to construct a variant of the “smoothing D^m -spline,” seen in previous sections, that will be discretized in a suitable piecewise-polynomial space. The use of such spline functions has been shown to be efficient in the context of geophysical applications such as Ground Penetrating Radar data analysis (Apprato et al., 2000) or the creation of Digital Elevation Models describing topography (Gout and Komatitsch, 2000).

Let us present in this section the theoretical aspects of the method. We first introduce a functional J_ε that we shall minimize, defined on $H^m(F)$ by

$$J_\varepsilon(v) = \|v - f\|_{0,F}^2 + \varepsilon |v|_{m,\Omega}^2, \tag{26}$$

where $\varepsilon |v|_{m,\Omega}^2$ is a smoothing term, $\varepsilon > 0$ being a classical smoothing parameter. The key idea

here is that the fidelity criterion to the data $\|v - f\|_{0,F}^2$ honors their continuous aspect. We

now need to numerically estimate this L^2 -norm, which is done using a quadrature formula.

In this respect, the approach is quite different from more classical techniques that usually simply make use of a large number of data points on F in order to solve the approximation problem. For any integer $j, j = 1, \dots, N$, and any $\eta > 0$, let $\{\zeta_i\}_{1 \leq i \leq L}$ be a set of $L = L(j)$ distinct

points $\zeta_i = \zeta_i(j) \in \overline{F}_j$ such that $\max_{1 \leq i \leq L-1} \delta(\zeta_i, \zeta_{i+1}) \leq \eta$, where δ is the Euclidean distance in

\mathbb{R}^2 . This relation implies that the distance between two consecutive ζ_i is bounded by η , it

also enables one to study the convergence of the approximation when η tends to 0. The $\{\zeta_i\}$

will also be the nodes of a numerical integration formula. Let us also introduce a set $\{\lambda_i\}_{1 \leq i \leq L}$ of real numbers (that will be the weights of a quadrature formula) such that

$\lambda_i = \lambda_i(j) > 0$, and let us define, for any $v \in C^0(\overline{F}_j), \forall \eta > 0$,

$$l_j^\eta(v) = \sum_{i=1}^L \lambda_i v(\zeta_i), \quad (27)$$

and for any $v \in C^0(\overline{F}_j)$,

$$l(v) = \sum_{j=1}^N l_j^\eta(v). \quad (28)$$

In what follows, we will suppose that, for any $v \in H^m(\overline{F})$ and any $\eta > 0$, there exists $C > 0$ such that

$$\left| l_j^\eta(v^2) - \|v\|_{0,F_j}^2 \right| \leq C\eta \|v\|_{m,\Omega}^2. \quad (29)$$

When this hypothesis is satisfied, one can consider l as a theoretical quadrature formula for $\|v\|_{0,F_f}^2$. Note that in some case ($N=1$), $l(v)$ is a quadrature formula for the curvilinear integral $\int_F v ds$. Note also that in most applications the F_j are polygonal curves, and one can,

therefore, use a classical quadrature formula (e.g., Arcangéli and Gout, 1976, or Gout and Guessab, 2001).

For the discretization, we proceed like in Section 2, using a finite element space.

3.2 Application to surface reconstruction from bathymetry ship track data in the Marianas trench

Detailed bathymetry maps are essential in several fields in geophysics, such as oceanography and marine geophysics. Historically, over the past decades, research vessels have collected a large number of depth echo soundings, also called SONAR (for "SONic Navigation And Ranging") bathymetry ship track data. Many of these measurements have been compiled to produce global bathymetry maps (e.g., Canadian Hydrographic Office in 1981). In recent years tremendous advances in satellite altimetry have enabled researchers to produce very detailed bathymetry maps independently of satellite gravity field measurements. However, long-wavelength variations of the depth of the ocean floor are difficult to constrain using satellite altimetry, and ship track data are still often used instead for that purpose. It is, therefore, of interest to address the issue of producing a bathymetry map from a given set of SONAR bathymetry ship tracks. Let us mention that SONAR ship tracks are typically acquired as a discrete set of measurement points, as opposed to continuous recording. However, the typical horizontal interval between measurement points is always small compared to expected bathymetry variations; therefore, in the context of this study the dataset can be considered as consisting of smooth continuous lines.

As presented in Url, 2009, in order to understand the shape of the seafloor, oceanographers go out on ships and collect sonar data. Sonar data are collected using echosounders and side-scan sonar systems. The digital data are then converted into maps and images.

How does this work?

Echosounders : Since World War II echosounders have been used to determine water depths of the oceans. Echosounders are usually attached to the hull of a ship. The echosounder sends an outgoing sound pulse into the water. The sound energy travels through the water to the ocean bottom where it is reflected back towards the source, received, and recorded.

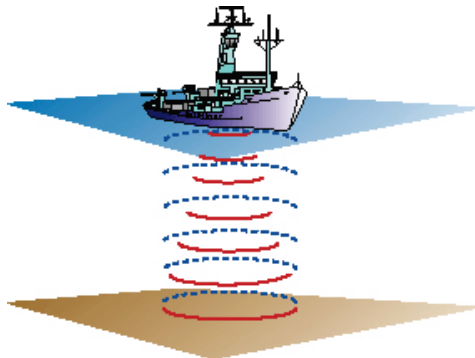


Fig. 6. Sound travels from the ship to the seafloor and is reflected back. The time it takes is converted into distance yielding water depth. (Credit : www.womenoceanographers.org)

The time that it takes for the sound to make the round trip to the seafloor is accurately measured. Water depth is determined from the travel time and the speed of sound in water. Water depth can be estimated simply by using an average sound speed and the following relationship: $\text{Distance} = \text{speed} \times \text{time}/2$, (the time is divided by 2 to take into account the round trip from the echosounder to the seafloor). The unique drawback of such an echosounder is that it will only give one depth at each time. That is why multibeam echosounder have been created...

How are water depths turned into a map?: As a ship steams ahead through the water, multibeam echosounders provide water depths for a swath of the seafloor. The water depths are located in space using satellite navigation. From these data, oceanographers can make maps of the seafloor that resemble topographic maps of land areas. In the early times, bathymetry maps were drawn by hand. Contours (lines) of equal water depth were drawn through a grid of numbers that had been plotted on a sheet of paper. Colors, put on by hand, indicated regions of equivalent water depth. Eventually computers took over and produced paper charts of the data, contoured and automatically colored. Now computer softwares enable individual scientists to process the data and display them on their own computer monitors. Maps can be imported into graphic software applications and annotations and other marks can be added.

"Side-scan"sonars: Similar to the multibeam echosounder, the sound transmitted by a side scan sonar instrument travels to the seafloor, bounces off the seafloor, returns to the instrument, and is recorded. In the case of a side-scan sonar, it is the intensity or strength of the returning acoustic signal that is recorded. This is controlled primarily by the slope of the seafloor and by what the seafloor is made of. A stronger return is received if the seafloor slopes down to the instrument. Also, the return is stronger if the seafloor is made of bare rocks. The strength of the return is much lower if the seafloor is covered by mud or sand. Volcanoes and other features that stick up above the surrounding seafloor will cast acoustic

shadows. These shadows are just like the shadow behind a person when a flashlight is shone on him.

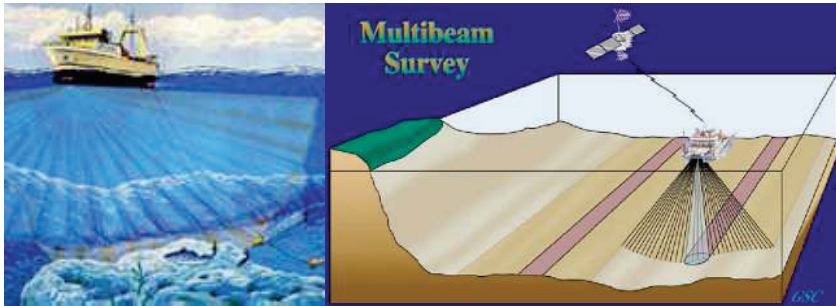


Fig. 7. Cartoon from a NOAA web site showing the swath of seafloor insonified by the multibeam echosounder. (Credit : NOAA and GSC)

Converting intensity into an image: The strength of the sound recorded by the side-scan sonar instrument is converted into shades of gray. A very strong return, say from bare rock, is white; a very weak return is black. The echo strengths that fall between these two extremes are converted into different shades of gray. Historically, side scan sonar data have been displayed on a hard copy paper recorder. The paper chart used to be the most convenient method for displaying and storing these data (as well as bathymetry data). Since the 1980s or so, software and hardware have been developed to process side scan data using computers and display the data on computer screens.

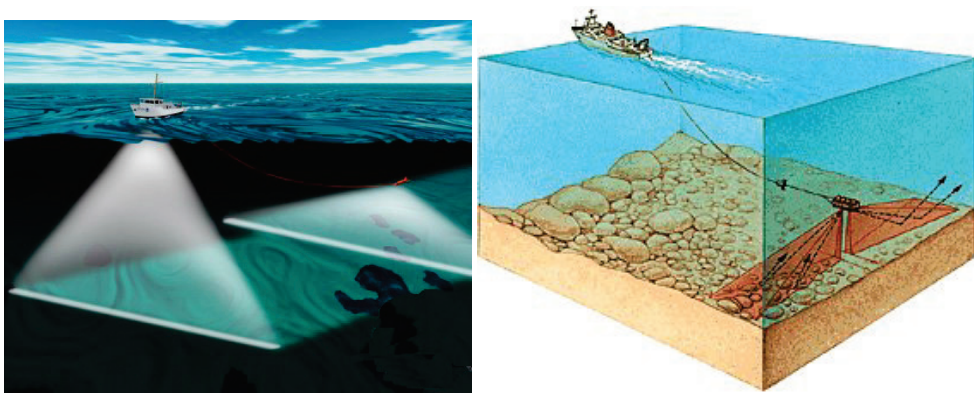


Fig. 8. Left : Hull-mounted multibeam sonar (left) and towed side scan sonar (right). NOAA - Right : The side scan sonar instrument is towed by the ship. Sound is transmitted into the water and images are made based on the strength of the recorded return. (credit : NOAA)

We select the region of the Marianas trench (Figure 9). The trench is located in the North Pacific ocean, east of the South Honshu ridge, parallel to the Mariana Islands. It corresponds to the subduction zone where the fast-moving Pacific plate converges against the slower moving Philippine plate. It is also the place on Earth where the oceans are the deepest, reaching a maximum depth of slightly more than 11 km in the so-called “Challenger Deep”

area (Figure 9). This region is ideal to test our surface approximation technique because it has been thoroughly studied; therefore, many ship track datasets are available. We select a 45×45 km area, corresponding to latitudes between 11.2^\pm and 11.6^\pm North, and longitudes between 142^\pm and 142.4^\pm East. We use 16 tracks from the database assembled by **David T. Sandwell** and coworkers at the University of California, San Diego (<http://topex.ucsd.edu>). Each individual track contains between 62 and 152 points giving depth for a given latitude and longitude. The total number of points in the whole dataset is 1576. The depth varies between 6779 and 10952 m. As can be seen on Figure 9, the ship track coverage of the area is nonuniform. Note in particular the lack of data in the north-east and south-east corners. Fortunately, data coverage is much better near the center in the deepest part of the trench. We create an approximant using 169 quadrangular Bogner-Fox-Schmit finite elements defined on a regular 13×13 grid in the horizontal plane in the area under study. As underlined in the previous section, these elements enable us to obtain an approximant with C^1 regularity. Figure 9 shows a 3D view of the final surface obtained, as well as the original set of ship tracks. For display purposes, the approximant has been evaluated on a regular 200×200 grid of points and a vertical exaggeration factor of 3 has been applied. By comparing with Figure 9 and with the ship tracks, one can see that the smooth surface obtained correctly reproduces the general characteristics of the bathymetry of the region, and behaves satisfactorily even in the areas where the data coverage is sparse.

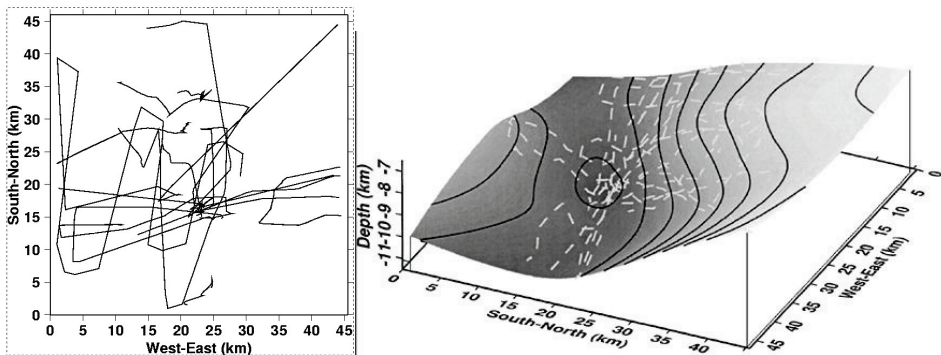


Fig. 9. **LEFT** : We focus on a 45×45 km region in the south-west of the Marianas trench. We use 16 bathymetry ship tracks, each containing between 62 and 152 points. The entire set of curves contains 1567 points. Each point gives depth for a given latitude and longitude. On this top view the coordinates have been mapped using the Universal Transverse Mercator (UTM) projection. The depth in the dataset varies between 6779 and 10952 m. One can see that the ship track coverage is nonuniform. For instance we have little information in the north-east and south-east corners of the area. **RIGHT** : We construct a bathymetry map from the set of 16 ship track data curves using a regular grid of 13×13 quadrangular Bogner-Fox-Schmit finite elements of class C^1 . For display purposes, the approximant obtained has been evaluated on a regular 200×200 grid of points, and a vertical exaggeration factor of 3 has been applied. The original 16 ship tracks are also shown (dashed lines) to illustrate the quality of the obtained surface. The isolines represent bathymetry every 500 m. By comparing with Figure 9 (right), one can see that we are correctly reproducing the general trends of the bathymetry of the area.

The quadratic error is equal to 3.29×10^{-5} , which is a very satisfactory result (unusually low in the context of surface approximation, e.g., Apprato *et al.* (2002); as a comparison, a usual D^m -spline (Arcangéli, 1989) applied to the same data set using the same finite-element grid gave an error of 6.4×10^{-4}).

3.3 Application to surface reconstruction from bathymetry ship track data and Lagrange data around Hawaiian hot spot

To demonstrate the efficiency of our method, we also give a numerical example from a set of 7049 data points (both bathymetry and Lagrange data) around the Big Island in Hawaii.



Fig. 10. Hawaiian Islands. From lower right to upper left, the Big Island (Hawaii), Maui, Kahoolawe, Lanai, Molokai, Oahu, Kauai, and Niihau islands all make up the state of Hawaii, which lies on more than 2,000 miles from any other part of the United States. The small red dot on the Big Islands southeastern side denotes a hot spot on Kilauea Volcanos southern flank. Kilauea has been erupting almost continuously since January 1983, and is one of the world's best studied volcanoes.

The maximum height of the big island is 4.7 km, and the depth of the seafloor reaches more than 4 km in several places. To get seafloor data, radar scan sonar are used. From the dataset, with the knowledge of the large variations, we have made a triangulation on each region using the software Mefisto. We give the C^1 approximant (see below). We also evaluate the approximant obtained at the 7049 data points of the dataset. To estimate the error quantitatively, we then evaluate the quadratic error on the dataset: we obtain a value of $5.65 \cdot 10^{-5}$, which is a satisfactory result.

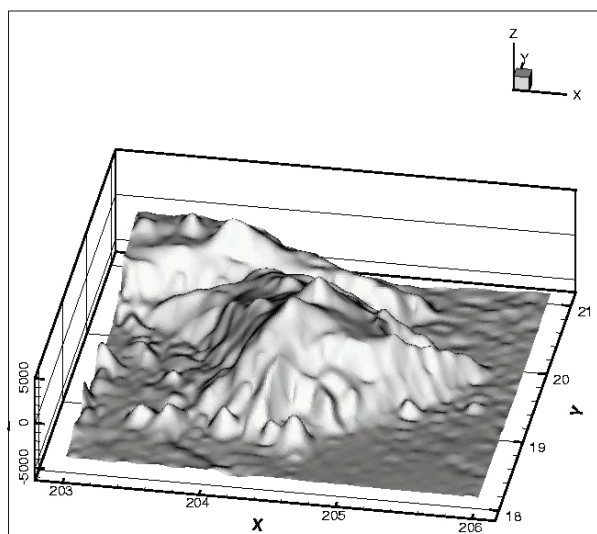


Fig. 11. A 3D view of the C^1 approximant of the Big Island, Hawaii.

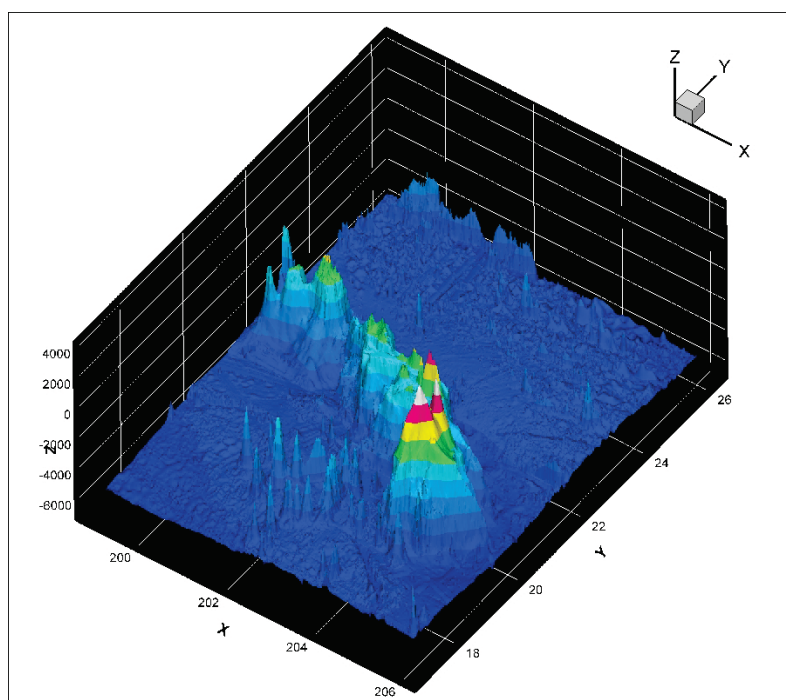


Fig. 12. A 3D view of the entire zone of the Hawaiian islands from a large dataset (one million data points).

4. Surface approximation from surface patches

The problem of constructing a surface from given patches on this surface appears, for instance, in geophysics or geology processes like migration of time-maps or depth-maps. The problem of surface approximation from surface patches can be posed as follows: from a finite set of open subsets ω_j , $j = 1, \dots, N$ (surface patches in our case) in the closure of a bounded nonempty open set $\Omega \subset \mathbb{R}^2$, and from a function f defined on $\Xi = \bigcup_{j=1, \dots, N} \omega_j$, construct a regular function Φ on Ω approximating f on Ξ , i.e.: $\Phi|_{\Xi} \cong f|_{\Xi}$.

The modelling corresponds to the one of the case of bathymetry dataset. The difference only rests on the fidelity criterion (see (23) for the case of bathymetry dataset) which is:

$$\|v\|_{0,\Xi} = \left(\sum_{j=1}^N \int_{\omega_j} v^2(x) dx \right)^{1/2}.$$

The functional (26) becomes: $J_\varepsilon(v) = \|v - f\|_{0,\Xi}^2 + \varepsilon \|v\|_{m,\Omega}^2$, and the introduced minimization problem has a unique solution (using Lax-Milgram lemma, see Gout (2002), or Apprato *et al.* (2000)). The main difficulty consists in approximating $\|v - f\|_{0,\Xi}^2$. It is done using quadrature formulae. The discretization is done using the finite element method. Convergence results and numerical examples are given in Apprato *et al.* (2000) and in Gout (2002).

5. Non regular Surface approximation: application in Geosciences

5.1 Modelling

The right approach to get a good approximant of a non regular surface consists in applying first a segmentation process to precisely define the locations of large variations and faults, and exploiting then a discrete approximation technique. To perform the segmentation step, we propose a quasi-automatic algorithm that uses a level set method to obtain from the given (gridded or scattered) Lagrange data several patches delimited by large gradients (or faults). Other approaches for fault detection can be found in Gutzmer and Iske 1997, or in Parra *et al.* 1996.

Then, with the knowledge of the location of the discontinuities of the surface, we generate a triangular mesh (which takes into account the identified set of discontinuities) on which a D^m -spline approximant is constructed. To show the efficiency of this technique, we will present the results obtained by its application to a dataset in Geosciences.

The main goal of this work is thus to give a quasi-automatic algorithm to determine the location of the large variations and the faults of the surface in order to use specific methods that rely on splines under tension with a nonconstant smoothing parameter near the identified set of discontinuities. Likewise, if one wants to use finite element methods in the discretization step, it is well known that to correctly reproduce the set of surface

discontinuities (both of the function—faults—and/or its derivatives—creases), there are some constraints regarding the triangulation of the domain of definition of the function: in particular, as shown by Arcangéli et al. , 1997, the edges of the triangles of the triangulation should not intersect the set of discontinuities, here denoted by D (Figure 13).

As a consequence, it is generally necessary to consider a number of different connected open subsets F_i , commonly called patches (Figure 13), and mesh them as done in Figure 13. Let us note that to improve the results, an *adaptive mesh refinement* can be made near the set D .

Then, it would be possible to use a finite element approximant. Unfortunately, because of difficulties linked to the geometry and the number of data points, finite element methods turn out to be hard to use. In this work we will use therefore a D^m -spline approximant whose definition takes into account the particular structure of the surface domain, as introduced in Arcangéli et al. , 1997. The algorithm is summarized in Diagram 1. To have more details about the segmentation method that will be exploited to locate the set of discontinuities D of the surface, please see Gout et al. 2008. To do the segmentation process, the input surface is converted into a grayscale image composed of pixels whose brightness values are given by the z -coordinate of the data points on each node of a regular grid. So, it is easy to apply segmentation tools developed in image processing (see Le Guyader et al., 2005, or Caselles et al., 1997, or Gout and Le Guyader, 2006 and 2008, or Gout et al. 2005, or Forcadel et al., 2008,) to surface approximation applications. As mentioned in Diagram 1, it would be possible to also work with random datasets on a surface (see Gout et al., 2008), but this paper primarily aims at showing the efficiency of the proposed strategy in the case of regularly distributed points. We then use a finite element method to mesh the surface taking into account the identified set of discontinuities D . The approximation operator and the convergence of the method when the number of data tends to infinity is discussed in Arcangéli et al. 1997.

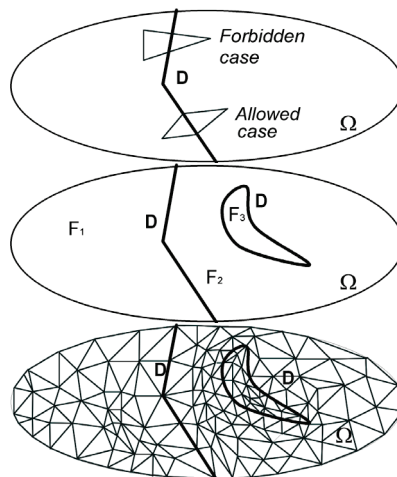


Fig. 13. **Top** : Examples of prohibited and allowed triangles in the domain triangulation: the identified set of discontinuities D must be taken into account. **Middle** : Example of a set of discontinuities D and the three different subsets F_1, F_2, F_3 it delineates. **Bottom** : Following the set of discontinuities D , a triangulation is made: no triangle intersects the set of discontinuities.

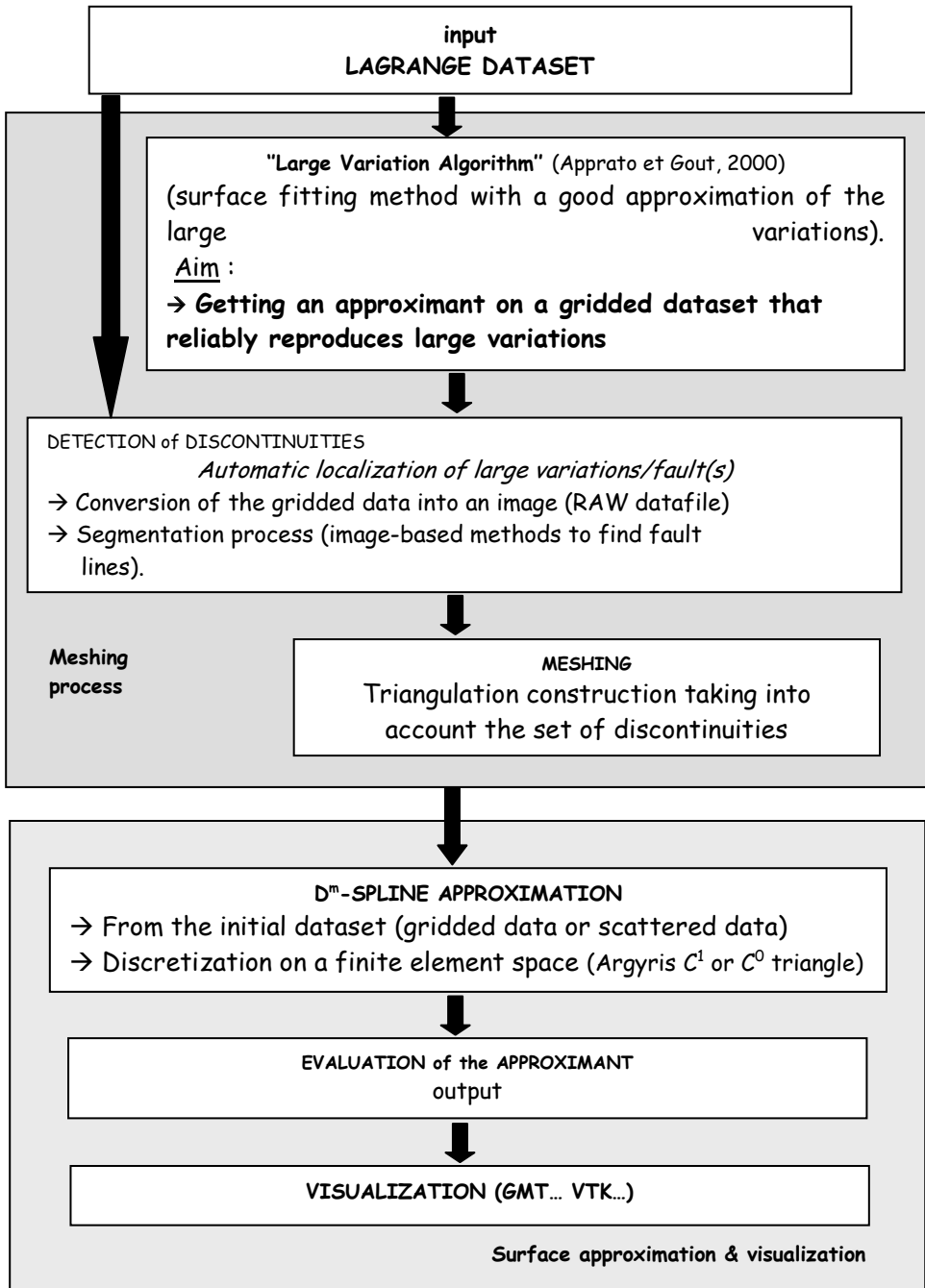


Diagram 1. Presentation of the modelling.

5.2 Numerical examples

In this section we give an example based on real data values which come from the Vallée d'Ossau, Pyrénées mountains, France. The ground penetrating radar (GPR) technique is based on the principle that high-frequency electromagnetic waves may be reflected at boundaries separating heterogeneous regions of the subsurface. This technique is a very high resolution geophysical tool, with a penetration depth of a few tens of meters (100 m in the best conditions), depending on the underground physical properties and on the radar wave frequency used. Usually, GPR surveys are conducted with a constant offset between transmitter and receiver and a single-fold coverage. The time unit is the nanosecond, the frequency range is between 10 MHz and 1 GHz. The propagation velocity range is between 0.01 and 0.3 m/ns, for example 0.12 m/ns in limestone, 0.07 m/ns in silts. In this study, we have used a frequency of 100 MHz, an offset of 1m and we have obtained a penetration depth of about 10 m.

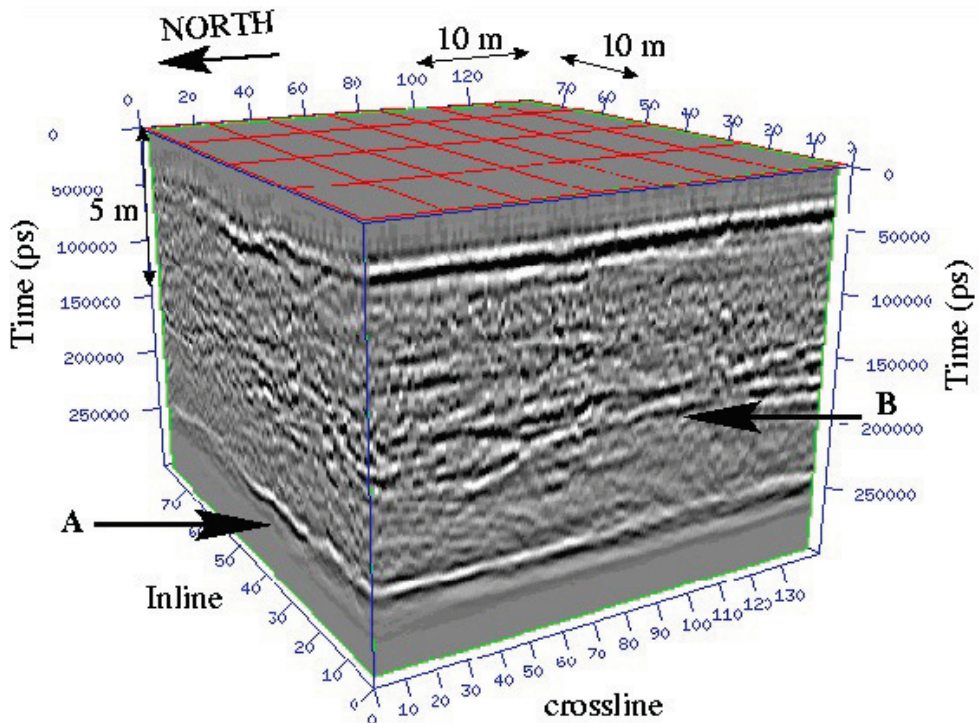


Fig. 14. View of the Three-dimensional data block. A and B correspond to two horizons. In this work we will use the dataset corresponding to Horizon A.

The studied area is located in the Western Pyrénées (30 km south of Pau, Béarn, France) in the Vallée d'Ossau which is an old glacial valley. 2D experiments were conducted and gave information on the sub-surface structures in the area. The interface between the limestone bedrock and the fluvio-glacial deposits has a depth varying between 2-3 and 12-13 m with a weak general dip from the north to the south (about 2). Above this interface, the fluvio-

glacial deposits show several sedimentary figures, with meter or decameter scale, which could correspond to old fluvial channels. However, this 2D acquisition cannot give us enough information to precisely describe the structures present on the site. The solution was to conduct a 3D GPR experiment on a part of the area, in order to obtain 3D information. So, a 3D GPR data acquisition on the area described above has been conducted. The single-fold GPR data were acquired along north-south profiles. The acquisition area corresponds to a rectangle of $38\text{m} \times 35\text{m}$. Throughout the whole acquisition work, a constant distance of 1m was maintained between the transmitter and receiver antennae, and both antennae were oriented perpendicularly to the profile direction. Each trace was vertically stacked 256 times in the field. The sampling rate was 0.676ns and the trace length 300ns. Figure 14 shows the 3D view of the cuboid. Continuity of reflectors is better in the inline direction because the number of traces is higher (141 for inline instead of 77 for crossline sections). A strong and continuous reflector (called Horizon A) appears at about 250 ns and is present on all the traces. According to the field observations and the first interpretation with the 2D profiles, this lowermost reflector can be interpreted as the interface between the limestone bedrock and the fluvio-glacial deposits. In order to test our algorithm, we have modified the dataset to create a fault. 2D and 3D views of the considered dataset are given in Figure 15 and 16 respectively.

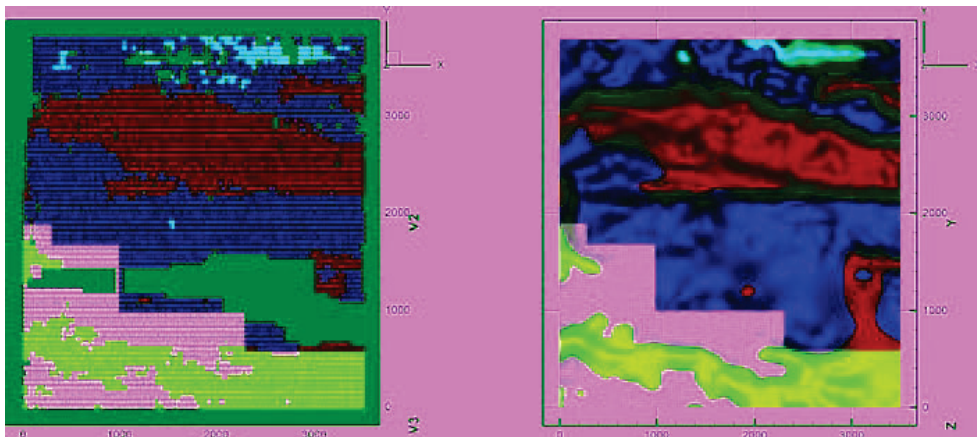


Fig. 15. *Left*: Two-dimensional view of the geological dataset (8949 data points) containing a fault in correspondence to the boundaries with the *white regions*. *Right*: Two-dimensional view of the locally C^1 approximant of the geological surface with the fault.

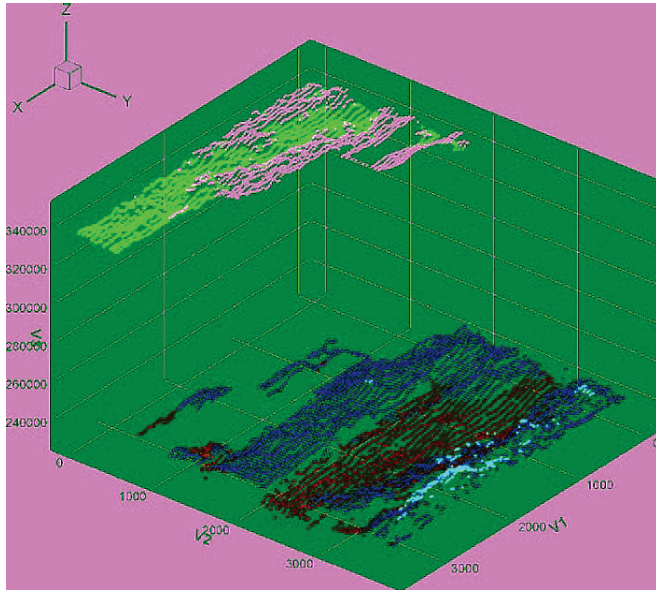


Fig. 16. Three-dimensional view of the dataset corresponding to the geological surface with the fault (8949 data points).

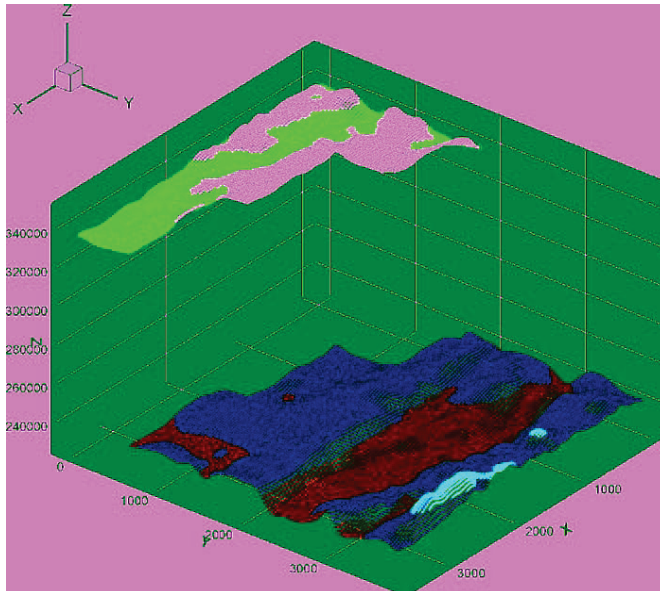


Fig. 17. Three-dimensional view of the locally C^1 approximant of the geological surface with the fault. The approximant has been evaluated on an evenly space grid made of 150×150 points. The same scale and colormap are used for each surface. The general agreement is excellent. No oscillations are present.

Here are the parameters used when running the proposed algorithm on this second real world example:

- The triangulation is made of 400 triangles;
- The adopted generic finite element is the Argyris triangle (class C^1);
- The smoothing parameter ε is chosen equal to 10^{-6} ;
- The evaluation grid is 150×150 .

The quadratic error on the 8949 data points turns out to be equal to $2.332 \cdot 10^{-3}$. The obtained approximant is depicted in Figures 15 (right) and 17 where a 2D and a 3D view are given respectively. As it appears, no oscillations are present and excellent reconstruction results can be achieved.

6. Conclusion

We have developed a novel and efficient algorithm for approximating non regular gridded data exhibiting large and rapid variations and/or complex fault structures. The main steps of the proposed strategy consist in (1) accomplishing a pre-processing phase to define the set of discontinuities of the surface, (2) generating a triangular mesh taking into account these discontinuities, (3) applying a specific (known) finite element domain decomposition method and using a spline operator that relies on scale transformations (which seems to be very useful in controlling the behavior of the surface in the presence of steep gradients not found by the segmentation process) to produce the final approximating surface. Compared with conventional data fitting methods that exist in the literature, the proposed algorithm is able to suppress or at least decrease the undesired oscillations that generally arise near steep gradients, thus ensuring a faithful and accurate representation. The presented numerical examples illustrate the efficacy of the method.

7. References

- Apprato, D., Gout, C. (2000). A result about scale transformation families in approximation: application to surface fitting from rapidly varying data. *Numer. Algorithms* **23**(2-3), 263-279.
- Apprato, D., Arcangéli, R., Manzanilla, R. (1987). Sur la construction de surfaces de classe C_k à partir d' un grand nombre de données de Lagrange *M2AN* **21**(4), 529-555.
- Apprato, D., Gout, C., Sénéchal, P. (2000). C_k reconstruction of surfaces from partial data. *Math. Geol.* **32**(8), 969-983.
- Apprato, D., Gout, C., Komatitsch, D. (2002). A new method for C_k approximation from a set of curves: application to ship track data in the Marianas trench. *Math. Geol.* **34**(7), 831-843.
- Arcangeli, R., (1989). Some applications of discrete D_m -splines, In *Mathematical Methods in Computer Aided Geometric Design*, (Edited by T. Lyche and L.L. Schumaker), pp. 35-44, Academic Press, New York.
- Arcangéli, R., Manzanilla, R., Torrens, J.J. (1997). Approximation spline de surfaces de type explicite comportant des failles. *Math. Model. and Numer. Anal.* **31**(5), 643-676 (1997)
- Arcangéli, R., Lopez de Silanes, M.C., Torrens, J.J.(2004). Multidimensional minimizing splines. Theory and Applications. ISBN: 1-4020-7786-6, Kluwer Academic Publishers, Boston.

- Arcangeli, R., Gout, J.L. (1976). Error estimates for Lagrange interpolation—Sur l'évaluation de l'erreur d'interpolation de Lagrange dans un ouvert de R^n . *Math. Model. and Numer. Anal.*, *RAIRO Anal. Numer.* **10**(3), 5–27.
- Bouhamidi, A., Le Méhauté, A. (1999). Multivariate interpolating (m, l, s)-splines. *Adv. Comput. Math.* **11**(4), 287–314.
- Caselles, V., Kimmel, R., Sapiro, G. (1997). Geodesic active contours, geodesic active contours. *Int. J. Comput. Vis.* **22**(1), 61–87.
- Ciarlet, P.G. (1977). The finite element method for elliptic problems. North Holland, Amsterdam.
- Franke, R. (1982). Smooth interpolation of scattered data by local thin plate splines. *Comp. Math. Appl.* **8**(4), 273–281.
- Forcadel, N., Le Guyader, C., Gout, C. (2008). Generalized fast marching method: applications to image segmentation. *Numer. Algorithms* **48**, no. 1-3, 189–211.
- Gout, C., Le Guyader, C. (2006). Segmentation of complex geophysical structures with well data. *Comput. Geosci.* **10**(4), 361–372.
- Gout, C., Komatitsch, D. (2000). Surface fitting of rapidly varying data using rank coding: application to geophysical surfaces. *Math. Geol.* **32**(7), 873–888.
- Gout, C., (2002). An algorithm for C^k surface approximation with large variations. *Int. J. of Comp. Math.* **79** (1), pp. 111–131.
- Gout, C., Le Guyader, C., Vese, L. (2005). Segmentation under geometrical conditions using geodesic active contours and interpolation using level set methods. *Numer. Algorithms* **39**(1–3), 155–173.
- Gout, C., Guessab, A., (2001). A new family of Extended Gauss Quadratures with an Interior Constraint, *J. of Computational and Applied Mathematics*, Vol 131/1-2, pp 35-53.
- Gout, C.; Le Guyader, C.; Romani, L.; Saint-Guirons, A.-G (2008). Approximation of surfaces with fault(s) and/or rapidly varying data, using a segmentation process, D -splines and the finite element method. *Numer. Algorithms* **48**, no. 1-3, 67 – 92.
- Gutzmer, T., Iske, A. (1997). Detection of discontinuities in scattered data approximation. *Numer. Algorithms* **16**(2), 155–170 (1997)
- Hsieh, H.-C., Chang, W.T. (1994). Virtual knot Technique for Curve Fitting of Rapidly Varying Data, *Computer Aided Geometric Design*, Vol. 11, 71-95.
- Issaks, E. H., and Srivastava, R. M. (1989). An introduction to applied geostatistics: Oxford University Press, Oxford.
- Le Guyader, C., Apprato, D., Gout, C. (2005). Using a level set approach for image segmentation under interpolation conditions. *Numer. Algorithms* **39**(1–3), 221–235.
- Le Guyader, C., Gout, C. (2008). Geodesic active contour under geometrical conditions: theory and 3D applications. *Numer. Algorithms* **48**, no. 1-3, 105 – 133.
- Manzanilla, R.: (1986). Sur l'approximation de surfaces définies par une équation explicite, Thèse, Université de Pau, Pau, France.
- Necas, J. (1967). Les méthodes directes en théorie des équations elliptiques. Masson, Paris.
- Nielson, G.M., Franke, R. (1984). A method for construction of surfaces under tension. *Rocky Mt. J. Math.* **14**, 203–221.
- Parra, M.C., Lopez de Silanes, M.C., Torrens, J.J. (1996). Vertical fault detection from scattered data. *J. Comput. Appl. Math.* **73**(5), 225–239.
- Salkauskas, K. (1974). C_1 splines for interpolation of rapidly varying data. *Rocky Mt. J. Math.* **14**, 239–250.

- Schoenberg, I.J. (1960). Contribution to the problem of approximation of equidistant data by analytic functions. *Q. Appl. Math.* **4**, 45-99 and 112-141.
- Torrens, J.J. (1991). Interpolacion de superficies parametricas con discontinuidades mediante Elementos Finitos, Aplicaciones, Thèse, Universidad de Zaragoza, Spain.
- Url, (2009). Internet link : <http://www.womenoceanographers.org>

Three-Dimensional Microwave Imaging using Synthetic Aperture Technique

Shi Jun, Zhang Xiaoling, Yang Jianyu, Liao Kefei and Wang Yinbo
*University of Electronic Science and Technology of China
Chengdu, P.R.China*

1. Introduction

With the ability of two-dimensional (2-D) microwave imaging, synthetic aperture radar (SAR) has been an important imaging tool for civilian and military applications. The basic idea of 2-D SAR is to synthesize a linear array by moving a high-range-resolution (HRR) radar along a straight path, and obtain the additional azimuthal resolution. To extract the height information from the 2-D SAR, interferometric SAR (InSAR) technique, which requires multiple antennas or repeated flight paths, has been developed and is widely used for remote sensing applications.

However, since the interferometric SAR technique is based on the 2-D SAR images, it will be invalid, when there is more than one scatterer projected in the same pixel of the 2-D SAR image. This disadvantage makes it difficult to be used in high-precision 3-D RCS measurement and topographical survey in urban region.

To improve the ability of microwave remote sensing, some new 3-D SAR systems, such as circle SAR, elevation circular SAR, curve SAR, and linear array SAR have been developed based on the synthetic aperture technique. The basic idea of them is to produce 2-D resolution by moving the HRR radar in 2-D / 3-D space and obtain the third dimensional resolution using pulse compression technique. This chapter will discuss the principle and imaging processing technique of 3-D SAR.

In section 2, an approach to calculate the oscillatory integral has been introduced, which could simplify the analysis of 3-D SAR ambiguity function. In section 3, the ambiguity function and spatial resolution of 3-D SAR are discussed. The backprojection method and experiment data processing are presented in section 4 and 5 respectively. The multiresolution approximation techniques that can reduce the computational cost of 3-D SAR are discussed in section 6.

2. Preliminary

Compared with the traditional SAR, the echo model of 3-D SAR is more complex. To simplify the analysis of 3-D SAR, the calculation of oscillatory integral using density function will be introduced in this section.

2.1 A Simple Example

At first, let's observe a simple example. Assume that there is a discrete function $f(n) = (0, 1/4, 2/4, 0, 1/4, 1/4)$, and we need to calculate the sum of $\exp(j \cdot 2\pi \cdot f(n))$. Obviously, we have:

$$\sum_n \exp(j \cdot 2\pi \cdot f(n)) = e^{j \cdot 2\pi \cdot 0} + e^{j \cdot 2\pi \cdot 1/4} + e^{j \cdot 2\pi \cdot 2/4} + e^{j \cdot 2\pi \cdot 0} + e^{j \cdot 2\pi \cdot 1/4} + e^{j \cdot 2\pi \cdot 1/4} \quad (1)$$

According to the commutative law of addition, eq. (1) can be rewritten as:

$$\sum_n \exp(j \cdot 2\pi \cdot f(n)) = 2 \cdot e^{j \cdot 2\pi \cdot 0} + 3 \cdot e^{j \cdot 2\pi \cdot 1/4} + 1 \cdot e^{j \cdot 2\pi \cdot 2/4} + 0 \cdot e^{j \cdot 2\pi \cdot 3/4} \quad (2)$$

where, coefficients 2, 3, 1 and 0 represent the frequency (in the sense of probability, which is defined as the number of times that value occurs in the data set) of every exponential term. Thus, by introducing the concept of density function $\mathcal{D}(i) = \{2, 3, 1, 0\}$, we have:

$$\sum_n \exp(j \cdot 2\pi \cdot f(n)) = \sum_{i=0}^3 \mathcal{D}(i) e^{j \cdot 2\pi \cdot i/4} \quad (3)$$

Obviously, eq. (3) matches the definition of Discrete Fourier Transform (DFT). Denoting the DFT of $\mathcal{D}(i)$ as $\widehat{\mathcal{D}}(k)$, we have:

$$\sum_n \exp(j \cdot 2\pi \cdot f(n)) = \widehat{\mathcal{D}}(1) \quad (4)$$

Similarly, we have:

$$\sum_n \exp(j \cdot 2\pi \cdot k \cdot f(n)) = \sum_{i=0}^3 \mathcal{D}(i) e^{j \cdot 2\pi \cdot k \cdot i/4} = \widehat{\mathcal{D}}(k) \quad (5)$$

This example indicates that the exponential sum of a finite discrete-time function can be calculated using its density function. However, since the commutative law of addition holds only when $f(n)$ is finite, and the concept of frequency is meaningful for finite set, a more precise definition of density function using Lebesgue measure and a theorem that extends eq. (4) to the continuous-time function will be presented in next subsection.

2.2 Calculation of the oscillatory integral using density function

According to measure theory, functions are divided into four classes, simple function, Bounded Function Supported on a set of Finite Measure (BFFM), non-negative function and integrable function (the general case). For the analysis of array whose size is finite, the BFFM assumption is sufficient.

Given a BFFM $f(t)$ with support F , analogous to the definition of Cumulative Density Function (CDF) in probability theory, define the Cumulative Density Function of $f(t)$ as:

$$C_f(y) \triangleq m(\mathbf{F}_y), \quad (6-1)$$

$$F_y \triangleq \{t : f(t) \leq y; y \in \mathbb{R}\}, \tag{6-2}$$

where, F_y is the subset of F . $m(F_y)$ denotes the Lebesgue measure of F_y , which describe the volume (area) of F_y . Especially, when F is finite set, $m(F_y)$ is the cardinality of subset F_y .

Then we define the Density Function (DF) of $f(t)$ as the derivative of $C_f(y)$:

$$\mathcal{D}_f(y) \triangleq dC_f(y)/dy \tag{7}$$

Obviously, $\mathcal{D}_f(y)$ satisfies:

1. $\mathcal{D}_f(y) \geq 0$;
2. $\int_{-\infty}^{+\infty} \mathcal{D}_f(y)dy = m(F) < +\infty$;

Properties 1 and 2 of $\mathcal{D}_f(y)$ indicate that $\mathcal{D}_f(y)$ is absolutely integrable, and its Fourier transform exists.

Using the concept of density function, we can calculate an oscillatory integral using the following theorem.

Theorem 1: Given a BFFM phase function $f(t)$ supported on a set E , we have:

$$\int_E e^{if(t)} dt = \widehat{\mathcal{D}_f}(1) \tag{8}$$

where, $\widehat{\mathcal{D}_f}(\bullet)$ denotes the Fourier transform of $\mathcal{D}_f(y)$

Proof:

The proof of theorem 1 includes two steps: firstly, we consider $f(t)$ as a simple function, and then extend the conclusion to BFFM function.

According to the definition in measure theory, a simple function is a finite sum of a group of characteristic functions:

$$f(t) = \sum_{k=1}^K a_k \chi_{E_k}(t) \tag{9-1}$$

$$\chi_{E_k}(t) = \begin{cases} 1 & t \in E_k \\ 0 & t \notin E_k \end{cases} \tag{9-2}$$

where, a_k is constant, E_k denotes a measurable subset of set E , $\chi_{E_k}(t)$ denotes the characteristic function of E_k .

● Case 1: rational number

Assume that a_k are all rational number, there exists an equal-interval infinite rational number set:

$$\mathbf{B} = \{-\infty, \dots, \frac{-1}{Q}, 0, \frac{1}{Q}, \dots, +\infty\}, Q \in \mathbb{N} \tag{10}$$

satisfying $a_k \in \mathbf{B}$ (in the example, $\{a_k\} = \{0, 1/4, 2/4\}$, $\mathbf{B} = \{0, 1/4, 2/4, 3/4\}$).

Using set \mathbf{B} , we can construct a group of characteristic function $\chi_{F_i}(t)$, and $f(t)$ could be written as:

$$f(t) = \sum_{i=-\infty}^{+\infty} b_i \chi_{F_i}(t) \tag{11}$$

where, $b_i \in \mathbf{B}$; $F_i = E_k$, when $b_i = a_k$; otherwise, $F_i = \Phi$.

According to the definition of Lebesgue integral, we have:

$$\int_E e^{jf(t)} dt = \sum_{i=-\infty}^{+\infty} e^{jb_i} m(F_i) = \sum_{i=-\infty}^{+\infty} e^{jb_i} \mathcal{D}_f(i) \tag{12}$$

Case 2: real number

Assume that a_k are all real number, according to the real analysis, for every real number a , there exists a sequence $\{a_n\}$ of rational numbers can approximate to it. Thus, there exists a sequence $\{\mathbf{B}_Q\}$ of rational number sets can approximate to all of the a_k , i.e.:

$$\int_E e^{jf(t)} dt = \lim_{Q \rightarrow +\infty} \sum_{i=-\infty}^{+\infty} e^{jb_i^Q} \mathcal{D}_f^Q(i), \tag{13}$$

with the increase of Q , the interval $1/Q$ of \mathbf{B} trends toward zero, and eq. (13) can be written in integral form as:

$$\int_E e^{jf(t)} dt = \int_{-\infty}^{+\infty} e^{jy} \mathcal{D}_f(y) dy. \tag{14}$$

Since the Fourier transform of $\mathcal{D}_f(y)$ exists, we have:

$$\int_E e^{jf(t)} dt = \widehat{\mathcal{D}}_f(1) \tag{15}$$

For a BFFM function $f(t)$ bounded by M and supported on a set E , there exists a sequence $\{f_n\}$ of simples functions, with each f_n bounded by M and supported on a set E , and such that:

$$f_n(t) \rightarrow f(t) \text{ for all } t. \tag{16}$$

Thus, eq. (8) holds for all BFFM functions.

□

Theorem 1 provides a method to calculate the oscillatory integral without any approximation. Compared with the principle of stationary phase (PSP), this method does not need $f(t)$ be derivable, and holds for all BFFM function.

Using theorem 1, we can obtain the following corollary directly by rewritten eq. (12) as:

$$\int_E e^{j-u \cdot f(t)} dt = \int_{-\infty}^{+\infty} e^{j-u \cdot y} \mathcal{D}_f(y) dy = \widehat{\mathcal{D}}_f(u) \tag{18}$$

Corollary 1:

$$\int_E e^{j-u \cdot f(t)} dt = \widehat{\mathcal{D}}_f(u) \tag{19}$$

As it will be seen in the next section, this corollary is crucial for the analysis on the ambiguity function of 3-D SAR.

3. Principle of 3-D SAR

3.1 Introduction on the typical 3-D SARs

In this subsection, a brief discussion on the typical 3-D SAR systems including circle SAR (CSAR), elevation circular SAR (E-CSAR), curve SAR, and linear array SAR (LASAR) will be proposed.

In 1999, Tsz-King Chan, Yasuo Kuga, and Akira Ishimaru proposed a novel method for radar topographical imaging which required the SAR platform to move in a circular orbit, and named it as circular SAR. In their experiment, the transmitting and receiving antennas were mounted on two separate wooden rings that were individually driven by stepping motors with an angular precision of approximately 0.02. Imaging result of a model helicopter of length 30 cm has been obtained and published (Tsz-King Chan; Kuga, Y.; Ishimaru, A., 1999). In 2001, at the Radar Division of Georgia Tech Research Institute (GTRI), a 3-D inverse synthetic aperture radar (SAR) system has been developed that performs synthetic aperture measurement via a linear motion of the radar in the elevation domain, and a circular (turntable) motion of the target in the range and cross-range domains, which was named as elevation circular SAR (E-CSAR) system. Its geometry is shown in Figure 1.

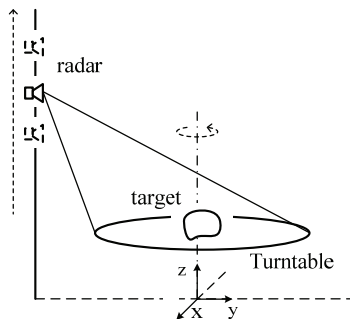


Fig. 1. Geometry of E-CSAR

The radar motion in elevation provides target coherent radar cross section (RCS) as a function of the elevation (or depression) angle. The target's circular motion yields the azimuthal look angle information. The imaging results of T-72 tank have been obtained and published (Bryant, M.L., Gostin, L.L., Soumekh, M. 2003). In fact, the concept "elevation circular SAR" could be extended by controlling the radar moving around the target in a helix trajectory which is shown in Figure 2. The cylindrical surface produces the 2-D resolution vertical (approximately) to the range resolution. Furthermore, to simplify the motion control, the helix trajectory could be composed by the circle motion of the radar and the rectilinear motion of the target, which is shown in Figure 3.

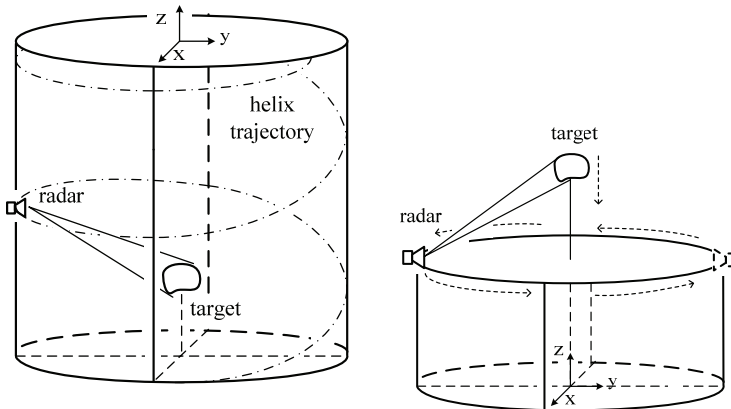


Fig. 2. Geometry of E-CSAR with helix trajectory Fig. 3. Geometry of E-CSAR composing helix trajectory by circle motion and rectilinear motion

The advantage of E-CSAR is that we can obtain the RCSes of the target in different elevation angle and azimuthal angle in one observation session; its disadvantage is that the target's size must be smaller than the diameter of the cylinder, which makes it difficult to be employed for large-size target. In fact, since the RCS of the target varies with the elevation angle and azimuthal angle, the synthetic aperture is a local region of the cylindrical surface, which is illustrated in Figure 4 (left).

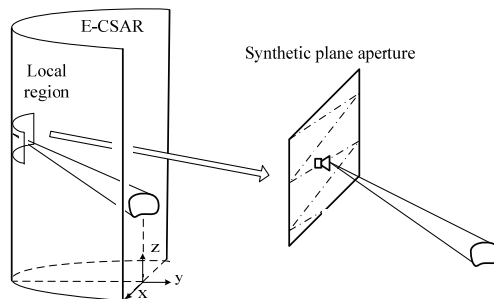


Fig. 4. Approximation of a local region of the E-CSAR using a synthetic plane aperture

Thus, we could approximate the local region as a plane and control the antenna phase centre moving in the plane by using a 2-D motion control platform or a linear array mounted on a 1-D motion control platform, which is shown in Figure 4 (right), and obtain the radar cross section (RCS) in one specific direction in one observation session. To obtain the RCSes in different elevation angle and azimuthal angle, one can just rotate the target or the platform. Compared with the E-CSAR, the size of the synthetic plane aperture is small and could be used in 3-D RCS measurement for large-size target.

Besides the E-CSAR, the curve SAR has also been researched in the radar community. In 1995, Jennifer L.H., Webb and David C. Munson, Jr. considered the problem of spotlight-mode synthetic aperture radar (SAR) imaging for an arbitrary radar path to reconstruct a 2-D image of 3-D surfaces. In 2004, Sune R. J. Axelsson researched the beam characteristics of 3-D SAR in curved or random paths in detail, and concluded that the SAR sidelobe suppression of a single circle path was worse than that of a circular antenna of similar size due to the fact that only a line boundary was used as SAR aperture. The spiral paths and random paths were discussed to improve the beam characteristic of 3-D SAR. The geometry of typical curve SAR is shown in Figure 5. The radar is mounted on a platform with curve path to synthesize a 2-D aperture.

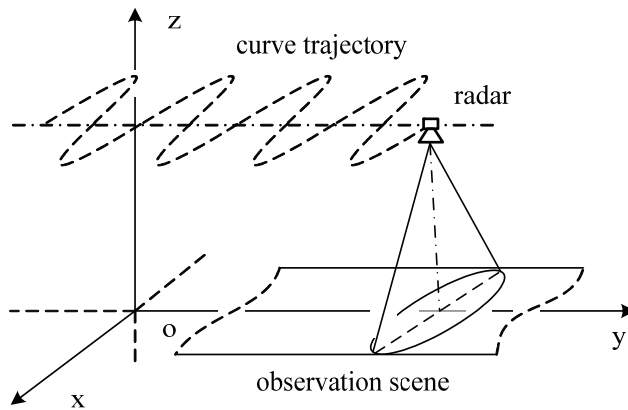


Fig. 5. Geometry of curve SAR

The advantage of curve SAR is that the size of the synthetic aperture could be far larger than the other 3-D SAR systems, which means high-resolution in the cross-track (x) direction; its disadvantage is that the motion control is too difficult to be implemented for the application of topographical survey in practice.

In 1996, Bassem R. Mahafza and Mitch Sajjadi proposed the concept “linear array SAR”, which mounted a linear array on a platform with rectilinear motion and synthesized a 2-D plane array. Its geometry is shown in Figure 6. In 2004, R. Giret, H. Jeul and, P. Enert conceived a millimeter-wave imaging radar onboard an UAV, and designed a 3-D millimeter-wave imaging plan. In their plan, a linear array was mounted above the ground (perpendicular to the ground plane) and vehicles passed under the system to obtain the 3-D image of the vehicles.

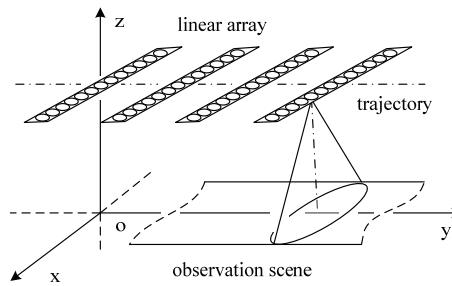


Fig. 6. Geometry of linear array SAR

Compared with the curve SAR, the motion control of linear array SAR is simpler. While, to achieve high cross-track resolution, the linear array must be rather long and the number of element is large, which is difficult and expensive to be implemented. To reduce the system complexity and cost, M. Weiß and J.H.G. Ender introduced the concept “MIMO radar” into the linear array SAR. With this concept, one can synthesize a sparse linear array SAR with relatively low cost, whose equivalent geometry is shown in Figure 7.

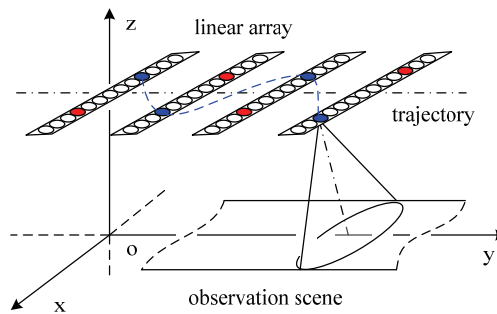


Fig. 7. Geometry of sparse LASAR, only the colored elements active at one pulse repetition period

The disadvantage of LASAR and sparse LASAR is that the cross-track resolution is determined by the length of the linear array. Since the length of the linear array is limited by the size of the platform, its cross-track resolution will be the bottleneck. Theoretically, the curve SAR could be considered as a kind of sparse LASAR (shown in Figure 7 in the blue dash-line).

In a word, the key problem of 3-D SAR is to vary the position of antenna phase centre (APC) in the 2-D / 3-D space. This work could be implemented mechanically (such as 2-D motion control platform and aircraft), or electrically (such as linear array). By moving HRR radar in 2-D plane using high precision motion control platform, we can build a low-cost 3-D RCS measurement device. The linear array SAR with MIMO technique might be the most feasible 3-D SAR system for the topographical survey application, though there are still some problems, such as, the balance between the length of linear array and the cross-track resolution and the compensation of motion measurement error.

3.2 General echo model of 3-D SAR

For the traditional SAR, the echo is always considered as a function of fast-time τ and slow-time t . While, for 3-D SAR, there might be more than one channel echo received at one pulse repetition period, such as LASAR and sparse LASAR, and it is not convenient to describe the 3-D SAR echo using slow-time t . In fact, the synthetic aperture technique produces additional resolution by moving the antenna phase centre in the spatial domain, and we should pay more attention on the change of the antenna phase centre in spatial domain rather than that in the time domain. Thus, a general echo model is built in this subsection, which can describe different 3-D SAR systems.

Given a scatterer with position $\bar{\mathbf{P}}_\omega$, its slant range to the antenna phase centre with position $\bar{\mathbf{P}}_{apc}$ is:

$$R(\bar{\mathbf{P}}_\omega, \bar{\mathbf{P}}_{apc}) \triangleq \|\bar{\mathbf{P}}_{apc} - \bar{\mathbf{P}}_\omega\|_2 \quad (20)$$

where, $\|\cdot\|_2$ denotes the 2-norm of vector.

Given the transmitted baseband signal $f(t)$, ignoring the radiation pattern, the scatterer's echo $\mathcal{D}(\tau; \bar{\mathbf{P}}_\omega; \bar{\mathbf{P}}_{apc})$ can be written as:

$$\mathcal{D}(\tau; \bar{\mathbf{P}}_\omega; \bar{\mathbf{P}}_{apc}) = \exp(j \cdot 2\pi \cdot 2R(\bar{\mathbf{P}}_\omega, \bar{\mathbf{P}}_{apc})/\lambda) \cdot f(\tau - 2R(\bar{\mathbf{P}}_\omega, \bar{\mathbf{P}}_{apc})/c) \quad (21)$$

where, τ denotes the fast time domain, λ denotes the wave length of the carrier. The first term in eq.(21) is the Doppler term arising from the relative position changes of the antenna phase centre with respect to the target. The second term is the fast-time term which causes the range resolution. Note that, in some cases, the transmitter and receiver might be operated independently, and term $2R(\bar{\mathbf{P}}_\omega, \bar{\mathbf{P}}_{apc})$ in eq. (21) should be rewritten as $R(\bar{\mathbf{P}}_\omega, \bar{\mathbf{P}}_{apc}^T) + R(\bar{\mathbf{P}}_\omega, \bar{\mathbf{P}}_{apc}^R)$, and the analysis should be modified correspondingly.

To describe the relative position changes, we introduce the concept "antenna phase centre set" \mathbf{P} denoting the collection of the positions of the antenna phase centre (note that the elements in the antenna phase centre set might be repetitive.).

For traditional 2-D SAR, its antenna phase centre set could be expressed as:

$$\mathbf{P} = \{ \langle x, y, z \rangle \mid x = x_0, y = v \cdot t, z = z_0; t \in \mathbf{T} \} \quad (22)$$

where, v denotes the speed of the platform, \mathbf{T} denotes the slow-time domain, z_0 denotes the height of the platform.

For E-CSAR, we have:

$$\mathbf{P} = \{ \langle x, y, z \rangle \mid x = \nu \cos(t), y = \nu \sin(t), z = v_h t; t \in \mathbf{T} \} \quad (23)$$

where, ν denotes the radius of the cylinder, v_h denotes the speed in the vertical direction.

For curve SAR, we have:

$$\mathbf{P} = \{ \langle x, y, z \rangle \mid x = x(t), y = v \cdot t, z = z_0; t \in \mathbf{T} \} \quad (24)$$

where, $x(t)$ and $v \cdot t$ compose the curve trajectory.

For linear array SAR, we have:

$$\mathbf{P} = \{ \langle x, y, z \rangle \mid x \in \mathbf{X}, y = v \cdot t, z = z_0; t \in \mathbf{T} \} \quad (25)$$

where, \mathbf{X} denotes the set of the x positions of the linear array, e.g., $\mathbf{X} = \{x \mid i \cdot d; i = 0, 1, \dots, N-1\}$, N denotes the element number of the linear array, d denotes the element interval.

For sparse LASAR, \mathbf{P} is a group of random positions, and we just simply denote it as \mathbf{P} .

Using the antenna phase centre set, we can easily express the echo of 3-D SAR as:

$$\mathbf{D}(r; \bar{\mathbf{P}}_\omega) = \{ \mathcal{D}(r; \bar{\mathbf{P}}_\omega; \bar{\mathbf{P}}_{apc}); \bar{\mathbf{P}}_{apc} \in \mathbf{P} \} \quad (26)$$

Note that, given r , \mathbf{D} is a set defined under the antenna phase centre set \mathbf{P} rather than a number.

Thought this echo model is more abstract than the classical one, as it will be seen in the next subsection, it will simplify the analysis of 3-D SAR ambiguity function. We could build the direct relationship between the antenna phase centre set \mathbf{P} (describes the shape of the synthetic aperture) and its ambiguity function, which make it easy for the 3-D SAR analysis and design.

3.3 Ambiguity function

Ambiguity function (AF) is one of the crucial concepts in the radar theory. For the pulse-Doppler (PD) radar, ambiguity function is a 2-D function of time delay and Doppler frequency. For imaging radar, it describes the interaction of different scatterers in the image space, which is also called point spread function. A well-designed imaging radar should have narrow mainlobe, low peak sidelobe ratio (PSLR) and low integrated sidelobe ratio (ISLR). In this section, we will discuss the ambiguity function of 3-D SAR.

Based on the echo model built in last subsection, the ambiguity function $\chi(\bar{\mathbf{P}}_\omega)$ of 3-D SAR can be defined as:

$$\chi(\bar{\mathbf{P}}_\omega) \triangleq \frac{\sum_{\mathbf{P}} \int \mathbf{D}(\tau; \bar{\mathbf{P}}_\omega) \cdot \mathbf{D}^*(\tau; \bar{\mathbf{0}}) d\tau}{\sum_{\mathbf{P}} \int |\mathbf{p}[\tau; \bar{\mathbf{0}}]|^2 d\tau} \quad (27)$$

where, superscript $*$ denotes complex conjugate, $\bar{\mathbf{0}}$ denotes the position of the reference point.

Since the integration with respect to the fast time τ in eq. (27) is the range-compression operation, eq. (27) can be rewritten as:

$$\chi(\bar{\mathbf{P}}_\omega) = \frac{1}{M} \sum_{\mathbf{p}} \exp(j \cdot 2\pi \cdot [R(\bar{\mathbf{P}}_\omega, \bar{\mathbf{P}}_{apc}) - R(\bar{\mathbf{0}}, \bar{\mathbf{P}}_{apc})] / \lambda) \cdot \chi^R(r - 2R(\bar{\mathbf{0}}, \bar{\mathbf{P}}_{apc})) \quad (28)$$

where, r denotes the range domain, $\chi^R(r)$ denotes the ambiguity function in the range direction, which is a sinc function.

Approximating $\chi^R(r)$ as the impulse function, the range AF in eq.(28) could be moved out of the summation by range migration adjustment during imaging processing, and eq. (28) could be rewritten as:

$$\chi(\bar{\mathbf{P}}_\omega) \approx \frac{1}{M} \left\{ \sum_{\mathbf{p}} \exp(j \cdot 2\pi \cdot [R(\bar{\mathbf{P}}_\omega, \bar{\mathbf{P}}_{apc}) - R(\bar{\mathbf{0}}, \bar{\mathbf{P}}_{apc})] / \lambda) \right\} \cdot \chi^R(r - 2R(\bar{\mathbf{0}}, \bar{\mathbf{P}}_0)) \quad (29)$$

where, $\bar{\mathbf{P}}_0$ denotes the centre of the synthetic aperture.

From eq.(29), the ambiguity function of 3-D SAR is the product of two terms: the first term in the curly brace is the Doppler term, which is caused by the synthetic aperture and produces resolution in the aperture direction(s); the second term is the fast-time term which produces range resolution. We define the synthetic aperture ambiguity function as:

$$\chi^p(\bar{\mathbf{P}}_\omega) \triangleq \frac{1}{M} \left\{ \sum_{\mathbf{p}} \exp(j \cdot 2\pi \cdot \Delta R(\bar{\mathbf{P}}_\omega, \bar{\mathbf{P}}_{apc}) / \lambda) \right\} \quad (30)$$

$$\Delta R^\omega(\bar{\mathbf{P}}_{apc}) \triangleq R(\bar{\mathbf{P}}_\omega, \bar{\mathbf{P}}_{apc}) - R(\bar{\mathbf{0}}, \bar{\mathbf{P}}_{apc})$$

where, $\Delta R^\omega(\bar{\mathbf{P}}_{apc})$ denotes the difference between $R(\bar{\mathbf{P}}_\omega, \bar{\mathbf{P}}_{apc})$ and $R(\bar{\mathbf{0}}, \bar{\mathbf{P}}_{apc})$.

And the AF of 3-D SAR could be written as the product of $\chi^R(r)$ and $\chi^p(\bar{\mathbf{P}}_\omega)$:

$$\chi(\bar{\mathbf{P}}_\omega) = \chi^p(\bar{\mathbf{P}}_\omega) \cdot \chi^R(r - 2R(\bar{\mathbf{0}}, \bar{\mathbf{P}}_0)) \quad (31)$$

Eq (31) indicates that the AF of 3-D SAR could be divided as a range AF and a synthetic aperture AF and analyzed independently. Since the range AF is a sinc function and independent to the antenna phase centre set, $\chi^p(\bar{\mathbf{P}}_\omega)$ should be paid more attention to.

For further analysis, we approximate $\Delta R^\omega(\bar{\mathbf{P}}_{apc})$ using the multivariable Taylor's theorem, and have:

$$\Delta R_{apc}^\omega \approx \bar{\mathbf{P}}_{apc} \cdot \bar{\mathbf{P}}_\omega^T / R(\bar{\mathbf{0}}, \bar{\mathbf{P}}_0) \quad (32)$$

where, superscript T denotes the transpose operator.

Rewrite $\bar{\mathbf{P}}_\omega$ in the spherical coordinates as $\bar{\mathbf{P}}_\omega = \gamma \cdot \hat{\boldsymbol{\zeta}}$, we have:

$$\Delta R_{apc}^{\gamma \hat{\boldsymbol{\zeta}}} = \theta \cdot \bar{\mathbf{P}}_{apc} \cdot \hat{\boldsymbol{\zeta}}^T \quad (33)$$

$$\theta = \gamma / R(\bar{\mathbf{0}}, \bar{\mathbf{P}}_0)$$

where, γ denotes the radius of $\bar{\mathbf{P}}_\omega$, $\hat{\boldsymbol{\zeta}}$ denotes the direction of $\bar{\mathbf{P}}_\omega$, θ is the ratio of γ to $R(\bar{\mathbf{0}}, \bar{\mathbf{P}}_0)$.

Substituting eq. (33) into eq. (30), we have:

$$\chi_\zeta^{\mathbf{P}}(\theta) = \frac{1}{M} \cdot \sum_{\mathbf{P}} \exp \left[j \cdot 2\pi \cdot \theta \cdot (\bar{\mathbf{P}}_{apc} \cdot \hat{\boldsymbol{\zeta}}^T) / \lambda \right] \tag{34}$$

Regarding $(\bar{\mathbf{P}}_{apc} \cdot \hat{\boldsymbol{\zeta}}^T) / \lambda$ as a phase function, it is the projection of antenna phase centre set \mathbf{P} onto the $\hat{\boldsymbol{\zeta}}$ direction. Using corollary 1, we have the AF in the $\hat{\boldsymbol{\zeta}}$ direction:

$$\chi_\zeta^{\mathbf{P}}(\theta) = \frac{1}{M} \cdot \int_{-\infty}^{+\infty} e^{j \cdot 2\pi \cdot \theta \cdot y} \mathcal{D}_\zeta(y) dy = \frac{1}{M} \cdot \widehat{\mathcal{D}}_\zeta(\theta) \tag{35}$$

The physical meaning of eq. (35) is shown in Figure 8. We can obtain the density function $\mathcal{D}_\zeta(i)$ by counting the number of elements whose projection on the $\hat{\boldsymbol{\zeta}}$ direction is in the neighbourhood of i , and the AF in the $\hat{\boldsymbol{\zeta}}$ direction is the Fourier transform of $\mathcal{D}_\zeta(i)$ approximately.

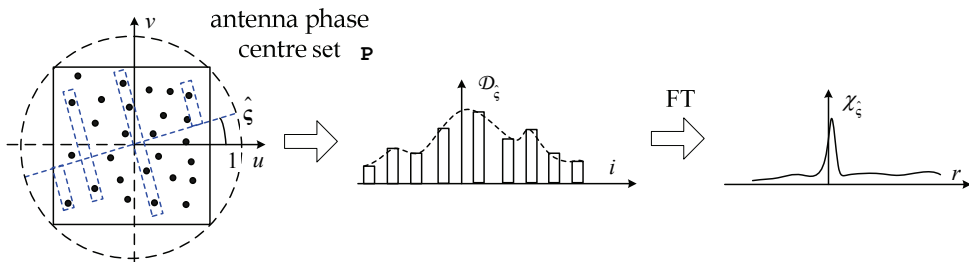


Fig. 8. Explanation on eq. (35), the AF in the $\hat{\boldsymbol{\zeta}}$ direction is the Fourier transform of $\mathcal{D}_\zeta(i)$ approximately.

Eq. (35) builds the direct relationship between the antenna phase centre set \mathbf{P} and the synthetic aperture ambiguity function $\chi_\zeta^{\mathbf{P}}(\theta)$. Then, the synthetic aperture ambiguity function of typical antenna phase centre sets will be discussed.

● **Z-shaped trajectory**

As a kind of simple continuous trajectory, Z-shaped trajectory is easy to be implemented using 2-D motion control platform and has been used in our experiments, which is shown in Figure 9-a. Figure 9-b and Figure 9-c are its ambiguity functions obtained by simulation and experiment respectively, and its grating lobe is high and dense.

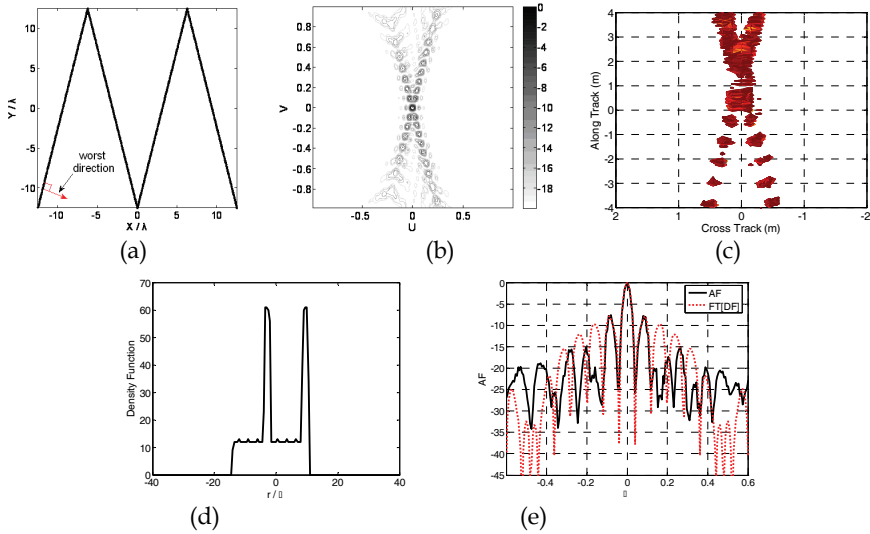


Fig. 9. (a) Z-shaped trajectory; (b) 2-D AF by simulation; (c) 2-D AF by experiment; (d) DF vertical to the edge, that two impulses are added on a rectangle function; (e) AF corresponding to Figure (d), whose grating lobes are quite high and dense.

This phenomenon could be explained using eq. (35). Observing Figure 9-a, we find that when the direction is vertical to the edge of the triangle function, all of the elements on one edge are projected on the same point, and there are two impulses are added on the density function (Figure 9-d). Consequently, the sidelobe of the directional AF (which is the Fourier transform of Figure 9-d and shown in Figure 6.e) is high, whose PSLR and ISLR are 7.60dB and -3.28 dB respectively.

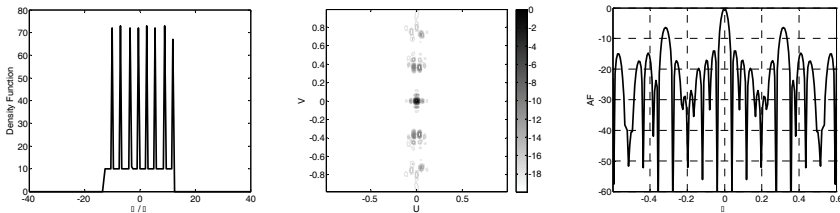


Fig. 10. (a) DF vertical to the edge with short period; (b) 2-D AF by simulation; (c) AF corresponding to Figure (a), whose grating lobes are sparse

High ISLR and PSLR mean that the sidelobe of strong scatterer will submerge the weak scatterer, and cause measurement error in the 3-D RCS measurement application. This problem could be solved by increasing the periodicity of the Z-shaped trajectory. With the increase of periodicity, the impulses in the density function increase correspondingly (Figure 10 a), and the grating lobe becomes sparse (Figure 10. b and c). Just as the grating lobe problem in the theory of antenna array, when the period of the Z-shaped trajectory is less than 1/2, the grating lobe will be eliminated completely.

● Dense square array

For LASAR, its synthetic aperture is a full-element square array, which is shown in Figure 11.a. Since the number of elements is equal for different x , its density function in the x direction is rectangle function, which is shown in Figure 11.b. The directional AF in the x direction is the Fourier transform of the rectangle function, which is a sinc function and shown in Figure 11.e. The black solid line is obtained by numerical simulation, the red dot line is the Fourier transform of Figure 11.b. The peak sidelobe ratio (PSLR) and integrated sidelobe ratio (ISLR) are -11.80 dB and -8.41 dB respectively, which is near to the sinc function (-13.30 dB and -10.16 dB respectively).

Then, let's observe the density function in the diagonal direction. From Figure 11.a, it is obvious that the number of elements increases linearly from one endpoint of the diagonal to the centre, and decreases linearly from the centre to the other endpoint. In consequence, the density function in the diagonal direction is a triangle function, which is shown in Figure 11.c. Figure 11.f is the AF in the diagonal direction. The black solid line is obtained by numerical simulation, the red dot line is the Fourier transform of Figure 11.c. The PSLR and ISLR are -22.86 dB and -20.71 dB respectively, which is near to the Fourier transform of triangular window (-26.82 dB and -22.02 dB respectively)

Figure 11.d is the 2-D AF of dense square array. We find that it has star-shaped AF.

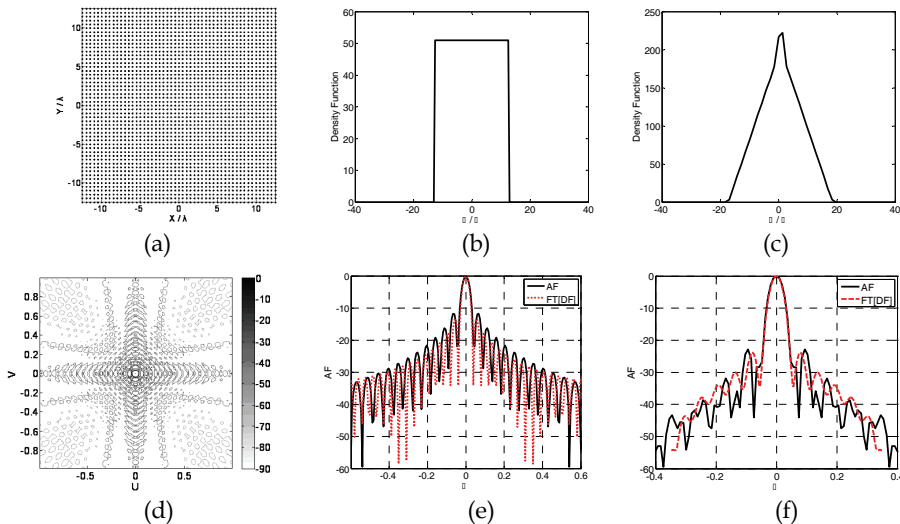


Fig. 11. (a) Dense square array; (b) DF in the x direction, which is a rectangle function; (c) DF in the diagonal direction, which is a triangle function; (d) 2-D AF of dense square array, whose sidelobes are distributed in the x and y directions mainly; (e) AF in the x direction, which is a sinc function; (f) AF in the diagonal direction, which is the Fourier transform of triangle function.

● Random sampling

For sparse LASAR, its synthetic aperture is a random sampling in the full-element square array, which is shown in Figure 12.a

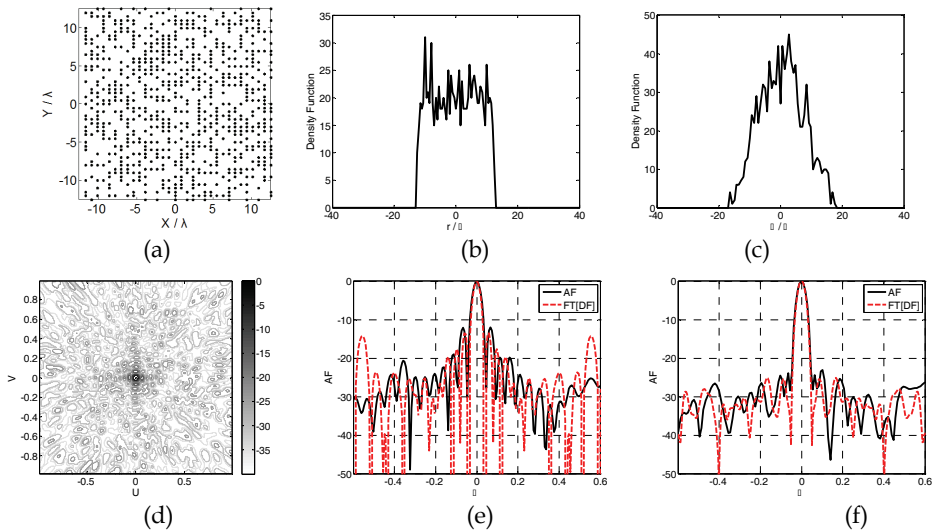


Fig. 12. (a) Random sampling square array; (b) DF in the x direction, which is a rectangle function with noise; (c) DF in the diagonal direction, which is a triangle function with noise; (d) 2-D AF; (e) AF in the x direction, which is similar to the Fourier transform of Figure b; (f) AF in the diagonal direction, which is similar to the Fourier transform of Figure c.

Figure 12. b and c are the density functions in the x and diagonal directions. Comparing with their counterparts of dense square array, we find that its density functions could be considered as the density functions of dense square array modulated by a noise. As a result, its PSLR in the x and diagonal directions (Figure 12.e and 12.f) are similar to those of dense square array. However, since the noise modulated on the density functions increases the high-frequency components, which are corresponded to the far-area sidelobes, its ISLR is higher than that of dense square array.

Figure 12.d is the 2-D AF of uniform distribution sparse array. Comparing with Figure 12.d, we find that its far-area sidelobe is higher than that of dense square array (Note the color bar). Energy leak is the main disadvantage of sparse array. One can improve the PSLR and ISLR by increasing the random sampling number, since its mainlobe energy is proportional to the square of the sampling number, and the sidelobe energy is proportional to the sampling number(Gauss distribution).

3.4 Spatial Resolution

The range resolution of 3-D SAR is produced by the pulse compression technique, and we can obtain the resolution formula directly as:

$$\rho_r = c/(2B) \tag{36}$$

where, c denotes the speed of light, B denotes the signal bandwidth.

The other two dimensional resolutions are produced using the array theory, and we can write the resolution formula as:

$$\rho = \lambda / (2\theta) \tag{37}$$

where, λ denotes the wave length, θ denotes the aperture angle. Remark that the “2” in eq. (37) indicates that the transmitter and receiver moving cooperatively. If the transmitter or receiver is fixed, the resolution formula should be approximately be $\rho = \lambda / \theta$.

Note that, the aperture angle θ is influenced by the size of the array, the beam angle of the T/R antenna and the scatterer angle (the angle in which the RCS could be considered as constant.), which are shown in Figure 13.

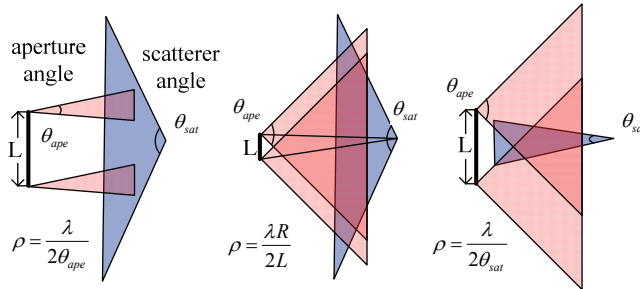


Fig. 13. Influenced of the array size, aperture angle and scatterer angle on the resolution
The resolution is restricted by the worst factor, i.e.:

$$\rho = \max\left(\frac{\lambda}{2\theta_{ape}}, \frac{\lambda R}{2L}, \frac{\lambda}{2\theta_{sat}}\right) \tag{38}$$

The first two factors could be optimized in the design of 3-D SAR system; the last factor arises from the scattering mechanism and is difficult to be reduced. It also means that we can not improve the resolution of 3-D SAR unlimitedly.

4. Backprojection Method

Backprojection (BP) algorithm is a 3-D SAR imaging algorithm based on the time domain correlation (TDC) technique, which coherently adds the data at the fast-time bin that corresponds to the location of a point for all synthetic aperture locations. The BP algorithm can be considered as the implementation of the definition of ambiguity function, and has been used in E-CSAR data processing.

The input of the BP operator is the raw data, antenna phase centre set and the scatterer’s position; the output is the RCS of the scatterer.

Let \mathbf{D}_{II} , \mathbf{P} and $\bar{\mathbf{P}}_{\omega}$ be the raw data after range compression, antenna phase centre set and the scatterer’s position, the BP operator can be expressed as:

$$\mathbf{C}[\mathbf{D}_{II}, \mathbf{P}, \bar{\mathbf{P}}_{\omega}] \rightarrow \sigma_{\omega} \tag{39}$$

The implementation of the BP operator $\mathbf{C}[\bullet]$ is presented in eq. (40):

$$\mathcal{C}[\mathbf{D}_{II}, \mathbf{P}, \bar{\mathbf{P}}_{inw}] \triangleq \sum_{\mathbf{p}} \mathbf{D}_{II} \cdot \exp(-j \cdot 2\pi \cdot 2R(\bar{\mathbf{P}}_{inw}, \bar{\mathbf{P}}_{apc}) / \lambda) \cdot \chi^R(r - 2R(\bar{\mathbf{P}}_{inw}, \bar{\mathbf{P}}_{apc})) \quad (40)$$

where, $\bar{\mathbf{P}}_{inw}$ denotes the pixel in the image space.

From eq. (40), we know that just like the 2-D BP algorithm, the 3-D BP algorithm can roughly be divided into four steps: range-compression, interpolation, resampling and coherent summation, whose block diagram is shown in Figure 14. Processing the 3-D image region one pixel by one pixel, we can obtain the 3-D RCS distribution finally.

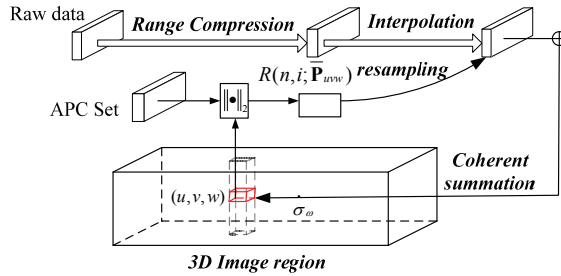


Fig. 14. Block diagram of 3D BP algorithm

From eq. (40), the computational cost Ξ_c of the single-scatterer compression operator $\mathcal{C}[\cdot]$ can easily be calculated as:

$$\Xi_c = M \cdot (\Xi_{int} + \Xi_{coh}) \quad (41)$$

where, M denotes the total element number of antenna phase centre set, Ξ_{int} and Ξ_{coh} denote the computational costs of the interpolation operation and the coherent summation operation respectively.

Ignoring the computational cost of the range-compression operation, for a 3-D image region with size $L \times W \times H$ (pixel³), the total computational cost of 3-D BP algorithm is:

$$\Xi_{BP} = L \cdot W \cdot H \cdot \Xi_c \quad (42)$$

Compared with 2-D BP, there are two factors that cause the computational cost of 3-D BP algorithm is far larger than that of 2-D BP: the increase of the acquired data and the extension of image region. The former factor can partially be solved using the sparse array technique. The latter factor can be solved using the multiresolution approximation technique, which will be introduced in the following section.

5. Imaging Processing of 3-D SAR

To verify the feasibility of “one-active” LASAR, a series of experiments have been carried out. The typical experiment plan is shown in Figure 15, and includes three parts: “one-active” LASAR, reference points and scene area.

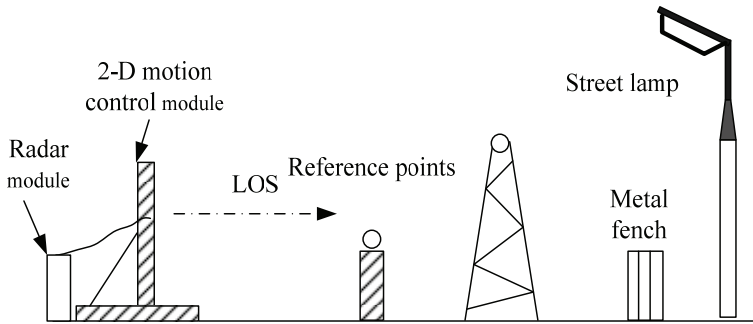


Fig. 15. Experiment plan of 3-D SAR

The “one-active” LASAR consists of two parts: radar module and motion control module. The radar module is used to transmit the LFM signal and receive the echo from the observation scene. The whole system works on the X-band with signal bandwidth about 120MHz and pulse repetition frequency 20Hz. The motion control module is used to control the transmitter and the receiver moving in a 2-D plane, and synthesize a virtual 2-D antenna array. The motion control module consists of a set of high-precision transfer device with effective length 2m x 2m and two high-precision motors, which can compose any continuous 2-D curve. The Z-shaped trajectory with period 1/5, 1/20 and 1/30 are used in the experiment.

Figure 16 is the experiment data of “one-active” LASAR after range-compression. The top black line is the echo of reference point after range compression; the bottom stripped area is the echo of scene area.

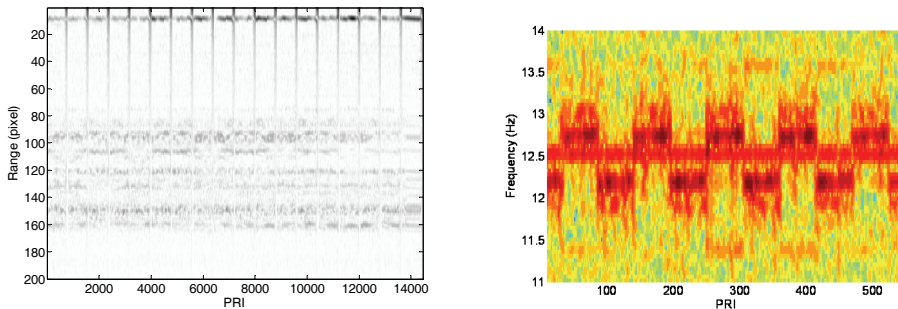


Fig. 16. Experiment data after range-compression Fig. 17. Time-frequency spectrum of single scatterer

To analyze the signal characteristic of “one-active” LASAR, we select one row data in one range-bin, and obtain its time-frequency spectrum by Short-Time Fourier transformation (STFT). Figure 17 is the typical time-frequency spectrum of single scatterer. We find that the time-frequency characteristic of 3-D SAR is more complex than that of traditional SAR (chirp signal), and contains more information on the target.



Fig. 18-a photo of the whole scene Fig. 19-a photo of fence and lamp Fig. 19-b imaging result of the fence and lamp

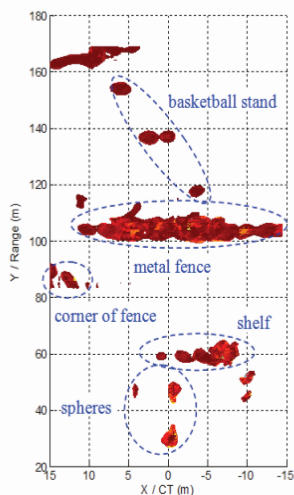


Fig 18-b

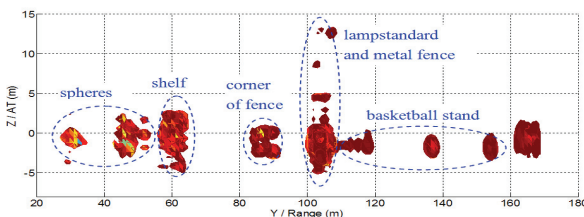


Fig 18-c

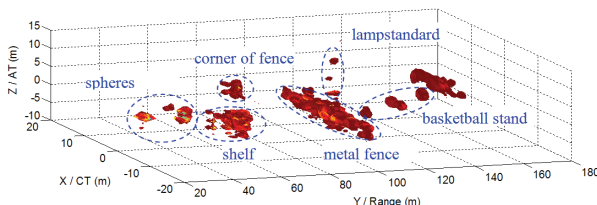


Fig 18-d

Fig. 18-b, c and d, imaging results of the whole scene in the top view, side view and 3-D view respectively.

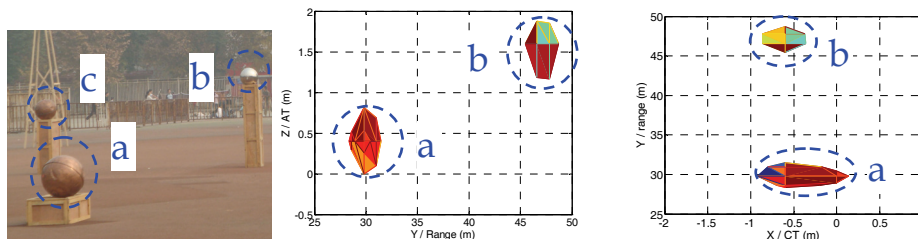


Fig. 20-a photo of metal spheres Fig. 20-b imaging result of spheres (side view) Fig. 20-c imaging result of spheres (top view)

Figure 18-a is the photo of observation scene, Figure 18-b, c and d show its imaging result in the top view, side view and 3-D view. And we find that the imaging result can depict the main features of the observation scene.

Figure 19-a is the photo of a local region of the observation scene, which contains a metal fence and a street lamp with metal lampstandard. Figure 19-b is its imaging result. Obviously, the imaging result can be divided into three parts, which correspond to the metal fence, corner of the metal fence (far area in the photo) and the metal lampstandard soundly. Especially, according to Figure 18-c, we can read the height of the lamp is about 14m, which matches the real height (14.5m) correctly.

Figure 20-a is the photo of another local region of the observation scene, which contains two copper spheres (a, c) and a stainless steel sphere (b). since its RCS is too low, the copper sphere c is not shown in Figure 20-b and c (could be found in Figure 18-b). Figure 20-b is the imaging result (side view) of spheres a and b. From it, we can read the relative height of spheres a and b is 1.2m, which matches the measurement value (1.09m) soundly. Figure 20-c is the imaging result (top view) of spheres a and b. From it, we can read the relative distance of spheres a and b is 15.8 m, which matches the measurement value (14.96m) soundly.

The above experiment results demonstrate the ability of 3-D SAR in the application of 3-D RCS measurement.

6. Multiresolution Approximation Techniques

Unlike the 2-D SAR, a great deal of 3-D image region contains no scatterer (such as atmosphere) or is shadowed by the other scatterers, and it is not necessary to compress all the pixels in the image region. Based on this idea, multiresolution approximation techniques could be employed to reduce the computational cost of 3-D SAR.

6.1 Scattering Model

Much excellent work has been done on the modeling radar backscatter for both naturally occurring terrain and man-made objects. One of the most popular models is the three-component scattering model developed by Anthony Freeman and Stephen L. Durden. In this model, the scattering mechanism of target is divided into three components, including the rough surface scattering, double-bounce scattering and canopy scattering.

The rough surface scattering component assumes that the backscatter is reciprocal, such as road and bare soil, whose mechanism is illustrated in Figure 21 (left).

The double-bounce scattering component is modeled by scattering from a dihedral corner reflector, where the reflector surface can be made of different dielectric materials, such as a ground-trunk interaction, whose mechanism is illustrated in Figure 21 (middle).

The canopy (volume) scattering component assumes that the radar echo is from a cloud of randomly oriented, very thin, cylinder-like scatterers, such as the forest canopy.

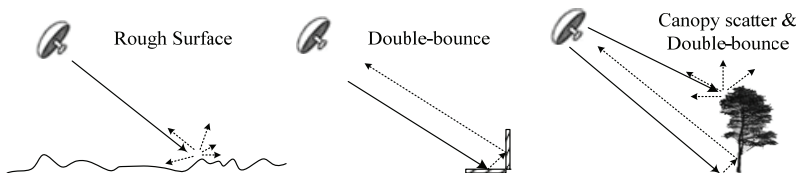


Fig. 21. Scattering mechanisms of three-component model: rough surface scattering (left), double-bounce scattering (middle) and tree scattering model (right) containing both canopy scattering and double-bounce scattering.

In the application of 3-D RCS measurement, the scatterers always concentrate in the local region of the 3-D image space which are shown in Figure 18. In the application of topographical survey, since the thickness of the scattering layer is far smaller than the height of the imaging region, it is convenient to consider the scattering layer as a surface in a low height-resolution level. And one can obtain the scattering layer by searching the neighborhood of the scattering surface.

6.2 Multiresolution approximation

In this subsection, we will introduce the basic concept on the multiresolution wavelet approximation, which is necessary for the design of 3-D LASAR imaging method via multiresolution approximation.

● Multiresolution approximation

The multiresolution approximation of $f(t)$ is defined as the orthogonal projection $P_{V_j}[f]$ on a multiresolution approximation subspace of $L^2(\mathbb{R})$. The multiresolution approximation of $L^2(\mathbb{R})$ is a sequence $\{V_j\}_{j \in \mathbb{Z}}$ of closed subspaces of $L^2(\mathbb{R})$ that obeys the following 5 properties:

$$\forall (j, k) \in \mathbb{Z}^2, f(t) \in V_j \Leftrightarrow f(t - 2^j k) \in V_j \quad (43-1)$$

$$\forall j \in \mathbb{Z}, V_j \subset V_{j+1} \quad (43-2)$$

$$\forall j \in \mathbb{Z}, f(t) \in V_j \Leftrightarrow f(2t) \in V_{j+1} \quad (43-3)$$

$$\lim_{j \rightarrow -\infty} V_j = \bigcap_{j=-\infty}^{+\infty} V_j = \{0\} \quad (43-4)$$

$$\lim_{j \rightarrow +\infty} V_j = \text{closure} \left(\bigcup_{j=-\infty}^{+\infty} V_j \right) = L^2(\mathbb{R}^2) \quad (43-5)$$

where, j denotes the approximation level, and the resolution at level j is 2^{-j} .

Property (43-1) means that V_j is invariant by any translation proportional to the scale 2^j .

The inclusion (43-2) is a causality property which proves that an approximation at a resolution 2^{j+1} contains all the information to compute an approximation at a coarser resolution 2^j . Recursive eq. (43-3) specifies the relationship between approximation subspaces. The property (43-4) implies that we lost all the details of f when the level goes to $-\infty$; on the other hand, when the level goes $+\infty$, property (43-5) imposes that the signal approximation converges to the original signal.

According to the approximation theory, the basis of V_j can be generated by dilating and translating a scaling function $\phi(t)$:

$$\phi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \phi\left(\frac{t-n}{2^j}\right) \quad (44)$$

Thus, the multiresolution approximation $\tilde{f}_j(t)$ of $f(t)$ at level j can be calculated as:

$$\tilde{f}_j(t) = \sum_{n=-\infty}^{+\infty} \langle f, \phi_{j,n}(t) \rangle \cdot \phi_{j,n}(t) \tag{45}$$

where, $\langle \cdot \rangle$ denotes the inner product.

● **Conjugate mirror filter**

To ensure that the multiresolution approximation can be conducted recursively, it is necessary to analyze the relationship between the approximations of $f(t)$ at level i and $i+1$. The multiresolution causality property (42-2) imposes that $V_j \subset V_{j+1}$. Since $\phi_{j+1,n}(t)$ is an orthonormal basis of V_{j+1} , we can decompose $\phi_{j,0}(t)$ as:

$$\phi_{j,0}(t) = \sum_{n=-\infty}^{+\infty} h[n] \cdot \phi_{j+1,n}(t) \tag{46}$$

With:

$$h(n) = \langle \phi_{j,0}(t), \phi_{j+1,n}(t) \rangle \tag{47}$$

where, eq. (47) is called two-scale relation, $h(n)$ denotes the conjugate mirror filter corresponding to the scaling function $\phi(t)$, provide that:

$$\hat{\phi}(\omega) = \prod_{p=1}^{+\infty} \frac{\hat{h}(2^{-p}\omega)}{\sqrt{2}} \hat{\phi}(0)$$

According to the fast orthogonal wavelet transform developed by Stephane. G. Mallat, the reconstruction of $\tilde{f}_{j+1}(t)$ can be implemented by a two-channel multirate filter bank:

$$\begin{aligned} \tilde{f}_{j+1}(n) &= \check{f}_j(n) * h(n) + \check{d}_j(n) * g(n) \\ \check{f}_j(n) &= \begin{cases} \tilde{f}_j(n) & n = 2p \\ 0 & n = 2p+1 \end{cases} \quad p = 0, 1, \dots \\ \check{d}_j(n) &= \begin{cases} \tilde{d}_j(n) & n = 2p \\ 0 & n = 2p+1 \end{cases} \quad p = 0, 1, \dots \end{aligned} \tag{48}$$

where, $\check{d}_j(n)$ denotes detail coefficients, $g(n)$ denotes the high-pass filter corresponding to $h(n)$, satisfying:

$$\hat{g}\left(\frac{\omega}{2}\right) = e^{-j\omega} \hat{h}^*(\omega + \pi) \tag{49}$$

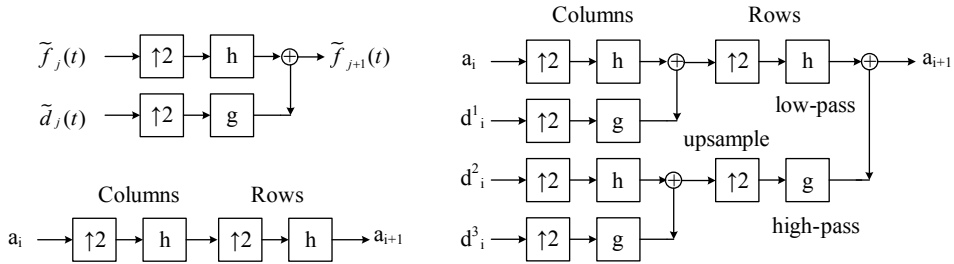


Fig. 22. Diagrams on wavelet interpolation, the left top, right and left bottom are the diagrams of 1-D fast wavelet transform, 2-D fast wavelet transform and 2-D wavelet interpolation respectively

The diagram of 1-D wavelet reconstruction is shown in Figure 22 (left top), where, upsample operator “ $\uparrow 2$ ” inserts zeros at odd-indexed elements.

● **2-D multiresolution surface approximation**

The properties of two-dimensional wavelet are essentially the same as in one dimension. A separable two-dimensional wavelet transform can be factored into one-dimensional wavelet transforms along the rows and columns.

Assume that the conjugate mirror filters corresponding to the one-dimensional wavelet transform are denoted as $h(n)$ and $g(n)$. According to the fast two-dimensional wavelet transform, the reconstruction of a two-dimensional function $\tilde{f}_{j+1}(n_1, n_2)$ can be implemented by the following equation:

$$\begin{aligned} \tilde{f}_{j+1}(n_1, n_2) = & \check{f}_j(n_1, n_2) * [h(n_1)h(n_2)] + \check{d}_j^1(n_1, n_2) * [h(n_1)g(n_2)] \\ & + \check{d}_j^2(n_1, n_2) * [g(n_1)h(n_2)] + \check{d}_j^3(n_1, n_2) * [g(n_1)g(n_2)] \end{aligned} \tag{50}$$

where, $\check{d}_j^1(n_1, n_2)$, $\check{d}_j^2(n_1, n_2)$ and $\check{d}_j^3(n_1, n_2)$ denotes the detail coefficients matrices. And the corresponding diagram is shown in Figure 22 (right). In the application of surface prediction, since there is no information on the detail coefficients matrices, we just consider that all of them are zero matrices, and the reconstruction formula can be simplified as:

$$\tilde{f}_{j+1}(n_1, n_2) = \check{f}_j(n_1, n_2) * [h(n_1)h(n_2)] \tag{51}$$

And the corresponding diagram is shown in Figure 22 (left bottom).

● **Typical conjugate mirror filters**

According to the discussion in above section, we conclude that the 2-D surface multiresolution prediction is specified by the conjugate mirror filter $h(n)$. In this subsection, we present some typical conjugate mirror filter as follows:

Shannon conjugate mirror filter:

$$\hat{h}(\omega) = \begin{cases} \sqrt{2} & \omega \in [-\pi/2, \pi/2] \\ 0 & \text{others} \end{cases} \quad (52)$$

Meyer conjugate mirror filter:

$$\hat{h}(\omega) = \begin{cases} \sqrt{2} & \omega \in [-\pi/3, \pi/3] \\ 0 & \omega \in [-\pi, -2\pi/3] \cup [-2\pi/3, \pi] \end{cases} \quad (53)$$

DB conjugate mirror filter:

$$|\hat{h}(\omega)|^2 = 2 \cos^{2p}\left(\frac{\omega}{2}\right) P(\sin^2\left(\frac{\omega}{2}\right)) \quad (54)$$

where, polynomial $P(y)$ satisfies:

$$(1-y)^p P(y) + y^p P(1-y) = 1$$

Spline biorthogonal conjugate mirror filter

$$\hat{h}(\omega) = \sqrt{2} \exp\left(-\frac{j\varepsilon\omega}{2}\right) \cos^p\left(\frac{\omega}{2}\right) \quad (55)$$

where, $\varepsilon = 0$ if p is even and $\varepsilon = 1$ for p is odd.

For the different applications, the conjugate mirror filter might affect the estimation error of the prediction operator, it is sensible to select the conjugate mirror filter according to the application.

6.3 Surface Tracing Technique

In the application of topographical survey, the scatterers combine a surface in the 3-D image space, we can trace the surface and focus those scatterers near it via specific searching method. Thus, the 3-D SAR imaging processing can be reduced to a 2-D imaging problem, and the computational cost will be reduced greatly. Those methods based on the above idea are named as surface-tracing-based 3-D imaging method (STB 3-D imaging method).

● Principle and steps

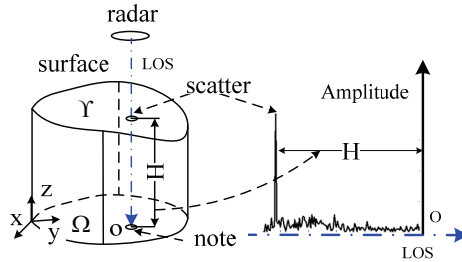


Fig. 23. Principle of STB 3-D SAR imaging technique

The principle of STB 3-D imaging method is illustrated in Figure 23. Assume that the line of radar sight (i.e. LOS) is parallel to the elevation (z) direction, and the 3-D scattering surface Υ in the observation scene can be expressed as following:

$$\Upsilon = \{(x, y, z) / z = h(x, y) \quad (x, y) \in \Omega \subset \mathbb{R}^2\} \tag{56}$$

where, (x, y) is called note, Ω is the note set, h denotes the elevation function.

For a given note (x_0, y_0) , the RCS distribution in z direction is a quasi-impulse function which is shown in Figure 23 (right). The index of the maximum is the elevation of note (x_0, y_0) . Searching the maximum of every note (x, y) in Ω , one can reconstruct the DEM of the scene eventually. Generally, the steps of STB 3-D imaging method are stated as follows:

Step-1 Initiation:

Compress in a subset Ω_0 of the note set Ω by 3-D BP imaging method, find the maximum and the associated index of every note, and obtain the initial subsurface Υ_0 ;

$$\Upsilon_0 = \{(x, y, z) / z = h(x, y) \quad (x, y) \in \Omega_0 \subset \Omega\} \tag{57}$$

Step-2 Prediction:

Expand the note set Ω_0 to Ω_1 ($\Omega_0 \subset \Omega_1 \subseteq \Omega$), predicate the surface on Ω_1 using the known surface Υ_0 by a surface prediction operator $\mathcal{P}[\cdot]$, and obtain the estimation of the surface on Ω_1 , denoted as $\hat{\Upsilon}_1$;

$$\mathcal{P}[\Upsilon_0] \rightarrow \hat{\Upsilon}_1 \tag{58}$$

Step-3 Searching:

Search in the neighborhood of $\hat{\Upsilon}_1$, and obtain the actual surface Υ_1 ;

$$\mathcal{S}[\hat{\Upsilon}_1] \rightarrow \Upsilon_1 \tag{59}$$

Step-4 : Recursion:

Replace Ω_0 and Υ_0 in step 2 by Ω_1 and Υ_1 , and repeat the step 2-4 until obtaining the 3-D surface Υ .

where, the most important steps are predication and searching, which will be discussed the in the rest of this section.

● **Surface prediction operator**

The aim of prediction operator $\mathcal{P}[\bullet]$ is to estimate the likely elevation of the surface using the known subsurface, which is necessary for the searching operator.

The input of the prediction operator is the known subsurface on a note set Ω_i ; the output is the estimation of the subsurface on the note set Ω_{i+1} ($\Omega_i \subset \Omega_{i+1}$).

Let Υ_i and Υ_{i+1} be the subsurfaces on the note sets Ω_i and Ω_{i+1} respectively, the prediction operator $\mathcal{P}[\bullet]$ can be expressed as:

$$\mathcal{P}[\Upsilon_i] \rightarrow \hat{\Upsilon}_{i+1} \tag{60}$$

And the prediction error can be defined as:

$$e = \mathbf{z} - \hat{\mathbf{z}} \tag{61}$$

where, $\hat{\mathbf{z}}$ and \mathbf{z} denote the predicated elevation and the actual elevation at given note respectively.

Mathematically, the prediction operator can be implemented using multivariate interpolation technique. According to the interpolation method, the prediction operator includes the polynomial prediction operator, ridge prediction operator, spline prediction operator and multiresolution (wavelet) prediction operator, etc.

In the viewpoint of predication strategy, the prediction operator can roughly be classified as two classes: local prediction operator and multiresolution prediction operator. The former operator starts from the local region of the note set and expands the known region from the edge, which is shown in Figure 24-a. The latter one starts from a coarse resolution scene and improves the resolution of the scene recursively, which is shown in Figure 24-b.

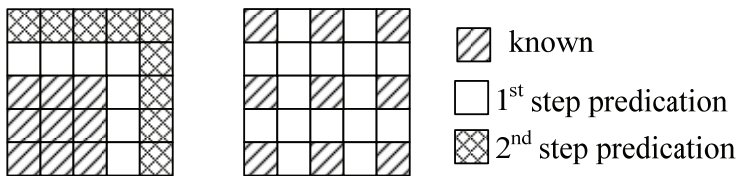


Fig. 24. Local prediction (left) and multiresolution prediction (right)

In the case of high-resolution 3D SAR imaging, since the elevations of the neighbor notes are affected by the fluctuation of ground greatly, the prediction error of the local prediction operator is larger than that of the multiresolution prediction operator. On the other hand, for an $N \times N$ scene, the recursion times of the local prediction operator are N , and the

recursion times of the multiresolution prediction operator are $\log_2(N)$. Less recursion times always mean less interpolation operation and less computational cost.

● **Searching operator**

The searching operator $\mathcal{S}[\bullet]$ is to find out the maximum and the associated index at every note.

The input of the searching operator includes the raw data, note, estimated elevation, threshold; the output includes the actual elevation and the associated RCS.

Let $\mathbf{D}, (x_0, y_0), \hat{Z}, \Theta, Z$ and σ be the raw data, note, estimated elevation, threshold, actual elevation and the associated RCS respectively, the searching operator can be expressed as:

$$\mathcal{S}[\mathbf{D}, (x_0, y_0), \hat{Z}, \Theta] \rightarrow [Z, \sigma] \tag{62}$$

The prediction operator can be implemented using the following equation:

$$\begin{cases} \sigma(x_0, y_0, z_{\max}) \geq \sigma(x_0, y_0, z_{\max} - 1) \\ \sigma(x_0, y_0, z_{\max}) \geq \sigma(x_0, y_0, z_{\max} + 1) \\ \sigma(x_0, y_0, z_{\max}) > \Theta \end{cases} \tag{63}$$

It indicates that the maximum is larger than the adjacent pixels and the detection threshold. The detection threshold Θ can be selected based on the constant false alarm rate (CFAR) criteria. Assume that the RCS and the noise both obey the normal distribution, Θ can be calculated as:

$$\Theta = \nu_0 \cdot \text{erfc}^{-1}(P_{false}) + \mu_0 \tag{64}$$

where, μ_0 and ν_0 denote the mean and variance of the signal respectively, and can be obtained in the step of initiation, $\text{erfc}^{-1}(\bullet)$ denotes the inverse complementary error function.

● **Numerical results**

In this subsection, some numerical experiments are conducted to demonstrate the procedures of the STB 3-D BP algorithm.

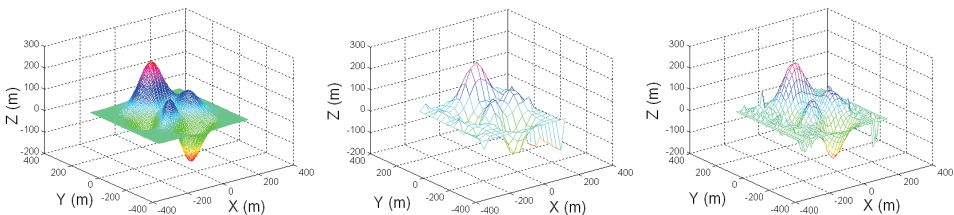


Fig. 25-a Terrain used in simulation, Fig. 25-b 1st iteration imaging result Fig. 25-c 2nd iteration imaging result

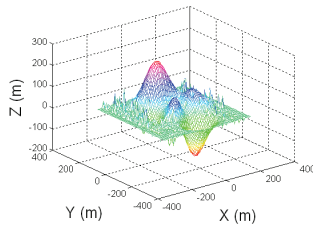
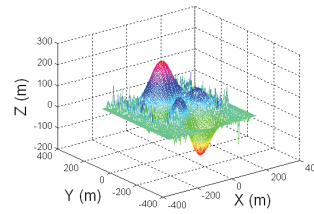
Fig. 25-d 3rd iteration imaging resultFig. 25-e 4th iteration imaging result

Figure 25-a is the terrain used in the numerical experiments. Figure 25-b, c, d and e are the imaging results of the STB 3-D BP algorithm with 1st, 2nd, 3rd, and 4th iterations respectively. From them, we find that the STB 3-D BP algorithm can reconstruct the DEM of the scene correctly. By further analysis (Shi Jun, Zhang Xiaoling, Jianyu Yang, Wang Yinbo, 2008), we find that, in general case, the total times to call the compress operator of the STB 3-D imaging algorithm is a dozen of times larger than that of 2-D BP algorithm generally, and is far smaller than that of the 3-D BP algorithm that is about several hundreds times larger than 2-D BP (determined by the height of the 3-D imaging scene).

6.4 Subaperture Approximation Technique

The STB based multiresolution approximation technique can only be used for the topographical survey application, which can reduce the size of image space. In fact, the multiresolution approximation technique can also reduce the number of antenna elements that need to be processed, since only a subaperture is necessary to obtain a low-resolution image. Based on this idea, we can obtain a low-resolution image using a subaperture, detect the interested regions with scatterers, and process the interested regions with a larger subaperture iteratively until obtain the fine-resolution image. This kind of multiresolution approximation technique is named as subaperture approximation technique, which can be used in both 3-D RCS measurement and topographical survey applications. Limited by the length of this chapter, this topic will be discussed hereafter.

7. Summary

Using synthetic aperture technique, we can obtain the 3-D RCS distribution of target. The precondition of synthetic aperture technique is that the target's RCS does not vary with time in one aperture. The synthetic aperture can be implemented mechanically (such as CSAR, E-CSAR and curve SAR), or electrically (such as linear array SAR). By moving HRR radar in 2-D plane using high precision motion control platform, we can build a low-cost 3-D RCS measurement device. The linear array SAR with MIMO technique might be the most feasible 3-D SAR system for the topographical survey application, though there are still some problems, such as, the balance between the length of linear array and the cross-track resolution and the compensation of motion measurement error.

Ambiguous function (AF) of 3-D SAR is the product of the range AF and the synthetic aperture AF, which can be analyzed independently. The range AF is a sinc function without any window function; the synthetic aperture AF could be analyzed using the theory of array antenna. The resolution in the synthetic aperture direction(s) is restricted by the size of the

array, the beam angle of the T/R antenna and the scatterer angle. Since the RCS varies in different elevation angle and azimuthal angle, we can not improve the resolution of 3-D SAR unlimitedly.

Backprojection method can be employed in 3-D SAR imaging processing. Its disadvantage is the high computational cost. Unlike the 2-D microwave image, the scatterers always concentrate in the local regions of 3-D image space. Based on this feature, the multiresolution approximation technique could be employed in imaging processing, which can reduce the computational cost greatly.

8. References

- Anthony Freeman, Stephen L. Durden. (1998). A three-component scattering model for polarimetric SAR Data. *IEEE Trans on Geoscience and Remote sensing*, Vol. 36, No. 3, pp. 953-973, ISSN: 0196-2892
- Bryant, M.L., Gostin, L.L., Soumekh, M. (2003). 3-D E-CSAR imaging of a T-72 tank and synthesis of its SAR reconstructions., *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 39, Issue 1, pp. 211 - 227, ISSN: 0018-9251
- Elias M. Stein and Rami Sharkarchi. (2005). *Real analysis: measure theory, integration, & Hilbert spaces*. Princeton University Press, ISBN-13: 978-0-691-11386-9, Princeton, New Jersey
- Jennifer L.H. Webb and David C. Munson, Jr. (1995). SAR image reconstruction for an arbitrary radar path. *Processing of Acoustics, Speech, and Signal*, pp. 2285-2288, ISSN: 07367791, Detroit, MI, USA, May, 1995, IEEE, Piscataway, NJ, United States
- M. Weiß, Ender, J.H.G. (2005). A 3D imaging radar for small unmanned airplanes - ARTINO. *Proceedings of EURAD 2005 Conference*, pp. 229-232. ISBN-13: 9782960055139, Paris, France, October, 2005, IEEE, Piscataway, NJ, United States
- Mahafza, B.R.; Sajjadi, M. (1996). Three-dimensional SAR imaging using linear array in transverse motion. *IEEE Trans. on aerospace and electronic system*, Vol. 32, Issue 1, pp. 499 - 510, ISSN: 0018-9251
- Mehrdad Soumekh. (1999). *Synthetic aperture radar signal processing with matlab algorithms*. A Wiley-Interscience publication, ISBN-10: 0471297062, Hoboken, NJ, USA
- Paul R. Halmos. (1974). *Measure Theory*. Published by Van Nostrand, ISBN-10: 0387900888, New York, USA
- Shi Jun, Zhang Xiaoling, Jianyu Yang, Wang Yinbo, (2008). Surface-Tracing-Based LASAR 3-D Imaging Method via Multiresolution Approximation. *IEEE Trans on Geoscience and Remote Sensing*, Vol. 46, Issue 11, Part 2, pp. 3719-3730, ISSN: 0196-2892
- Stephane G. Mallat. (1989). A Theory for Multiresolution Signal Decomposition: the Wavelet Representation. *IEEE Trans on pattern analysis and machine intelligence* Vol. 11, No. 7 pp. 674-693, ISSN: 0162-8828
- Stephane Mallat. (1999). *A wavelet Tour of signal processing*. 2nd Edition, Academic Press, ISBN: 7-111-12768-4, Salt Lake City UT 84105 USA
- Sune R.J. Axelsson. (2004). Beam characteristics of the three-Dimensional SAR in curved or random paths. *IEEE Trans. On Geosciences and remote sensing*, vol. 42 No 10. pp. 2324-2334. ISSN: 0196-2892

- Tsz-King Chan; Kuga, Y.; Ishimaru, A. (1999). Experimental studies on circular SAR imaging in clutter using angular correlation function technique. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37, Issue 5, Part 1, pp. :2192 - 2197, ISSN: 0196-2892
- Walter Rudin (1976). *Principles of mathematical analysis*. 3rd Edition, McGraw-Hill Companies, Inc, ISBN Number: 0070856133, New York, USA
- Ward Cheney Will light. (2004). *A course in approximation theory*. China Machine Press, ISBN:7111135121., Beijing, China

Corn Monitoring and Crop Yield Using Optical and Microwave Remote Sensing

Jesus Soria-Ruiz¹, Yolanda Fernandez-Ordóñez² and Heather McNairn³

¹*National Institute of Research for Forestry, Agriculture and Livestock (INIFAP), Mexico.*

²*Postgraduate College in Agricultural Sciences (COLPOS), Mexico.*

³*Agriculture and Agri-Food, Canada.*

1. Introduction

Remote sensing (RS) is the practice of deriving information about the Earth's land and water surfaces using images acquired from an overhead perspective, using electromagnetic radiation in one or more regions of the electromagnetic spectrum, reflected or emitted from the Earth surfaces (Campbell, 2006). Using various sensors, we remotely collect data that may be analyzed to obtain information about the object, areas or phenomena being investigated. There are many forms in which the data are acquired, including variations in force distributions, acoustic wave distributions, or electromagnetic energy distributions.

Optical RS makes use of visible, near infrared and short-wave infrared sensors to form images of the Earth's surface by detecting the solar radiation reflected in these wavelengths from targets on the ground. Different materials reflect and absorb energy differently at these visible and infrared wavelengths. Thus, targets can be differentiated by their spectral reflectance signatures captured in the remotely sensed images. Optical RS systems are classified into the following types, depending on the number of spectral bands used in the imaging process: panchromatic imaging systems (i.e. Ikonos pan, Spot HRV-Pan), multispectral imaging systems (i.e. Landsat MSS, Landsat ETM, Spot HRV-XS, Ikonos MS), super spectral imaging systems (i.e. Modis & Meris), hyperspectral imaging systems (i.e. Hyperion on EO1 satellite).

In contrast, radar (Radio Detection and Ranging) sensors operate in the microwave portion of the electromagnetic spectrum beyond the visible and thermal infrared regions (Henderson & Lewis, 1998). Radars have long been exploited for communication and navigation purposes. More recently, Synthetic Aperture Radar (SAR) sensors have become an increasingly important source of information to support agriculture and natural resources monitoring and management. Operating in the microwave region of the electromagnetic spectrum improves signal penetration within vegetation and soil targets. Unlike optical sensors, the longer wavelengths of a radar imaging system are not affected by cloud cover or haze, permitting data acquisition independent of atmospheric conditions. Radar systems transmit microwave signals at specific wavelengths or frequencies according to their design specifications.

SAR sensors transmit microwave energy, illuminating the terrain, and measuring the amount of energy scattered by the target or surface. This response (also known as radar return or backscatter) is recorded by the SAR sensor. The greater the amount of energy scattered back to the sensor, the brighter the response recorded in the radar image. Active microwave sensors provide their own source of electromagnetic energy and are therefore capable of operating independent of sunlight. SARs can therefore acquire data day or night. Radars offer a variety of advantages for geoscientists and agronomists. These sensors are unaffected by adverse atmospheric conditions and because they operate independent of solar illumination, are available to acquire imagery 24 hours a day. Radar data provide a unique perspective of the landscape and many opportunities for quantitative terrain analysis.

Optical RS has been used for monitoring the state of the world's agricultural production, including identifying and differentiating most of the major crop types and conditions. However for agricultural regions under frequent cloud cover, the use of this technology for crop monitoring can be unreliable. In contrast, radar RS data are sensitive to vegetation biomass and structure and as a result these sensors are an attractive option for crop monitoring. Radar data and visible and infra-red wavelengths provide complementary information related to different target properties (Brisco and Brown, 1998). The synergy associated with data acquired by SAR and optical sensors has led to intensive research activities towards the application of RS technologies. Used together, optical and radar data provide a valuable information source for agricultural applications. Results have been very promising for a wide range of specific applications including crop type identification, crop condition, crop monitoring and crop yield.

Since the 1980s optical imagery from sensors such as Landsat and more recently Ikonos and Quickbird, has been used consistently to determine corn cultivated areas in Mexico, where over 7 million hectares of this staple crop are sown every year. Corn yield prediction is an information service provided by the National Institute of Research for Forestry, Agriculture and Livestock (INIFAP) to the Ministry of Agriculture, where it is used as a decision making aid. Techniques to combine information from optical and radar sensors have been proposed to detect and separate vegetation targets. However further development is needed to improve these techniques to increase accuracies for crop condition and crop monitoring. Information with respect to productivity is required as far ahead of harvest time as possible. Land use, based on intensive and diversified agricultural production, is integral to the economy in many countries. The heterogeneity of corn-growing conditions in developing countries makes accurate data for yield prediction difficult to obtain. Accuracy can be increased for a particular crop by integrating the information provided from optical and radar satellite images.

Government agencies require the best accuracy for production plots in order to relate these statistics to the general agricultural regional or nationwide productivity. More accurate information will support (a) timely responses and better decision making, (b) production risk reduction and (c) increased efficiency in crop management and production.

This chapter will first discuss the interaction of SAR microwaves with agricultural targets by considering the system and target parameters which influence the radar backscattering process. Then, the approaches for crop type identification, the first step in a monitoring program, will be discussed. This will be followed by a review of crop condition, crop monitoring and crop yield estimation using optical and radar data. The chapter ends with

comments on present and future research for agricultural applications using optical and microwave remote sensing.

2. Microwave Interaction with Agricultural Targets

Remote sensing observations have been used for identification and monitoring of agricultural targets since the late 19th century when balloons first started carrying photographic cameras and other instruments over the ground. All optical sensors are limited by solar illumination, cloud cover and haze. In spite of this, optical remote sensing has seen many useful if limited applications in agriculture and other areas. The use of radar sensors for agricultural applications has been intensively studied since 1970. The all weather, day or night data acquisition capability of radar systems, provides a more reliable data source. However, the interaction of the radar signal with agricultural targets is affected by a variety of factors. From this perspective, it is convenient to separate the discussion into radar system parameters which affect radar backscatter - such as frequency, polarization and incidence angle - and target parameters which influence the scattering process.

Target parameters can be related to the dielectric and geometrical properties of the material in question. Dielectric properties are very closely associated with the water content of the material while leaf shape and size (with respect to wavelength) are examples of geometrical characteristics (Brisco & Brown, 1998). The brightness of features in a radar image is dependent on the portion of the transmitted energy that is returned back to the radar (hence the term backscatter) from targets on the surface. The magnitude or intensity of this backscattered energy is dependent on how the radar energy interacts with the surface, which is a function of several variables or parameters. These parameters include the particular characteristics of the radar system and the generated image products as well as the characteristics of the incident surface (land cover type, topography, roughness, etc.).

Frequency or wavelength, incidence angle, and polarization are the primary system parameters which define a radar sensor and its data gathering characteristics. Other important system parameters which influence the type of product to use in an application include range and azimuth resolution, swath width, pulse length, transmitter power, and bandwidth.

2.1. The effect of frequency

With respect to the effect of frequency on microwave interaction with an agricultural target, the magnitude of the radar backscatter is dependent upon it (or wavelength) due to: differences in the dielectric constant of water content as a function of frequency; and to the relationship between wavelength and plant part size and/or penetration depth. The degree of moisture content affects the dielectrical properties of an object or medium. Changes in the electrical properties influence the absorption, transmission, and reflection of microwave energy. Thus, the moisture content will influence how targets and surfaces reflect energy from the radar signal and how they will appear on an image. In general, reflectivity (image brightness) increases with increased moisture content.

Since agricultural targets are composed of significant and varying amounts of water this frequency dependence on the dielectric constant is very important in the interaction process. As frequency decreases the signal penetration into crops and/or soil increases and the sizes

of the target components (i.e., leaves, stems, etc.) relative to the wavelength are smaller leading to a "smoother" target.

The radar return from each resolution cell of an agricultural target is the vector sum of electromagnetic (EM) fields scattered from the elements of the vegetation canopy, and those scattered from the soil beneath. Individual scattered contributions are determined by the scattered dimensions to wavelength ratio. Radar returns from an ensemble of scatterers are determined by the population of those scatterers, by the scatterer dimension function, and by the EM reflection coefficient of each scatterer. When the scatterer dimension is approximately the size of the wavelength the shape of the scatterer becomes very important in determining the backscattered EM fields and the detailed calculations and interpretations are complex.

Lower frequency radars are better suited for soil moisture estimation, especially when vegetations are present, while the higher frequency systems emphasize the crop component. When large amounts of vegetation are present L-band or lower frequencies are preferred to minimize the crop contribution to the backscatter (Brown *et al.*, 1992). The higher frequencies are generally preferable for crop type mapping, but this can change regionally depending on the crop mix and seasonally as a function of crop development.

When discussing microwave energy propagation and scattering, the polarization of the radiation is an important property. For a plane EM wave, polarization refers to the locus of the electric field vector in the plane perpendicular to the direction of propagation. The length of the vector represents the amplitude of the wave, the rotation rate of the vector represents the frequency of the wave and the polarization refers to the orientation and shape of the pattern traced by the tip of the vector. The linear combination of horizontal (H) and vertical (V) polarization states for the transmitted and received signals (with transmit denoted first) give HH, VV, HV and VH combinations. HV and VH are the cross-polarizations while HH and VV are the like polarizations. With respect to the effect of polarization on microwave interaction with agricultural targets, polarization can be a useful discriminant in SAR image analyses.

Most SAR systems are designed to transmit microwave radiation that is either horizontally polarized (H) or vertically polarized (V). If a SAR transmits and receives two orthogonal polarizations (such as H and V), and records both, and during processing retains the phase between these two polarization, then any transmit-receive polarization can be synthesized. It is the analysis of these transmit and receive polarization combinations that constitute the science of radar polarimetry. Systems that transmit and receive both of these linear polarizations are commonly used.

Radar systems can have one, two or all four of these transmit/receive polarization combinations. Examples include the following types of radar systems:

- HH or VV (or possibly HV or VH) – Single polarization
- HH and HV, VV and VH, or HH and VV – Dual polarization
- HH, VV, HV, and VH with phase information retained – Polarimetric

Quadrature polarization and fully polarimetric can be used as synonyms for "polarimetric". The relative phase between channels is measured in polarimetric radars and is essential for polarization synthesis, for generating a range of polarimetric parameters and for image decomposition.

Ferrazzoli's (2002) work to retrieve crop variables considered three main steps in the process: i) identification of a convenient radar configuration, ii) modeling and iii) solution of the inverse problem. We now discuss aspects relevant to applying these steps to crop monitoring.

SAR systems operate in different wavelength ranges or bands. The choice of wavelength is dependent upon the remote sensing application. L-band radars operate at a wavelength of 15-30 cm and a frequency of 1-2 GHz. S-band operates at a wavelength of 8-15 cm and a frequency of 2-4 GHz. At these wavelengths S-band microwaves are not easily attenuated, making these sensors useful for near and far range weather observation. C-band radars operate at 4-8 cm wavelengths 4-8 GHz frequencies. These frequencies are well suited for many marine applications, in particular ice detection and monitoring. At smaller X band wavelengths (2.5-4 cm and a frequency of 8-12 GHz), microwaves are more sensitive to small scale changes, and these sensors have been used for studies on cloud development and to detect light precipitation. X band microwaves are very easily attenuated making them well suited for very short range weather observation. K band sensors operate at a wavelength of 0.75-1.2 cm or 1.7-2.5 cm and a corresponding frequency of 27-40 GHz or 12-18 GHz.

Microwave scattering from vegetation is dependent upon both the SAR frequency and polarization. Therefore, radar imagery collected using different polarization and wavelength combinations may provide different and complementary information. Multi-polarization combinations permit an image interpreter to infer more information about the agricultural surface characteristics. With polarimetric sensors any linear, circular or elliptical polarization can be synthesized, in addition to other polarimetric information including for example, co-polarization phase statistics or polarization signatures. Polarization signatures are three-dimensional plots which assist in the interpretation of the scattering behavior of the target. The polarization signature of the target provides a convenient way of visualising a target's scattering properties. The signatures are also called "polarization response plots". An incident electromagnetic wave can be selected to have an electric field with ellipticity between -45° and $+45^\circ$, and an orientation between 0 and 180° . These variables are used as the x- and y-axes of a 3-D plot portraying the polarization signature. For each of these possible incident polarizations, the strength of the backscatter can be computed for the same polarization on transmit and receive (the co-polarized signature) and for orthogonal polarizations on transmit and receive (the cross-polarized signature). The strength is displayed on the z-axis of the signatures.

From experimental airborne SAR systems and the SIR-C (shuttle) mission SAR polarimetry has provided data to researchers who have studied a number of applications. It has been shown that the interpretation of a number of features in a scene is indeed facilitated when the radar is operated in polarimetric mode. The launch of RADARSAT-2 has made polarimetric data available on an operational basis, and uses of such data can be expected to become more routine and more sophisticated. Some agriculture applications for which polarimetric SAR has already proved useful include crop type identification, crop condition monitoring, soil moisture measurement and soil tillage and crop residue identification.

2.2. The effect of the incidence angle

With respect to the effect of incidence angle on microwave interaction with agricultural targets, the relationship between viewing geometry and the geometry of the surface features

plays an important role in how the radar energy interacts with targets and affects the corresponding brightness recorded on an image. Variations in viewing geometry will accentuate and enhance topography and relief in different ways, such that varying degrees of foreshortening, layover, and shadow may occur depending on surface slope, orientation, and shape.

The effect of the incident angle (θ) on radar backscatter has proven to be difficult to study with airborne SARs because of the rapid change of θ across the swath and the large dynamic range in backscatter many targets exhibit with varying θ . This challenge is largely overcome on satellite platforms for which the change in θ across the swath is much smaller. Nevertheless, little research has been undertaken to either correct for incidence angle change or to exploit target information as a function of differences in this angle. Mohan and Mehta (1987) used multiple incidence angles in the analysis of SIR-B data. They concluded that microwave radar response at L-Band (HH) at 25.6° and 45.2° for various land cover features is indeed a function of the incidence angle. For crops differences in radar response are also related to the imaging wavelength as well as to the crop type and its development stage. Shallower incident angles increase the pathlength through the vegetation maximizing response from the crop canopy itself and reducing the contribution from the soil. The radar signal strength decreases exponentially as the canopy depth increases due to both microwave absorption and scattering. Shallower incident angles increase the extinction of the radar signal also due to pathlength. The extinction coefficient is a function of both absorption and multiple scattering losses. At small incidence angles, (<30°) the backscatter is dominated by the direct scattering from the soil while for large incidence angles (>30°), the backscatter is dominated by the direct scattering from the canopy. As a result, small incidence angles are favored for soil moisture applications since roughness effects and vegetation attenuation are minimized at these angles (Daughtry *et al.*, 1991). In general, crop discrimination based on crop-canopy backscatter, is optimal at larger incidence angles.

Much of the information required for crop monitoring can be provided by satellite radar systems operating at L and C band, at linear co- and cross-polarizations and at an intermediate incident angle θ ranges (30° - 40°). Retrieval techniques based on multi-temporal data and assimilation of RS information in crop models appears promising. Significant further research and development is required to understand the information provided by polarimetric SARs and to develop the methods and models to extract soil and crop information from these advanced sensors.

Crop type and crop growth stage define the geometry of the canopy and the size, shape and orientation of the canopy constituents which influence microwave attenuation and scattering. The difference in radar backscatter between grain crops and broad-leafed crops is largely explained by the significant difference in the geometry of these canopies, as well as the increased biomass and water content of the broad-leafed crops. A larger backscatter response is associated with broad-leaf crops. In general, for broad-leafed crops like corn C- and L-band backscatter increases rapidly with plant growth, saturating early in the growing season. Little change in backscatter occurs during the rest of the growing season (Bouman, 1988). This saturation effect does not occur for grain crops such as wheat and barley. For these lower biomass crops, very dynamic temporal variations in radar backscatter are observed throughout the growing season. These variations are largely due to changes in crop structure (such as the emergence of grain heads) and canopy moisture changes (as occur during the period of senescence).

Research using C-HH or C-VV configurations has demonstrated that as fields are tilled, the increase in soil surface roughness results in an increase in backscatter (CCRS, 2008). Fields that are covered with significant post-harvest crop residue also experience an increase in backscatter. The increase in multiple scattering associated with rough tilled fields results in the depolarization of the microwave signal and induces a high cross-polarized backscatter. When residue cover is present on the fields, the increase in volume scattering has a similar effect and also results in higher cross-polarized responses (CCRS, 2008).

Plants and soil parameters, also affect microwave interaction with agricultural targets. Several parameters show consistent significant correlations with backscatter. These parameters include plant height, leaf area index (LAI), plant biomass, and plant water content (Brisco & Brown, 1998). Soil type affects the radar backscatter through the soil water holding characteristics and the relative amounts of bound and free water. Organic matter, salinity, sodium content and other soil properties can affect backscatter although these effects are less pronounced relative to the impacts observed from soil roughness and water content.

3. Agricultural Applications Using Optical and Microwave RS

The intensity of vegetation reflectance is commonly greater than from most inorganic materials. Consequently, vegetation appears bright in the near-IR wavelengths due mostly to the sensitivity of these wavelengths to internal plant pigmentation. Radar sensors are able to capture plant structure and soil moisture content. As a result, both optical and radar sensors can contribute to measuring and monitoring crop condition at different phenological stages, supporting the estimation of crop yields.

For optical and radar RS, the classification of crops can be challenging as the difference in reflectance or backscatter can be small among crop types. In addition, differences in crop condition among fields of the same crop type, can cause confusion in separating crop by type. Multiple scattering within a canopy can be useful for discrimination of crops using radar RS. Research using multiple dates of radar data has demonstrated that radar RS could play a very important role in agricultural applications (Zhang, 1999). In addition to the sensitivity of radar backscatter to crop canopy characteristics, given their all weather capability, SAR sensors provide a reliable option for crop monitoring. Nevertheless, much research remains in order to advance the use of SAR for operational applications.

3.1. Crop type and crop condition

In order to successfully apply RS technologies for crop classification, reflectance and backscatter signatures must be well defined for each crop type. Difficulty arises when signatures among crops are not sufficiently unique or when the variance in the signature within a single crop class is too large. Integration of optical and radar imagery is an attractive option. Both technologies offer complementary information about the crop canopy, and SAR sensors can fill the gap for optical acquisitions during periods of persistent cloud cover. End of season crop maps are of value, but provision of early season crop area estimates provide additional value as they support in-season crop production and yield forecasting. The heterogeneity of corn-growing conditions in many countries makes accurate data for yield prediction difficult to obtain. Small agricultural plots, irregular shapes, different sowing seasons and variations in crop cultivars are contributing factors to

classification errors. Accuracy can be increased for this particular crop through combinations of the information obtained from optical and radar satellite images.

Research studies have demonstrated that timing of image acquisition is very important to the success of crop mapping with optical imagery. Unless optical imagery is available during key stages of crop development and when field data are collected, these images alone will not provide the information necessary for operational field-level crop monitoring. Acquisition of SAR data during key phenological stages is more reliable and consequently, these data are an important information source for crop monitoring system. SAR or SAR-optical solution for crop monitoring have been explored in different regions of the world, and these studies are now reviewed.

Through integration of both optical and SAR imagery, McNairn *et al.* (2008) demonstrated that multi-temporal satellite are successful in the classification of crops for a variety of cropping systems. McNairn *et al.* (2009) indicate that multitemporal TerraSAR-X data can provide a classification accuracy of 84%; using a post-classification filter to remove noise in the map product final accuracies of 95% were obtained.

Many studies have reported on the use of airborne optical multispectral imagery to estimate crop parameters such as leaf area index, canopy temperature, and plant height. These studies examined the relationship between crop condition and spectral response to determine whether these images could be used to estimate various crop condition parameters. A number of statistically significant correlations exist between the image reflectance and the crop condition parameters and these correlations vary as a function of crop type, time of year, and crop condition. The results suggest that in many cases, multi-spectral optical imagery can be used to monitor variations in crop condition parameters across the growing season for a variety of crop types (Cloutis *et al.*, 1996).

On the other hand and as already explains, SAR investigations have confirmed that microwaves are sensitive to both soil and crop characteristics. Results using multi-temporal RADARSAT-1 imagery have confirmed that C-HH backscatter can detect differences in crop type, crop growth stage and crop indicators like crop height, biomass and leaf area index. Active microwave systems have a significant advantage over optical systems, particularly for crop monitoring, since SAR acquisitions are not impeded by cloud cover. The multi-beam modes associated with RADARSAT-1 also provide significant flexibility related to the timing, spatial resolution and incidence angle of the acquired imagery (McNairn *et al.*, 2000). The availability of multi-polarization data from a number of SAR sensors operating at different frequencies (X-Band from TerraSAR-X, C-Band from ASAR and RADARSAT-2 and L-Band from ALOS PALSAR) has significantly advanced the use of SAR for agriculture and land cover mapping. The multi-polarized configurations provide more information related to crop structure and crop condition. Using simulations of data in preparation for the availability of RADARSAT-2 data, the Canada Centre of Remote Sensing (CCRS) gathered airborne polarimetric imagery over several Canadian sites in 1998 and 1999. These data were used to evaluate the sensitivity of multi-polarized SAR data to characteristics of corn, wheat and soybean crops (McNairn *et al.*, 2000). Multiple polarizations provided a significant advantage for crop identification relative to the use of a single or dual polarization. The most important polarization for crop classification was the linear cross polarization (HV or VH). Cross polarization responses are a result of multiple scattering from within a crop canopy. Differences in canopy architecture due to differences in crop type result in unique cross polarization signatures. In a study using C-HH RADARSAT-1 data, multiple dates of

RADARSAT-1 imagery provided information on crop type and condition, with or without the integration of multi-spectral optical imagery. Regression analysis established that some indicators of crop vigor - in particular Leaf Area Index and crop height - were correlated with backscatter. The success of this RADARSAT-1 study was attributed to the acquisition of the SAR data during the critical reproduction and seed development crop growth stages. (McNairn et al., 2002).

3.2 Corn monitoring and crop yield

The prediction of final yield or determination and monitoring of crop condition throughout the growing season has considerable economic value for the agronomic community. Yields vary considerably from year to year. Consequently, considerable effort has been devoted to the estimation of final crop yield using remotely sensed data. Many studies have reported on the use of optical RS for crop monitoring and yield prediction using optical RS (Soria-Ruiz & Fernandez-Ordóñez, 2003; Ferencz *et al.*, 2004; Soria-Ruiz *et al.*, 2004; Calera *et al.*, 2004; Fang *et al.*, 2008). Although the successful use of optical RS has been demonstrated, implementation of a reliable operational approach dependent upon optical imagery is difficult in regions prone to continuous cloud coverage during the growing seasons. Some results for central Mexico are shown in Figures 1 and 2.

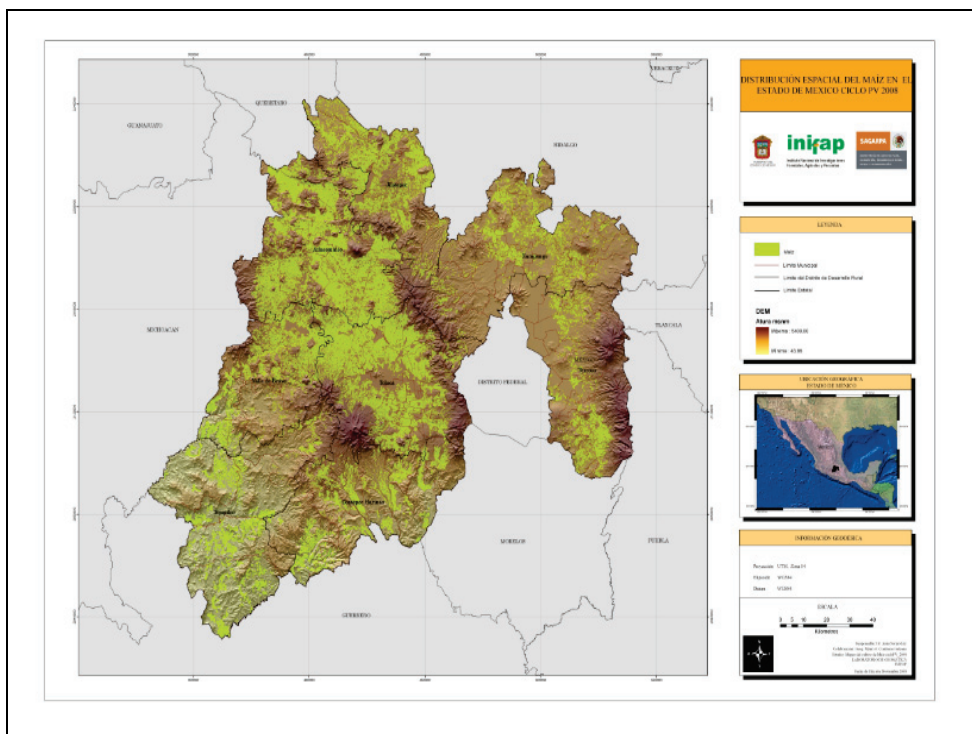


Fig. 1. Corn monitoring in 2008 using Spot Images. State of Mexico. Mexico.

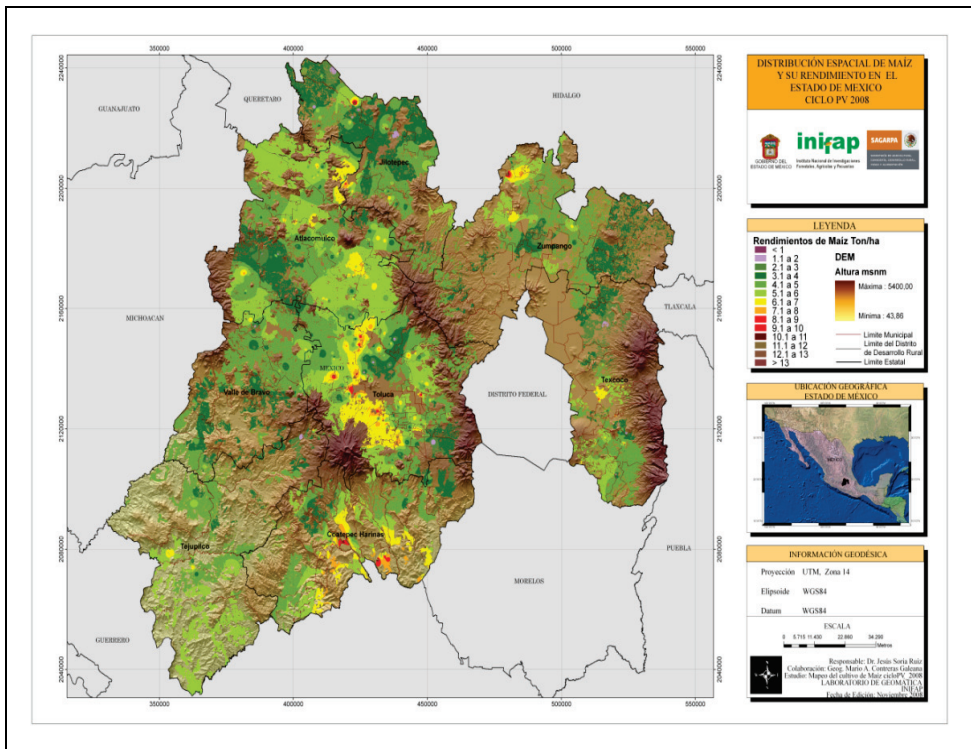


Fig. 2. Corn yield estimation using LAI and Spot images during 2008. State of Mexico, Mexico.

The dielectric constant of water is very large compared to the values of most other materials or targets. Consequently there is a strong dependence of the radar backscatter on the amount of water present in vegetation. However, in order to use the radar backscatter to assess potential crop yield directly (for example using regression analysis) or through a yield model it is necessary to relate the backscatter to vegetation parameters indicative of crop productivity. Leaf area index (LAI) along with the intensity of the solar radiation determines the amount of energy available to the plant for photosynthesis, which in turn drives the plant development and subsequent yield. LAI is related to whole plant biomass, light interception and loss of water through evapotranspiration. From LAI, Major *et al.*, (1986) defined LAI duration which provides a good indication of biomass throughout the season and of the total photosynthetic rate. Consequently, establishing a link between LAI and SAR backscatter would assist with the estimation of crop productivity and yield.

Airborne optical multi-spectral and C-band HH-polarized SAR imagery were acquired in conjunction with ground-based measurements of various crop conditions (Leaf Area Index, canopy temperature, plant height) at a test site in southern Alberta, Canada in July 1994. Data were acquired for a variety of crops (wheat, canola, peas and beans) and irrigation practices. A number of crop condition-imagery relationships were examined to determine whether the imagery could be used to estimate the various crop condition parameters. A

number of statistically significant correlations were found between the imagery and the crop condition parameters, and these correlations varied as a function of crop type, sensor and crop condition parameter. The results suggested that airborne remote sensing is well suited for measuring variations in crop conditions and that C-band SAR and multi-spectral imagery provided complementary information (Cloutis, 1999).

Several methods to estimate crop yield over large hilly areas that include high spatial resolution satellite imagery have been applied. These approaches incorporated QuickBird imagery with a production efficiency model (PEM) to estimate crop yield. The results indicated that QuickBird imagery can improve the accuracy of predicted results relative to the Landsat TM image. The predicted yield approximated well with the data reported by the farmers ($r^2 = 0.86$; $n = 80$). The spatial distributions of crop yield derived also offers valuable information to manage agricultural production and understand ecosystem functioning (Gang *et al.*, 2009). In order to attain better accuracy, Soria-Ruiz *et al.*, (2007) have applied optical and microwave RS data for corn monitoring and crop yield estimation under the heterogeneous corn-growing conditions in Mexico. Fusion of Landsat ETM+ and RADARSAT-1 provided better results than using optical data alone, for identifying crop and other land covers (Soria-Ruiz *et al.*, 2008). These results are summarized in Figure 3.

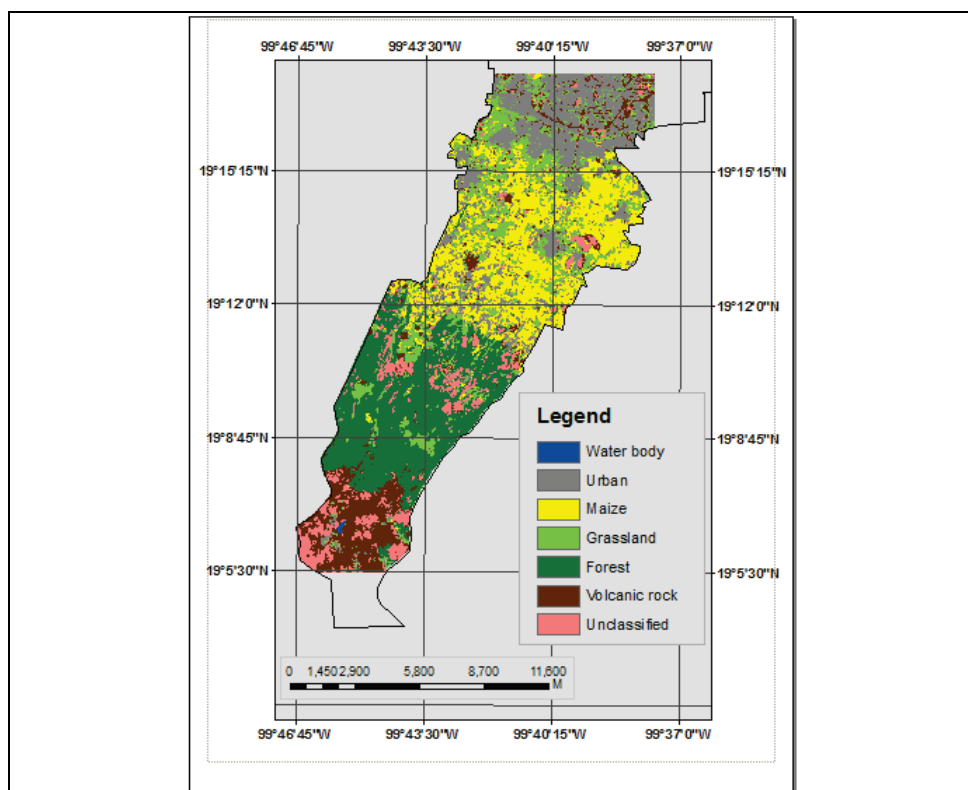


Fig. 3. Land-cover map obtained of data fusion from Landsat ETM and RADARSAT - 1 (Soria-Ruiz *et al.*, 2008).

4. Present and Future Research

Recent research to assess relative classification accuracies of multi-polarized combinations for target crops using airborne data has been reported. In addition to identifying crop type and variety, identifying crop growth stage is valuable. Crop condition, loosely defined as the vigor or health of a crop in a particular growth stage, is related to crop productivity and yield; however, the relationship is complex. Main crop condition indicators include biomass, height, leaf area and contents of plant water, chlorophyll and nitrogen. Crop-type and crop-condition mapping are among the applications that are expected to benefit the most from the technical enhancements embodied by RADARSAT-2. The potential of RADARSAT-1 data for these applications has been rated as "limited", whereas for RADARSAT-2 data this potential is anticipated to be "strong". The Science and Operational Applications Research for RADARSAT-2 Program (SOAR) is promoting the evaluation of SAR capabilities by providing images to our project: N° 2657 RADARSAT-2 for Corn Monitoring and Crop Yield in Mexico (Soria-Ruiz *et al.*, 2007).

Within this project, we are researching a) the use of RADARSAT-2 data, SPOT and Ikonos data to determine cultivated areas and monitor crop condition; b) relating polarization signatures from RADARSAT-2 data to corn Leaf Area Index and photosynthetic active radiation (PAR) parameters. The expected benefits of this project are: to obtain knowledge about crop type, crop condition and crop yield with better accuracy than with current methodologies; to support national corn farmers associations; to support the design of agriculture related policies within state agriculture plans; to support the corn product industry and aid government decision making. Relevant results and economical impact will imply operational usage of RADARSAT-2 data in the agricultural sector in Mexico (Soria-Ruiz *et al.*, 2007).

Satellite imagery is an efficient method for mapping crop characteristics over large spatial areas and tracking temporal changes in soil and crop conditions. Some SAR sensors such as RADARSAT-1 acquire imagery with a single transmit-receive polarization, providing a single radar image. Therefore, more than one acquisition date is usually required to estimate meaningful crop information. With RADARSAT-2 several new features are expected to prove beneficial to the agricultural sector. These advancements include the availability of dual-polarization and quad-polarization modes, enabling the simultaneous acquisition of multiple polarizations on transmit and receive. In the quad-polarized mode four polarization channels are acquired. Valuable crop information can be extracted from one RADARSAT-2 image, particularly if these data are integrated with optical or SAR data acquired at complementary (X and L-band) frequencies.

Crop type and crop condition mapping are among the applications together with crop yield that are expected to benefit the most from access to advanced sensors such as RADARSAT-2. The applications potential for RADARSAT-2 data is anticipated to be strong (van der Sanden, 2004). Images acquired in the polarimetric and ultra-fine resolution modes are expected to contain moderately improved information in support of crop-yield mapping. For crop condition mapping, the improved potential of the polarimetric and ultra-fine resolution data products for crop yield mapping can be explained by the increased sensitivity to crop structure and the capacity to obtain within-field zonal information.

5. Conclusion

Crop yield is a key element in rural development and an indicator of national food security. Optical and radar RS have been used separately in most cases for agriculture applications. Increased exploitation of SAR data is expected as these data become more readily accessible and as users become more familiar with the processing and interpretation of these data. In addition significant research is still required to advance methods and models to derive meaningful crop information from SAR data. Recent advances in the integration of optical and SAR data for agriculture applications are shedding more light on the communities understanding of how best to exploit both imagery sources. These advancements will assist in securing more accuracy results to support day-to-day decision making. Optical and radar RS are based on different physical principles. Radar data are sensitive to water content in the vegetation and the large scale structure of the canopy. Optical wavelengths respond largely to the internal leaf structure and pigmentation. SAR data do not directly measure plant parameters, such as chlorophyll, important for plant photosynthesis. However parameters indicative of plant production, such as leaf area index, influence radar backscatter.

Vegetation type identification has been successful when multi-dimensional approaches have been applied, often with accuracies at or above operationally effective goals of 90% classification accuracy. As with optical imagery, quantification of crop condition is more challenging for SAR data, particularly because radar backscatter also includes scattering contributions for the soil. Nevertheless, the integration of SAR and optical imagery for crop condition and productivity estimation appears promising.

6. References

- Brisco, B. & R.J. Brown. (1998). Agricultural applications with radar, In: *Principles and Applications of Imaging Radars (Manual of Remote Sensing Vol. 2)*, Henderson, F.M. and Lewis, A.J. (Ed), pages (381-403), John Wiley & Sons. ISBN: 0-471-29406-3, USA.
- Brown, R.J., M.J. Manore & S. Poirier. (1992). Correlations between X, C and L-Band imagery within and agricultural environment. *International Journal of remote Sensing*, Vol 13, No. 9, pp. (1645-1661, ISSN 0143-1161.
- Bouman, B.A.M.(1988). Microwave backscatter from beets, peas, and potatoes throughout the growing season. Spectral signatures of objects in remote sensing, Aussois, France, January 18-22, pp. 25-30.
- Calera, A., J. Gonzalez-Piqueras & J. Melia. (2004). Monitoring barley and corn growth from remote sensing data at field scale. *International Journal of Remote Sensing*, Vol. 25, No. 1, pp. (97-109), ISSN: 0143-1161.
- Campbell, J.B. (2006). Introduction to Remote Sensing, The Guilford Press, Fourth Edition, ISBN-13: 978-15938-53198. Printed in the United States of America.
- CCRS. (2008). Agricultural remote sensing research in preparation for RADARSAT-2 Soil tillage and crop residue cover. Accessed in June 2009 at: http://cct.rncan.gc.ca/radar/spaceborne/radarsat2/sim/spaceborne/tillage_e.php
- Cloutis, E.A., D.R. Connery, D.J. Major & F.J. Dover. (1996). Airborne multi-spectral monitoring of agricultural crop status: effect of time of year, crop type and crop

- condition parameter, *International Journal of Remote Sensing*, Vol. 17, No. 13, pp (2579 – 2601), ISBN: 1366-5901.
- Cloutis, E.A. (1999). Agricultural crop monitoring using airborne multi-spectral imagery and C-band synthetic aperture radar. *International Journal of Remote Sensing*, Vol. 20, No. 4, pp. (767-787), ISSN: 0143-1161.
- Daughtry, C.S.T., K.J. Ranson & L.L. Biehl. (1991). C-band backscattering from corn canopies. *International Journal of Remote Sensing*, Vol. 12, No. 5, pp. (1097-1109), ISSN: 1366-5901.
- Fang, H., S. Liang, G. Hoogenboom & M. Cavigelli, (2008). Corn yield estimation through assimilation of remotely sensed data into the CSM-CERES-Maize model. *International Journal of Remote Sensing*, Vol. 29, No. 10, pp (3011-3032). ISSN:0143-1161.
- Ferrazzoli, P. (2002). SAR for agriculture: advances, problems and prospects. *Proc. 3rd International Symposium "Retrieval of Bio- and Geophysical Parameters from SAR Data for Land Applications"*, pp. 47-56, Sheffield, UK, 11-14 Sept. 2001.
- Ferencz, C., P. Bognár, J. Lichtenberger, D. Hamar, G. Tarcsai, G. Timár, G. Molnár, S. Pásztor, P. Steinbach, B. Székely, O. Ferencz & I. Ferencz-Árkos. (2004). Crop yield estimation by satellite remote sensing. *International Journal of Remote Sensing*, Vol. 25, No. 20 (4113-4149), ISSN: 0143-1161.
- Gang, P., G.J. Sun & F.M. Li (2009). Using QuickBird imagery and a production efficiency model to improve crop yield estimation in the semi-arid hilly Loess Plateau, China. *Environmental Modelling & Software*, Vol. 24, No. 4, pp. (510-516), ISSN: 1364-8152.
- Henderson, F.M. & A.J. Lewis. (1998). *Principles and Applications of Imaging Radars (Manual of Remote Sensing Vol. 2)*, John Wiley & Sons. ISBN: 0-471-29406-3, USA.
- Major, D.G., G.B. Schaalje, A. Asrar & E.T. Manemosu (1986). Estimation of whole plant biomass and grain yield from Spectral reflectances of cereals. *Canadian Journal of Remote Sensing*, Vol. 12, No. 1, pp. (47-54), ISSN: 0703-8992.
- McNairn, H., J. Shang, C. Champagne & X. Jiao. (2009). TERRASAR-X and RADARSAT-2 for crop classification and acreage estimation. IGARSS-2009. 12-17 July 2009, Cape Town, South Africa, paper 1115.
- McNairn, H., C. Champagne, J. Shang, D. Holmstrom & G. Reichert. Integration of optical and Synthetic Aperture Radar (SAR) imagery for delivering operational annual crop inventories. *ISPRS Journal of Photogrammetry & Remote Sensing*, in press, (doi:10.1016/j.isprsjprs.2008.07.006). (2008)
- McNairn, H., J.J. van der Sanden, R.J. Brown & J. Ellis. (2000). The potential of Radarsat -2 for crop mapping and assessing crop condition. Second International Conference on Geospatial Information in Agriculture and Forestry. Lake Buenavista, Florida. Vol. 2, pp. 81-88.
- McNairn, H., J. Ellis, J.J. van der Sanden, T. Hirose & R.J. Brown (2002). Providing crop information using Radarsat-1 and satellite optical imagery. *International Journal of Remote Sensing*, Vol. 23, No. 5, pp. 851-870, ISSN: 1366-5901.
- Mohan, S. & R.L. Mehta. (1987). Multiple incidence angle SIR-B data analysis over parts of the a plains, *Journal of the Indian Society of Remote Sensing*, Vol. 15, No. 2 (37-41), ISBN: 1-59454-326-7.

- Soria-Ruiz, J., & Y. Fernandez-Ordonez. (2003). Prediction of corn yield in Mexico using Vegetation Indices from NOAA-AVHRR satellite images and Degree-days. *Geocarto International*, Vol, 18, No. 4, pp. (33-42), ISSN: 1010-6049.
- Soria-Ruiz, J., Y. Fernandez-Ordonez & R. Granados-Ramirez. (2004). Methodology for prediction of corn yield using remote sensing satellite data in Central Mexico. *Investigaciones geograficas*, No. 55, pp. (61-78), ISSN: 0188-4611.
- Soria-Ruiz, J., Y. Fernandez-Ordonez, H. McNairm & J. Bugden-Storie. (2007). Corn monitoring and crop yield using optical and Radarsat-2 images. Geoscience and Remote Sensing Symposium, IGARSS 2007. 23-28 July 2007, pp.(3655-3658), Barcelona, Spain, ISSN: 978-1-4244-1211-2
- Soria-Ruiz, J., Y. Fernandez-Ordonez & I. Woodhouse. Land cover classification using radar and optical images: a case study in Central Mexico. *International Journal of Remote Sensing*, (Accepted in 2008), 2008-0572.
- Zhang, W.T. (1999). Analysis of Advantage on Radar Remote Sensing for Agricultural Application. Remote Sensing Center of China Agricultural University Beijing 100094, China.
- van der Sanden, J.J. (2004). Anticipated applications potential of Radarsat - 2 data. *Canadian Journal of Remote Sensing*, Vol. 30, No. 3, pp. (360-379), ISSN: ISSN 1712-7971.

Radargrammetric SAR image processing

Stéphane Méric, Franck Fayard and Éric Pottier
*Institute of Electronics and Telecommunications of Rennes,
European University of Brittany
France*

1. Introduction

Throughout history, humans have tried to represent what they see through images. Mapmakers have always sought ways in which to represent both the location and the three dimensional shape of land. At the beginning, the way to obtain a 3D representation of land was to measure planimetry and height (as we can identify later by longitude, latitude and height) using basic measuring devices. Nowadays, the improvements of airborne and spatial instruments make it possible to produce images by sensing the electromagnetic radiation from the Earth. So, we can distinguish two classes of remote sensors: optical sensors and radar sensors. Optical sensors, such as Landsat or SPOT 5, operate around the visible spectrum and provide images with a fine resolution (less than 5 meters for SPOT 5). Thus, these kinds of sensors become very useful for civilian applications (cartography, elevation map, agriculture, hydrography, management of natural hazards, meteorology, geology, deforestation and so on). Considering the subject of this chapter, the extraction of terrain elevation by stereoscopic images can give digital elevation models with an error of about 5 meters (Toutin, 2000). However, optical sensors could be critically useless because of weather conditions or lack of light (i.e. sun). Thus, the use of radar sensors is a good way to overcome the limitations of optical sensors: not very sensitive to rain, considered as active sensors (because they have their own source of energy). Thanks to the signal processing applied to radar signal (pulse compression and synthetic aperture), radar systems can provide images with a very high resolution (for example, Radarsat-2 has an ultra-high resolution mode of about 3 meters for resolution). So, radar images are considered as additional information to optical images. With regard to these properties, one can estimate that radar images are used to get elevation terrain. The more intuitive way to extract depth information from remote sensing images is stereogrammetry. As the brain operates on optical images from eyes, the technique of radargrammetry is applied to SAR (Synthetic Aperture Radar) stereo data and provides digital elevation models (DEM). Considering this preamble to the radargrammetric world, this chapter examines one way to produce digital elevation models (DEM) from a mountainous area (the French Alps) and the way to improve the accuracy of the DEM. So, we will organize the discussion in three parts. In part 1, in order to better understand the stereo computation, we need to explain the basic characteristics of a radar image, which is particularly important to be considered during the radargrammetric processing. Thus, a radar image can be seen as a distribution of reflected electromagnetic energy on the ground. So, each element (i.e. a pixel) of an image is described by its size along the azimuth and range axis. Also, specific characteristics of a radar image are described as layover, shadowing and foreshortening. Because radargrammetric processing is

based on fitting images, we need to establish a common reference to radar images and to set up geographical coordinates for each image. Considering the position of the sensor, we can establish rigorous radar projection equations that can be compared to the so-called photogrammetric equations. As the radiometry is important to interpret a radar image, we consider the main radiometric models and the speckle phenomenon considered as noise in the SAR image. In part 2, considering a radar image, we will present the basic operations of extraction from satellite radar data. There are several methods to reconstruct elevation model from radar images. These images are essentially described as 2D information. So, one has to extrapolate 3D information from 2D description (as DEM). There are different methods to do this: clinometry, stereoscopy, interferometry and polarimetry. Since any sensor, system or method has its own advantages and disadvantages, the choice of a radargrammetric technique depends on the sensors and the means used during image acquisition. For the stereoscopic method, the capability of radar image pairing to achieve radargrammetric processing depends on geometric configuration in relation with the radar trajectory. Considering this radar trajectory, one can define the radar stereo base, the intersection angle and the parallax. We propose to review different ways to process the matching operations. These ways are correlation operations based upon searching for match points as area correlation methods or elementary correlation. After that, we will expose some improvements in the matching process (pyramidal scheme, speckle filtering). Part 3 will deal with the description of radargrammetric applications on real data (from SIRC shuttle mission) and the different steps to obtain a DEM. First of all, we describe the radar image and especially the relations between the satellite route and the ground radar image. This step is crucial in order to efficiently match the stereo radar images. Also, we explain the significance of using ground control points (GCPs) to rectify radar images. The next step is the matching operation between the two stereo SAR images. It consists in determining the point co-ordinates inside the secondary image for each point in the reference image, which is called the corresponding pixel. The computation of the 2D normalized cross-correlation coefficient is used on SAR images. At this step, we use a hierarchical strategy to reduce process time and use a filter to get the high accuracy disparity map. Then, we apply the rigorous radar stereo intersection problem and compute the stereo radargrammetric equations. Using the solutions, we obtain a DEM from the stereo radar images. This DEM is compared with a reference DEM. At the end, we move on to the point of improvement of the DEM: obvious improvements (correction of incoherent points) and further improvements in progress (use of adaptive correlation windows or polarimetric parameters).

2. Radargrammetric sensors

2.1 Introduction

As the acronym RADAR means "Radio Detection and Ranging", the basic principles are to detect and range objects located in front of the radar system. In the context of remote sensing, a scene (i.e. the terrain) is considered to be imaged by transmitting an incident electromagnetic wave from the radar, reflecting towards the radar (monostatic consideration) and receiving the reflected wave. The radar signal is obtained through the conversion of an electrical current on the antenna surface induced by an electromagnetic field around this antenna and vice-versa. Thus, the received signal contains information about the scene such as dielectric properties. Firstly, we can describe the received power P_r through the radar equation:

$$P_r = \frac{P_t \cdot G^2 \cdot \lambda_c^2}{(4\pi)^3 R^4} \sigma$$

where P_t is the transmitted power, G is the gain of the transmitted and received antenna, λ_c is the wavelength of the transmitted wave, R represents the distance between the radar and the scene and σ is the radar cross section. This parameter depends on many parameters such as the frequency and polarisation state of the emitted wave, the dielectric nature of the object, geometrical body of the object and so on. For example, buildings forming a corner with the ground or other buildings, correspond to high reflected energy. Conversely, roughness surfaces diffuse the incident energy and correspond to low reflected energy.

2.2 Signal processing and radar imaging

The side looking aperture radar (see figure 1) makes it possible to get radar images of the ground by emitting pulses of electromagnetic waves. The platform (aircraft or satellite) of

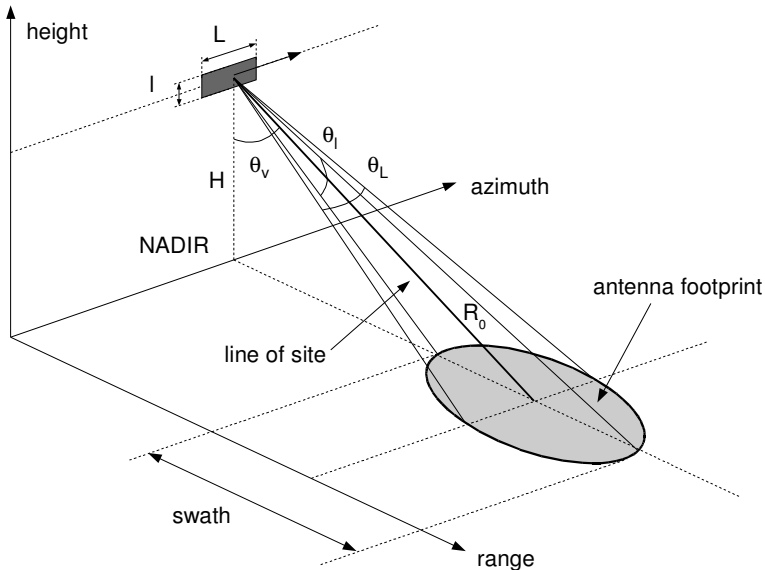


Fig. 1. Configuration of side-looking

such a radar travels forward in the flight direction or along-track (azimuth axis) with the nadir directly beneath the platform which is at the height H . The range axis refers to the across-track dimension perpendicular to the flight direction. The microwave beam is transmitted obliquely (elevation angle θ_v to the direction of flight illuminating a swath. The side looking geometry is necessary to avoid the Doppler ambiguity. Some configurations exhibit a squint angle rather than an antenna pointing perpendicularly to the flight direction. The footprint of the antenna is defined through the line of sight of the main beam of the antenna and the aperture angles (along the range and azimuth axis) of this antenna. This aperture angle refers to the physical dimension of the antenna (respectively l and L). Swath width refers to the strip of the Earth's surface from which data is collected by the radar. The longitudinal extent of the swath is defined by the motion of the aircraft with respect to the surface, whereas the swath width is measured perpendicularly to the longitudinal extent of the swath.

by a matched filter that fine tunes the range resolution δ_d :

$$\delta_d = \frac{c}{2.B_p}$$

Thus, the range resolution is inversely equal to the bandwidth of the emitted signal. Therefore, using the parameters of the SIR-C mission and especially the value of B_p (10 Mhz), we can get a range resolution of about 15 meters.

2.2.3 Azimuth resolution

Crossrange resolution is naturally achieved by use of an antenna with a narrow beam and specified by θ_L . If the beamwidth along the crossrange axis is given approximately by $\theta_L \approx \lambda/L$ where λ is the wavelength of the transmitted signal, the corresponding azimuth resolution δ_a at range R_0 is then $\delta_a = \lambda.R_0/L$. Considering the SIR-C mission again, the azimuth resolution would be about 30 kilometers, which is also unacceptable. The synthetic aperture processes the received signal by using the fact that the radar views the scene from slightly different angles. These different views (at each emitted pulse) are obtained because the radar moves through its synthetic aperture. Considering the response of one point on the ground, the reflected signal from this point can be seen as a frequency modulated signal (Doppler frequency). Also, a matched filtering operation is applied along the azimuth axis under certain assumptions (width of Doppler spectrum and duration of the seen point), we write the azimuth resolution δ_a as

$$\delta_a = \frac{L}{2}$$

which gives an azimuth resolution of 6 meters considering the characteristics of the antenna of the shuttle (SIR-C).

2.2.4 Radar image corrections

The values of resolution given above are usually better than those obtained by the real system. Also, the signal processing must take into account undesirable effects that affect the performances of the radar. Concerning our discussion about radargrammetry, we can note among these effects:

- the range migration that can be modelled by the parabolic variation of the distance between the target point on the ground and the radar along the synthetic aperture (this point is corrected by different processing methods (Carrara et al., 1995)),
- the radiometric variations due to the change of received signal power from the beginning of the swath (near range) to the end of the swath (far range) for each position of the radar (using well-known ground points as RCS references can correct this effect),
- the motion compensation that corrects the deviation of the antenna from its nominal flight path.

Despite the corrections, some errors such as bad localization of pixels can still be found on the radar image. These errors can finally be eliminated by making use of ground control points such as buildings, cross-roads, mountain tops and so on.

2.3 Geometric interpretation of a SAR image

Actually, the importance of geometry for the interpretation of radar images recurs throughout this chapter. As we wrote before, the radar system can be considered as an 'all-weather' system and contrary to optical imagery, does not need ambient light or an external source of energy to obtain images. However, upon comparing a SAR image and an optical image, we can assume that certain properties of an optical image are not included in the radar image. For example, this phenomenon is clearly visible when looking at pixels farther from the radar, which appear smaller along the range axis than pixels closer to the radar. Although the cross range resolution is not affected by the radar imagery process, we suppose that the relief of the terrain will induce radiometric and, especially, geometric distortion. Thus, if we consider a ground point with a height h and located at a range R from the radar at the height H , the position x_{sol} along the range axis is given by:

$$x_{sol} = \sqrt{R^2 - (H - h)^2}$$

and means that a single radar image doesn't give the altitude of a pixel but must be associated to a height model of the terrain. This is one of the tricky points about the interpretation of a radar image.

2.3.1 Distortion of a radar image

The projection of a terrain slope on the slant range of the radar induces well-known distortion that can be expected as regards the planimetry (see figure 3). And, the values of resolution

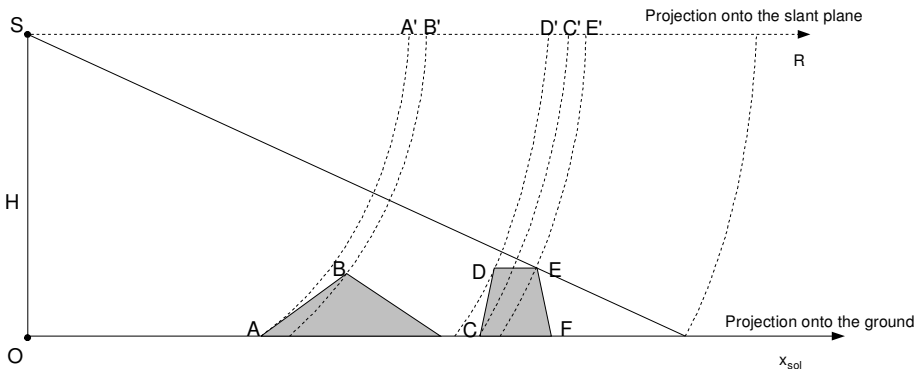


Fig. 3. Geometrical distortion occurs in the slant radar image.

given above are usually better than those obtained by the real system. Also, the signal processing must take into account undesirable effects that affect the performances of the radar. Concerning our discussion about radargrammetry, we can note among these effects the foreshortening effect, the layover effect and the shadowing effect which result from relief displacement.

2.3.2 Foreshortening effect

The foreshortening effect occurs when the radar beam reaches the base of a slope tilted towards the radar before the top of this same slope. The straight segment [AB] and its image [A'B'] onto the slant range illustrate this effect in figure 3. Thus, the radar measured distance

seems to be shorter than the real one and this effect is maximum when the radar beam is perpendicular to the mountain slope.

2.3.3 Layover effect

The layover effect occurs when the radar beam reaches the top of a mountain or a hill before its base. The straight segment [CD] and its image [C'D'] onto the slant range illustrate this effect in figure 3. Also, a terrain slope towards the radar produces a viewing permutation between the top and the base of a mountain on a radar image.

2.3.4 Shadowing effect

The shadowing effect occurs when the radar beam is not able to illuminate the radar scene. This effect that can be seen in figure 3 considering the straight segment from the point E', image of the point E, to the end of the swath. Also, the radar shadow is considered as an optical shadow and induces a black area on the radar image because no reflected wave comes from this kind of region (for example, point F is not seen on the radar image). All these effects are quite severe in order to understand a radar image well and especially in mountainous areas. Moreover, the incidence angle of the radar beam is another important parameter to estimate the influence on the interpreted radar image. So, the efficiency of the radargrammetric processing must take into account these characteristics.

2.3.5 Geometrical model of the radar position

The capabilities to link each pixel of a radar image to a real position on the terrain is one of the most important steps of the radargrammetric processing because correction, rectification, resizing and superimposition processings of the image need to know the geometrical position of a pixel. The model of the platform (e.g. in our study a satellite) flight path is described in figure 4 provides relation between radar image indexes and the terrain (Girard, 2003) thanks to

- radar parameters (frequency, size of the antenna, incidence angle ...),
- instantaneous position and motion of the radar platform,
- an ellipsoidal model of the Earth.

For the last item, the figure 4 gives several parameters to describe the model as

- angles λ and ϕ which are respectively the longitude position and the latitude position,
- Earth's referential (G, i, j, k) which is established by the centre of the Earth G, the i-axis towards the Greenwich meridian, the k-axis coinciding with the Earth's axis of rotation and the j-axis forming a right-handed system with i-axis and k-axis instantaneous position and motion of the radar platform,
- referential of satellite (S, l, r, t) linked to the satellite and described by the position S of the satellite, the l-axis colinear to the vector \vec{GS} , the t-axis simultaneously perpendicular to the l-axis and the vector \vec{S} and the r-axis forming a right-handed system with l-axis and t-axis.

As described in (Dhond & Aggarwal, 1989), stereoscopic processing needs to know several parameters which corresponds, for radargrammetry processing, to:

- the wavelength λ_c of the transmitted wave,
- the azimuth resolution δ_a and the range resolution δ_d ,

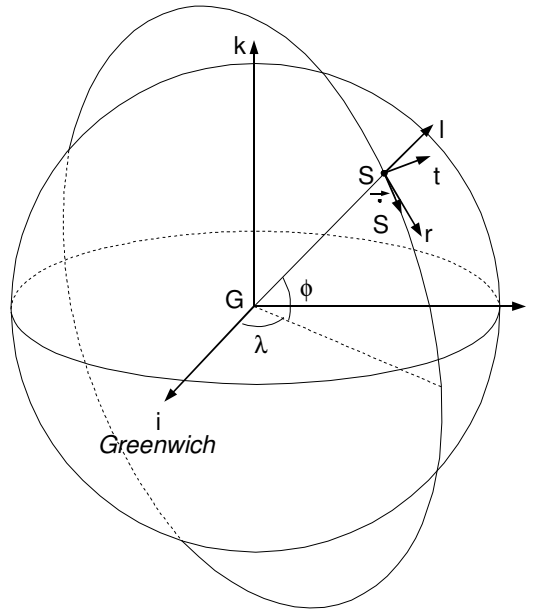


Fig. 4. Position and motion of a satellite

- the central Doppler frequency f_D of the received signal,
- the time t_0 for which the values of the position and the velocity of the satellite are known,
- the initial time t_{init} of the beginning of the radar image,
- the range distance r_0 given for a reference line of the image,
- parameters that make it possible to calculate the behaviour of the satellite (position, orientation, velocity) for each value of time.

Actually, the position and velocity of the satellite are known at specific values of time which are called ephemerides. Thus, we have to interpolate the path of the satellite in order to have all the position and velocity of the satellite along the flight path.

2.3.6 Geographic coordinates of a radar image

Thanks to the parameters describing the flight path of the satellite, it is possible to give geographic information for each pixel of the radar image. In order to establish this relation and to measure locations accurately, some references of coordinates are used (Dufour, 2001). In this chapter, we use the global coordinate system which has been described before (see figure 4). The ellipsoidal height h of a point is the vertical distance of the point in question above the reference ellipsoid. The reference ellipsoid is described by the WGS84 system (geodetic) and the significant parameters defined by

- the semi-major axis $a = 6378137.0$ meters,
- the semi-minor axis $b = 6356752.3$ meters.

Considering a point M defined by its height h and its geocentric coordinates (x, y, z) in the (G, i, j, k) reference, we can write the above expression:

$$\frac{x^2 + y^2}{(a + h)^2} + \frac{z^2}{(b + h)^2} = 1$$

2.3.7 Radar coordinates and image coordinates

In the radar reference, each pixel of the image gives information about the range distance r and the time t elapsed since the beginning of the recorded raw data. Another way to describe a radar image refers obviously to the azimuth u and range r coordinates. Also, a data transformation is feasible via the number of looks N_f used to establish the radar image (Curlander, 1991) and the spatial sampling frequency f_e along the range axis:

$$\begin{cases} t &= \frac{N_f}{f_r} \cdot u + t_{init} \\ r &= \frac{c}{2f_e} \cdot v + r_0 \end{cases}$$

We have to note that the values of u and v are immediately obtained from the radar image. At this time, we have to set up the coordinates t and r in the defined Earth’s reference.

2.3.8 Range sphere and Doppler cone

We can define the range sphere as the constant distance r of a point M from the radar located at the position S:

$$|\vec{SM}| = r$$

Moreover, the Doppler cone is the cone of equal Doppler frequency and has its apex located at the centre of the range sphere:

$$f_D = \frac{2}{\lambda_c} \cdot \frac{\vec{S} \cdot \vec{SM}}{|\vec{SM}|}$$

In the case of side-looking radar, the centroid Doppler frequency f_D is equal to zero, which means the cone becomes a plane perpendicular to the velocity vector \vec{SM} . Considering the coordinates

- (x, y, z) of the point M on the radar image,
- (X_S, Y_S, Z_S) of the position S of the radar,
- $(\dot{X}_S, \dot{Y}_S, \dot{Z}_S)$ of the velocity of the radar,

the equations 2.3.7 and 2.3.8 establish a system of 2 equations of 3 unknowns (x, y, z) whose solutions describe a circle called Doppler circle (see figure 5). The Earth’s model as defined before and raised of height h_e finally makes it possible to get two solutions of the given system. One of these can be eliminated considering the line of site (LOS) (figure 6). Unfortunately, the different slopes of terrain above the Earth’s ellipsoid that we described before and the associated effects (especially in foreshortening areas) on the radar image result in more than one solution.

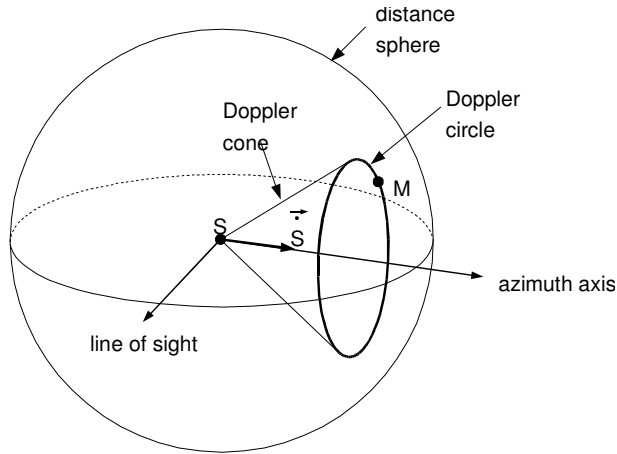


Fig. 5. Description of the distance sphere, Doppler cone and Doppler circle

2.4 Radiometric phenomena in an SAR image

The first remark concerns the main difference between the radar image and the optical image. The Earth's surfaces reflecting strong energy towards the radar correspond to very bright pixels on the radar image (and can appear dark on an optical image). The radar scene reflects a certain amount of radiation according to its geometrical and physical characteristics. This part will deal with radiometric phenomena that occur on the ground and which essentially depend on the electrical properties of the soil and the roughness of the area. Moreover, as we have seen before, the geometric shape of an area or an object on the ground mainly determines the radiometry of a pixel and the brightness of a feature could be a combination with other objects. Another important parameter is the wavelength of the incident radiation wave and the electromagnetic interaction falls with either surface interaction or volume interaction. Also, we can separate the interactions into two main topics:

- smooth surfaces that reflect (nearly) all the incident waves towards to a particular direction: specular reflection. If the surface is tilted towards the radar, the corresponding radar image appears very bright. Conversely, if the surface is not turned towards the radar (e.g. calm water or paved roads), the surface appears dark on the radar image;
- rough surface that scatters the incident wave in many directions: diffuse reflector.

In order to determine the degree of roughness of a surface, we use to establish (Beckman & Spizzichino, 1987) a relation between the state of the surface quantified by the average height variation h , the wavelength of the wave λ_c and the local incidence angle θ_i (see figure 7). This relation is known as the Rayleigh criterion:

$$\begin{cases} h < \frac{\lambda_c}{8 \cos \theta_i} \text{ lorsque } \lambda_c \gg h \\ h < \frac{\lambda_c}{32 \cos \theta_i} \text{ lorsque } \lambda_c \simeq h \end{cases}$$

Let us consider the local incidence angle: an incidence angle is the angle between the radar beam and the target object. The value of this angle determines the radar appearance of this

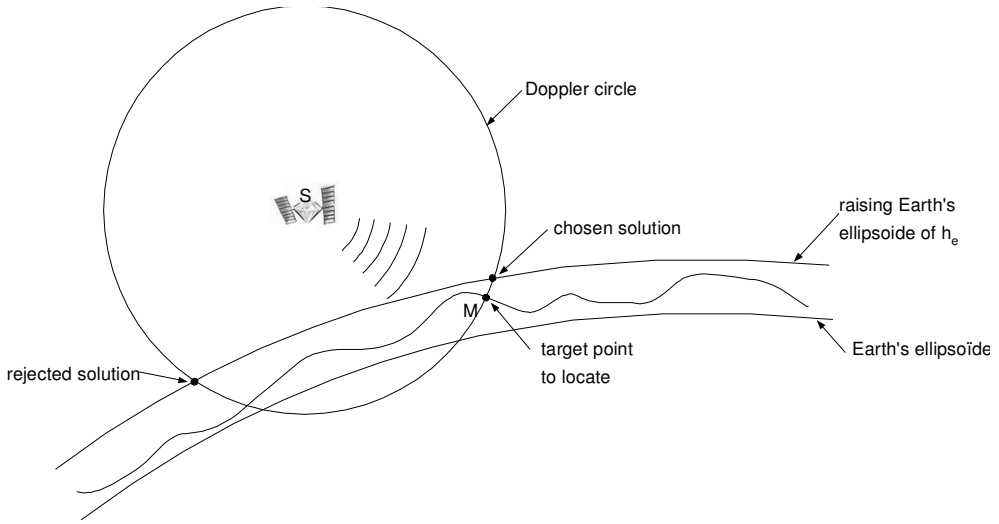


Fig. 6. Intersection of the Doppler circle and the Earth's surface

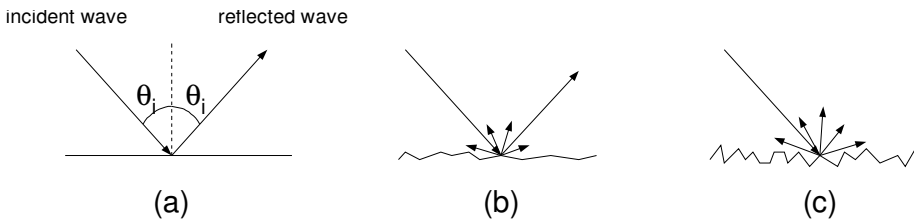


Fig. 7. Rayleigh criterion: (a) smooth surface, (b) low roughness surface and (c) high roughness surface

target on the radar image. Moreover, we can attach to each pixel of the radar image a local incidence angle so that we can notice variations in pixel brightness concerning one target object (rocks, trees, grass, buildings). Finally, we can note that the variation of incidence angles is less for a satellite radar than an airborne radar because of the height of the platform. Among the natural Earth's surfaces, we can characterize (Ulaby, 1981) three kinds of surface

- bare surface where simple reflections occur and the amount of energy towards the radar depends on the roughness of the soil,
- farmed surface where reflections are quite complex and depend on the crops, the moisture, the direction of the parcels and so on,
- vegetation surface where the reflection phenomena essentially depend on the wavelength. For example, the waves of the radar band X are only reflected by the top of the canopy. Lower wavelength waves penetrate the canopy and volume scattering has to be considered. Finally, some features on the ground can be considered as close targets, which means these features have two (or more) surfaces (generally smooth) forming a right angle and cause double (or more) bounce reflections (figure 8).

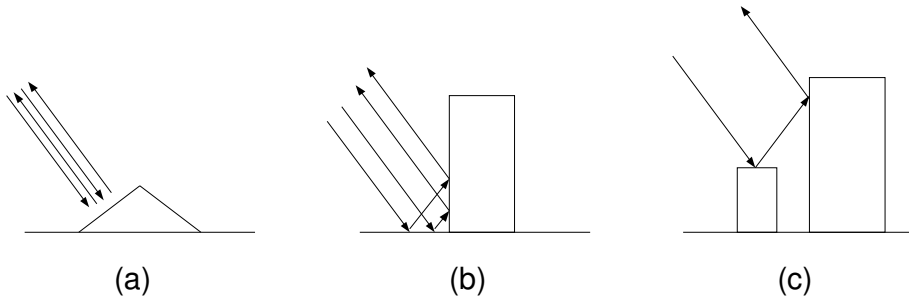


Fig. 8. Reflection phenomena: (a) from slope towards the radar, (b) from corner reflector (double bounce reflection) and (c) multiple bounce reflections.

The typical occurrence of this phenomenon is the corner reflection. Corner reflectors are very common in urban sites and show up as very bright targets on the radar image.

2.4.1 Speckle phenomena

As the radar image is created through a radar coherent wave, a particular effect modifies the radiometry of pixels as a noise-like effect inherent in coherent imaging systems. This effect is obviously visible on large covered-grass areas and looks like a "salt and pepper" texture. This texture is due to the chaotic response of multiple small targets on the ground whose global response is seen as a constructive or destructive random process. Thus, this kind of process randomly produces bright and dark pixels: the radar image is speckled. Many articles are dedicated to the study of the speckle phenomena (Goodman, 1976). Even it could be considered as information for special applications, the speckle effect is seen as a multiplicative noise and degrades the quality of a radar image.

3. Radargrammetric operations

3.1 State-of-the-art

The definition of radargrammetry has been stated by Leberl (Leberl, 1990): "Radargrammetry is the technology of extracting geometric information from radar images". To extract the geometrical characteristics of the ground, four different techniques are implemented: stereoscopy, clinometry (Horn, 1975), interferometry (Massonet & Rabaute, 1993) and polarimetry (Schuler et al., 1996). These are usually combined with SAR systems which have been briefly presented in this paper. Because the aim of this chapter is only to expose the radargrammetry as a radar stereoscopic method, the other ones will not be more developed. The first works on radargrammetry began after the Second World War and the first principles were defined by La Prade (La Prade, 1963). These works were completed by several mathematical developments (Gracie et al, 1970) and fully developed by numerous researchers (Rosenfeld, 1968) (Leberl, 1990) (Polidori, 1997). All of these developments were tested and improved thanks to several operational measurements both airborne (for example (Azevedo, 1971) mapping the world's tropical belt) and spatial (for example (Schrier, 1993) geocoding radar images from ERS-1 mission). Since the 1980s with the Shuttle Imaging Radar (SIR-A, SIR-B and especially SIR-C), the European satellite (ERS-2 and ENVISAT), the Canadian sensor (RADARSAT-1 and 2), the number of researchers working on the radargrammetric topic has increased and data

analysis has become more sophisticated (various incident angles, various frequencies and polarisations of the wave and so on).

3.2 Basics of radargrammetry as a radar stereoscopic method

3.2.1 Principle

Stereoscopy is a viewing method that forces our eyes to see, at the same time, two images taken from different angles. This technique allows us to see in three dimensions as it reinforces physiological indicators. The indicators used by stereoscopic method are parallax and convergence angle and can be defined as follows:

- the parallax P of an observed point is a parameter that is directly connected to the point elevation and it increases with the altitude of the point,
- the convergence angle $\Delta\theta_v$ is defined by the intersection of the two lines of sight of the radar and this angle increases as the baseline B_s rises.

In figure 9, the same-side stereoscopic configuration is exposed and the description of the parallax P , the base-line B_s and the intersection angle $\Delta\theta_v = \theta_{v1} - \theta_{v2}$ is given. The latter parameters have an important function as regards the quality and the accuracy of the terrain reconstruction.

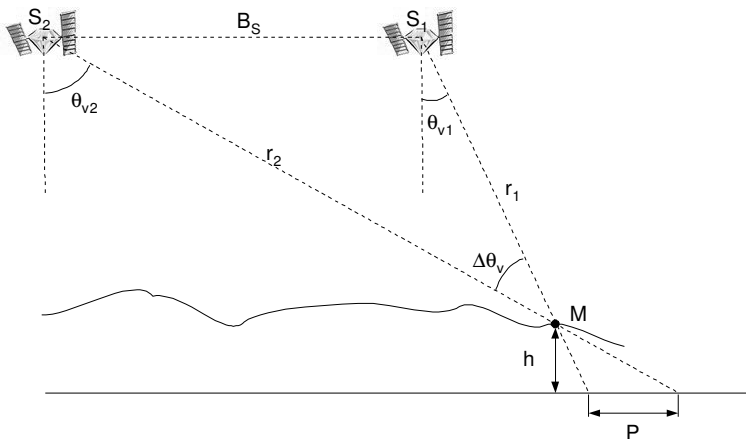


Fig. 9. One radar stereoscopic configuration.

3.2.2 Matching step

Stereoscopic techniques applied to radars are influenced by optical techniques (we can compare the baseline B_s in the radargrammetry configuration and the vertex in the human description), except that SAR images replace optical systems images. But, the main difficulty is to get used to new and unnatural radar viewing (as we exposed before) and especially when both geometric and radiometric disparities are large. However, radar images can be viewed in stereo after training. The point of radargrammetry is to match two radar images by a “registration” processing. The registration step aligns two images containing the same radar scene but viewed from different positions. The aim of the matching step is to get a dense description in order to achieve the accuracy of image registration. The main difficulty of the registering

operations comes from the dissimilarities between the pair of images that are caused by different imaging configurations. The identification of corresponding image points is the main feature of the processing. This step is generally achieved by using several methods and we will present two of them:

- grey-level image matching,
- edge-based method.

The first one is generally computed with the normalised cross-correlation coefficient (Leberl et al., 1994) and many improvements such as the use of the sum of mean normalised absolute difference or the least squares solutions are investigated. The second one is based on the fact that an object or a structure may look quite similar in both images whatever the radar position (Marr & Hildreth, 1980). However, this method needs some preprocessing (e.g. filtering operations) in order to be really efficient and the application to, for example, a mountainous area is not possible because of the small area of edges relatively to the total area of images. Thus, the combination of both methods can achieve good results (Paillou & Gelautz, 1999).

3.2.3 Disparity measurement and terrain reconstruction

For each pair of images, we get one map of disparities along both the azimuth axis and the range axis. In the case of a flat Earth, no disparity along the azimuth axis should occur when radar images come from parallel flight paths. But, because of the lack of precision of the radar trajectories, azimuth disparities exist and the way to eliminate these is to resample images into an epipolar geometry. At the end of the radargrammetric processing, the computation of a disparity map obtained under the flight conditions produces the terrain elevation which is called DEM (e.g. Digital Elevation Model). The calculated height of each pixel on the image agrees with the different equations describing the geometry of the flights of path. Moreover, in order to get a better DEM, the use of ground control points is essential to correct the geometric model of the terrain and to set up the best stereomodel as regards the solution of the stereo geometry.

3.3 Radargrammetric processing

As the radargrammetric method was briefly described in the late section, we intend to expose more precisely all the steps required to reach a terrain elevation thanks to a pair of stereo radar images.

3.3.1 Acquisition of stereo images

An important radar stereoscopic issue is the way measurements have to be made. Two main configurations can be considered: same-side (the radar is located on the same side considering the position of the two radars) and opposite-side (the scene is located between the two radars) viewing. Considering the same-side configuration (see figure 10), a large baseline (e.g. a large intersection angle) makes it possible to achieve good geometry for stereo plotting because of the increase in parallax values. And the higher the parallax value is, the more accurate the elevation reconstruction is. Conversely, the matching processing needs to manipulate images as closely identical as possible in order to succeed in stereo viewing. That implies a small intersection angle. The opposite-side configuration (figure 11) provides a large baseline and thus precise stereo plotting. Moreover, we can see in figures 11 and 10 the consequence of a range estimation error (the real point M migrates to the point M_e that is located by processing) that is less significant in the opposite-side case than the same-side one. But, the radiometric differences are so important in the case of opposite-side configuration that the matching operation

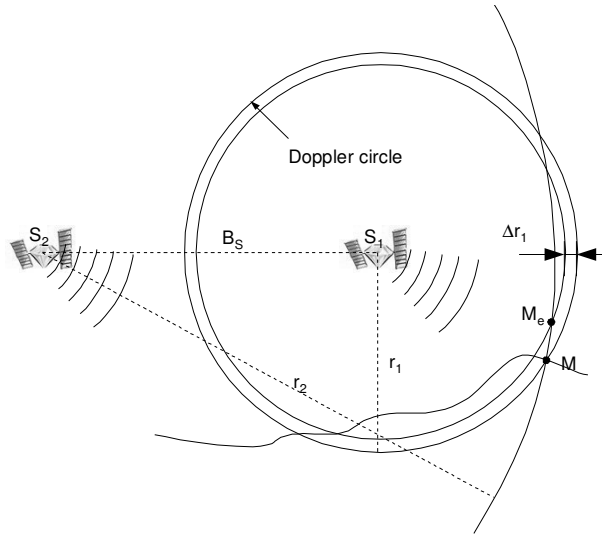


Fig. 10. Same-side configuration and range error consequence

is almost impossible without a preprocessing of images (for example, radiometric inversion). However, some studies (Toutin & Gray, 2000) demonstrate that we can have conflicting conclusions about theory developments and image applications. Anyway, the choice of the pairs of stereoscopic images comes up regarding the capability to get the parallax values and the accuracy of the height reconstruction. Thus, a compromise has to be reached between these two topics and concerns the baseline B_s to the height H of the platform ratio. This ratio can vary from 0.25 to 2. For example, a study about RADARSAT measurements (Sylvander et al., 1997) suggests an intersection angle of about 8° that corresponds to a value of B on H ratio equal about 0.3.

3.3.2 Correlation matching operation

The most common image matching method is area correlation. For a given area in the primary image, the matching computation has to detect the closest one in the secondary image by searching for the best matched area. The difference of position is the value of the parallax or disparity. The classical method of finding match areas is to use an analytical metric comparison and the zero-mean normalized cross-correlation (ZNCC) can be applied to searching for windows of radar images. These windows are usually squared and the size is $(2n+1)$ by $(2n+1)$ pixels, so a centre pixel can be defined. The ZNCC is often used because of robustness on the radiometric variations of the radar image and the result is given by the cross-correlation coefficient ρ . This coefficient ρ can be stated as follows:

$$\rho = \frac{E[I_1 I_2] - E[I_1]E[I_2]}{\sqrt{V(I_1)V(I_2)}}$$

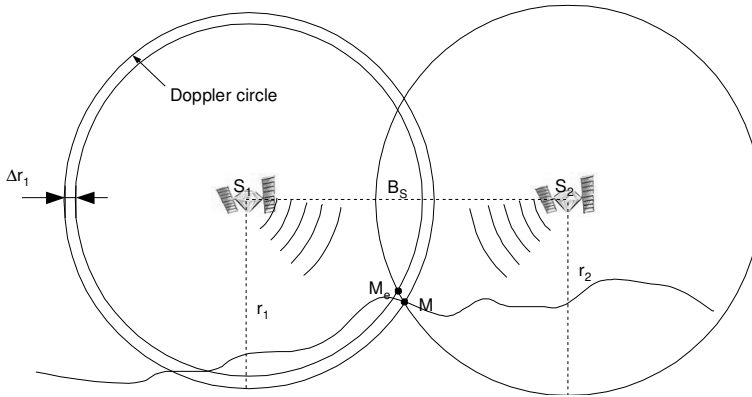


Fig. 11. Opposite-side configuration and range error consequence

where I_1 and I_2 represents the amplitude value of the pixels of the window. The mean or mathematical expectation $E[I_i]$ is calculated thanks the following expression:

$$E[I_{1,2}] = \frac{1}{N} \sum_{k=1}^N I_{1,2}^k \tag{1}$$

where N represents the number of pixels inside the window. Moreover, the variance expression $V(\cdot)$ about the window I_i is given by:

$$V(I_{1,2}) = E[(I_{1,2} - E[I_{1,2}])^2] \tag{2}$$

The value of ρ is bounded by (-1) and (+1) and the windows are considered matched for the maximum value of ρ . The coefficient ρ is calculated for each position (az_s and rg_s) of the researching window in the researching area. Also, we get a correlation surface obtained with the values of the coefficient ρ and the maximum of this surface gives the disparity $disp_{az}$ along the azimuth axis

$$disp_{az} = |az_s(\max) - az_r|$$

and the disparity $disp_{rg}$ along the range axis

$$disp_{rg} = |rg_s(\max) - rg_r|.$$

This step is carried out for each point of the primary image in order to get the disparity map. The figure 12 illustrates the correlation computation applied for one pixel inside the primary image. Considering the assumptions of radiometric distortions in a radar image, the cross-correlation computation does not work very well on such degraded images (shadowing effect for example). That is the reason why the choices of the viewing configuration and the value of B_S are very important. Especially in mountainous areas, a large part of unmatched pixels can occur because of the shortening and layover effects. Finally, the choice of the greatest value of ρ for a given correlation computation is not necessarily the optimum criterion but must be considered with other parameters. Several methods can be applied to improve the matching operation.

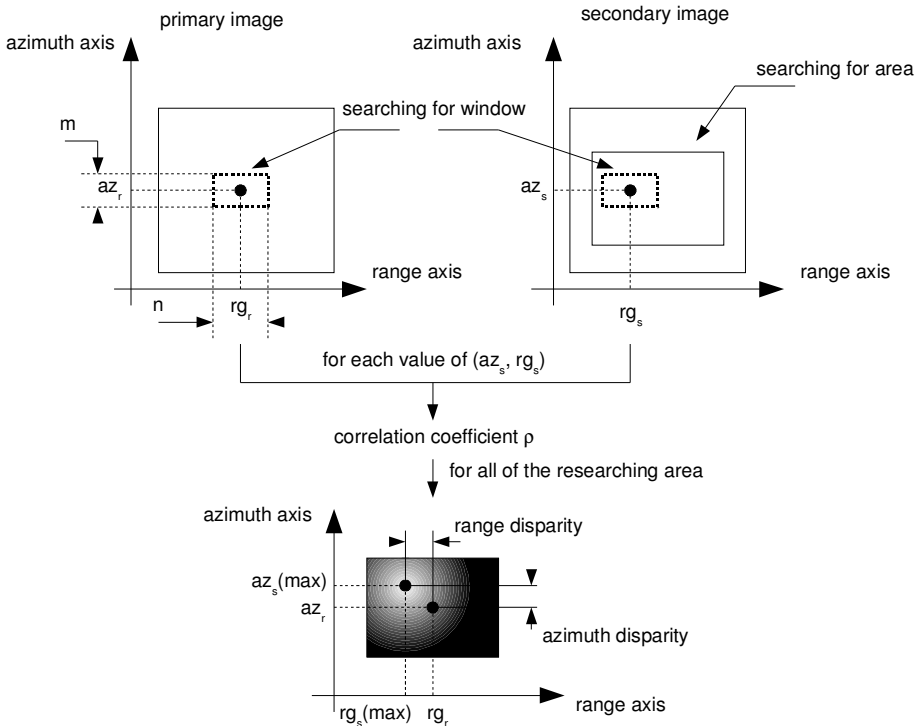


Fig. 12. Matching operations between primary image and secondary image.

3.3.3 Epipolar geometry

The use of smaller correlation windows is one way to limit the false matching result. For example, an epipolar constraint (Zhang et al., 1995) can be applied and reduce the research of the matched window along the azimuth axis. Considering parallel flight paths at a constant altitude and using the epipolar geometry, we can reduce the search area assuming that for a given point in an image, the corresponding point is located on the same azimuth line. Ideally, the search area can be reduced on a thin strip of one pixel thickness on the epipolar line. Practically, it is better to have a reduced search area one to 3 pixels wide along the azimuth axis because the estimation errors can lead to mistaken parameters. Finally, the epipolar geometry considerably reduces the size of the search area and also reduces computing time. Moreover, it limits false matching because for one pixel to match, there are fewer candidates on the other images than a larger window. The second way uses a partial knowledge of the terrain elevation that limits the research along the range axis: knowing the minimum and maximum elevation of the area, we compute the minimum and the maximum disparities along the range axis.

3.3.4 Pyramidal procedure

Another way can be considered as a hierarchical strategy used to reduce processing time and to make it possible to work with large images (Denos, 1992). The principle is quite simple:

from the original image, we build an image pyramid. At each level, the image size is reduced by a factor 2^k corresponding to the k th-iteration step. The images are reduced by transforming the pixels gray levels: in the reduced image, each pixel value corresponds to other pixels in the previous image. There are several possibilities for the transformation law: a simple one i.e the average of 4 pixels to get one pixel (see figure 13) in the reduced image or a more elaborated law i.e a Gaussian filter (Burt & Adelson, 1983) whose impulse response is given as follows:

$$w_k(u, v) = \frac{1}{2^k \sigma_I \sqrt{2\pi}} \exp\left(-\frac{u^2 + v^2}{2^{2k} \sigma_I^2}\right)$$

where σ_I is the standard deviation of the image $I(u, v)$. For each iteration, the matching pro-

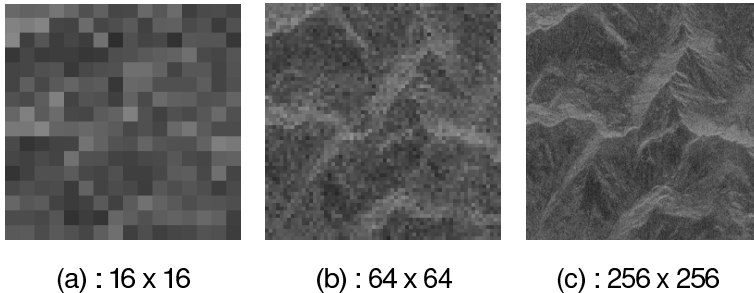


Fig. 13. Radar images with growing resolutions: from the first step (a) to the final step (c)

cess makes it possible to establish an approximate disparity map. Thus, we are able to predict the disparity offsets at the next level of the hierarchical process, reducing computation time and speckle errors. With increasing interaction, we obtain better accuracy for each level. At the final step, the last disparity map is used to produce the Digital Elevation Model. In this way, some DEM have been produced by using very large areas such as the one computed thanks to the RADARSAT-1 data about 8,000 by 8,000 pixels.

3.3.5 Speckle filtering

As previously developed, the speckle phenomenon affects the interpretation of a radar image and is undesirable for radargrammetric applications. Speckle reduction is required prior image analysis in order to improve the use of radar images. The reduction operations called speckle filtering may be very subtle because we have to get rid of the speckle effect but not of the edges and structures in the image (figure 14). Several studies (Denos, 1992) (Jacquis, 1997) prove that speckle filtering could be efficient in order to improve radargrammetric processing. But, other works about the needs to remove the speckle effect (Dowman et al., 1993) demonstrate that speckle filtering does not improve the results of radargrammetric computation. Anyway, speckle reduction can be achieved in two ways:

- multi-look processing that refers to the division of the radar beam in N_f narrow sub-beams and the result is independent as regards the speckle effect. The N_f images are summed and averaged to form the final image (Porcello et al., 1976). However, this simple method degrades the azimuth resolution by a factor of N_f ,
- filtering techniques applied to the SAR image (Frost et al., 1982) (Lee, 1981) (Kuan et al., 1985) (Wu & Maître, 1990).

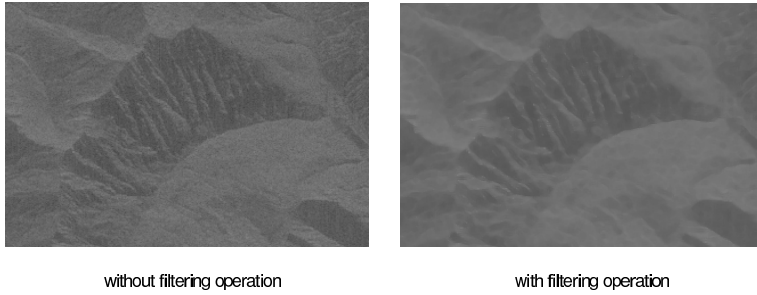


Fig. 14. Speckle filtering

The consequence of the filters on the radargrammetric performances depends on the correlation method. In our case, the computation of correlation matching based on the radar image radiometry can be improved thanks to median or Lee filters. As an overall conclusion, the filtering step is not essential to set up a radargrammetric tool kit but the application of a speckle filter to specific areas of the radar image could be beneficial in order to cancel the bad matching operations.

3.3.6 Computation of the radar stereo model

The objective of this step is to extract three-dimensional geometric data from radar stereo pairs of images by using the coordinates (position and velocity) of the satellite along the flight path. The results of such a computation is to calculate the coordinates (x, y, z) in the chosen reference as described in part 2.3.5. In the case of monocular observations, the height information h is known and we have to get the position of this point. Therefore, we can establish the system given the coordinates (x, y, z) according to the value of h of one point and the corresponding position (X_i, Y_i, Z_i) and the velocity $(\dot{X}_i, \dot{Y}_i, \dot{Z}_i)$ of the satellite indexed by $i \in 1, 2$:

$$\begin{cases} (x - X_i)^2 + (y - Y_i)^2 + (z - Z_i)^2 & = r_i^2 \\ (x - X_i)\dot{X}_i + (y - Y_i)\dot{Y}_i + (z - Z_i)\dot{Z}_i & = 0 \\ \frac{x^2 + y^2}{(a + h)^2} + \frac{z^2}{(b + h)^2} & = 1 \end{cases} \quad (3)$$

Alternatively, the binocular observations use the diversity of the vision angle to get the coordinates of the point (stereoscopic method). In the radar image, a pixel is referenced by its range and azimuth indexes. On the one hand, the range distance locates the point on a range sphere that the centre is the radar position: this is the range sphere. On the other hand, the azimuth position of a pixel can give the Doppler cone which is replaced by a plane in our case because of the null Doppler frequency at the perpendicular direction of the radar beam. The intersection of the range sphere and the Doppler plane provides two solutions but only one is obviously the right one according to the direction of the radar beam. The solution (x, y, z) of the search point satisfies the following equations system

$$\begin{cases} (x - X_1)^2 + (y - Y_1)^2 + (z - Z_1)^2 & = r_1^2 \\ (x - X_1)\dot{X}_1 + (y - Y_1)\dot{Y}_1 + (z - Z_1)\dot{Z}_1 & = 0 \\ (x - X_2)^2 + (y - Y_2)^2 + (z - Z_2)^2 & = r_2^2 \\ (x - X_2)\dot{X}_2 + (y - Y_2)\dot{Y}_2 + (z - Z_2)\dot{Z}_2 & = 0 \end{cases} \quad (4)$$

where the position $(X_{1,2}, Y_{1,2}, Z_{1,2})$ and the velocity $(\dot{X}_{1,2}, \dot{Y}_{1,2}, \dot{Z}_{1,2})$ of the radar are required to obtain a solution. Mathematically speaking, the above system is oversized because we have 3 unknowns for 4 equations. Thus, one of the 4 equations seems to be useless. However, the choice of the unused equation is not arbitrarily made but we must base our judgement on the practical measurements (see the next part 4.4.4 and especially the *Stereoscopic localisation in the geocentric reference* section)

3.3.7 Using the disparity map

In order to obtain the relief of the scene which corresponds to the height h of each pixel of the radar image (see figures 1 and 9), we can use the disparity map which has been set up for the correlation step for a pixel which is located at the value of r_g along the range axis. Generally, we can consider the baseline B_S described by the co-ordinate B_{S_r} along the range axis and B_{S_h} along the height axis. The expression of the disparity p which is also the value of parallax is given by (Leberl, 1990):

$$p = \sqrt{r_g^2 + (H - h)^2 - H^2} - \sqrt{(r_g - B_{S_r})^2 + (H + B_{S_h} - h)^2 - (H + B_{S_h})^2} - B_{S_r}$$

where the parallax p depends on the value of r_g for a given height h . Thus, the expression of h is the root of a quadratic degree equation. In the case of parallel flight paths with the same height H of the two flight paths (e.g. B_{S_h} is null or $B_{S_r} = B_S$), the expression of h can be exhibited as:

$$h = \frac{2 H B_S + 2 H p - \sqrt{4 H^2 B_S^2 + p \Delta}}{p + B_S}$$

with specifying that

$$\Delta = 8 B_S (H^2 - r_g^2 + r_g B_S) + p (4 B_S^2 + p^2 + 4 p B_S) + 4 p (H^2 - r_g^2 + r_g B_S)$$

This expression can be more simple in the case of a plane front wave, which means the height of the radar H is much greater than the height h of the point and also than the parallax p :

- considering the parallax along the ground range:

$$h = \frac{p}{\cot \theta_{v1} \pm \cot \theta_{v2}}$$

- considering the parallax along the slant range:

$$h = \frac{p}{\cos \theta_{v1} \pm \cos \theta_{v2}}$$

where the sign (-) is about the same-side configuration and the sign (+) the opposite-side one. The latter expressions are used for the SIR-C configuration and are available for altitudes less than 3,000 meters. Finally, the results of a DEM can exhibit empty or inconsistent areas because of the nature of the terrain (for example low radiometric levels). In order to improve the reconstruction of an elevation model, some operations such as interpolation could be applied to known areas (for example, to constrain flatness in the case of lakes).

4. Radargrammetric experimental results

4.1 Introduction

This part is dedicated to the application of the radargrammetric operations described in the latter parts, on raw data recorded by the shuttle Endeavour during the SIR-C mission (Evans, 2006). We obtained first results by using preprocessed radar images (Fayard et al. 2006) (Fayard et al., 2007a) (Fayard et al., 2007b). Therefore, we will present in this section the DEM of a mountainous area (French Alps) obtained through radargrammetric processing.

4.2 Description of SIR-C images

For our studies, we have several images obtained by the SIR-C mission during the month of April, 1994. The interesting area is around the French and Italian Alps. For obvious reasons, we prefer to deal with mountain areas in order to get elevation information rather than urban or lake areas. Thus, the stereoscopic pair of radar images is the PR17310 and PR17429 part of flight as described in figure 15. This part is also very interesting because we can obtain

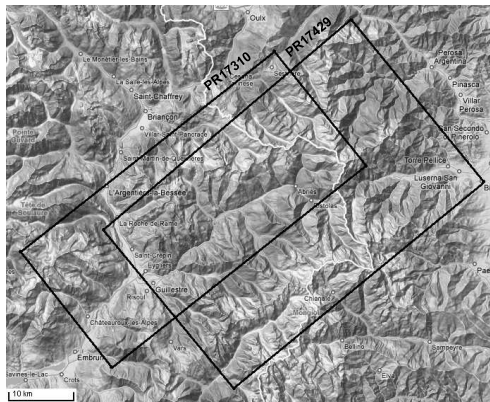


Fig. 15. Elevation map of the interesting area get from Google Maps

elevation information thanks to the IGN maps published about this region. The two flight paths are close as regards the time consideration (PR17429 on the 10th of April 1994 at 6h31 and PR17310 on the 12th of April 1994 at 5h34) so the radiometric difference due to season modifications (snow) are not present as we can see in figure 16. Moreover, the SIR-C raw data

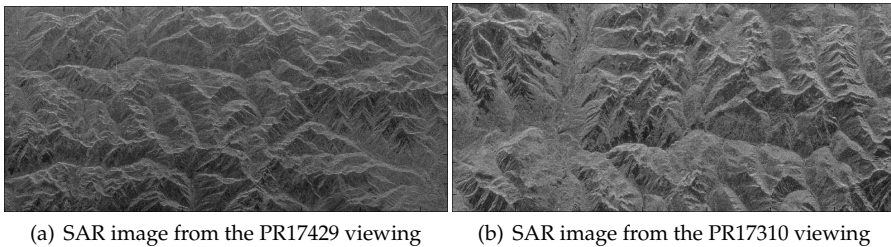


Fig. 16. SAR images of the interesting area

is also recorded with the viewing parameters which are quite important for radargrammetric processing.

4.3 Preprocessing images

As it was mentioned before (part 2.3.5 about the geometrical model of the radar position), we have to describe images taking into account the co-ordinates in order to apply matching parameters. This description requires precise information about satellite trajectory.

4.3.1 Parameters of satellite tracks

In a previous section (in the part 3.3.6 about the computation of the stereo model), we drew the reader's attention to the importance of knowing the position and the velocity of the satellite during the viewing flight in order to resolve the equations (3) and (4). We cannot use the flat Earth model or strictly parallel flight in the case of raw data. Therefore, it is possible to evaluate all the positions and velocities of the satellite along its track thanks to certain viewing parameters:

- time duration τ_i defined by the time t_{init} of the beginning and the time t_{end} of the end of the recorded data,
- data sets giving the position and the velocity of the satellite at three moments t_{DS_1} , t_{DS_2} and t_{DS_3} (these moments are 4.5 seconds apart).

Thus, the interpolation of the satellite track is possible in order to link, for each pixel of the radar image, a value of the position and the velocity of the satellite along the azimuth axis. Moreover, because this interpolation is not sufficient in order to get the absolute position of radar pixels, the geocentric co-ordinates of each corner of the radar images are used to refer images to the geocentric reference. The co-ordinates of these points, latitude and longitude, are given considering the null height:

- P_{NRET} (e.g. Near Range Early Time),
- P_{NRLT} (e.g. Near Range Last Time),
- P_{FRET} (e.g. Far Range Early Time),
- P_{FRLT} (e.g. Far Range Last Time).

The figure 17 describes the geometry of the viewing path and the corresponding parameters. Also, the definition of an absolute reference for radar images is essential to get the height of the pixels and to apply epipolar transformation on radar images.

4.3.2 Epipolar resampling

In the section (part 3.3.3 about the epipolar geometry), we moved on to the epipolar procedure that reduces the execution time for matching computation. This procedure makes it possible to limit to a thin width of az_s pixels (az_s is equal to one in theory) the search in the secondary image of the corresponding point of p_r (which is in the reference image) as can be seen in the figure 18. There are two steps to put the radar images in the epipolar geometry: forward localisation and backward localisation. For each point p_r in the reference image (e.g. #1), forward localisation is set up by using the system described by (3) and a given set of values of the height h . The result of this forward localisation is a set of points which are the solutions (x, y, z) of (3) for each value of h and a given value of r_1 . We have to note that the value of r_1 is calculated thanks to the image co-ordinates az_1 (along the azimuth axis) and rg_1 (along the range axis) of a pixel and the position of the satellite corresponding to this pixel hence

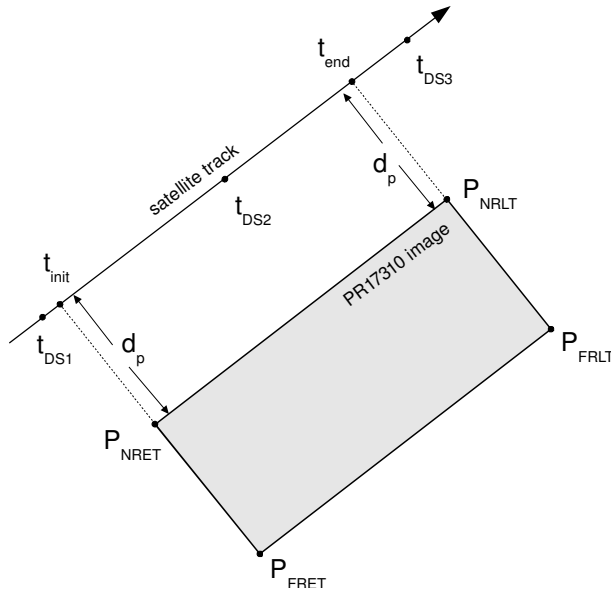


Fig. 17. Parameters of the viewing track path.

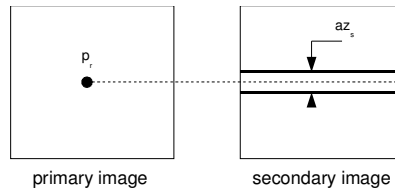


Fig. 18. Effect of epipolar geometry.

the importance of determining the satellite track parameters. Inversely, for the secondary image (e.g. #2) and from the knowledge of the height h and the co-ordinates (x, y, z) of a given point, we search for the least value of the solution r_2 of the system (3) according to the position and velocities of the satellite related to the image #2. This solution r_2 also gives the azimuth position az_2 of the corresponding pixel (because the radar beam is perpendicular to the flight path) and the calculation of the co-ordinate rg_2 is easy thanks to the value of r_2 and the radar position (X_2, Y_2, Z_2) . This step is repeated for each point in the image #1 and thus the corresponding points establish the epipolar line P_s in the image #2. To obtain the epipolar line P_r in the reference image #1 from the epipolar line P_s in the secondary image #2, we have to apply the same operations i.e. forward localisation then backward localisation except that for the forward localisation from a given point of P_s , the corresponding point is calculated for only one height h_{mean} . All these operations are summarized in figure 19. In order to illustrate the achieving epipolar lines, we propose an example of epipolar line in the working area which is shown in the figure 20 from a specific point: the peak of Agrenier. This point is located in the working area by its co-ordinates az_r and rg_r in the radar image reference. For the mentioned

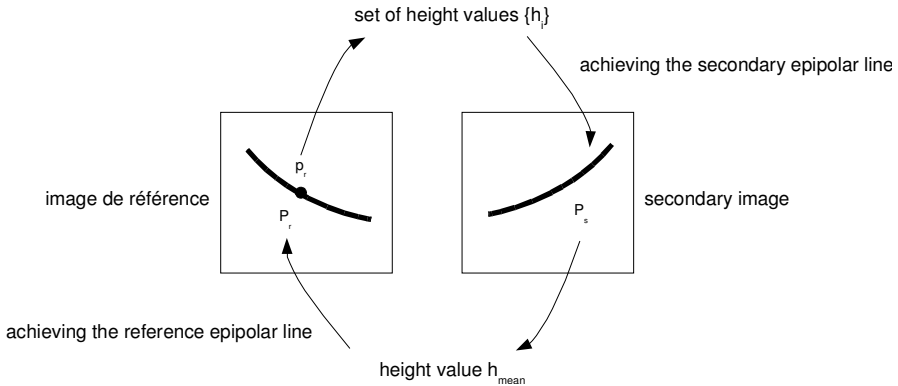


Fig. 19. Achievement the epipolar lines (reference and secondary).

Input data			Output results			
Image co-ordinates		Height	Range	Geocentric co-ordinates		
az_r (index)	rg_r (index)	h (m)	r_1 (m)	x (m)	y (m)	z (m)
1841	614	500	346,181.72	4,501,176.05	536,781.38	4,472,535.72
		1,000		4,501,729.01	537,104.25	4,472,653.93
		1,500		4,502,281.18	537,426.02	4,472,773.02
		2,000		4,502,832.56	537,746.71	4,472,892.99
		2,500		4,503,383.16	538,066.31	4,473,013.85
		3,000		4,503,932.97	538,384.84	4,473,135.57
		3,500		4,504,482.01	538,702.30	4,473,258.16
		4,000		4,505,030.27	539,018.69	4,473,381.61

Table 1. Forward localisation applied on the peak of Agrenier.

area, the IGN map gives approximately a set of heights from $h_{min} = 500$ meters to $h_{max} = 4,000$ meters. The step increment of height Δh is set to 500 meters thus we obtain 8 points for each value of h by the forward localisation. These points are described in the geocentric reference with the values (x, y, z) (see table 1) Also, for each output result described in table 1, we obtain the solutions r_2 and the corresponding points identified by image co-ordinates (see table 2). The output results describe the epipolar line in the secondary image (e.g. image #2) and this line is drawn in the working area of the PR17429 image in figure 21. Considering this figure, we notice the following:

1. the calculated corresponding point is on the epipolar line,
2. the calculated epipolar line does not pass through the actual corresponding point i.e. the peak of Agrenier.

The result is that the corresponding point is correctly found on the epipolar line and the accuracy of the localisation is not sufficient to retrieve the right corresponding point. Also, this inaccuracy must be corrected in order to set up the right disparity map.

4.3.3 Use of ground control points (GCP)

Because of the geode model inaccuracy, the quality of the terrain elevation reconstruction will be low. Also, we have to refine the stereo model parameters and some GCPs are required. In

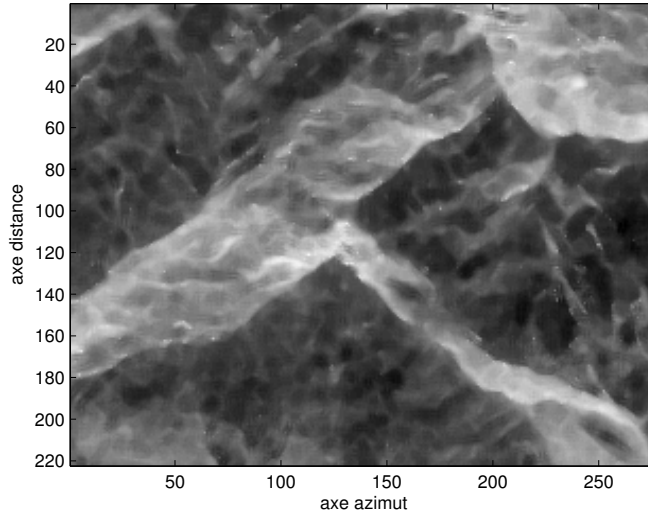


Fig. 20. Working area: PR17310 image extract

our studies, we choose 8 GCPs which cover the full terrain elevation range and are located almost at the border of the image. These GCPs are listed in table 3 just as the difference between the actual and the calculated positions of the GCPs. This comparison can be made thanks to the height information of GCPs and the forward and backward localisation operations. We note an average difference along the azimuth axis of about 6.25 pixels with a standard deviation value of 0.46 pixels and respectively 3.25 and 0.89 pixels along the range distance. So, the global correction which is applied to the secondary image reference makes it possible to recalculate the epipolar line (figure 22) that is passed through the actual corresponding point. In figure 22, we can see the search for an area about 3 pixels wide.

Input data				Output results		
Geocentric co-ordinates			Height	Range	Image co-ordinates	
x (m)	y (m)	z (m)	h (m)	r_2 (m)	az_s (index)	rg_s (index)
4,501,176.05	536,781.38	4,472,535.72	500	272,640.34	1233	239
4,501,729.01	537,104.25	4,472,653.93	1,000	272,459.73	1,232	225
4,502,281.18	537,426.02	4,472,773.02	1,500	272,279.67	1,232	212
4,502,832.56	537,746.71	4,472,892.99	2,000	272,100.16	1,231	198
4,503,383.16	538,066.31	4,473,013.85	2,500	271,921.20	1,230	185
4,503,932.97	538,384.84	4,473,135.57	3,000	271,742.78	1,229	172
4,504,482.01	538,702.30	4,473,258.16	3,500	271,564.90	1,228	158
4,505,030.27	539,018.69	4,473,381.61	4,000	271,387.57	1,228	145

Table 2. Backward localisation applied on the peak of Agrenier.

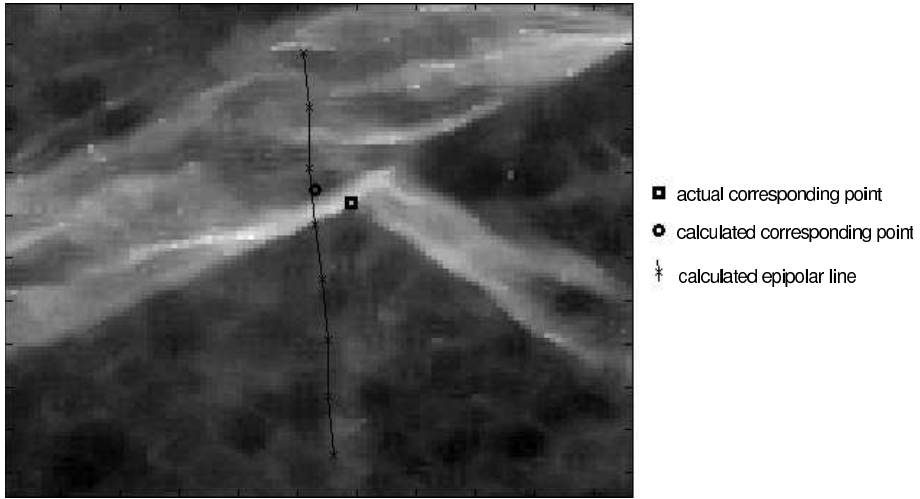


Fig. 21. Drawing the epipolar line in the PR17429 image

<i>name of the GCP</i>	<i>difference</i> Δaz_s	<i>difference</i> Δrg_s
Peak of Agrenier	6	3
Les Ourgières	6	3
Peak of Clapouse	7	4
Dent du Ratier	6	3
East of Col Garnier	7	5
Peak of Fond Queyvras	6	3
SE peak of Rochebrune	6	2
Top of Assan	6	3

Table 3. Difference of the co-ordinates of actual GCPs and their calculated corresponding points.

4.4 Radargrammetric processing

At this step of the entire processing, we obtain preprocessed images to which the specific radargrammetric processing will be applied: matching processing, disparity map and terrain elevation.

4.4.1 Confidence in correlation coefficient

After computing the matching operation which is described in part 3.3.2 (see the section *Correlation matching operation*), we obtain the disparity map. However, the values of disparity should be considered according to the confidence in correlation coefficient. The highest value inside a correlation surface can be perfectly detected and the corresponding position is obvious: this corresponds to a high confidence of correlation. But, this maximum position cannot clearly be obtained so the confidence correlation is considered as low (see figure 23). For this case, additional noise can modify the results of the disparity map and so applying the speckle reduction and pyramidal procedure should strengthen the correlation results.

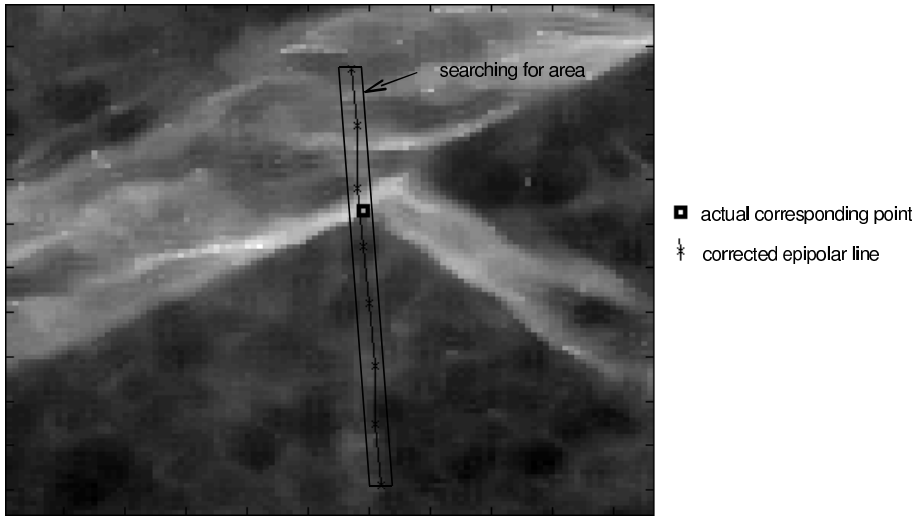


Fig. 22. Drawing the corrected epipolar line in the PR17429

4.4.2 Speckle filtering

In our application, we use two methods to reduce speckle effect. The first one is the multi-look technique which has been described before (see part 3.3.5 about the speckle filtering) and the value of N_f is equal to 4 in order not to degrade the azimuth resolution regarding the value of SIR-C parameters. Moreover, a Lee filter is applied to the radar images so the edges are preserved, which could be important considering the mountainous area. Several tests are done and the best results are obtained by using a 5 by 5 pixel window (that seems to be correct as regards the heterogeneous area). Although the speckle reduction improves the quality of the terrain reconstruction, it is not sufficient for certain areas.

4.4.3 Pyramidal computation

This method has been developed in the above section 3.3.4 and the results of this procedure will now be exposed. Firstly, we obtain the disparity map of our working area without the pyramidal steps within 50 minutes of computation using a 1.8 GHz workstation with 1GB of RAM. The resulting disparity map is described in figure 24. After that, we apply the pyramidal approach to the radar images and the resulting disparity map is obtained within 24 minutes of computation using the same workstation as before. This first consequence speaks in favour of the pyramidal scheme. Moreover, the quality of disparity map described in figure 25 is obviously better than the one in figure 24. Also, we can note two advantages of applying the pyramidal steps: computation time reduction and disparity map quality improvement.

4.4.4 Stereoscopic localisation in the geocentric reference

Thanks to the disparity map, we can reconstruct the terrain elevation by resolving the system (4). We remember this system is oversized because of 3 unknowns described by 4 equations. So, we have to choose the equation to be removed by studying the sensitivity of induced errors. This sensitivity corresponds to a correlation success when errors of about plus or minus 10 pixels are applied to the actual location of corresponding points along the azimuth axis or

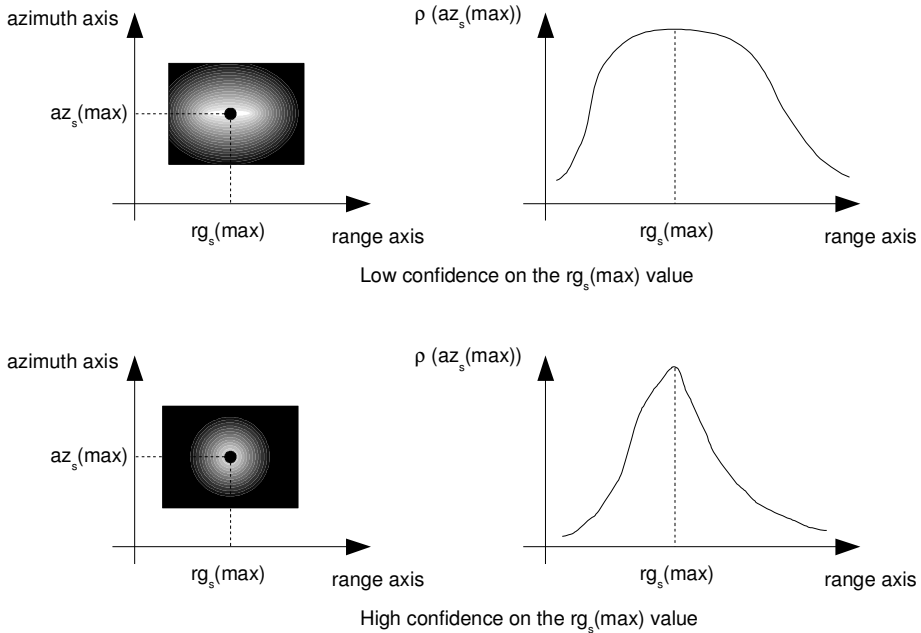


Fig. 23. Different values of confidence in correlation coefficients

	configuration #1		configuration #2	
	azimuth(m)	range(m)	azimuth(m)	range(m)
longitude	340	0.14	13.5	19.4
latitude	438	0.13	9.8	23.5
height	404	0.8	16	37

Table 4. RMS errors (in meters) resulting from a one pixel error in the disparity map along the azimuth axis or the range axis and considering the two configurations of the binocular system.

the range axis. Thus, by resolving 3 of the 4 equations of 4, we obtain the co-ordinates (x, y, z) which are described in the geocentric reference as latitude ϕ , longitude λ and height h and compared with the actual terrain model. The resulting error is calculated as a root-mean square operation applying to all the pixels of the working area. The results for a location error of one pixel are summarized in table 4. Two configurations of an undersized system are studied: the first one (configuration #1) uses the two iso-Doppler equations and one iso-range equation and the second one (configuration #2) uses two iso-range equations and one iso-Doppler equation. The result is obvious: it is better to chose the second configuration because an error of one pixel along the azimuth axis induces an error of less than one meter regarding the height reconstruction although the sensitivity along the range axis seems to be less in the first configuration. Another conclusion from this study is that the minimum of the correlation surface does not occur at a null shift along the azimuth and the range axis. That means this shift induce errors in the localisation and in the height reconstruction. These errors are calculated thanks to the GCPs which are used for the correction of the image indexes (see

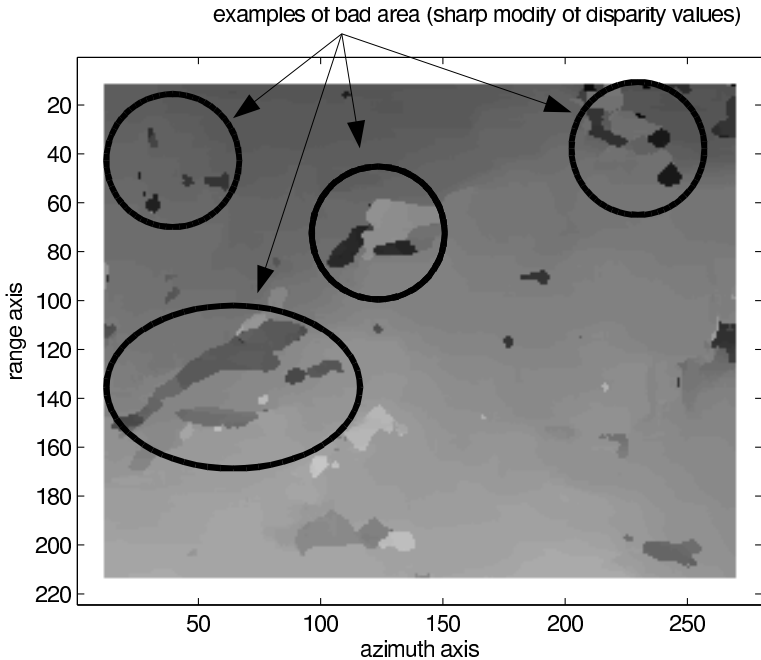


Fig. 24. Disparity map without the pyramidal procedure

section 4.3.3 about the use of ground control points). The conclusion is that the errors are less than the resolution values both for the localisation (latitude and longitude) reconstruction and for the height reconstruction.

4.4.5 Post processed DEMs

Thanks to the transformation applied at this step, we can reconstruct the terrain elevation of the working area which is seen in figure 20 by resolving the system described through configuration #2. In order to quantify the accuracy of our elevation reconstruction, we compare it with the SRTM (Shuttle Radar Topography Mission) DEM (see figure 26). We need to apply a resampling operation to our DEM because its resolution is higher than that of the SRTM . In this way, the DEM we obtain (which we can called the raw DEM) and the comparison with the SRTM DEM are shown in figure 27. The first results of the comparison are described in table 5 and show that an error of height reconstruction of less than 50 meters occurs for only 46.4 percent of pixels. Moreover, only 80 percent of pixels exhibit an error less than 200 meters. These results mean that post processing must be applied to the raw DEM. This post processing consists in removing the obvious errors which are detected by a comparison between neighbouring areas. The choice of the worked area is done thanks to an eye examination and the connected disparity is not computed to obtain the DEM. Also, the calculated DEM is not complete but more accurate than the raw one and the corresponding errors are shifted to a blank pixel (see figure 28). After removing these bad disparities, we can compare this corrected and post processed DEM with that of the STRM and the results of the comparison are described in table 5. The examination of the results shows us that more than 98 percent of pixels present

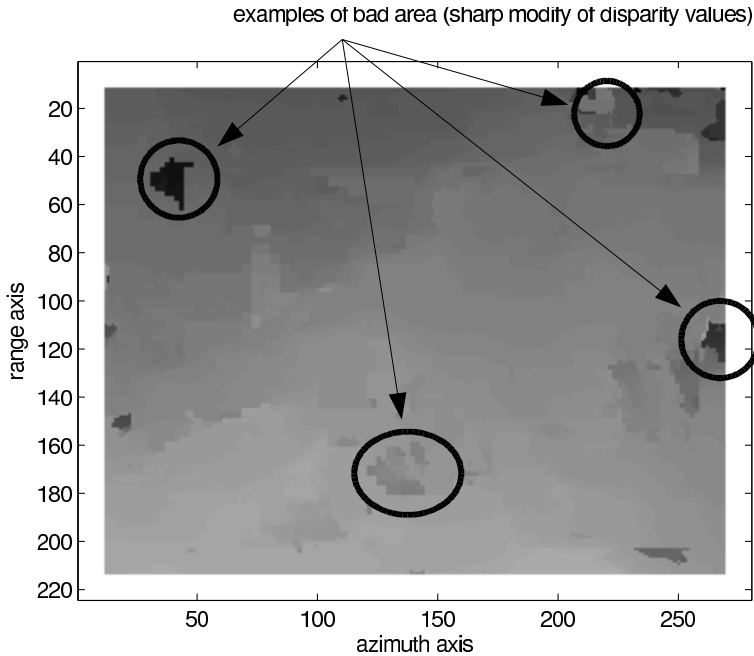


Fig. 25. Disparity map with the pyramidal procedure by using three levels of resolution

nature of DEM	number of considered points	consideration of height errors							
		< 20 m		< 50 m		< 100 m		< 200 m	
		%	ϵ_{moy}	%	ϵ_{moy}	%	ϵ_{moy}	%	ϵ_{moy}
raw DEM	2938	21.9	9.8	46.4	22.9	65.9	37.4	80.0	55.2
corrected DEM	2126	29.5	9.8	61.6	22.7	85.5	36.3	98.7	49.6

Table 5. Percent of errors and average errors ϵ_{moy} of the calculated DEMs.

an error of less than 200 meters (in comparison with the 80 percent without post computation) and the pixels whose height error is less than 50 meters are more than 61 percent (46.4 percent before). Considering the relief type and the resolution values, these results are close to the results obtained by other satellites (Toutin, 2000) (Toutin & Gray, 2000).

5. Conclusion and further developments

This chapter has dealt with the relevance of using stereoscopic radar images in order to retrieve the relief of terrain. Firstly, the basic characteristics of the radar image (SAR image) were described and the parameters which were different from those of an optical image were pointed out especially the image resolution and set up in the slant plane. Other characteristics such as the geometric and radiometric distortions were described in the rest of the section. These distortions have to be taken into account in radar stereoscopic applications in order to determine the better viewing parameters and avoid the consequences of specific radar image geometry (for example, foreshortening) and radiometry (for example, speckle effect). In the second part, we presented the radargrammetric method applied to radar images and how

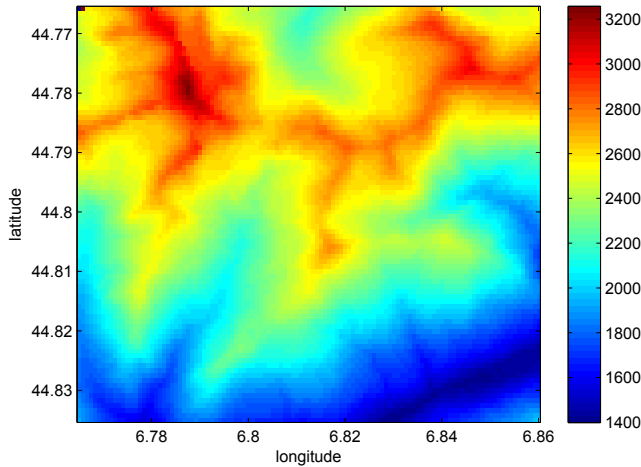
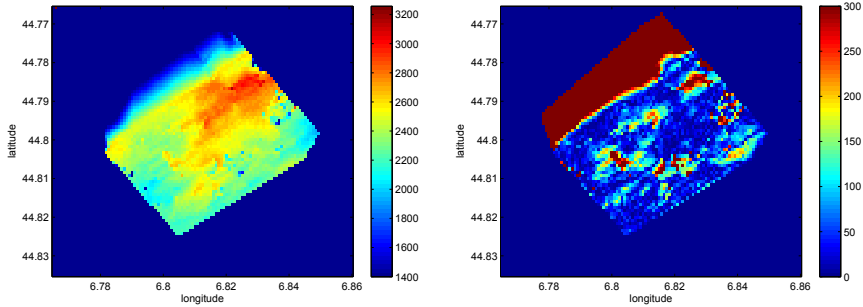


Fig. 26. SRTM DEM of the working area

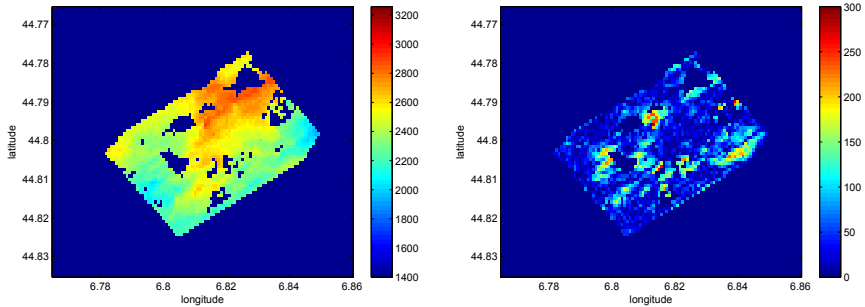
they can be compared to the optical stereoscopic method. The aim of radargrammetry is to extract the height information of a radar scene from a stereo pair of radar images. Compared with the optical method, radargrammetry is based on the geometry of the visualisation flight path over the scene and the parallax induced by two views of a point characterized by its elevation. This parallax is also called the disparity between a primary image and a secondary image. The disparity is defined for each pixel in the radar image and is determined by matching computation in order to set up a disparity map. This disparity map of all the radar scenes is essential to reconstruct the height elevation by resolving a stereo-model which is described by range sphere and Doppler circle equations for each position of the radar. The accuracy of the terrain reconstruction depends on the quality of the disparity map and also on the success of the matching operation. This operation can be improved by several processing steps and especially the reduction of the speckle effect and the pyramidal approach. We can note that the geometry of the viewing scene also influences the achievement of the 3D co-ordinates of the terrain. At the end of the discussion, we illustrated radargrammetric processing by using SIR-C data over the French Alps. We showed all the steps required to obtain an acceptable DEM: from the registration of each pixel of the radar image regarding the satellite path (position and velocity) to post processing the DEM by removing the obvious bad reconstruction to choosing the better stereo-model and to using GCPs in order to refine the radar images. The resulting DEM of our radargrammetric processing is almost identical to the DEM which can be obtained thanks to specific matching and filtering operations. One of the advantages of our method is the simplicity with which an acceptable DEM is obtained.

However, it is possible to apply new methods to further improve the crucial matching step and this is what we will be working on next. We will investigate the improvement of the radargrammetric tool kit along two axes. The first one deals with the opportunities to apply some optical methods during the correlation step. Especially, the work will deal with stereo matching algorithm with an adaptive window in an SAR context. Depending on the statistical behaviour of the radar signal, we can manage the size of the correlation window in order



(a) Raw DEM described in the geocentric reference (b) Difference between the raw DEM and the SRTM DEM

Fig. 27. Quantification of the raw DEM



(a) Corrected DEM (b) Difference between the corrected DEM and the SRTM DEM

Fig. 28. Quantification of the corrected DEM.

to improve the confidence of the correlation during the matching computation. The second method concerns the registration of the different areas of the image considering polarimetric parameters. Because certain areas inside an SAR image are not cooperative to the matching cooperation (e.g. shadowed or foreshortened areas), these kinds of areas could be matched together regarding the polarimetric parameters of the areas.

6. References

- Beckman, P. and Spizzichino, A. (1987), *The scattering of electromagnetic waves from rough surfaces*, Artech House, 1987.
- Burt, P. and Adelson, E. (1983), The laplacian pyramid as a compact image code, *IEEE Transactions on Communications*, Vol. COM-31, No. 4, pages 532–540, 1983.
- Carrara, W.G., Goodman, R.S. and Majewski, R.M. (1995), *Spotlight Synthetic Aperture Radar*, Norwood, MA: Artech House, 1995.

- Curlander, J.C. (1991). *Synthetic Aperture Radar, Systems and Signal Processing*. J.A. Kong, Wiley, 1991.
- Denos, M. (1992), A pyramidal scheme for stereo matching SIR-B imagery, *International Journal of Remote Sensing*, Vol. 13, No. 2, pages 387 - 392, 1992.
- Dhond, U.R. and Aggarwal J.K. (1989). Structure from stereo—a review. *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 19, No. 6, pages 1489–1510, November 1989.
- Dowman, I., Pu-Huai, C., Clochez, O. and Saundercock, G. (1993), Heighting from stereoscopic ERS-1 data, in *proceedings Second ERS-1 Symposium*, pages 609–614, 1993.
- Dufour, J.P. (2001). *Introduction to geodesy*. Hermès, 2001.
- Evans, D.L. (2006), Spaceborne imaging radar-C/X-band synthetic aperture radar (SIR-C/X-SAR): a look back on the tenth anniversary, *IEE Proceedings on Radar, Sonar and Navigation*, Vol. 153, No. 2, pages 81–85, 2006.
- Fayard, F., Méric, S. and Pottier, E. (2006), First studies on a radargrammetric tool kit, *European conference on Synthetic Aperture Radar, EUSAR'06, Dresden*, 2006.
- Fayard, F., Méric, S. and Pottier, E. (2007a), Matching stereoscopic SAR images for radargrammetric applications, *IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2007*, pages 4364 - 4367, 2007.
- Fayard, F., Méric, S. and Pottier, E. (2007b), Mise en appariement d'images SAR stéréoscopiques, *Journées Nationales des Micro-ondes, JNM'07, Toulouse*, 2007
- Frost, V.S., Stiles, J., Shanmugan, K. and Holtzman, J. (1982). A model for radar images and its application to adaptive digital filtering of multiplicative noise, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-4, No. 2, pages 157 - 166, March 1982.
- Girard, M.C. (2003). *Processing of remote sensing data*, Taylor and Francis, France.
- Goodman, J.W. (1976). Some fundamental properties of speckle. *Journal of the optical society of America*, Vol. 66, No. 11, pages 1145–1150, November 1976.
- Gracie, G. et al, (1970). Stereo Radar Analysis, *US Engineer Topographic Laboratory, Report No FTR-1339-1*.
- Horn, B. (1975), *The psychology of computer vision—Chap. 4: obtaining shape from shading information*, Mac-Graw Hill, 1975.
- Jacquis, F. (1997), *Techniques de corrélation pour la radargrammétrie — Filtrage et détection de structures — Application à des images satellites ROS-ERS1*, PhD dissertation, Joseph Fourier university, Grenoble, 1997.
- Kuan, D.T., Sawchuk, A.A., Strand, T.C. and Chavel, T. (1987), Adaptive restoration of images with speckle, *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-35, No. 3, pages 373–383, 1987.
- Leberl, F. (1990). *Radargrammetric image processing*, Artech House, Norwood, MA.
- Lee, J.S. (1981), Refined filtering of image noise using local statistics, *Computer Graphics and Image Processing*, No. 15, pages 380 - 389, 1981.
- La Prade, G. (1963). An analytical and experimental study of stereo for radar. *Photogrammetric Engineering*, Vol. 29, No. 2, pages 294-300.
- Marr, D., Poggio, T. (1980), A theory of edge detection, *Proc. Royal Society of London*, pages 283–287.
- Massonnet, D. and Rabaute, T. (1993), Radar interferometry: limits and potential, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 31, No. 2, pages 455–464, 1993.

- Paillou, Ph. and Gelautz, M. (1999), Relief reconstruction from SAR stereo pairs: the “optimal gradient” matching method, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37, No. 4, pages 2099–2107.
- Polidori, L. (1997). *Cartographie radar*, Gordon and Breach Science, France.
- Porcello, L., Massey, N., Innes, R. and Marks J. (1976). Speckle reduction in synthetic aperture radar, *Journal of the Optical Society of America*, Vol. 66, No. 11, pages 1305–1311, 1976.
- Rosenfield, G.H., Stereo radar techniques, *Photogrammetric Engineering*, Vol. 34, pages 586–594.
- Schreier, G (1993). *SAR Geocoding: data and systems*, (18 papers) Wichmann Verlag, Karlsruhe, Germany.
- Schuler, D.L., Lee, J.S. And De Grandi, G. (1996), Measurement of topography using polarimetric SAR images, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 34, No. 5, pages 1266–1277, 1996.
- Sylvander, S., Cousson, D. and Gigord, P. (1997), Etude des performances géométriques des images RADARSAT, *Bulletin de la société Française de Photogrammétrie et de Télédétection* 148, pages 57–65.
- Toutin, T. (2000), Evaluation of radargrammetric DEM from RADARSAT images in high relief areas, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 38, No. 2, pages 782–789, 2000.
- Toutin, T. and Gray, L. (2000), State-of-the-art of elevation extraction from satellite SAR data, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 55, pages 13–33, 2000.
- Toutin, T. (2006). Generation of DSMs from SPOT-5 in-track HRS and across-track HRG stereo data using spatiotriangulation and autocalibration. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 60, No. 3, May 2006, pages 170–181.
- Ulaby, F.T. (1981), *Microwave remote sensing, active and passive*, Artech House, 1981.
- Wu, Y. and Maître, H. (1990), A speckle suppression method for SAR images using maximum homogeneous region filtering, in *IGARSS'90*, Vol. 3, pages 2413–2416, 1990.
- Zhang, Z., Deriche, R., Faugeras, O. and Luong, Q.T., A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry, *Artificial intelligence*, Vol. 78, pages 87–119, 1995.

MAP Classification of a Reference Image Using Auxiliaries Images with Different Prevalent Classes

^{1,2}Orlando Alves Máximo and ²David Fernandes

¹*Instituto de Estudos Avançados (IEAv)-Comando- Geral de Tecnologia Aeroespacial*

²*Instituto Tecnológico de Aeronáutica (ITA)-Comando -Geral de Tecnologia Aeroespacial
Brazil*

1. Introduction

The observation of a scene by an imaging sensor produces an image that is a function of the sensor characteristics and the sensor ability to interact with the targets in the scene. In this process the true classes in the scene can be merged or become very close in the image generated by the sensor in such way that each image, one for each sensor, can present different number of prevalent classes. As an example, Figure 1 shows a six classes scene that was observed by two sensors in two different ways. The first sensor generates an image that has three prevalent classes (green, blue and orange) and the second sensor generates an image with four prevalent classes (green, black, purple and red).

There are several mathematical approaches that can be used to classify the scene using one or more images of the scene. These approaches include Support Vector Machine (Bruzzone et al., 2006; Camps-Valls et al., 2007), Artificial Intelligence (Liu et al., 2008), Decision Trees (Pal & Mather, 2003), New Nearest Neighbor Approaches (Zhu & Basir, 2005; Samaniego et al., 2008) and the most used statistical approach (Valet et al., 2001). In the context of the statistical classification, the Bayesian approach is widely used for the classification error mitigation (Fukunaga, 1990). In the statistical classification method based in the Maximum A Posteriori (MAP) the image set to be classified must contain all images with the same number and type of classes (Schowengerdt, 1997).

In this Chapter is presented an extension for the classical MAP classification for the case in which each image in a set of images can have different numbers and types of prevalent classes. In this extension, one image is chosen as the reference image to be classified according with its dominant classes and the others images are used as additional information (Máximo & Fernandes, 2008).

In the example shown in Figure 1 if the image 1 is selected as the reference image with three prevalent classes and image 2 as the complementary information, the perfect classification is given in Figure 2a. If the image 2 is selected as the reference, the ideal classification is given in the Figure 2b.

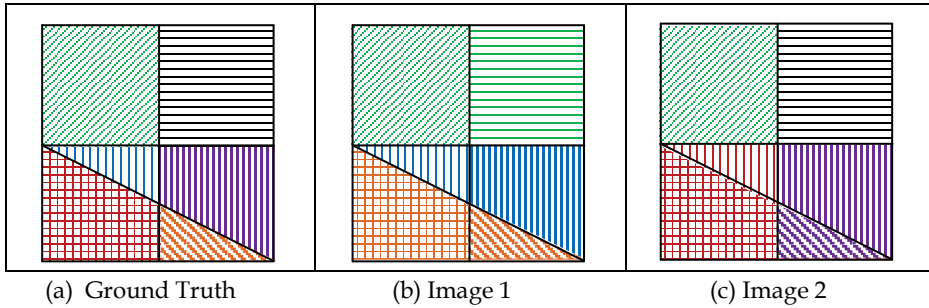


Fig. 1. Classes definition: (a) Ground truth with six classes, (b) Three dominant classes (green, blue and orange) in the image generated by the first sensor and (c) Four dominant classes (green, black, purple and red) in the image generated by the second sensor

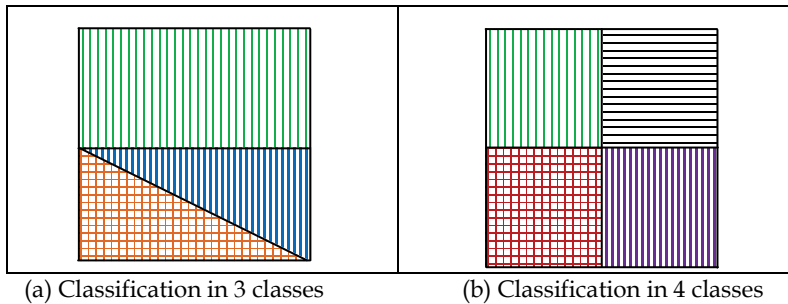


Fig. 2. The ideal classification

In section 2 is presented the MAP classification and the proposed extension that consider one image in a set of images as the reference and that each image may have different numbers and types of prevalent classes. Section 3 shows the application of the proposed extended MAP classification method in comparison with the classical MAP classification by using simulated Single-Look Synthetic Aperture Radar (SLC-SAR) images (Oliver & Quegan, 1998). The classification performance comparison is due by the Kappa coefficient estimation (Rosenfield & Fitzpatrick-Lins, 1986).

2. The classification decision rule

2.1 The MAP decision rule

Given a set of N images where each image has M classes represented by $w_m, m = 1, 2, \dots, M$, the Bayes Risk is defined as (Sharf, 1991; Swain, 1978):

$$L_{\bar{x}}(w_m) = \sum_{k=1}^M \lambda_{km} p(w_k | \bar{x}) \tag{1}$$

where $\bar{x} = (x^{(1)}, x^{(2)}, \dots, x^{(N)})$ is the observation set associated to the random variables (r.v.) $X = (X^{(1)}, X^{(2)}, \dots, X^{(N)})$, with $X^{(n)}$ a r.v. representing a pixel in the n -th image, λ_{km} is the

classification loss function related to the observation \tilde{x} be classified in the class w_k when in the reality it belongs to the class w_m and $P(w_k | \tilde{x})$ is the w_k class *a posteriori* probability (conditional probability after the observation). Choosing

$$\lambda_{km} = \begin{cases} 0 & \text{for } k = m \\ 1 & \text{for } k \neq m \end{cases} \quad (2)$$

the Bayes Risk becomes:

$$L_{\tilde{x}}(w_m) = 1 - p(w_m | \tilde{x}) \quad (3)$$

and the minimum Risk is reached when the *a posteriori* conditional probability is maximum. Then we can state a Maximum A Posteriori (MAP) decision rule as:

$$\tilde{x} \mapsto w_m \Leftrightarrow F_m^o(\tilde{x}) = F_k^o(\tilde{x}) \quad (4)$$

where $\tilde{x} \mapsto w_m$ represent the association of the observation \tilde{x} with the class w_m and $F_m^o(\tilde{x})$ is the Global Membership Function (GMF) (Lee et al, 1987) given by:

$$F_m^o(\tilde{x}) = P(w_m | \tilde{x}) = \frac{P(w_m)p_X(\tilde{x} | w_m)}{p_X(\tilde{x})} \quad (5)$$

where $p_X(\tilde{x} | w_m)$ is the probability density of X given that the w_m class is true (conditional probability). Since $p_X(\tilde{x})$ is common to all GMF with $m = 1, 2, \dots, M$, it can be neglected in the decision processes. Therefore the GMF can be simplified as

$$F_m^{os}(\tilde{x}) = P(w_m)p_X(\tilde{x} | w_m) \quad (6)$$

If we consider independent observations $p_X(\tilde{x} | w_m) = \prod_{n=1}^N p_X(x^{(n)} | w_m)$ then

$$F_m^{os}(\tilde{x}) = P(w_m) \prod_{n=1}^N p_X(x^{(n)} | w_m) \quad (7)$$

Equation (7) can be used in the decision rule as stated in (4) changing $F_m^o(\tilde{x})$ to $F_m^{os}(\tilde{x})$:

$$\tilde{x} \mapsto w_m \Leftrightarrow F_m^{os}(\tilde{x}) = F_k^{os}(\tilde{x}) \quad (8)$$

2.2 The MAP decision rule for a reference image

It will be considered the general case in which every n -th image, $n = 1, 2, \dots, N$, has M_n classes represented by $w_{m_n}^{(n)}$, $m_n = 1, 2, \dots, M_n$, and also that the first observation $x^{(1)}$ is the reference image, in the sense that we want to classify only the first image in its prevalent

classes and having the others images in the set as auxiliary or complementary information. In this case the Bayes Risk can be written as

$$L_x(w_{m_1}^{(1)}) = \sum_{m_a=1}^{M_a} \lambda_{m_a m_1} p(w_{m_a}^{(1)} | \bar{x}) = 1 - p(w_k^{(1)} | \bar{x}) \quad (9)$$

and the decision rule by

$$x^{(1)} \mapsto w_{m_1}^{(1)} \Leftrightarrow F_{m_1}(\bar{x}) = F_{m_a}(\bar{x}) \quad (10)$$

where $F_{m_1}(\bar{x})$ is given by

$$F_{m_1}(\bar{x}) = P(w_{m_1}^{(1)} | \bar{x}) = \frac{p_X(\bar{x}, w_{m_1}^{(1)})}{p_X(\bar{x})} \quad (11)$$

Again $p_X(\bar{x})$ is independent of the class and can be neglected and the GMF $F_{m_1}(\bar{x})$ can be written as

$$\begin{aligned} F_{m_1}^s(\bar{x}) &= p_X(\bar{x}, w_{m_1}^{(1)}) = \sum_{m_2=1}^{M_2} \sum_{m_3=1}^{M_3} \dots \sum_{m_N=1}^{M_N} p_X(\bar{x}, w_{m_1}^{(1)}, w_{m_2}^{(2)}, \dots, w_{m_N}^{(N)}) \\ &= \sum_{m_2=1}^{M_2} \sum_{m_3=1}^{M_3} \dots \sum_{m_N=1}^{M_N} p_X(\bar{x}, W_1^N) \\ &= \sum_{m_2=1}^{M_2} \sum_{m_3=1}^{M_3} \dots \sum_{m_N=1}^{M_N} p_X(\bar{x} | W_1^N) P(W_1^N) \\ &\stackrel{\text{Chain rule}}{=} P(w_{m_1}^{(1)} | \sum_{m_2=1}^{M_2} \sum_{m_3=1}^{M_3} \dots \sum_{m_N=1}^{M_N} p_X(\bar{x} | W_1^N) \prod_{n=2}^N P(w_{m_n}^{(n)} | W_1^{n-1})) \\ &\stackrel{\text{Independent observations}}{=} P(w_{m_1}^{(1)}) \sum_{m_2=1}^{M_2} \sum_{m_3=1}^{M_3} \dots \sum_{m_N=1}^{M_N} p_X(x^{(1)} | W_1^N) \prod_{n=2}^N p_X(x^{(n)} | W_1^N) P(w_{m_n}^{(n)} | W_1^{n-1}) \end{aligned} \quad (12)$$

where:

$$W_1^n = \{w_{m_1}^{(1)}, w_{m_2}^{(2)}, \dots, w_{m_n}^{(n)}\} \quad (13)$$

Equation (13) can be used in the decision rule as stated in (10) changing $F_{m_1}(\bar{x})$ to $F_{m_1}^s(\bar{x})$:

$$x^{(1)} \mapsto w_{m_1}^{(1)} \Leftrightarrow F_{m_1}^s(\bar{x}) = F_{m_a}^s(\bar{x}) \quad (14)$$

In the particular case where all images have the same classes $M_n=M$ and $w_m^{(n)} = w_m$ we have that $F_{m_a}^s(\bar{x}) = F_{m_a}^{os}(\bar{x})$.

We can define now the decision space $S_{m_1} \subset R$ in such way that

$$x^{(1)} \mapsto \omega_{m_1}^{(1)} \Rightarrow x^{(1)} \in S_{m_1} \quad (15)$$

2.3 The MAP decision rule for a set of classes

Let $\Psi_1^N(r_1, r_2, \dots, r_N)$ be a set of any N classes, one class in each image, represented by

$$\Psi_1^N(r_1, r_2, \dots, r_N) = \{w_{m_1=r_1}^{(1)}, w_{m_2=r_2}^{(2)}, \dots, w_{m_N=r_N}^{(N)}\}, \quad r_n \in \{m_1, m_2, \dots, m_{M_n}\}, \quad n = 1, 2, \dots, N \quad (16)$$

The classification of a pixel \tilde{x} in this special classes group can be made by the decision rule

$$\tilde{x} \mapsto \Psi_1^N(r_1, r_2, \dots, r_N) \Leftrightarrow F_{r_1, r_2, \dots, r_N}(\tilde{x}) = \max_{k_1, k_2, \dots, k_N} F_{k_1, k_2, \dots, k_N}(\tilde{x}) \quad (17)$$

where the GMF function $F_{r_1, r_2, \dots, r_N}(\tilde{x})$ is given by:

$$\begin{aligned} F_{r_1, \dots, r_N}(\tilde{x}) &= p_X(\tilde{x}, \Psi_1^N(r_1, r_2, \dots, r_N)) = P(\Psi_1^N(r_1, r_2, \dots, r_N)) p_X(\tilde{x} | \Psi_1^N(r_1, r_2, \dots, r_N)) \\ &= P(w_{r_1}^{(1)}) p_X(\tilde{x} | \Psi_1^N(r_1, r_2, \dots, r_N)) \prod_{n=2}^N p(w_{r_n}^{(n)} | \Psi_1^{n-1}(r_1, r_2, \dots, r_{n-1})) \\ &\stackrel{\text{independent observations}}{=} p(w_{r_1}^{(1)}) p(x^{(1)} | \Psi_1^N(r_1, r_2, \dots, r_N)) \prod_{n=2}^N p(x^{(n)} | \Psi_1^N(r_1, r_2, \dots, r_N)) p(w_{r_n}^{(n)} | \Psi_1^{n-1}(r_1, r_2, \dots, r_{n-1})) \end{aligned} \quad (18)$$

We can again define the decision space $S_{r_1, r_2, \dots, r_N} \subset R^N$ in such way that

$$\tilde{x} \mapsto \Psi_1^N(r_1, r_2, \dots, r_N) \Rightarrow \tilde{x} \in S_{r_1, r_2, \dots, r_N} \quad (19)$$

The equation (18) can be introduced in (12) in such way that for independent images

$$F_{m_1}^s(\tilde{x}) = \sum_{m_2=1}^{M_2} \sum_{m_3=1}^{M_3} \dots \sum_{m_N=1}^{M_N} F_{m_1, \dots, m_N}(\tilde{x}) \quad (20)$$

The space $S_{m_1} \subset R$ in (15) and the space $S_{r_1, r_2, \dots, r_N} \subset R^N$ in (19) are related by

$$S_{m_1} = \bigcap_{r_2=0}^{M_2} \bigcap_{r_3=0}^{M_3} \dots \bigcap_{r_N=0}^{M_N} S_{m_1, r_2, \dots, r_N} \quad (19)$$

The decision rule (10) and (19) can also be related by

$$x^{(1)} \mapsto \omega_{m_1}^{(1)} \Leftrightarrow \tilde{x} \mapsto \Psi_1^N(m_1, r_2, \dots, r_N), \quad r_n \in \{m_2, \dots, m_{M_n}\}, \quad n = 2, \dots, N \quad (20)$$

An equivalent form of the decision rule (20) can be state as

$$x^{(1)} \in S_{m_1} \Leftrightarrow \tilde{x} \in S_{m_1, r_2, \dots, r_N}, \quad r_n \in \{m_2, \dots, m_{M_n}\}, \quad n = 2, \dots, N \quad (21)$$

Expression (20) can be generalized for any image k as a reference image, in such way that

$$F_{m_k}^{S}(\tilde{x}) = \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \dots \sum_{m_{k-1}=1}^{M_{k-1}} \sum_{m_{k+1}=1}^{M_{k+1}} \dots \sum_{m_N=1}^{M_N} F_{m_1, m_2, \dots, m_{k-1}, m_{k+1}, \dots, m_N}(\tilde{x}) \quad (22)$$

and the decision rule becomes

$$x^{(k)} \mapsto w_{m_k}^{(k)} \Leftrightarrow F_{m_k}(\tilde{x}) = \max_{m_a} F_{m_a}(\tilde{x}), \quad k = 1, 2, \dots, N \quad (23)$$

or in terms of decision space:

$$x^{(k)} \in S_{m_k} \Leftrightarrow \tilde{x} \in S_{r_1, r_2, \dots, r_{k-1}, m_k, r_{k+1}, \dots, r_N}, \quad r_n \in \{m_1, m_2, \dots, m_{k-1}, m_{k+1}, \dots, m_{M_n}\}, \quad n = 2, \dots, N \quad (25)$$

with

$$S_{m_k} = \bigcap_{r_1=0}^{M_1} \bigcap_{r_2=0}^{M_2} \dots \bigcap_{r_{k-1}=0}^{M_{k-1}} \bigcap_{r_{k+1}=0}^{M_{k+1}} \dots \bigcap_{r_N=0}^{M_N} S_{r_1, r_2, \dots, r_{k-1}, m_k, r_{k+1}, \dots, r_N} \quad (26)$$

3. Simulation example

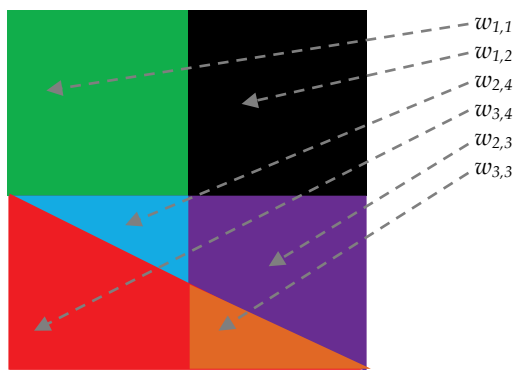
3.1 The simulated SAR images

Four set of two One-Look SAR images were simulated according to Table 1. The simulated SAR images have 512x512 pixels and are Rayleigh distributed (Oliver & Quegan, 1998). All of them have six classes and were smoothed by a $K \times K$ mean filter ($K = 3, 5$ and 7). The filtering causes that the distribution $p_X(x^{(n)} | w_{m_1}^{(1)}, w_{m_2}^{(2)}, \dots, w_{m_n}^{(N)})$ or $p_X(x^{(n)} | w_m^{(n)})$ fit with a Gaussian distribution.

		Mean values of the Rayleigh r.v. for the classes					
Classes:		$w_{1,1}$	$w_{1,2}$	$w_{2,3}$	$w_{2,4}$	$w_{3,3}$	$w_{3,4}$
Set 1	Image 1a	20.6	19.4	24.4	25.6	29.4	30.6
	Image 2a	30	35	39.4	55.6	40.6	54.4
Set 2	Image 1a	20.6	19.4	24.4	25.6	29.4	30.6
	Image 2b	30	45	59	106	61	104
Set 3	Image 1b	21	19	34	36	49	51
	Image 2a	30	35	39.4	55.6	40.6	54.4
Set 4	Image 1b	21	19	34	36	49	51
	Image 2b	30	45	59	106	61	104

Table 1. Images sets with six classes parameters

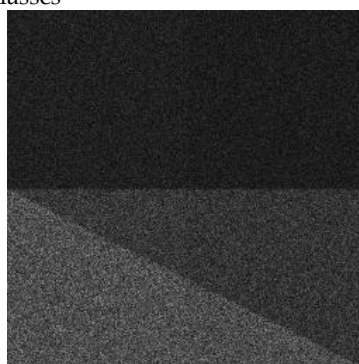
Figure 3(a) shows the ground truth of a scene with 6 classes with the labels: $w_{1,1}$, $w_{1,2}$, $w_{2,4}$, $w_{3,4}$, $w_{2,3}$ and $w_{3,3}$. The scene is observed by two different SAR sensors which generate two different and independent images with the mean values defined in Table 1. The different simulated images are shown in Figure 3(b)-3(e).



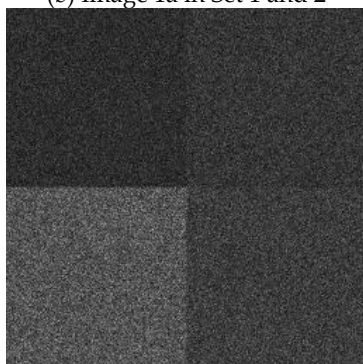
(a) Ground Truth with six classes



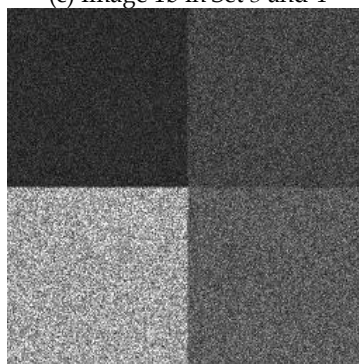
(b) Image 1a in Set 1 and 2



(c) Image 1b in Set 3 and 4



(d) Image 2a in Set 1 and 3



(e) Image 2b in Set 2 and 4

Fig. 3. Simulated SAR images with six classes

It is been considered that the sensors have some particular characteristics such that the first image generated by the sensor 1 (image 1a or 1b) has three prevalent classes and the second image (image 2a or 2b) generated by the sensor 2 has four classes. In all sets the Images 1a and 1b have three predominant classes each of them composed by two different classes:

$$\begin{aligned}w_1^{(1)} &= f_1^{(1)}(w_{1,1}, w_{1,2}) \\w_2^{(1)} &= f_2^{(1)}(w_{2,3}, w_{2,4}) \\w_3^{(1)} &= f_3^{(1)}(w_{3,3}, w_{3,4})\end{aligned}\quad (27)$$

and in all sets the images 2a and 2b have four predominant classes each of them composed as follow:

$$\begin{aligned}w_1^{(2)} &= f_1^{(2)}(w_{1,1}) \\w_2^{(2)} &= f_1^{(2)}(w_{1,2}) \\w_3^{(2)} &= f_3^{(2)}(w_{2,3}, w_{3,3}) \\w_4^{(2)} &= f_4^{(2)}(w_{2,4}, w_{3,4})\end{aligned}\quad (28)$$

Image 1a has its predominant classes with the mean values of its components classes very close (difference equal to 1.2 - set 1 and 2). In image 1b the predominant classes are more separated (difference equal to 3.0 - set 3 and 4). Image 2a has its predominant classes with the mean values of its components classes very close (difference equal to 1.2 - set 1 and 3). In image 2b its predominant classes are more separated and more different (difference equal to 3.0 - set 2 and 4). Due to this simulated characteristics the classification become easier and therefore more precise from set 1 to set 4.

The *a priori* probabilities of the classes in the ground truth are given in Table 2 and the conditional probability is presented in Table 3.

Classes:	$w_{1,1}$	$w_{1,2}$	$w_{2,3}$	$w_{2,4}$	$w_{3,3}$	$w_{3,4}$
Probabilities:	1/4	1/4	3/16	1/16	1/16	3/16

Table 2. Six classes a priori probabilities

Classes:	$m_2=1$	$m_2=2$	$m_2=3$	$m_2=4$
$m_1=1$	1/2	1/2	0	0
$m_1=2$	0	0	3/4	1/4
$m_1=3$	0	0	1/4	3/4

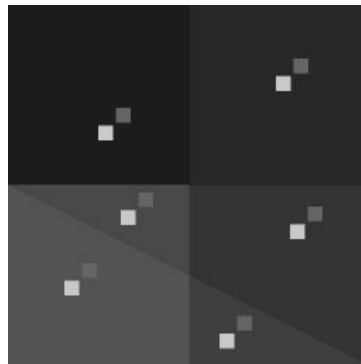
Table 3. Conditional probabilities $P(w_{m_1}^{(1)} | w_{m_2}^{(2)})$

3.2 The classification results

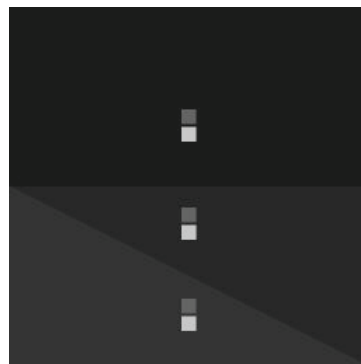
It was performed a supervised image classification of the six classes images in set 1, 2, 3 and 4 in three ways:

- a) Image 1 was classified by equation (8) in six classes as a single image ($N = 1$);
- b) Image 1 and 2 were classified in six classes by equation (8) as a bidimensional observation process ($N = 2$) and
- c) Image 1 is selected as the reference image having the classes $w_1^{(1)}$, $w_2^{(1)}$ and $w_3^{(1)}$ and image 2 as the complementary image with the classes $w_1^{(2)}$, $w_2^{(2)}$, $w_3^{(2)}$ and $w_4^{(2)}$ in the classification structure given by equation (14).

In the supervised classification processes $p_x(x^{(1)} | W_1^N)$ and $p_x(x^{(n)} | W_1^{n-1})$ were estimated in a neighborhood of 20x20 pixels in each class and classes combinations. The error matrix (Congalton, 1991) for the Kappa estimation was also calculated in a different neighborhood also of 20x20 in each class. Figure 4 shows the samples used for the probability density estimations and for the error matrix calculation. The Kappa estimated values and its RMS error (Congalton & Green, 1999) considering the smooth filter with $K \times K = 3 \times 3, 5 \times 5$ and 7×7 for the three classifications processes are shown in Tables 4 to 6.



(a) Estimation and test sample for the six classes classification



(b) Estimation and test samples for three classes classification

Fig. 4. Samples distributions in the images: □ estimation window and ■ test window

The classification, as could be expected, is better with the use of the two images and become also better from the set 1 to 4. The classification performance considering the reference image is better than the others including the most critical situation that is given by the images in the set 1.

Classification in 6 classes (equat. 8)	Set 1	Set 2	Set 3	Set 4
Image 1	0.17±0.01	0.17±0.01	0.33±0.01	0.33±0.01
Image 1 and 2	0.32±0.01	0.51±0.01	0.53±0.01	0.71±0.01
Classification in 3 classes (equat. 14)	Set 1	Set 2	Set 3	Set 4
Reference Image 1	0.50±0.02	0.53±0.02	0.79±0.01	0.81±0.01

Table 4. Kappa values for the classification (3x3 smooth window)

Classification in 6 classes (equat. 8)	Set 1	Set 2	Set 3	Set 4
Image 1	0.26±0.01	0.26±0.01	0.43±0.01	0.43±0.01
Image 1 and 2	0.55±0.01	0.73±0.01	0.78±0.01	0.91±0.01
Classification in 3 classes (equat. 14)	Set 1	Set 2	Set 3	Set 4
Reference Image 1	0.74±0.02	0.78±0.01	0.94±0.01	0.96±0.01

Table 5. Kappa values for the classification (5x5 smooth window)

Classification in 6 classes (equat. 8)	Set 1	Set 2	Set 3	Set 4
Image 1	0.34±0.01	0.34±0.01	0.52±0.01	0.53±0.01
Image 1 and 2	0.73±0.01	0.84±0.01	0.90±0.01	0.98±0.01
Classification in 3 classes (equat. 14)	Set 1	Set 2	Set 3	Set 4
Reference Image 1	0.86±0.01	0.88±0.01	0.99±0.01	0.99±0.01

Table 6. Kappa values for the classification (7x7 smooth window)

Figures 5 and 6 show the classification results for the original SAR images smoothed by a 5x5 mean filter. In these figures it can be seen that the classification becomes better from the set 1 to 4 and the classification with the reference image get the best classification results.

3.3 The classification in three classes

We will now consider an extreme case in which the ground truth in Fig 3(a) has only three classes as shown in Fig 7(a) in such way that:

$$\begin{aligned}
 w_1 &= w_{1,1} \cup w_{1,2} \\
 w_2 &= w_{2,3} \cup w_{2,4} \\
 w_3 &= w_{3,3} \cup w_{3,4}.
 \end{aligned}
 \tag{29}$$

It is also considered the maximum possible classes separation in Table 1 resulting in images with Rayleigh r.v. mean values shown in Table 7. Figures 7(b)-7(e) show the four sets of images in the simulation.

The Kappa classification results using equation (8) with the smoothing filter with $K \times K = 3 \times 3, 5 \times 5$ and 7×7 are shown in Tables 8 - 10. The values in the last line of Tables 8 - 10 can be considered for each smooth filter as the limit value for the MAP classification in the ideal case for the simulated images. Comparing these values with the values in the last line of Tables 4 - 6, respectively, we conclude that for the worst case (set 1 and set 1o) the classification of the set 1 with a reference image has its Kappa:

75.8% of the value of the best classification in set 1o for 3×3 smooth window,
80.4% of the value of the best classification in set 1o for 5×5 smooth window, and
86.9% of the value of the best classification in set 1o for 7×7 smooth window.

Figures 8(a)-(c) and 9(a)-(c) show for the 5×5 smooth filter the classification results for the three classes images according to the images in the sets 1o, 2o, 3o and 4o. Figure 8(d)-(e) and Figure 9(d)-(e) repeat the three classes classification for the image 1a and 1b as reference images shown in Figure 5(d)-(e) and Figure 6(d)-(e), respectively.

Classes:		Mean values of the Rayleigh r.v. for the classes		
		w_1	w_2	w_3
Set 1o	Image 1oa	19.4	25.6	30.6
	Image 2oa	30	40.6	55.6
Set 2o	Image 1oa	19.4	25.6	30.6
	Image 2ob	30	61	106
Set 3o	Image 1ob	19	36	51
	Image 2oa	30	40.6	55.6
Set 4o	Image 1ob	19	36	51
	Image 2ob	30	61	106

Table 7. Images sets with three classes parameters

Classification in 3 classes: (equat. 8)	Set 1o	Set 2o	Set 3o	Set 4o
Image 1o	0.39±0.02	0.39±0.02	0.77±0.01	0.77±0.01
Image 1o and 2o	0.66±0.02	0.93±0.01	0.85±0.01	0.95±0.01

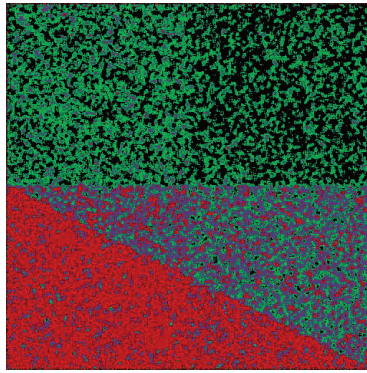
Table 8. Kappa values for the classification in three classes (3×3 smooth window)

Classification in 3 classes: (equat. 8)	Set 1o	Set 2o	Set 3o	Set 4o
Image 1o	0.61±0.02	0.61±0.02	0.93±0.01	0.93±0.01
Image 1o and 2o	0.92±0.01	0.99±0.01	0.98±0.01	0.99±0.01

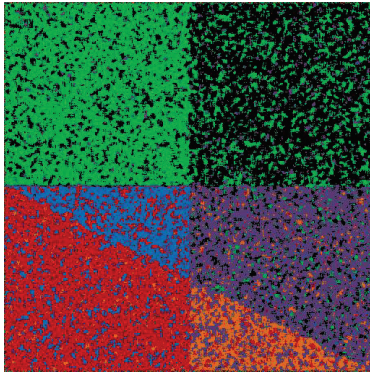
Table 9. Kappa values for the classification in three classes (5×5 smooth window)

Classification in 3 classes: (equat. 8)	Set 1o	Set 2o	Set 3o	Set 4o
Image 1o	0.81±0.02	0.81±0.01	0.99±0.01	0.99±0.01
Image 1o and 2o	0.99±0.01	1.00	1.00	1.00

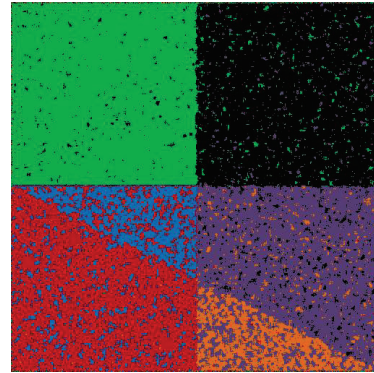
Table 10. Kappa values for the classification in three classes (7×7 smooth window)



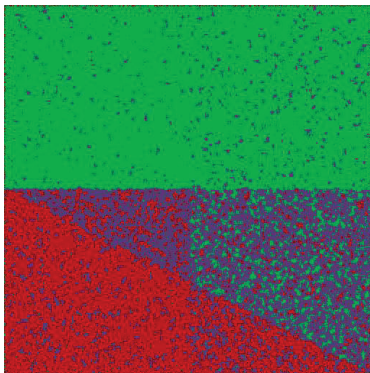
(a) Image 1a in set 1 and in set 2
Six classes classification



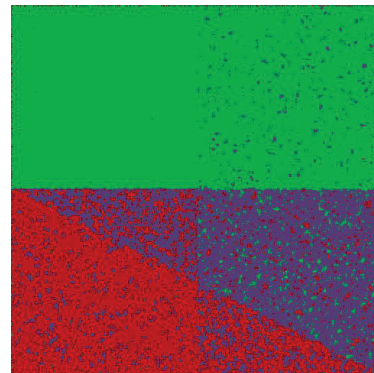
(b) Image 1a and 2a in set 1
Six classes classification



(c) Image 1a and 2b in set 2
Six classes classification



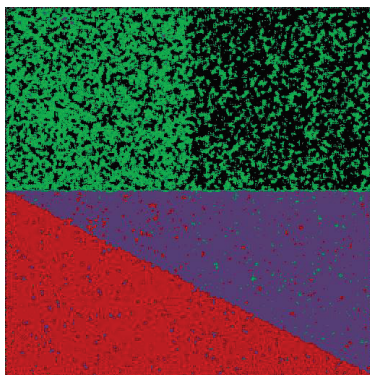
(d) Image 1a as reference in set 1
Three classes classification



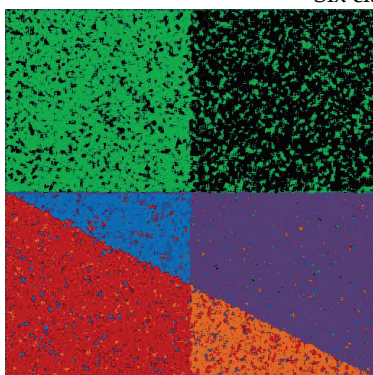
(e) Image 1a as reference in set 2
Three classes classification

Six classes: $w_{1,1}$: ■ $w_{1,2}$: ■ $w_{2,3}$: ■ $w_{2,4}$: ■ $w_{3,3}$: ■ $w_{3,4}$: ■
 Three classes: $w_1^{(1)}$: ■ $w_2^{(1)}$: ■ $w_3^{(1)}$: ■

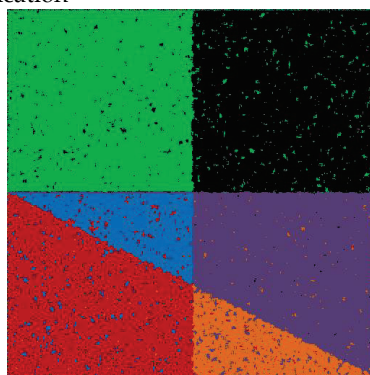
Fig. 5. Classification results in set 1 and set 2 (5x5 smooth filter)



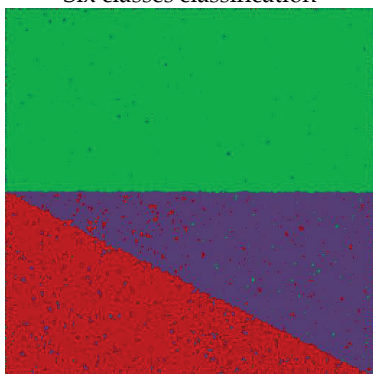
(a) Image 1b in set 3 and in set 4
 Six classes classification



(b) Image 1b and 2a in set 3
 Six classes classification



(c) Image 1b and 2b in set 4
 Six classes classification



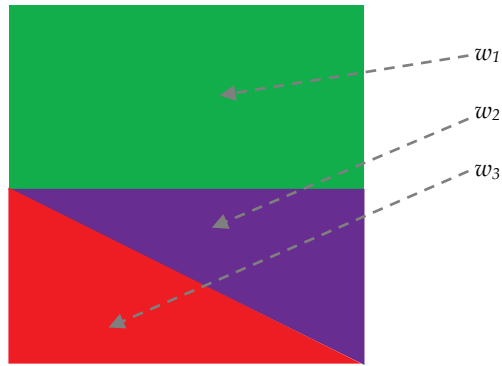
(d) Image 1b as reference in set 3
 Three classes classification



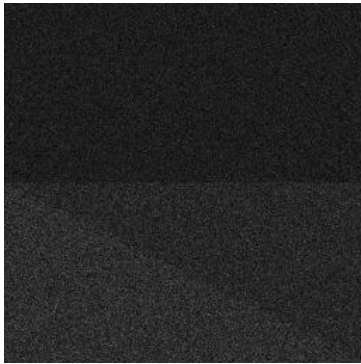
(e) Image 1b as reference in set 4
 Three classes classification

Six classes: $w_{1,1}$: ■ $w_{1,2}$: ■ $w_{2,3}$: ■ $w_{2,4}$: ■ $w_{3,3}$: ■ $w_{3,4}$: ■
 Three classes: $w_1^{(1)}$: ■ $w_2^{(1)}$: ■ $w_3^{(1)}$: ■

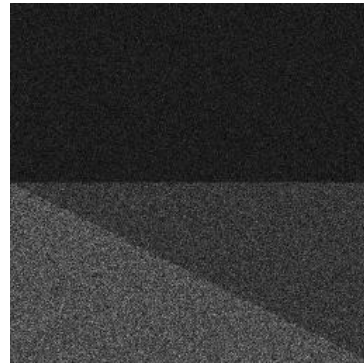
Fig. 6. Classification results in set 3 and set 4 (5x5 smooth filter)



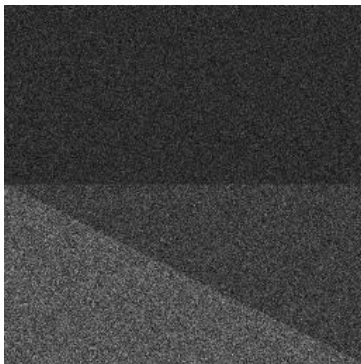
(a) Ground Truth with three classes



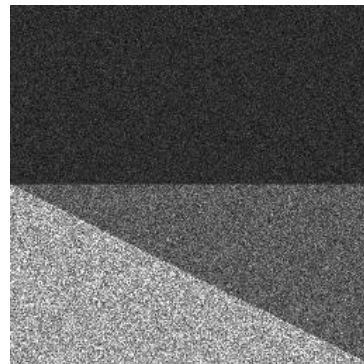
(b) Image 10a in Set 1o and 2o



(c) Image 10b in Set 3o and 4o

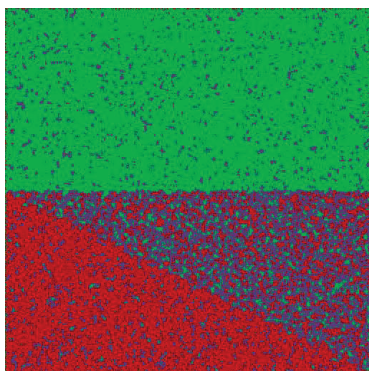


(d) Image 20a in Set 1o and 3o

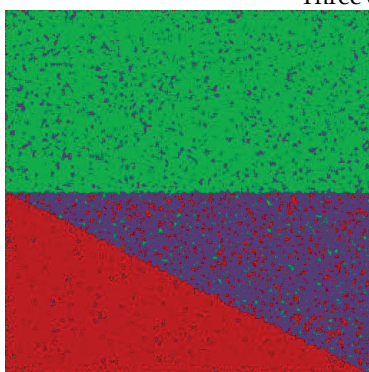


(e) Image 20b in Set 2o and 4o

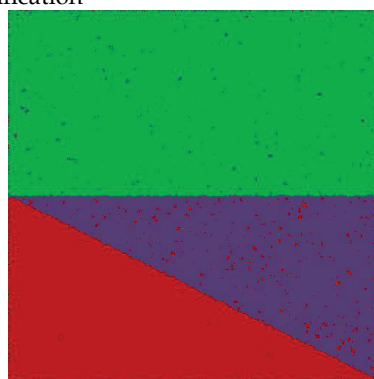
Fig. 7. Simulated images with three classes



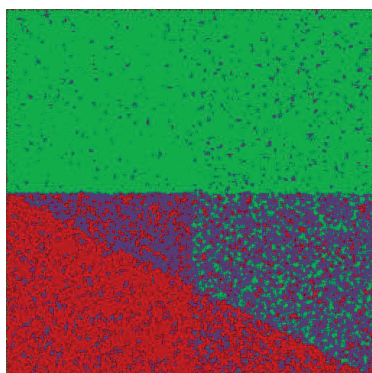
(a) Image 10a in set 1o and in set 2o
Three classes classification



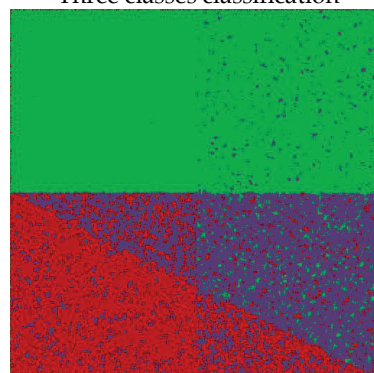
(b) Image 10a and 20a in set 1o
Three classes classification



(c) Image 10a and 20b in set 2o
Three classes classification



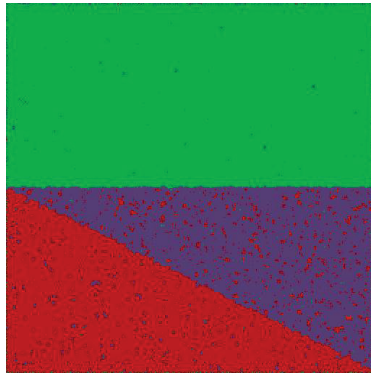
(d) Image 1a as reference in set 1
Three classes classification



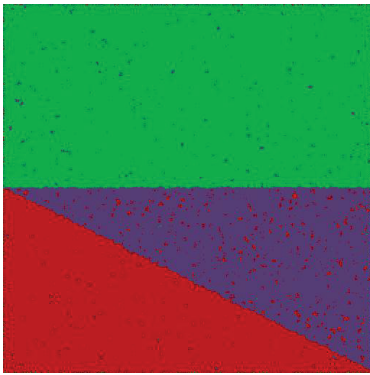
(e) Image 1a as reference in set 2
Three classes classification

Three classes: $w_1^{(1)}$:  $w_2^{(1)}$:  $w_3^{(1)}$: 

Fig. 8. Classification results in set 1o and set 2o (5x5 smooth filter)



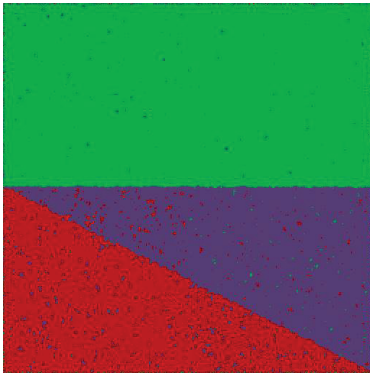
(a) Image 1ob in set 3o and in set 4o
Three classes classification



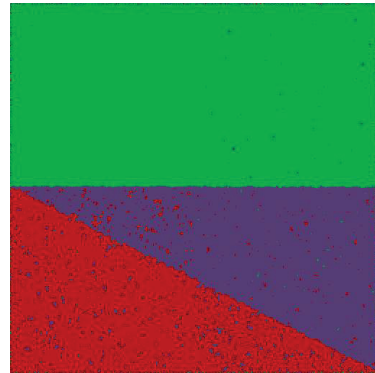
(b) Image 1ob and 2oa in set 3o
Three classes classification



(c) Image 1ob and 2ob in set 4o
Three classes classification



(d) Image 1b as reference in set 3
three classes classification



(e) Image 1b as reference in set 4
three classes classification

Three classes: $w_1^{(1)}$:  $w_2^{(1)}$:  $w_3^{(1)}$: 

Fig. 9. Classification results in set 3o and set 4o (5x5 smooth filter)

4. Conclusions

In the MAP classification context, it was shown a classification rule that consider a reference image to be classified and a set of complementary images as additional information. The images in the classification process can have different numbers and types of prevalent classes. The different prevalent classes are due to the imaging sensor characteristics (frequency, polarization, resolution etc) and its interaction with the scene (incidence angle, geometry, reflectivity etc).

It was presented a simulation example considering four sets of two amplitude SAR images of a scene. The reference and the complementary images have three and four dominant classes respectively. The original Rayleigh distributed images were filtered by a moving average filter and the new Gaussian distributed images had their parameters estimated for the MAP classification. The performance of the classification was evaluated by the error matrix Kappa coefficient.

In general, to apply the classification process that has a reference image to be classified and a set of images as complementary information one has to follow these steps to calculate the GMF $F_{m_1}^s(\tilde{x})$ given by equation (12):

a) Each image, including the reference image, must be observed or raw classified in order to be established a raw map of all possible classes in the scene and in the images. In the presented simulation the scene has six classes $\{w_{1,1}, w_{1,2}, w_{2,3}, w_{2,4}, w_{3,3}, w_{3,4}\}$. The images have the classes $\{w_1^{(1)}, w_2^{(1)}, w_3^{(1)}\}$ and $\{w_1^{(2)}, w_2^{(2)}, w_3^{(2)}, w_4^{(2)}\}$ that are compositions of the former classes. The classes in each image and in the scene (considered as the ground truth) must have an association as shown in equations (27) and (28).

b) It must be estimated the *a priori* probability $P(w_{m_1}^{(1)})$ and the conditional probabilities, $P(w_{m_n}^{(n)} | W_1^{n-1})$ with $W_1^n = \{w_{m_1}^{(1)}, w_{m_2}^{(2)}, \dots, w_{m_n}^{(n)}\}$. The classes $w_{m_n}^{(n)}$ that don't have an association with the classes in W_1^n , through the former classes in the scene, have as result $P(w_{m_n}^{(n)} | W_1^{n-1}) = 0$. In the simulation, the classes $w_1^{(1)} = f_1^{(1)}(w_{1,1}, w_{1,2})$ and $w_3^{(2)} = f_3^{(2)}(w_{2,3}, w_{3,3})$ don't have intersection and, therefore, $P(w_3^{(2)} | w_1^{(1)}) = 0$.

c) It must be estimated the conditional distribution parameters of $p_X(x^{(1)} | W_1^N)$ and $p_X(x^{(n)} | W_1^N)$ for all classes combinations that have intersections. In the simulation, the classes $w_1^{(1)}$ and $w_3^{(2)}$ don't have intersection and, therefore, $p_X(x^{(n)} | w_1^{(1)}, w_3^{(2)}) = 0$. If the images are not independent, the conditional distribution $p_X(\tilde{x} | W_1^N)$ must be calculated instead $p_X(x^{(n)} | W_1^N)$.

d) Given $F_{m_1}^s(\tilde{x})$ for each class $w_{m_1}^{(1)}$, $m_1=1, 2, \dots, M_1$, the decision rule (14) can be applied to classify only the reference image $x^{(1)}$ using the $x^{(2)}, x^{(3)}, \dots, x^{(N)}$ images as auxiliary information in the MAP decision processes.

5. References

- Bruzzone, L.; Chi, M. & Marconcini, M. (2006). A novel transductive svm for semisupervised classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, n. 11, November, pp. 3363-3373.
- Camps-Valls, G.; Marsheva, T. & Zhou, D. (2007). Semi-supervised graph based hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, n. 10, October, pp. 3044-3054.
- Congalton, R. & Green, K. (1999). *Assessing the accuracy of remotely sensed data: principles and practices*. Lewis Publishers, Boca Raton.
- Congalton, R. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, vol. 37, n. 1, pp. 35-46.
- Fukunaga, K. (1990). *Introduction to statistical pattern recognition, 2 ed.* Academic Press, San Diego.
- Lee T.; Richards, J & Swain, P. H. (1987). Probabilistic and evidential approaches for multisource data analysis. *IEEE Transactions on Geoscience and Remote Sensing*, GE-25, n. 3, May, pp. 283-293.
- Liu, X.; Li, X.; Liu, L.; He, J. & Ai, B. (2008). An innovative method to classify remote-sensing images using ant colony optimization. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, n. 12, December, pp. 4198-4208.
- Máximo, O. A. & Fernandes, D. (2009) *Classificação de imagens de diversas fontes de informação com o uso de controladores de influência para as imagens e suas classes* - PhD Thesis - Instituto Tecnológico de Aeronáutica, Brasil - In Publishing Process.
- Oliver, C. & Quegan, S (1998). *Understanding Synthetic Aperture Images*, Artech House, London.
- Pal, M. & Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, vol. 86, pp. 554-565.
- Rosenfield G. & Fitzpatrick-Lins, K. (1986). A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing*, vol. 52, n. 2, February, pp. 223-227.
- Samaniego, L.; Bárdossy, A. & Schulz, K. (2008). Supervised classification of remotely sensed imagery using a modified k-nn technique. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, n. 7, July, pp. 2112-2125.
- Schowengerdt, R. A. (1997) *Remote Sensing - Models and Methods for Image Processing*, Academic Press, San Diego.
- Sharf, L. L. (1991). *Statistical signal processing - detection, estimation and time series analysis*. Addison-Wesley Publishing Company, Massachusetts.
- Valet, L.; Mauris, G. & Bolon, P. (2001). A statistical overview of recent literature in information fusion. *IEEE Aerospace and Electronic Systems Magazine*, vol 16, n. 3, March, pp. 7-14.
- Zhu, H. & Basir O. (2005). An adaptive fuzzy evidential nearest neighbor formulation for classifying remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, n. 8, August, pp. 1874-1889.

Optical Satellite Volcano Monitoring: A Multi-Sensor Rapid Response System

Kenneth A. Duda¹, Michael Ramsey², Rick Wessels³ and Jonathan Dehn⁴
¹SGT, contractor to U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota; work performed under USGS contract 08HQC�0005. ²Department of Geology & Planetary Science, University of Pittsburgh, and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Science Team. ³USGS Alaska Volcano Observatory. ⁴University of Alaska Fairbanks and Alaska Volcano Observatory, United States of America.

1. Introduction

In this chapter, the use of satellite remote sensing to monitor active geological processes is described. Specifically, threats posed by volcanic eruptions are briefly outlined, and essential monitoring requirements are discussed. As an application example, a collaborative, multi-agency operational volcano monitoring system in the north Pacific is highlighted with a focus on the 2007 eruption of Kliuchevskoi volcano, Russia. The data from this system have been used since 2004 to detect the onset of volcanic activity, support the emergency response to large eruptions, and assess the volcanic products produced following the eruption. The overall utility of such integrative assessments is also summarized.

The work described in this chapter was originally funded through two National Aeronautics and Space Administration (NASA) Earth System Science research grants that focused on the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) instrument. A skilled team of volcanologists, geologists, satellite tasking experts, satellite ground system experts, system engineers and software developers collaborated to accomplish the objectives. The first project, *Automation of the ASTER Emergency Data Acquisition Protocol for Scientific Analysis, Disaster Monitoring, and Preparedness*, established the original collaborative research and monitoring program between the University of Pittsburgh (UP), the Alaska Volcano Observatory (AVO), the NASA Land Processes Distributed Active Archive Center (LP DAAC) at the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center, and affiliates on the ASTER Science Team at the Jet Propulsion Laboratory (JPL) as well as associates at the Earth Remote Sensing Data Analysis Center (ERSDAC) in Japan. This grant, completed in 2008, also allowed for detailed volcanic analyses and data validation during three separate summer field campaigns to Kamchatka Russia. The second project, *Expansion and synergistic use of the ASTER Urgent Request Protocol (URP) for natural disaster monitoring and scientific analysis*, has expanded the project to other volcanoes around the world and is in progress through 2011.

The focus on ASTER data is due to the suitability of the sensor for natural disaster monitoring and the availability of data. The instrument has several unique facets that make it especially attractive for volcanic observations (Ramsey and Dehn, 2004). Specifically, ASTER routinely collects data at night, it has the ability to generate digital elevation models using stereo imaging, it can collect data in various gain states to minimize data saturation, it has a cross-track pointing capability for faster targeting, and it collects data up to $\pm 85^\circ$ latitude for better global coverage. As with any optical imaging-based remote sensing, the viewing conditions can negatively impact the data quality. This impact varies across the optical and thermal infrared wavelengths as well as being a function of the specific atmospheric window within a given wavelength region. Water vapor and cloud formation can obscure surface data in the visible and near infrared (VNIR)/shortwave infrared (SWIR) region due mainly to non-selective scattering of the incident photons. In the longer wavelengths of the thermal infrared (TIR), scattering is less of an issue, but heavy cloud cover can still obscure the ground due to atmospheric absorption. Thin clouds can be optically-transparent in the VNIR and TIR regions, but can cause errors in the extracted surface reflectance or derived surface temperatures. In regions prone to heavy cloud cover, optical remote sensing can be improved through increased temporal resolution. As more images are acquired in a given time period the chances of a clear image improve dramatically. The Advanced Very High Resolution Radiometer (AVHRR) routine monitoring, which commonly collects 4-6 images per day of any north Pacific volcano, takes advantage of this fact. The rapid response program described in this chapter also improves the temporal resolution of the ASTER instrument.

ASTER has been acquiring images of volcanic eruptions since soon after its launch in December 1999. An early example included the observations of the large pyroclastic flow deposit emplaced at Bezymianny volcano in Kamchatka, Russia. The first images in March 2000, just weeks after the eruption, revealed the extent, composition, and cooling history of this large deposit and of the active lava dome (Ramsey and Dehn, 2004). The initial results from these early datasets spurred interest in using ASTER data for expanded volcano monitoring in the north Pacific. It also gave rise to the multi-year NASA-funded programs of rapid response scheduling and imaging throughout the Aleutian, Kamchatka and Kurile arcs. Since the formal establishment of the programs, the data have provided detailed descriptions of the eruptions of Augustine, Bezymianny, Kliuchevskoi and Sheveluch volcanoes over the past nine years (Wessels et al., in press; Carter et al., 2007, 2008; Ramsey et al., 2008; Rose and Ramsey, 2009).

The initial research focus of this rapid response program was specifically on automating the ASTER sensor's ability for targeted observational scheduling using the expedited data system. This urgent request protocol is one of the unique characteristics of ASTER. It provides a limited number of emergency observations, typically at a much-improved temporal resolution and quicker turnaround with data processing in the United States rather than in Japan. This can speed the reception of the processed data by several days to a week. The ongoing multi-agency research and operational collaboration has been highly successful. AVO serves as the primary source for status information on volcanic activity, working closely with the National Weather Service (NWS), Federal Aviation Administration (FAA), military and other state and federal emergency services. Collaboration with the

Russian Institute of Volcanology and Seismology (IVS)/Kamchatka Volcanic Eruption Response Team (KVERT) is also maintained. Once a volcano is identified as having increased thermal output, ASTER is automatically tasked and the volcano is targeted at the next available opportunity. After the data are acquired, scientists at all the agencies have access to the images, with the primary science analysis carried out at the University of Pittsburgh and AVO. Results are disseminated to the responsible monitoring agencies and the global community through e-mail mailing lists.

2. Overview

Few natural hazards have the devastating impact of large volcanic eruptions (Figure 1). These events can affect scales from the local citizen to the rare global impact where large Plinian eruptions can alter global climate for years to decades (Yang and Schlesinger, 2002). Eruptions can present significant and varying levels of threat to public safety and health by way of explosive forces that launch rocks and ash, destructive pyroclastic lava flows and gaseous emissions, disruption of transportation and communication, introduction of disease and the loss of life. In addition to threats on the ground, aircraft and passengers are at considerable risk if travelling in affected areas (Miller and Casadevall, 2000). This is a particular concern in the north Pacific where numerous flight routes pass over active volcanoes (Figure 2).



Fig. 1. An ash-rich volcanic cloud from the 1989-1990 eruption of Redoubt volcano (courtesy J. Warren, April 21, 1990). Large eruption clouds such as these are a major hazard to commercial aviation in the busy north Pacific corridor.

Therefore, advance warning of incipient volcanic activity can lead to better and more accurate eruption prediction and help to save lives. High-quality and detailed data are needed in order to assess the relative risks posed by any one of the potentially active volcanoes along the Aleutian-Kamchatka-Kurile arcs. These data can include seismic

monitoring, deformational analysis, studies of the emitted gas, visual and thermal observations, as well as the use of orbital remote sensing. The volcanoes can be monitored using in situ equipment, airborne and satellite imaging systems, or some combination of all of these.

Ground-based seismic and deformation measurements commonly provide the earliest indications of renewed activity and subsurface magma movement at a particular volcano (Stephens and Chouet, 2001; Fournier et al., 2009; Lu et al., 2007). Increased seismic activity and a determination of the frequency, depth, and type of earthquakes below a volcano can be leading indicators of later eruptions. Field- or space-based measurements of volcanically induced deformation together with increased levels of heat and gas emissions can further verify the seismic results and lead to a more complete picture of the changes with time. Changes in the alignment of surface features are noted through precise measurements and the collection of spectral, thermal, and GPS data can serve to validate the satellite-based observations.

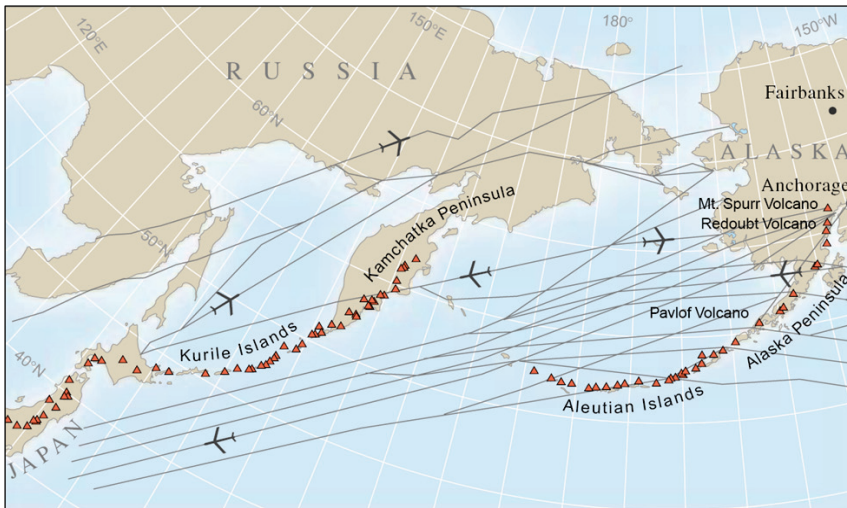


Fig. 2. North Pacific flight route map and the locations of the active volcanoes of the Alaska Peninsula, the Aleutian Islands, the Kamchatka Peninsula, the Kurile Islands, and into Japan (Courtesy USGS Fact Sheet 030-97).

In addition to ground-based measurements, satellite-based observations of volcanoes are of value. Satellite-based observations are useful in the northern Pacific and at many other active volcanoes around the world, where limited resources do not allow extensive ground-based monitoring. Large regions of the world and widely distributed and commonly remote targets can be monitored frequently and economically using satellite data. These images can reveal changes in thermal or gas emission, in the composition of the surface rocks, and in the deformation over time. They can thus provide additional insight into the evolving status of volcanoes. Fixed-wing aircraft and helicopters can also be used where resources permit to carry sensors such as thermal cameras, gas sensors, and cameras, which obtain higher

resolution data of the volcanic activity. The integrated analysis of all available data offers an enhanced perspective on the subtle and unique differences among volcanoes and on the likelihood of impending eruptive activity at a given site.

The scientists and engineers assembled to carry out the previously described NASA-funded research programs have the goal of improving volcanic monitoring in the north Pacific region in order to minimize the subsequent risks and increase scientific knowledge of these dynamic geologic processes. Incorporating the finer spatial resolution, multispectral ASTER data in the analysis process allowed this possibility and added greater clarity to the characterization of many remote volcanoes in the north Pacific region.

The work involved developing and implementing algorithms and tools to detect new activity, establish protocols for escalating response activity, create system linkages between AVO, the LP DAAC and JPL to enable semi-automated transmission of satellite tasking requests, develop the tools and procedures to identify ASTER overpass opportunities and control scheduling requests, secure sensor tasking authorization from the ASTER Science Team, employ existing systems to capture, downlink and process ASTER data, and finally create new data distribution mechanisms to ensure the timely availability of the acquired data. New data analysis procedures are then employed to assess current conditions and issue alerts when needed.

3. Sensors

The critical imaging requirements (i.e., optimum temporal frequency, spatial resolution, wavelengths, etc.) must be understood in order to use sensors on Earth-orbiting spacecraft to monitor volcanoes. The most commonly used instruments for volcano monitoring in the north Pacific include AVHRR, Geostationary Operational Environmental Satellite (GOES), Multifunctional Transport Satellites (MTSAT-1R), Moderate Resolution Imaging Spectroradiometer (MODIS), the Landsat Enhanced Thematic Mapper Plus (ETM+), and ASTER (Table 1). Sub-meter visible data from several commercial satellites (e.g. QuickBird, WorldView, IKONOS and GeoEye) have also recently become integrated as another useful tool for volcano monitoring.

Current sensors such as AVHRR, MODIS, MTSAT, and GOES provide the frequency necessary to detect the onset of large thermal anomalies in near real-time (Dehn et al, 2000; Wright et al., 2002, Schneider et al., 2000). In comparison, high spatial resolution instruments such as ASTER and Landsat ETM+ provide data at a much improved spatial scale ideal for scientific analysis, damage assessment, and smaller scale monitoring, but at the expense of rapid repeat times (Harris et al., 1998; Ramsey and Dehn, 2004). In order to obtain frequent status updates where attempting to initially identify thermal anomalies, sensors offering quick revisits (i.e., GOES, AVHRR, MODIS) are used (Figure 3). The disadvantage is that such data are collected at a coarse spatial resolution allowing only very large or very hot anomalies to be detected, whereas the non-eruptive or small-scale activity is missed completely at many of the remote volcanoes. Higher spatial resolution data (i.e., ASTER, ETM+) are employed in order to obtain more detailed information of lava flows, thermal anomalies, and gas emissions. However, nominally ASTER can only revisit a site once every

16 days at the equator viewing nadir. This temporal frequency is improved with off-nadir pointing and/or at higher latitudes. For example, at the higher latitudes of Kamchatka the temporal frequency can be shortened to 13 hours with off nadir pointing.

The ASTER URP has been used in support of North Pacific volcano monitoring to improve the collection probability and the speed of ASTER data availability. Through this approach, data are typically available within six hours after acquisition. This combination of frequent coarse spatial resolution change detection and less frequent detailed higher resolution imaging has proven very valuable scientifically and operationally. The three bands of ASTER VNIR data span from 0.52 to 0.86 micrometers at a ground resolution of 15 m. These data have been used to generate digital elevation models, map the eruption deposits and changing surface characteristics and to detect ground temperatures in excess of 800 °C. Thermal anomaly and gas emissions identification is accomplished primarily using sensor bands in the SWIR and TIR wavelengths. The six ASTER SWIR bands cover wavelengths from 1.6 to 2.43 micrometers at 30-m ground resolution. Unfortunately, the ASTER SWIR subsystem is no longer operational after 2008, but the data were commonly used to detect alteration minerals on the surface and temperatures between 100 °C and 460 °C. The ASTER TIR subsystem has five TIR bands from 8.125 to 11.65 micrometers at 90-m ground resolution. These data have been used to extract temperatures less than 100 °C, map silicate and carbonate minerals, model the vesicle content of lavas, and estimate the thermal inertia of the surfaces. A brief history of the ASTER mission, including key orbital and data characteristics, is provided in the text box entitled The ASTER Mission.

Criteria	GOES / MTSAT	AVHRR	MODIS	LANDSAT ETM+	ASTER
Revisit frequency	Geostationary	2 passes per day	1-2 days	16 days	16 days, less off-nadir
Ground resolution	At 60°N latitude: 2 km VIS 8 km IR	1.1 km (LAC)	250m (B1-2), 500m (B3-7), 1000m (B8-36)	30m (B1-5, 7) 60m (B6) 15m (B8)	15m (B1-3), 30m (B4-9), 90m (B10-14)
Spectral coverage	5 bands: 0.55 to 12.5 μm	6 bands: 0.58 to 12.4 μm	36 bands: 0.405 to 14.385 μm	8 bands: 0.45 to 12.5 μm	14 bands: 0.52 to 11.65 μm
Swath Width	Full Earth view	2,399 km	2,330 km	185 km	60 km
Orbit Altitude	35,800 km	833 km	705 km	705 km	705 km

Table 1. Comparison of key characteristics for sensors used in the operational north Pacific monitoring system.

4. Ground Systems and Data Acquisition

The infrastructure needed to support ASTER urgent request tasking, data collection, processing and eventual distribution is summarized in Figure 4. The five key steps of the volcano monitoring sequence are shown in the center of the figure. Corresponding contributions by participating groups are shown in the outer circle, occurring in a sequence that commences at the arrow and proceeds clockwise. The existing ASTER mission procedures and systems were used for final tasking, collection, and data processing.

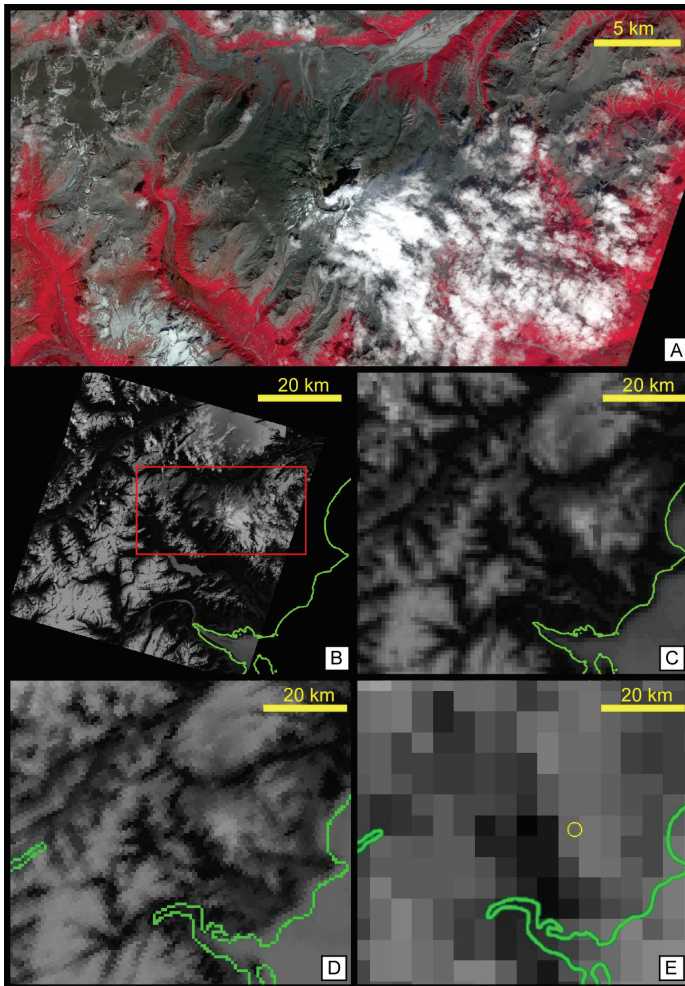


Fig. 3. Comparison of satellite data at several spatial resolutions over Redoubt volcano, Alaska during daylight hours between 21:30 and 22:01 UTC on June 6, 2009. **(A)** False-color subset of ASTER 15 m VNIR bands 3,2,1 as R,G,B. Figures B-E show TIR bands (centered on about 11 micrometers) from four different sensors: **(B)** ASTER 90 m TIR band 14, **(C)** MODIS 1 km TIR band 31, **(D)** AVHRR 1.1km TIR band 4, and **(E)** GOES 8 km TIR band 4.

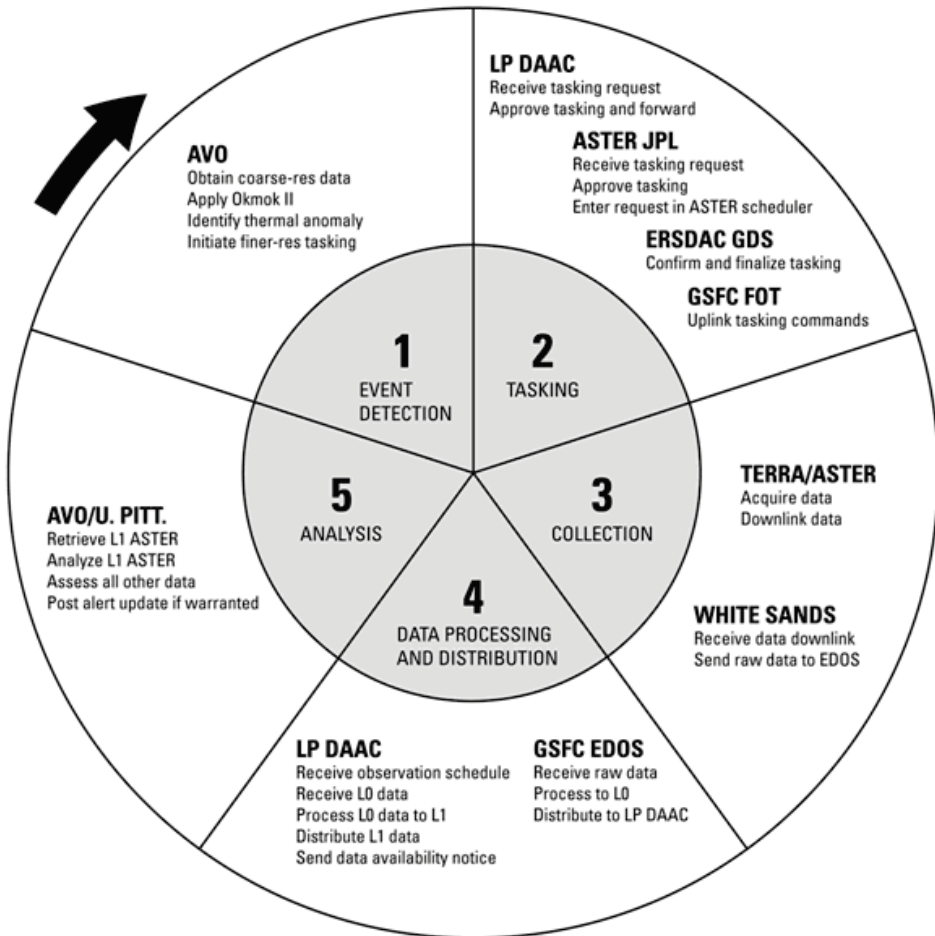


Fig. 4. Volcano monitoring actions and participant involvement. Size of each portion is not scaled to the actual work level involved. Acronyms: Alaska Volcano Observatory (AVO); Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER); Jet Propulsion Laboratory (JPL); Earth Remote Sensing Data Analysis Center (ERSDAC); Ground Data System (GDS); Goddard Space Flight Center (GSFC); Flight Operations Team (FOT); Earth Observing System Data and Information System (EOSDIS) Data and Operations System (EDOS); Land Processes Distributed Active Archive Center (LP DAAC); University of Pittsburgh (U. Pitt).

Initial thermal anomaly event detection in AVHRR or MODIS data triggers the ASTER ground systems, scheduling, and eventual data acquisition. The methods used to detect anomalous events include a detailed set of screening algorithms designed to minimize false positives prior to tasking ASTER. Once ASTER is tasked, a series of procedures are used to determine timing, scheduling support, and the eventual distribution of ASTER data. This

entire system of cross-satellite integration and supporting systems is located at AVO; LP DAAC; JPL; ERSDAC; Goddard Space Flight Center (GSFC); White Sands, New Mexico; and includes the ASTER instrument on the Terra spacecraft. The work of this project has contributed new techniques in event detection and tasking of a complex instrument.

4.1 Event Detection

The core of the triggering mechanism for the ASTER emergency acquisition requests is based on the Okmok algorithm (Dean et al., 1998). This algorithm uses a time series of AVHRR data to detect thermal anomalies above an expected average seasonal background temperature. Deviations in temperature may signal increased thermal emission and an impending eruption at a particular volcano. The algorithm scanned a small subsector of data over each volcano for the warmest IR band 3 (3.5-3.9 micrometers) radiant temperature. This value along with all other values in the data subset are recorded and tested based on a simple criterion: if the AVHRR band 3 temperature was greater than 35 °C, then an e-mail was generated and sent to all AVO staff. This process worked well, however its overall effectiveness was limited due to the poor geolocation accuracy of the AVHRR instrument and the generation of numerous false alarms due to noise.

In conjunction with the NASA funded projects, the Okmok algorithm has been significantly refined and applied to other sensors, such as MODIS. It has also been revised to better constrain the AVHRR data stream. The new algorithm, called Okmok II, focuses on decreasing the number of false alarms through a series of additional logic tests (Dehn and Harris, in press). However, it does not sacrifice the sensitivity to low grade thermal anomalies, which are notoriously hard to detect. Okmok II is built around seven levels of data processing and is interfaced with a new Web-based visualization tool (Figure 5). If triggered, an e-mail alert is sent directly to the ASTER Emergency Scheduling Interface and Control System (AESICS) database to immediately initiate the ASTER scheduling process.

4.2 ASTER Tasking and Collection

In support of the ASTER project, and for use with other applications, the LP DAAC created the ASTER Overpass Predictor to simplify the determination of future ASTER tasking opportunities. The data entry page includes the scene center latitude and longitude, and desired forecast window (Figure 6). Returned results identify possible dates, day or night opportunities, and which ASTER subsystems would be available. In addition to supporting the north Pacific volcano monitoring applications, this tool is now routinely used in support of other emergency response and research activities.

After an anomalous thermal event is detected, AVO initiates an ASTER tasking request and notifies the LP DAAC. These incoming requests are logged in a database included in AESICS. AESICS is a Web-based tool that was created to simplify tasking request submission and to control these requests (Figure 7). New tasking requests can be entered either manually or through the automated procedure, which begins through processes in place at AVO. After new requests are logged in the AESICS database, e-mail notifications are immediately forwarded to the ASTER team at JPL for the actual scheduling using pre-

existing mission protocols. Approvals to task the sensor occur at each stage of the process to ensure the best use of available resources and compliance with mission requirements and international agreements. For example, if there are multiple AVHRR alerts over a particular volcano in a given day, the system of checks guarantees that the ASTER scheduling system is not overloaded.

Following final approvals, and after confirming the absence of scheduling conflicts, the ASTER JPL team accesses the ASTER scheduling system at ERSDAC in Japan and the tasking request is uploaded. The final schedule is determined at ERSDAC and provided to the GSFC Flight Operations Team (FOT) for uplink to the Terra spacecraft. This entire process can be as quick as several hours. Once the data are acquired, they are stored in the Terra solid state recorder until downlink to receivers at White Sands, New Mexico. Raw data are transferred from White Sands via network to GSFC EDOS for initial processing from the raw image data format and then sent to the LP DAAC.

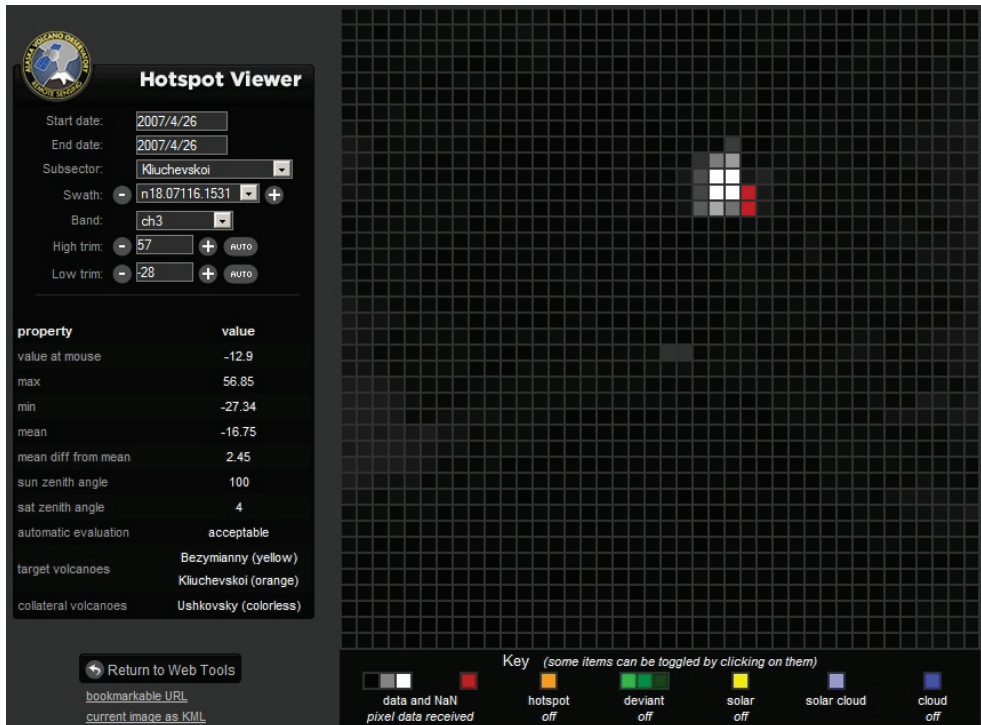


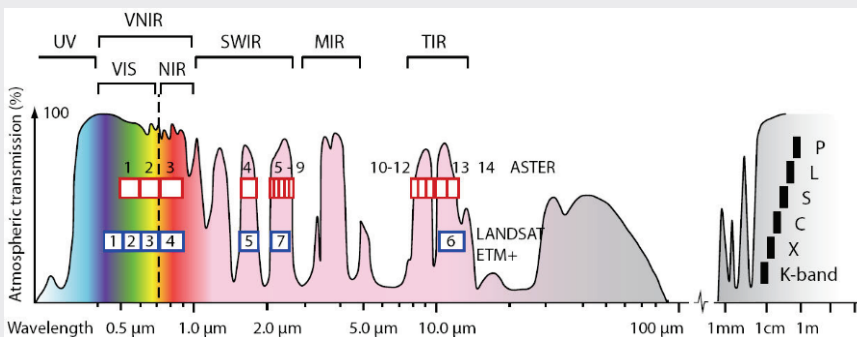
Fig. 5. AVO Hotspot Viewer Web interface for detection of AVHRR thermal anomalies. The AVHRR pixels are reformatted into 1-km grid cells with user-selectable tools for data enhancement. This tool is used to screen for actual anomalies and trigger an ASTER observation. Shown here are the data collected on April 26, 2007 from AVHRR on NOAA-18 for Kliuchevskoi volcano. This data viewer is also used by AVO in their routine volcano monitoring operations. (http://avo-animate.images.alaska.edu/auto_obs_viewer.php)

The ASTER Mission

ASTER is a joint endeavor involving NASA, Japan's Ministry of Economy, Trade and Industry, and other organizations. The ASTER sensor was launched in 1999 on the Terra spacecraft as part of NASA's Earth Observing System with the goal of conducting a global land mapping mission. ASTER has been accomplishing this goal very successfully, having already acquired over 1.5 million scenes. Though the instrument design life has been exceeded, ASTER continues to acquire approximately 450 new 60 by 60 km images of Earth's land surfaces daily, providing Earth land surface information useful for a wide variety of applications.

Terra orbits Earth once every 98.88 minutes at an elevation of 705 km in a sun-synchronous orbit at an inclination of 98.3 degrees. The descending orbit has a 10:30 AM equatorial crossing time and the revisit time is 16 days at the equator (less at higher latitudes or when off-nadir pointing is used). ASTER is tasked through scheduled observations at an 8% duty cycle, and has three imaging subsystems: VNIR, SWIR and TIR. Ground resolution is 15m for VNIR, 30m for SWIR, and 90m for TIR. ASTER has 14 spectral bands ranging from 0.52 micrometers to 11.65 micrometers, including a back-looking VNIR band 3 that enables the generation of digital elevation models. Data are distributed in the HDF-EOS format by the LP DAAC in the USA and by ERSDAC in Japan. More information is contained in Yamaguchi et al. (1998).

Level-1A reconstructed unprocessed instrument data are archived, and other products are generated from these data at the request of customers. Additional products include registered radiance at the sensor, surface reflectance, brightness temperature, surface kinetic temperature, surface emissivity, decorrelation stretch, polar surface and cloud classification, orthorectified, and digital elevation model (Abrams, 2000). The file size of the registered radiance product is approximately 118 MB, and contains data from each subsystem that are geometrically co-registered and radiometrically calibrated.



The Earth's atmospheric windows with the spectral coverage of various instruments shown. ASTER (shown in red), Landsat ETM+ (shown in blue), and SAR (shown in black). (Kaab, 2005)

Enter Latitude and Longitude	
Latitude	<input type="text" value="North"/> deg <input type="text"/> min <input type="text"/> sec <input type="text"/> OR decimal degrees <input type="text"/>
Longitude	<input type="text" value="West"/> deg <input type="text"/> min <input type="text"/> sec <input type="text"/> OR decimal degrees <input type="text"/>
<small>Do not include a "-" sign with the numeric value. Determine Latitude and Longitude for Geographic Features in U.S. & Territories</small>	
Select Criteria for ASTER Overpass Predictor	Maps & Weather
Predict Start Date: <input type="text" value="21"/> <input type="text" value="Sep"/> <input type="text" value="2009"/> <small>Default date is set at 2 days from current date.</small> Predict for: <input type="text" value="14"/> days	<input type="button" value="Location Map"/> <input type="button" value="Topographic Map (US Only)"/> <input type="button" value="Weather Forecast"/> Set spatial window: <input type="text"/> degrees
<input type="button" value="Forecast Possible ASTER Collection Times"/>	
<small>For Full Mode (VNIR, SWIR, TIR), require "Peak Elev" 81.5 °. For VNIR Only (wide point), require "Peak Elev" 66.0 °. Time over target = "Peak" UTC Time. In "Vis", Day = "DDD" and Night = "NNN, NNV, and NVV".</small>	

Fig. 6. ASTER Overpass Predictor Web page.
 (http://igskmncnw001.cr.usgs.gov/aster/estimator/reference_info.asp)

4.3 ASTER Data Processing and Distribution

Following standard ASTER mission protocols, raw ASTER data from the Terra spacecraft are processed by EDOS to Level-0. Level-0 data are then transferred via network to LP DAAC. LP DAAC receives the Level-0 data and processes it to Level-1 using executable code and observation schedule information received from ERSDAC. Upon completion of Level-1 processing, data are staged for retrieval via FTP. LP DAAC created a Recent Expedited Production Web site for this project to simplify and speed initial data assessment and downloading of ASTER expedited data. Data are also made available via AESICS, an FTP site, and through the Warehouse Inventory Search Tool (WIST) data search and order mechanism.

Employing the ASTER expedited approach greatly assists in the timely availability of image data for analysis. ASTER expedited data are typically available within six hours after collection, whereas the standard data product availability takes several days after acquisition. This lag time was actually several weeks when the URP volcano monitoring program was first conceived.

ASTER expedited products are very similar to standard products but there are some minor differences. Expedited data do not contain the back-looking band 3, so stereo data are not initially available for digital elevation model (DEM) generation. DEMs can be created later once the standard Level-1 products become available. In addition, short-term calibration for

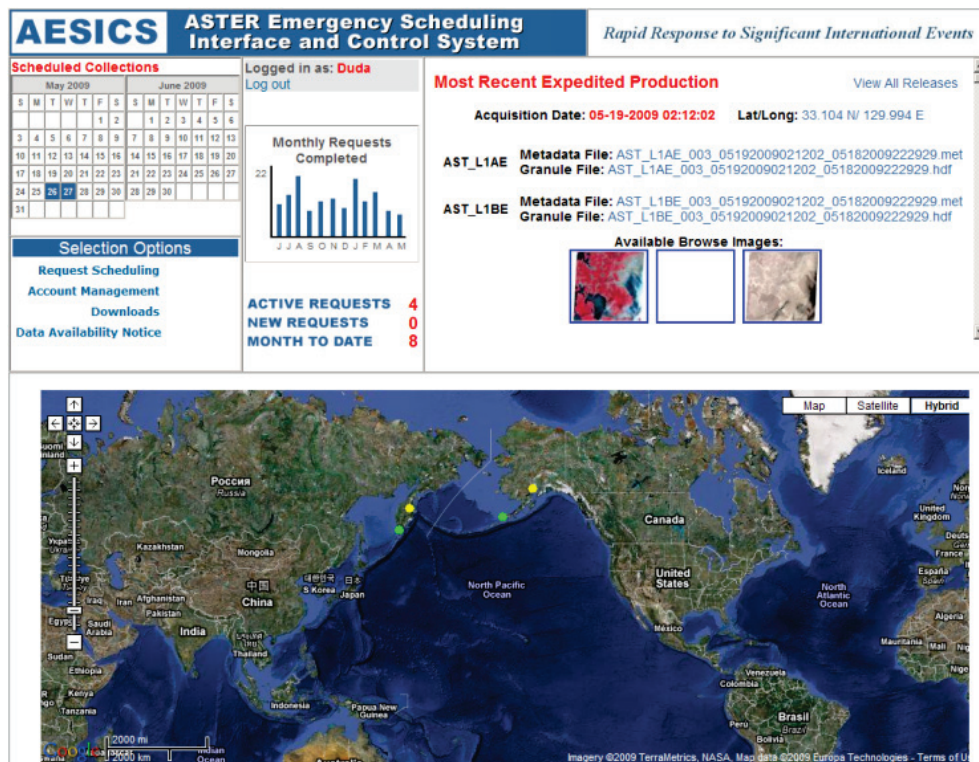


Fig. 7. AESICS Web-based interface. Users can rapidly access statistics on the number of ASTER urgent requests in the last month and year, thumbnail images of the most recent acquisitions, and a map-based display of all ASTER tasking requests in the past 30 days. Based on the Google Earth application, this interface allows users to see which ASTER tasks are new (red dot), approved (yellow dot), and completed (green dot). Each target can also be queried for more information.

the TIR is not available so long-term calibration is used. As a result, expedited TIR data quality is expected to be somewhat reduced. Inter-telescope registration quality is also slightly lower since numerous adjacent scenes are not available as is normally the case. Finally, expedited processing uses raw spacecraft ephemeris data, so the geometry is slightly different than for the data produced using standard processing, which uses refined (post-processed) ephemeris data. Even with these caveats, the image data are still of excellent quality and commonly used to report the latest activity and thermal output values.

4.4 ASTER Data Analysis and Alert Status Updates

After ASTER Level-1 data are processed, image analysts and volcanologists at AVO and the University of Pittsburgh obtain the images for inspection to gain further insight on the current state of the volcano. This may include a visual assessment of the available bands, a characterization of temperature conditions, areal extent of the anomaly, time series analysis,

documentation of flow patterns, and plume extent and content. New data are compared with historical satellite data and other information to identify trends. If warranted, volcano status alerts are updated to notify government authorities and other interested individuals. The volcano monitoring plan is also then updated to ensure that continued observations occur as needed.

5. The 2007 Eruption of Kliuchevskoi Volcano, Russia

One of the most volcanically active regions in the world is Kamchatka, Russia, which is located in the northwestern region of the Pacific Ocean. Several of the more than two dozen active volcanoes on this peninsula are commonly erupting at any given time. These eruptions can produce hazards for the sparse local population living near the volcanoes. But of far more consequence are the eruptions that produce larger ash columns, which are carried by the easterly winds into the routes of approximately 200 aircraft and 20,000 people overflying the region each day (Miller and Casadevall, 2000).

Kliuchevskoi is the highest (> 4800 m) and one of the most active volcanoes in Kamchatka. In the last century, summit eruptions have increased in frequency averaging one every 1-2 years. These eruptions are commonly caused by lava interaction with melting snow/ice and start with increased fumarolic activity within the summit crater. This thermal activity is typically followed by Strombolian explosions, the effusion of blocky lava flows, and the generation of hot avalanches and lahars. Less common are the paroxysmal eruptions, which last occurred in 1994. That eruption included a large convecting column, pyroclastic flows, lahars and lava flows. The largest explosive eruption reached 18 km above sea level and travelled approximately 1,000 km southeast into the north Pacific air traffic routes (Miller et al., 1994).

Ground-based techniques for monitoring the remote volcanoes of Kamchatka include seismic and visual observations. In 2007, there were 33 seismic stations deployed in Kamchatka (Chebrov, 2008). Human observations, reports, and a Web-based video camera system are also used. The Web camera has been installed in the town of Klyuchi 30 km north of Kliuchevskoi. Using the Web-based system, the height of the eruption column has been correlated with the level of seismic activity, thereby allowing seismic signals to predict the height of the eruption plume at night or in times of bad weather (McNutt, 1994; Roach et al., 2004). In addition to these data, high temporal/low spatial resolution orbital remote sensing monitoring of these remote volcanoes has been used for nearly two decades by volcanologists in Russia and the United States. ASTER has been an integral part of this monitoring since it was launched in 1999. Some of the first scientific images collected by ASTER in early 2000 were of the large eruption deposits of Bezymianny volcano (Ramsey and Dehn, 2004). Beginning in 2004, the ASTER rapid response/urgent request system has been linked to the routine remote sensing monitoring done by AVO and KVERT. The ASTER data from this collaborative program have provided the basis for enhanced monitoring efforts, new discoveries, and numerous scientific results (Carter et al., 2007, 2008; Ramsey et al., 2004; Rose and Ramsey, 2009).

During the period between 2000 and 2008, Bezymianny volcano, 10 km south of Kliuchevskoi, was nearly continuously active with approximately two large eruptions per year (Ramsey and Dehn, 2004; Carter et al., 2007). Therefore, this volcano produced numerous AVHRR thermal alerts and subsequent ASTER urgent request images (after the URP system was active). When both Bezymianny and Kliuchevskoi are active, the poor geolocation accuracy of AVHRR makes it difficult to discern which volcano is responsible for producing the thermal anomaly (Dehn et al, 2000). Similarly, both volcanoes commonly appear in one ASTER 60-km scene. It is therefore common for ASTER observations targeting one volcano to capture activity at the other. Typically, low-level thermal activity is seen long before visual or even AVHRR spaceborne observations detect that activity. The better radiometric accuracy, higher spatial resolution, and more precise geolocation of ASTER compared to AVHRR makes the TIR data ideal for detection of the very early stages of new activity at a volcano. However, the poorer temporal frequency commonly limits this important aspect of ASTER. Future TIR instruments with a temporal frequency of hours to days will provide a critically important new dataset for volcano monitoring and eruption prediction.

The detection of renewed activity at Kliuchevskoi first occurred in 2005 in an ASTER TIR scene collected to observe the waning stages of an eruption at Bezymianny (Rose and Ramsey, 2009). A similar situation occurred nearly two years later when observations of the 2006 eruptive activity at Bezymianny also showed a very slight increase in thermal output (~ 5 °C above background) at the summit of Kliuchevskoi as early as November 2, 2006. This activity was noted but did not raise concern due to its very low level and no other detectable activity from visual or seismic observations. On November 27, 2006 the activity remained unchanged. It was not until the first detection by AVHRR on December 14, 2006 that a rapid increase in activity was noted. Two AVHRR pixels between 10 and 30 °C above the average background temperature were detected at the summit, which likely meant that very hot gases and/or small amounts of lava had reached the surface. By December 22, 2006 the activity had further increased enough to trigger an ASTER urgent request image, which was scheduled for January 4-5, 2007 (Figure 8). No activity was detected in the subsequent ASTER VNIR data although a pixel-integrated brightness temperature of 332 °C was extracted from the SWIR data. By January 12, 2007 the color-code for Kliuchevskoi was raised from green to yellow indicating that heightened activity was taking place. However, this activity was low-grade enough that the first ground-based visible observations of a vigorous steam plume and the presence of lava at the summit were not confirmed until February 16, 2007.

Commonly, the winter weather in Kamchatka is clear and cold. Minimal to no cloud cover can be the norm for long stretches between December and May in this region, thereby making optical remote sensing an excellent tool. In the spring and summer months, low-level thicker clouds typically form after 9am and can persist the entire day. These clouds can hinder ground-based visual observations and make visits to the summits of the volcanoes difficult. However, the clouds typically do not extend high enough to obscure the summits of the taller volcanoes such as Kliuchevskoi. For these periods, it is not uncommon that optical remote sensing is the only form of monitoring possible. Higher altitude thin clouds, thin volcanic plumes, and jet contrails can also be problematic for optical remote sensing. These clouds are visually hard to detect, but can negatively-impact the extraction of accurate

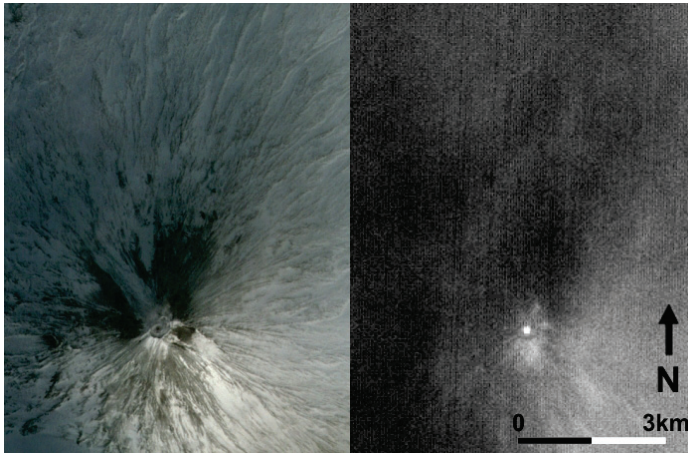


Fig. 8. ASTER data collected on January 4, 2007 and centered on the summit of Kliuchevskoi volcano. The VNIR color composite image (left) shows almost no signs of activity other than a small darker spot in the center of the summit crater (indicating snowmelt) and a minor amount of steam. The band 9 SWIR image (right) however shows a distinct thermal anomaly centered over the dark spot in the summit crater. The maximum integrated brightness temperature derived from the SWIR data was 332 °C indicating the presence of hot gases and/or magma very close to the summit.

surface composition and temperature. For example, this can be seen in the lower surface temperatures derived during periods of thin cloud cover for Kliuchevskoi (Figure 9). From mid-January through mid-April, ASTER continued to collect regularly scheduled and urgent request data of the volcanic activity at Kliuchevskoi. During this period, ground-based observations were hindered at times by the presence of low-level cloud cover. However, these clouds were commonly lower than the summit of Kliuchevskoi allowing the activity to be observed with ASTER and AVHRR. The summit activity continued to increase during this time, producing more thermally elevated pixels and raising the brightness temperature enough to saturate the ASTER SWIR data by early April (Figure 9). This was caused by the presence of a large amount of non-crusted lava in the summit crater from either a small actively overturning lava lake or very vigorous Strombolian eruption activity. On April 9, 2007 the clouds had diminished and photographs from the ground confirmed the Strombolian explosions and the presence of lava that was flowing down the northern slope at a location nearly identical to the 2005 flow (Rose and Ramsey, 2009).

ASTER captured another day/night pair of urgent request images on April 26-27, 2007 (Figure 9). By this time, the active lava flow had been present for over a month and the SWIR and TIR data were commonly saturated in many locations due to the high temperatures. These lava flows were producing numerous lahars that were emplaced 10-15 km further down the northern slope from the base of the lava flow. During this period, larger Vulcanian-style eruptions were also common at the summit crater. These short-duration explosions produced ash columns 5-15 km above the summit, which then commonly drifted S-SE over the peninsula and into the northern Pacific Ocean.

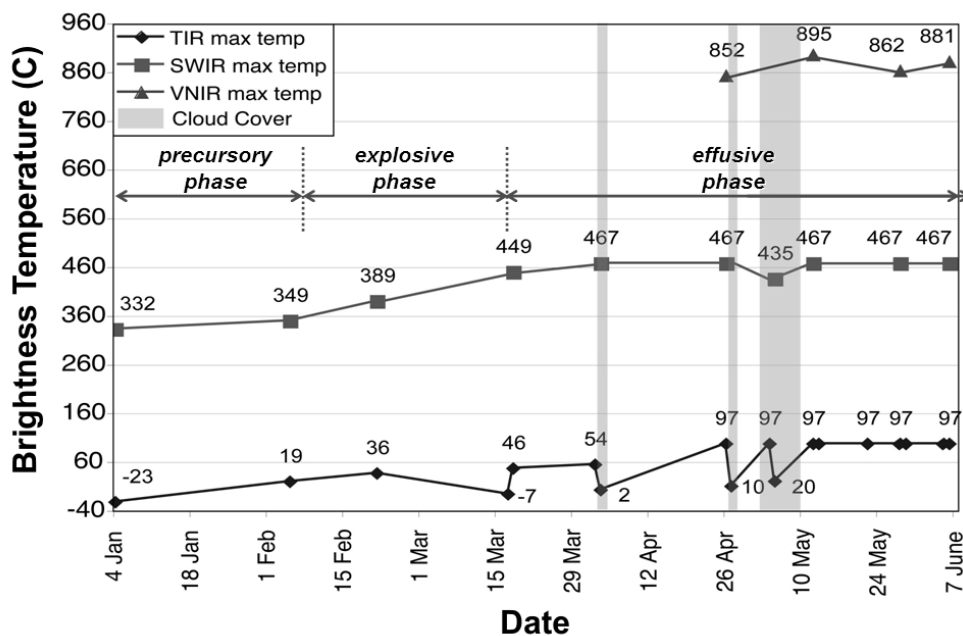


Fig. 9. Time-temperature plot of the 2007 eruption of Kliuchevskoi volcano showing the extracted pixel-integrated brightness temperatures for the three wavelength regions and the three phases of eruptive activity.

The plumes also produced proximal airfall deposits easily seen in contrast to the underlying snow (Figure 10). The April 26-27, 2007 ASTER image pair was unique in several other ways. The data were collected several hours before an AVHRR overpass, which had triggered the next ASTER urgent request (later acquired on May 4-5, 2007). However, because the AVHRR data are displayed on the AVO Web site (see Figure 5) within minutes to hours after collection, this image was seen before the ASTER data and initial descriptions of the activity on April 26 were based on the AVHRR data. The collection of both these datasets within hours of each other provided a rare opportunity to compare detailed ground data/features at the 15- to 90-m scale to what was imaged by the AVHRR instrument at 1-km spatial scale. AVHRR had 23 thermally-elevated pixels, five of which were saturated. The poor spatial resolution did not allow the discrimination of the two lava flows; however, the wide area of hot AVHRR pixels in conjunction with two recovery pixels at the far eastern edge of the thermally elevated area indicated that a new active lava flow was likely present, which was then later confirmed in the ASTER data. Recovery pixels are defined as areas with anomalously low temperatures that commonly occur on the down-scan margin of very high temperature thermal features (Higgins and Harris, 1997). This adverse response of the detectors to a zone of high radiance is produced by a slight delay during which time the previously saturated detector elements equilibrate and no data are collected.

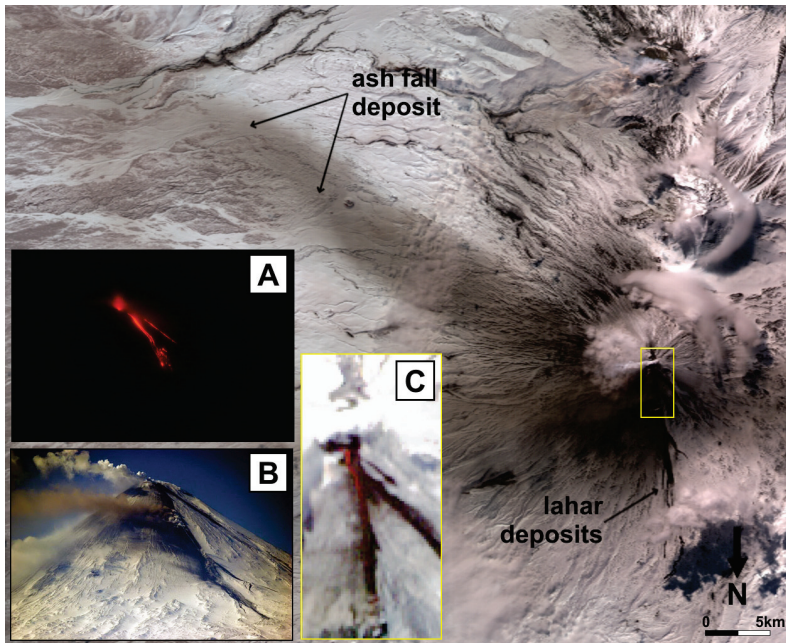


Fig. 10. Kliuchevskoi eruption captured April 24-26, 2007. The base figure, an ASTER VNIR image collected on April 26, 2007, shows a wider area around the volcano and is displayed with North down (opposite to the other ASTER images shown). This allows a similar view as the field photographs (**A** and **B**) taken from the town of Klyuchi by Y. Demyanchuk. In the ASTER image, an ash fall deposit to the SE, numerous lahar deposits to the N, and a vigorous steam plume at the summit are all visible. (**A**) Nighttime photograph of the summit taken on April 24, 2007 showing the Strombolian activity and two distinct lava flows. (**B**) Daytime photograph taken one day later (and one day before the ASTER image) showing both lava flows and the dendritic patterns of the lahar deposits all of which are clearly visible in the ASTER image. (**C**) Enhanced linear stretch of the ASTER VNIR data of the summit region (denoted by the yellow rectangle). The active incandescent lava flows are seen and the maximum brightness temperature extract from these data was 852 °C.

A cold plume extending to the NE in the AVHRR image (and not seen in the ASTER image) indicated the volcano was continuing to produce larger eruptions every few hours during the lava flow emplacement phase. These plumes were carried in different directions (SE in ASTER, NE in AVHRR) depending on the local wind at the time. Once the ASTER data were available, it was confirmed that lava was now being emplaced in a new direction and that the previous lava flow was beginning to cool. This new lava flow was active and large enough to be seen in the 15m VNIR data (Figure 10C). At that point the open channel was 15-30 m wide, 3 km long and had a pixel-integrated brightness temperature of 852 °C. To extract brightness temperatures in daytime VNIR or SWIR data, the solar reflected contribution in each pixel had to be removed. This was done by calculating that amount in each wavelength band using non-thermally-elevated pixels from a nearby region under similar lighting conditions (Rose and Ramsey, 2009). The flow was being emplaced to the

north, in the large volcano shadow because of the low sun angle at this latitude and this time of year. Therefore, the solar correction was minimal and the extracted brightness temperatures from the VNIR and SWIR data were much more accurate.

This dataset also marked the start of approximately two months of nearly cloud-free ASTER data where the locations of the lava flows and their VNIR-derived temperatures were tracked and used for detailed monitoring of the eruption (Figures 9-11). The extracted pixel-integrated brightness temperatures were divided into three phases of activity, which also correlated with visible observations and seismic data. In the precursory phase (November 2006 to February 2007), TIR temperatures began to rise above the average background temperature (approximately -40°C in January) and became hot enough to be detected by the SWIR data. This phase was dominated by fumarolic degassing and minor Strombolian activity at the summit. In the explosive phase (February 2007 to March 2007), degassing became more intense and Strombolian explosions at the summit were nearly constant. However, the amount of lava was not large enough to saturate either the TIR or SWIR data. Beginning in mid-March 2007 and continuing until mid-June 2007, the effusive phase was ongoing with the emplacement of three large basaltic andesite lava flows. Each of these flows contained an active center channel for long periods of time resulting in the saturation of both the TIR and SWIR data and allowing temperatures to be extracted from the 15-m VNIR data. These pixel-integrated temperatures are slightly lower than similarly derived temperatures for Hawaiian lava flows indicating that the actual temperature of the lavas were between 1050 and 1100°C (Ramsey and Wessels, 2007).

From mid-April to mid-June, Kliuchevskoi produced three new lava flows, small-scale Strombolian summit activity, and larger explosive eruptions resulting in ash plumes that extended to the east hundreds of kilometers. On May 29, 2007 the nighttime ASTER TIR image clearly showed the three N-NW lava flows and a weak TIR signal extending approximately 400 m to the SE of the summit crater. This linear thermal anomaly was predicted to be the start of new lava flow direction but was not confirmed until the next ASTER image pair on June 6-7, 2007 (Figure 11). The prediction of this new flow direction was initially discounted by most observers/scientists because of the high level of activity ongoing to the north and lack of historical lava flows emanating from the summit in this direction. The data collected on June 6, 2007, verified that the northern lava flows were no longer active. Their temperatures (between 10 - 20°C above the background temperature) had cooled well below the detection threshold for ASTER VNIR and SWIR but could still be discerned in the TIR image. The most obvious change was the presence of two new SE trending active lava flows in the exact direction as the linear thermal anomaly seen in the May 20, 2007, image. The larger flow was 3.1 km long and was emplaced at an average rate of 16 m/hr, which was nearly the same flow rate as the northern flows. However, the flow rate increased significantly near the end of the effusive phase of the 2007 eruption. The flow rate was calculated by examining changes in the small flow south of the larger SE flow (Figure 11B). ASTER collected a nighttime image 13 hours after the data shown in Figure 11 and the advancement of this flow was easily seen. Using the digital elevation model derived from the VNIR image (Figure 11A), the slope in this region was calculated and used to derive a flow rate of 26 m/hr. This flow rate continued for the rest of June, after which the

effusive phase ended. The lava flows continued to cool over the next several months and no new explosive activity was observed. The eruption was officially declared over in July 2007.

During the six month long 2007 eruption of Kliuchevskoi, ASTER provided unprecedented views of the summit activity. The eruption began in the winter, making ground or helicopter observations nearly impossible because of the harsh conditions. As the weather warmed, cloud cover further limited ground views of the summit. Thirty-two ASTER images were acquired during the day and night over this time period, which was an average of one scene every six days, although many images were day/night pairs collected 13 hours apart. Of this total, three day/night pairs (January 4-5, March 17-18, May 4-5) were collected as a result of the automatic triggering of the ASTER urgent request system. An additional seven image pairs were also acquired using the manual tasking of this urgent request system. The manual tasking was used during cloudy periods where AVHRR data were limited or to augment the volume of automatically tasked images. The ASTER URP program therefore produced a 63% increase in the data volume during the eruption of Kliuchevskoi, making it an invaluable tool for monitoring. The twelve additional ASTER images collected during this time period were part of the routine volcano monitoring performed by ASTER for all active volcanoes around the world (Pieri and Abrams, 2004).

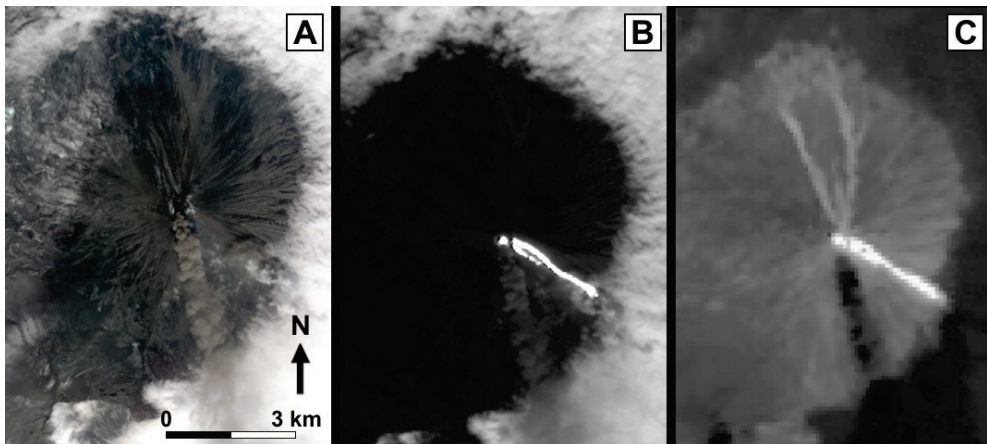


Fig. 11. ASTER urgent request data collected on June 6, 2007 over the summit region of Kliuchevskoi volcano. (A) VNIR image now showing the lack of snow on the upper flanks of the volcano and an ash-rich plume from the summit crater and extending southward. Note the low level cloud deck surrounding the volcano at ~2500 m. Such clouds are common starting in the spring and completely obscure ground observations. ASTER provided the only detailed record of eruptive activity during this period. (B) SWIR band 4 image covering the same area. The summit crater and SE trending lava flow are clearly visible, as is a secondary breakout flow to the south of the larger flow. Using the ASTER-derived DEM of the summit and the difference in the flow lengths between the day and nighttime images ($\Delta T = 13$ hours), the flow velocity was calculated to be 26 m/hour. This was significantly faster than the previous lava flows to the N and NW. (C) TIR band 10 image showing the active lava flow, the colder plume, and the two previous lava flows (N and NW) that were still cooling but no longer visible in the VNIR or SWIR data.

The numerous datasets provided by ASTER as part of the collaborative rapid-response program in conjunction with the relatively clear weather and the summit elevation and high latitude of Kliuchevskoi resulted in the largest and most comprehensive multispectral/multispacial high resolution dataset of a volcanic eruption. These data provide another means to monitor and characterize eruptive activity of this region of the globe. Precursors of eruption onset and new behavior can be better recognized using high resolution image data across the wavelength region and thus minimize future risks.

6. Conclusions

The value of using remote sensing assets in geoscience applications has been demonstrated by many authors and detailed here for the north Pacific volcano monitoring program. Numerous successes have been realized using this programmatic approach for capturing higher temporal frequency data, and a more rapid dissemination of critical information has been established for integration with future higher resolution datasets. The work described enabled the incorporation of higher spatial resolution ASTER data into an existing coarser resolution volcano monitoring initiative. New techniques were developed for event detection, satellite tasking, data distribution, and data analysis. These resulted in more rapid data availability, more detailed information, greater scientific insight on geologic processes, and more reliable alerts to the communities involved.

Specifically for Kamchatka, several beneficial factors have combined resulting in nearly 1,400 ASTER images of the five most thermally active Kamchatka volcanoes (Bezymianny, Karimsky, Kliuchevskoi, Sheveluch and Tolbachik). These factors include the orbital alignment of Terra, the high latitude of the peninsula, and the persistent activity in this region. From the inception of the automated rapid response program in 2004, an additional 350 scenes have been acquired over these volcanoes, many soon after larger eruptions. These data have produced valuable quantitative information on the small-scale activity and larger eruptions. A detailed example of the 2007 eruption of Kliuchevskoi described here enabled fundamental lava flow parameters to be determined. Numerous eruptions have been observed in Kamchatka by ASTER, which have displayed varying volcanic styles including basaltic lava flow emplacement, silicic lava dome growth, pyroclastic flow emplacement, volcanic ash plume production, fumarolic activity, and geothermal emission. The high spatial resolution and moderate spectral resolution of the data are ideal for deriving the energy flux from both high and low temperature systems, mapping chemical and textural changes of the volcanic products, and for imaging and understanding recent volcanic deposits.

The international collaboration developed for this work created professional relationships and infrastructure that will prove valuable in future work. The focus of further research will be on specific eruptions of the Kamchatka and Alaska volcanoes, the science results stemming from those data, expansion plans for global ASTER urgent request data and support for other types of events. The current ASTER rapid response program in Kamchatka and Alaska has produced a large archive of data, which has only been sampled to a small degree. These data offer a source for both the timely completion of current studies and new scientific analysis. It has also improved the timeliness and reliability of resulting hazard

notifications to responsible authorities and affected communities. However, it should be noted that the ASTER instrument has long exceeded its initial design life. The SWIR subsystem has now failed and the sensor could suffer further catastrophic losses at any time. Therefore, there is a critical need for a continuing series of sensors with similar characteristics to ASTER in order to support such geoscience applications.

7. Acknowledgements

The URP project described here was made possible by the efforts and support of the ASTER Science Team and through NASA funding to M. Ramsey (grants: NNG04GO69G and NNX08AJ91G). The authors would like to thank S. Rose for her detailed research on the Kliuchevskoi ASTER data and helping to create Figure 9. The authors greatly appreciate technical review comments received from Zhong Lu and Russell Rykhus, and editorial assistance from Tom Adamson and Aleksandar Lazinica. ASTER data are courtesy of NASA, GSFC, Japan's Ministry of Economy and Industry, ERSDAC, Japan Resources Observation System and Space Utilization Organization (JAROS), the U.S./Japan ASTER Science Team, and LP DAAC at the USGS EROS Center.

8. References

- Abrams, M. (2000). The Advanced Spaceborne Thermal Emission and Reflectance Radiometer (ASTER): data products for the high spatial resolution imager on NASA's Terra platform. *International Journal of Remote Sensing*, 21, 847–859.
- Carter, A.J., Ramsey, M.S., and Belousov, A.B. (2007). Recent crater formation at Bezymianny Volcano lava dome: Significant changes observed in satellite and field data, *Bull. Volc.*, doi: 10.1007/s00445-007-0113-x.
- Carter, A.J., Girina, O., Ramsey, M.S., and Demyanchuk, Y.V. (2008). ASTER and field observations of the 24 December 2006 eruption of Bezymianny Volcano, Russia, *Rem. Sens. Environ.*, 112, 2569–2577, doi: 10.1016/j.rse.2007.12.001.
- Chebrov, V. (2008). Complex seismological and geophysical investigation of Kamchatka and Commander Islands. *Annual report of KB GS RAS*. Ed. V. Chebrov. pp. 268. (in Russian).
- Dean, K., Servilla, M., Roach, A., Foster, B., and Engle, K. (1998). Satellite monitoring of remote volcanoes improves study efforts in Alaska, *EOS, Trans. Amer. Geophys. Union*, 79, 413, 422–423.
- Dehn, J., Dean, K.G., and Engle, K. (2000). Thermal monitoring of north Pacific volcanoes from space, *Geology*, 28, 755–758.
- Dehn, J., Harris, A.J.L. (in press). Thermal Anomalies at Volcanoes. in: *Monitoring Volcanoes in the North Pacific: Observations from Space*, Dean K.G. and Dehn J. editors, Praxis/Springer.
- Fournier, T., J. Freymueller, and P. Cervelli (2009). Tracking magma volume recovery at Okmok volcano using GPS and an unscented Kalman filter, *J. Geophys. Res.*, 114, B02405, doi:10.1029/2008JB005837.
- Harris, A.J.L., Flynn, L.P., Keszthelyi, L., Mouginis-Mark, P.J., Rowland, S.K., and Resing, J.A. (1998). Calculation of lava effusion rates from Landsat TM data, *Bull. Volc.*, 60, 52–71.

- Higgins, J. and Harris, A. (1997). VAST: A program to locate and analyse volcanic thermal anomalies automatically from remotely sensed data, *Computers & Geosci.*, v. 23. n. 6, pp. 621-645.
- Kaab, A. (2005). *Remote sensing of mountain glaciers and permafrost creep*, Physical Geography Series, Zurich, 48, 266 pages, ISBN 3 85543 244 9.
- Lu, Z., D. Dzurisin, C. Wicks, J. Power, O. Kwoun, and R. Rykhus (2007). Diverse deformation patterns of Aleutian volcanoes from satellite interferometric synthetic aperture radar (InSAR), in *Volcanism and Subduction: The Kamchatka Region* (edited by J. Eichelberger et al.), American Geophysical Union Geophysical Monograph Series 172, 249-261.
- McNutt, S.R. (1994). Volcanic tremor amplitude correlated with the Volcanic Explosivity Index and its potential use in determining ash hazards to aviation. *Acta Vulcanol.* 5. pp193-196.
- Miller, T.P., Kirianov, V.Y., Kelley, H.L. (1994). Klyuchevskoy Fact Sheet. *U.S. Geological Survey Fact Sheet. 94-067*, pp. 4. Also online (<http://eq.giseis.alaska.edu/volcanoes/klyu/klyufact.html>).
- Miller, T.P., and Casadevall, T.J. (2000). Volcanic Ash Hazards to Aviation. In: Sigurdsson, H., Houghton, B., McNutt, S.R., Rymer, H. and Stix, J., Editors, 2000. *Encyclopedia of Volcanoes*, Academic Press, San Diego, CA, pp. 915-930.
- Pieri, D.C., and Abrams, M.J. (2004). ASTER watches the world's volcanoes: a new paradigm for volcanological observations from orbit. *Journal of Volcanology and Geothermal Research.* 135: 13-28.
- Ramsey, M.S. and Dehn, J. (2004). Spaceborne observations of the 2000 Bezymianny, Kamchatka eruption: The integration of high-resolution ASTER data into near real-time monitoring using AVHRR, *J. Volc. Geotherm. Res.*, 135, issue 1-2, 127-146.
- Ramsey, M.S., Dehn, J., Wessels, R., Byrnes, J., Duda, K., Maldonado, L., and Dwyer, J. (2004). The ASTER emergency scheduling system: A new project linking near-real-time satellite monitoring of disasters to the acquisition of high-resolution remote sensing data, *Eos Trans. AGU*, 85(47), Fall Meet. Suppl., Abstract SF23A-0026.
- Ramsey, M.S., and Wessels, R.L. (2007). Monitoring changing eruption styles of Kilauea Volcano over the summer of 2007 with spaceborne infrared data, *Eos Trans. AGU*, 88(52): Fall Meet. Suppl., Abstract V51H-07.
- Ramsey, M.S., Anderson, S., and Wessels, R. (2008). Active dome and pyroclastic flow deposits of Sheveluch Volcano, Kamchatka: Unique thermal infrared and morphologic field observations, *IAVCEI, General Assem. Prog.*, p. 67.
- Roach, A.L., Benoit, J.P., Dean, K.G., McNutt, S.R. (2004). The combined use of satellite and seismic monitoring during the 1996 eruption of Pavlof volcano, *Alaska. Bulletin of Volcanology.* 62 (6-7): 385-399.
- Rose, S.R. and Ramsey, M.S. (2009). The 2005 eruption of Kliuchevskoi volcano: Chronology and processes derived from ASTER spaceborne and field-based data, *J. Volc. Geotherm. Res.*, doi:10.1016/j.jvolgeores.2009.05.001.
- Schneider, D., Dean, K.G., Dehn, J., Miller, T.P., and Kirianov, V. Yu. (2000). Monitoring and analyses of volcanic activity using remote sensing at the Alaska Volcano Observatory: case study for Kamchatka, Russia, December 1997. *Remote Sensing of Active Volcanism*, Geophysical Monograph 116, 65-85.

- Stephens, C. D., and Chouet, B. A. (2001). Evolution of the December 14, 1989 precursory long-period event swarm at Redoubt Volcano, Alaska: *Journal of Volcanology and Geothermal Research*, v. 109, n. 1, p. 133-148.
- Wessels, R.L., Schneider, D.J., Coombs, M.L., Dehn, J., and Ramsey, M.S. (in press). High-resolution Satellite and Airborne Thermal Infrared Imaging of the 2006 Eruption of Augustine Volcano, Alaska, in Power, J.A., Coombs, M.L., and Freymueller, J.T., eds., *Investigations of Augustine Volcano, Alaska after the 2006 eruption*, U.S. Geological Survey, Professional Paper.
- Wright, R., Flynn, L., Garbeil, H., Harris, A., and Pilger, E. (2002). Automated volcanic eruption detection using MODIS, *Rem. Sens. Environ.*, 82, 135-155.
- Yamaguchi, Y., Kahle, A. B., Tsu, H., Kawakami, T., & Pniel, M. (1998). Overview of Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). *IEEE Transactions on Geoscience and Remote Sensing*, 36, 1062-1071.
- Yang, F. and Schlesinger, M.E. (2002). On the surface and atmospheric temperature changes following the 1991 Pinatubo volcanic eruption; a GCM study, *J. Geophys. Research*, 107; 7-8.

The Extended Integral Equation Model IEM2M for topographically modulated rough surfaces

Jose Luis Alvarez-Perez

*Department of Signal Theory and Communications, Universidad de Alcala (UAH)
Spain*

1. Introduction

Remote sensing of terrain and ocean surfaces is circumscribed in the physical domain of electromagnetic scattering by rough surfaces. The development of accurate models has gathered a great deal of efforts since the 80's. Until that moment there were two classical approaches to be applied to two different asymptotic cases: the surfaces with small roughness and those having long correlation length. The first situation was dealt successfully via the small perturbation method (SPM) whereas the second one was the target of the Kirchhoff approximation (KA). In effect, the abundance of models in the last two decades has made it very difficult for the Earth Observation practitioner to properly classify them and choose between them. The most important effort to that purpose was made by Tanos Elfouhaily in Elfouhaily & Guerin (2004), and we refer to his work for those interested in having a comprehensive account of the available methods for the problem. We focus here on the model that has arguably awakened the largest share of interest within the remote sensing community, that is, the Integral Equation Model (IEM) presented by Fung and Pan in Fung & Pan (1986) and later corrected in a long series of amendments by the same authors Fung (1994); Hsieh et al. (1997); Chen et al. (2000); Fung et al. (2002); Chen et al. (2003); Fung & Chen (2004); Wu & Chen (2004); Wu et al. (2008). In effect, there has been a number of issues that made the model theoretically inconsistent, even if each amendment was accompanied by properly suiting numerically simulated results. In 2001 the author of this chapter carried out a complete revision of Fung's work and proposed a corrected IEM that successfully achieved one of the objectives of the rough surface scattering models developed so far: to unify in a single equation both the SPM and the KA in the most general case of bistatic scattering. This corrected IEM was named IEM with proper inclusion of multiple scattering at second order or IEM2M.

This chapter aim is twofold: on the one had a quick summary of the IEM2M is given and on the other an extension of it is proposed to include those surfaces comprising both a zero-mean height, random component and a deterministic component that we call here "topographical".

2. Summary of the IEM2M for surfaces with zero height mean

The rationale of the IEM and therefore of the IEM2M is to perform a second iteration in the integral equations describing the rough surface electromagnetic scattering problem, as given in Poggio and Miller Poggio & Miller (1973). The first iteration corresponds to the KA, where each point on the surface is locally surrounded by neighbouring points lying on a flat surface, which is equivalent to the assumption of a low curvature. As a matter of fact, the proper in-

clusion of this second or complementary term coming from a second iteration bridges the gap between SPM and KA since it includes the local effects due to these neighbouring points to the extent which is necessary to meet the SPM limit. Second order effects describe the interaction of points on the surface, considered in pairs, just like third order effects would include interactions among sets of points taken in triads. This second-order contribution happens to contribute to the first-order, KA term with a non-zero addend when the limit of two points approaching to each other is taken. Even if full detail of IEM2M is given in Alvarez-Perez (2001), we summarize here the results regarding the complete first-order model that includes the KA term plus aforementioned correction coming from the limit of the second-order where pairs of point approach to one another. Unlike in Alvarez-Perez (2001), this first-order IEM2M is spelled out in a completely explicit form that eases its direct implementation in a computer code. Thus, we have for the first-order scattering coefficient the following formula, which contains new terms over the KA owing to the limit phenomena explained above

$$\sigma_{qp}^o = \frac{1}{2} k_1^2 e^{-\sigma^2(k_{sz}-k_z)^2} \times \sum_{n=1}^{\infty} \frac{\sigma^{2n}}{n!} \left| I_{qp}^{(n)} \right|^2 W_1^{(n)}(k_{sx} - k_x, k_{sy} - k_y) \tag{1}$$

where

$$I_{qp}^{(n)} = (k_{sz} - k_z)^n f_{qp} + \frac{1}{4} [i_1 + i_2 + i_{1'} + i_{2'} + i_{3'} + i_{4'}] \tag{2}$$

with

$$\begin{aligned} i_1 &= (k_{sz} + k_z)^{n-1} F_{qp}^1(k_x, k_y, -k_z) e^{-\sigma^2(k_{sz}+k_z)^2} \\ i_2 &= [-(k_{sz} + k_z)]^{n-1} F_{qp}^1(k_{sx}, k_{sy}, -k_{sz}) e^{-\sigma^2(k_{sz}+k_z)^2} \\ i_{1'} &= (k_{sz} - k_z^{(2)})^{n-1} F_{qp}^2(k_x, k_y, k_z^{(2)}) \\ &\quad \times e^{-\sigma^2[k_z^{(2)2} - (k_{sz}+k_z)k_z^{(2)}]} e^{-\sigma^2k_{sz}k_z} \\ i_{2'} &= (k_{sz} + k_z^{(2)})^{n-1} F_{qp}^2(k_x, k_y, -k_z^{(2)}) \\ &\quad \times e^{-\sigma^2[k_z^{(2)2} + (k_{sz}+k_z)k_z^{(2)}]} e^{-\sigma^2k_{sz}k_z} \\ i_{3'} &= (k_{sz}^{(2)} - k_z)^{n-1} F_{qp}^2(k_{sx}, k_{sy}, k_{sz}^{(2)}) \\ &\quad \times e^{-\sigma^2[k_{sz}^{(2)2} - (k_{sz}+k_z)k_{sz}^{(2)}]} e^{-\sigma^2k_{sz}k_z} \\ i_{4'} &= [-(k_{sz}^{(2)} + k_z)]^{n-1} F_{qp}^2(k_{sx}, k_{sy}, -k_{sz}^{(2)}) \\ &\quad \times e^{-\sigma^2[k_{sz}^{(2)2} + (k_{sz}+k_z)k_{sz}^{(2)}]} e^{-\sigma^2k_{sz}k_z} \end{aligned} \tag{3}$$

and

$$W_1^{(n)}(k_{sx} - k_x, k_{sy} - k_y) = \frac{1}{2\pi} \int d\xi d\eta \rho^n(\xi, \eta) e^{-j[(k_{sx}-k_x)\xi + (k_{sy}-k_y)\eta]} \tag{4}$$

$$\begin{aligned} k_z^{(2)} &= (k_2^2 - k_x^2 - k_y^2)^{1/2} \\ k_{sz}^{(2)} &= (k_2^2 - k_{sx}^2 - k_{sy}^2)^{1/2} \end{aligned} \tag{5}$$

The symbols in equation (1) are: $\vec{k}^i = (k_x, k_y, k_z)$ represents the incident wave vector upon the scattering surface, $\vec{k}^s = (k_{sx}, k_{sy}, k_{sz})$ is the scattering wave vector, k_1 is the wave number of the incident medium (above the surface), k_2 is the wave number of the scattering medium (below the surface), σ is the standard deviation of the surface height and ρ is the correlation function of the surface height. The F_{qp} coefficients are given in Alvarez-Perez (2001). They, in turn, depend on some coefficients named as $C_i(\vec{k}^i, \vec{k}^s, \vec{l}_m^{(r)})$; $i = 1, \dots, 4$, where $\vec{l}_m^{(r)}$ represents the effective interaction vector of a second-order scattering event, with r representing its upwards (+1) or downwards (-1) character and m the medium through which the second-order interaction takes place. For the first-order reduction IEM2M this vector $\vec{l}_m^{(r)}$ reduces to a few possible values, as explained in Alvarez-Perez (2001). These C coefficients are provided in Alvarez-Perez (2001) in a very formal way that may pose a difficulty for those not familiar with surface geometry. Therefore, a more user-friendly version is given in Appendix A at the end of this chapter. Also some remarks on its implementation by other authors are given.

3. IE2M Scattering Coefficient for Topographical Surfaces

3.1 Average Coherent Power

The average coherent power density over an ensemble of statistically equivalent surfaces is the modulus of the Poynting vector for the coherently scattered field

$$S_{qp}^c = \frac{1}{2} \text{Re}\{1/\eta_1\} \langle \vec{E}_{qp}^s \rangle \langle \vec{E}_{qp}^{s*} \rangle \tag{6}$$

where η_1 is the impedance of the incident medium. It is common to assume far-zone fields to have a plane wave front. Although this is a valid approximation for incoherent scattering, it is now more convenient to replace the usual approximation

$$\frac{e^{jk_1|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} \simeq \frac{e^{jk_1r}}{r} e^{-jk_1\hat{r}\cdot\vec{r}'} \tag{7}$$

by

$$\frac{e^{jk_1|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} \simeq \frac{e^{jk_1r}}{r} e^{-jk_1\hat{r}\cdot\vec{r}'} e^{j\frac{r'^2}{2r}} \tag{8}$$

in the derivation of the Stratton-Chu-Silver integral. The reason to include the second order term in r'^2 in the phase of the spherical wave function is the higher sensitivity of a coherent interference to the wave front shape. Likewise, it is appropriate to assume a spherical incident front from the source of the incident field

$$\frac{e^{jk_1|\vec{r}_s-\vec{r}'|}}{|\vec{r}_s-\vec{r}'|} \simeq \frac{e^{jk_1r_s}}{r_s} e^{-jk_1\hat{r}_s\cdot\vec{r}'} e^{j\frac{r_s'^2}{2r_s}} \tag{9}$$

where \vec{r}_s is the position vector of the source. We will assume that the incident field is Gaussian modulated along the direction given by \vec{r}_s , according to the window

$$w_G(x, y) = e^{-g_0^2(x^2 \cos^2 \theta + y^2)}$$

$$g_0 = \frac{1}{r_s \beta_0} \tag{10}$$

where β_0 is the one-sided beamwidth of the transmitter. By placing the origin of coordinates on the plane to which the average rough surface belongs but far from the illuminated area, the following approximation can be made both in (8) and (9)

$$r'^2 = x'^2 + y'^2 + h'^2(x', y') \simeq x'^2 + y'^2 \quad (11)$$

With the inclusion of these changes plus the introduction of a shadowing function (see next section) and assuming $r_s = r$, the Kirchoff far-zone scattered field can be written as

$$(E_{qp}^s)_k = \frac{jk_1 E_0}{4\pi} \frac{e^{jk_1 r}}{r^2} \int_S \hat{f}_{qp} e^{jk_1 \frac{(x'^2 + y'^2)}{2r}} e^{-g_0^2(x'^2 \cos^2 \theta + y'^2)} e^{-j[(\vec{k}^s - \vec{k}^i) \cdot \vec{r}']} dx' dy' \quad (12)$$

where we have “dressed” the factor f_{qp} to include the shadowing function

$$\hat{f}_{qp} = \mathcal{S}(\hat{k}^i, \hat{k}^s) f_{qp} \quad (13)$$

Then, the coherently scattered power takes the form

$$S_{qp}^c = \frac{1}{2} \text{Re}\{1/\eta_1\} \left(\frac{k_1 E_0 \hat{f}_{qp}}{4\pi r^2} \right)^2 \left| \int_S e^{jk_1(x'^2 + y'^2)/2r} e^{-g_0^2(x'^2 \cos^2 \theta + y'^2)} e^{-j[(k_{sx} - k_x)x' + (k_{sy} - k_y)y']} \langle e^{-j[(k_{sz} - k_z)z']} \rangle dx' dy' \right|^2 \quad (14)$$

To calculate the averages comprised in the integrand of (14), we compute

$$\langle e^{-j(k_{sz} - k_z)z'} \rangle = e^{-j(k_{sz} - k_z)\bar{z}(x', y')} e^{-(k_{sz} - k_z)^2(\sigma^2/2)} \quad (15)$$

Hence,

$$S_{qp}^c = \frac{1}{2} \text{Re}\{1/\eta_1\} \left(\frac{k_1 E_0 \hat{f}_{qp}}{\pi r^2} \right)^2 e^{-(k_{sz} - k_z)^2 \sigma^2} |W_0(k_{sx} - k_x, k_{sy} - k_y)|^2 \quad (16)$$

where

$$W_0(k_{sx} - k_x, k_{sy} - k_y) = \int e^{-j2[(k_{sx} - k_x)x' + (k_{sy} - k_y)y']} e^{x'^2(jk_1/2r - g_0^2 \cos^2 \theta) + y'^2(jk_1/2r - g_0^2)} e^{-j(k_{sz} - k_z)\bar{z}(x', y')} dx' dy' \quad (17)$$

Integral W_0 has the shape of a Gabor transform, that is, of a Fourier transform with a Gaussian window included in the integrand.

3.2 Average Incoherent Power

The average incoherent power density over an ensemble of statistically equivalent surfaces is the modulus of the Poynting vector for the diffuse field

$$S_{qp}^d = \frac{1}{2} \text{Re}\{1/\eta\} \left(\langle \vec{E}_{qp}^s \vec{E}_{qp}^{s*} \rangle - \langle \vec{E}_{qp}^s \rangle \langle \vec{E}_{qp}^{s*} \rangle \right) \quad (18)$$

where $\text{Re}\{1/\eta_1\}$ is the real part of the inverse of the magnetic permeability in the incidence medium and $*$ is the symbol for complex conjugate. Separating the scattered field into the Kirchhoff and complementary terms, we obtain

$$\begin{aligned} S_{qp}^d = & \frac{1}{2} \text{Re}\{1/\eta\} \left\{ \langle E_{qp}^{sk} E_{qp}^{sk*} \rangle - \langle E_{qp}^{sk} \rangle \langle E_{qp}^{sk*} \rangle \right. \\ & + 2 \text{Re}\{ \langle E_{qp}^{sc} E_{qp}^{sc*} \rangle - \langle E_{qp}^{sc} \rangle \langle E_{qp}^{sc*} \rangle \} \\ & \left. + \langle E_{qp}^{sc} E_{qp}^{sc*} \rangle - \langle E_{qp}^{sc} \rangle \langle E_{qp}^{sc*} \rangle \right\} \quad (19) \end{aligned}$$

The analysis of (19) will be carried out by considering separately three terms, namely, the Kirchhoff term, the complementary term and the “interference” term between both, which will be named the cross term.

To perform the averages in (19), we need to know the statistics of the ensemble of surfaces. We select the ensemble of surfaces such that it follows a joint Gaussian distribution with a constant variance across the surface. This assumption greatly simplifies the computation of the averaging. However, the random surfaces included in the aforementioned ensemble will be allowed to have nonzero means at each point.

3.2.1 Kirchhoff Incoherent Power

Once the shadowing effects are included, the Kirchhoff diffuse power density can be written as

$$\begin{aligned} S_{qp}^{dk} = & \frac{1}{2} \text{Re}\{1/\eta_1\} \left\{ \langle E_{qp}^{sk} E_{qp}^{sk*} \rangle - \langle E_{qp}^{sk} \rangle \langle E_{qp}^{sk*} \rangle \right\} \\ = & \frac{|K E_o \hat{f}_{qp}|^2}{2} \text{Re}\{1/\eta_1\} \left(\left\langle \int_S e^{-j(\hat{k}_s - \hat{k}_i) \cdot (\vec{r}' - \vec{r}'')} dx' dy' dx'' dy'' \right\rangle \right. \\ & \left. - \left| \left\langle \int_S e^{-j(\hat{k}_s - \hat{k}_i) \cdot \vec{r}'} dx' dy' \right\rangle \right|^2 \right) \quad (20) \end{aligned}$$

The averages in (20) are readily evaluated

$$\langle e^{-jp_z z'} \rangle = e^{-jp_z \bar{z}(x', y')} e^{-p_z^2 (\sigma^2/2)} \quad (21a)$$

$$\langle e^{-jp_z (z' - z'')} \rangle = e^{-jp_z (\bar{z}(x', y') - \bar{z}(x'', y''))} e^{-p_z^2 \sigma^2 [1 - \rho(x' - x'', y' - y'')]} \quad (21b)$$

$$p_z = k_{sz} - k_z$$

Substituting now (21a) and (21b) into (20) and using the integration variables $\xi = x' - x''$ and $\eta = y' - y''$ instead of x' and y'' , we have

$$S_{qp}^{dk} = \frac{|K E_o \hat{f}_{qp}|^2}{2} \text{Re}\{1/\eta_1\} e^{-p_z^2 \sigma^2} \iint d\xi d\eta (e^{p_z^2 \sigma^2 \rho(\xi, \eta)} - 1) D_1(\xi, \eta; p_z) e^{-j(p_x \xi + p_y \eta)} \quad (22)$$

where $p_x = k_{sx} - k_x$, $p_y = k_{sy} - k_y$ and $D_1(\xi, \eta; p_z)$ is

$$D_1(\xi, \eta; p_z) = \iint dx'' dy'' e^{-jp_z[\bar{z}(x'' + \xi, y'' + \eta) - \bar{z}(x'', y'')]} \quad (23)$$

and represents the autocorrelation of the phase $e^{-jp_z \bar{z}(x'', y'')}$ over the surface.

3.2.2 Cross Incoherent Power

The incoherently scattered power for the cross term is given by

$$\begin{aligned} S_{qp}^{dkc} &= \text{Re}\{1/\eta_1\} \text{Re}\left\{\langle E_{qp}^{sc} E_{qp}^{sk*} \rangle - \langle E_{qp}^{sc} \rangle \langle E_{qp}^{sk*} \rangle\right\} \\ &= \frac{|KE_0|^2}{8\pi^2} \text{Re}\{1/\eta_1\} \sum_{m=1,2} \text{Re}\left\{\hat{f}_{qp}^* \int_{\mathbb{R}^2} du dv \int_{S^3} dx' dy' dx'' dy'' dx''' dy''' \right. \\ &\quad e^{j[u(x' - x'') + v(y' - y'')] } e^{-j[k_{sx}(x' - x''') + k_{sy}(y' - y''')] } e^{j[k_x(x'' - x''') + k_y(y'' - y''')] } \\ &\quad \cdot \left[\langle e^{-jk_{sz}(z' - z''')} e^{jk_z(z'' - z''')} e^{jq_m|z' - z''}| \hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, \vec{g}_m) \rangle \right. \\ &\quad - \langle e^{-jk_{sz}z'} e^{jk_z z''} e^{jq_m|z' - z''}| \hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, \vec{g}_m) \rangle \\ &\quad \left. \left. \cdot \langle e^{j(k_{sz} - k_z)z''} \rangle \right] \right\} \end{aligned} \quad (24)$$

where factors F_{qp}^m have been “dressed” to include the shadowing function

$$\hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, \vec{g}_m) = \mathcal{S}_m(\vec{k}^i, \vec{g}_m, \vec{k}^s) F_{qp}^m(\vec{k}^i, \vec{k}^s, \vec{g}_m) \quad (25)$$

On the other hand, factors \hat{F}_{qp}^m have been included within the averages since they depend on $(z' - z'')/|z' - z''|$. To compute these averages we will make use of the invariance of the formalism under the change

$$G_m(\vec{r}', \vec{r}'') = G_m^{\text{retarded}}(\vec{r}', \vec{r}'') \longrightarrow G_m^*(\vec{r}', \vec{r}'') = G_m^{\text{advanced}}(\vec{r}', \vec{r}'') \quad (26)$$

The Weyl representation of the retarded Green's function is given by

$$\begin{aligned} G_m^{\text{retarded}}(\vec{r}', \vec{r}'') &= \frac{j}{2\pi} \iint_{\mathbb{R}^2} e^{j[u(x' - x'') + v(y' - y'')] } \frac{e^{-jq_m|z' - z''|}}{q_m} du dv \\ q_m &= \begin{cases} (k_m^2 - u^2 - v^2)^{1/2} & \text{if } k_m^2 \geq u^2 + v^2 \\ -j(u^2 + v^2 - k_m^2)^{1/2} & \text{if } k_m^2 \leq u^2 + v^2 \end{cases} \end{aligned} \quad (27)$$

Therefore, the invariance under the change (26) is equivalent to

$$q_m \longrightarrow \begin{cases} -q_m & \text{if } q_m \in \mathbb{R} \\ q_m & \text{if } q_m \in \mathbb{I} \end{cases} \quad (28)$$

or, more formally, $q_m \rightarrow -q_m^*$. However, the damped cylindrical waves given by imaginary values of q_m have been neglected and therefore the invariance holds under the transformation

$$q_m \rightarrow -q_m$$

This symmetry permits the calculation of (24) by using

$$\langle \psi(q_m) \rangle = \frac{1}{2} \left(\langle \psi(q_m) \rangle + \langle \psi(-q_m) \rangle \right)$$

where ψ is any of the functions in (24) to be averaged. Thus, there are two averages to be computed, namely,

$$\begin{aligned} & \langle e^{-jk_{sz}(z'-z''')} e^{jk_z(z''-z''')} \hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, \vec{g}_m) e^{jq_m|z'-z''|} \rangle \\ &= \langle e^{-jk_{sz}(z'-z''')} e^{jk_z(z''-z''')} \frac{1}{2} \left[\hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, u, v, \Phi_{z'z''} q_m) e^{jq_m|z'-z''|} \right. \\ & \quad \left. + \hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, u, v, -\Phi_{z'z''} q_m) e^{-jq_m|z'-z''|} \right] \rangle \end{aligned} \tag{29}$$

and

$$\begin{aligned} & \langle e^{-jk_{sz}z'} e^{jk_zz''} \hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, \vec{g}_m) e^{jq_m|z'-z''|} \rangle \\ &= \langle e^{-jk_{sz}z'} e^{jk_zz''} \frac{1}{2} \left[\hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, u, v, \Phi_{z'z''} q_m) e^{jq_m|z'-z''|} \right. \\ & \quad \left. + \hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, u, v, -\Phi_{z'z''} q_m) e^{-jq_m|z'-z''|} \right] \rangle \end{aligned} \tag{30}$$

There are two types of addends in these averages: terms dependent on $\Phi_{z'z''} q_m$ and terms dependent on q_m^2 or completely independent of q_m . Only the former are functions of the space coordinates through $\Phi_{z'z''}$. Therefore, we have to compute the following quantities

$$\begin{aligned} & \langle e^{-j[k_{sz}(z'-z''')-k_z(z''-z''')]} e^{jq_m|z'-z''|} \rangle \\ &= \langle e^{-j[k_{sz}(z'-z''')-k_z(z''-z''')]} \cos(q_m|z'-z''|) \rangle \\ &= \frac{1}{2} \left(\langle e^{-j[k_{sz}(z'-z''')-k_z(z''-z''')]} e^{jq_m(z'-z'')} \rangle \right. \\ & \quad \left. + \langle e^{-j[k_{sz}(z'-z''')-k_z(z''-z''')]} e^{-jq_m(z'-z'')} \rangle \right) \end{aligned} \tag{31a}$$

and similarly

$$\begin{aligned} \langle e^{-j(k_{sz}z'-k_zz'')} e^{jq_m|z'-z''|} \rangle &= \frac{1}{2} \left(\langle e^{-j(k_{sz}z'-k_zz'')} e^{jq_m(z'-z'')} \rangle \right. \\ & \quad \left. + \langle e^{-j(k_{sz}z'-k_zz'')} e^{-jq_m(z'-z'')} \rangle \right) \end{aligned} \tag{31b}$$

$$\begin{aligned} & \langle e^{-j[k_{sz}(z'-z''')-k_z(z''-z''')]} e^{jq_m|z'-z''|} \Phi_{z'z''} q_m \rangle \\ &= \frac{q_m}{2} \left(\langle e^{-j[k_{sz}(z'-z''')-k_z(z''-z''')]} e^{jq_m(z'-z'')} \rangle \right. \\ & \quad \left. - \langle e^{-j[k_{sz}(z'-z''')-k_z(z''-z''')]} e^{-jq_m(z'-z'')} \rangle \right) \end{aligned} \tag{31c}$$

$$\begin{aligned} \langle e^{-j(k_{sz}z' - k_z z'')} e^{jq_m|z' - z''|} \Phi_{z'/z''} q_m \rangle &= \langle e^{-j(k_{sz}z' - k_z z'')} \Phi_{z'/z''} jq_m \sin(q_m|z' - z''|) \rangle \\ &\quad - \langle e^{-j(k_{sz}z' - k_z z'')} e^{-jq_m(z' - z'')} \rangle \end{aligned} \quad (31d)$$

Hence, we compute again the averages

$$\langle e^{-j[k_{sz}(z' - z''') - k_z(z'' - z''')]} e^{jq_m|z' - z''|} \rangle = \frac{1}{2} \left(e^{jw_1} e^{-\sigma_{w_1}^2} + e^{jw_2} e^{-\sigma_{w_2}^2} \right) \quad (32a)$$

$$\langle e^{-j(k_{sz}z' - k_z z'')} e^{jq_m|z' - z''|} \rangle = \frac{1}{2} \left(e^{jw_3} e^{-\sigma_{w_3}^2} + e^{jw_4} e^{-\sigma_{w_4}^2} \right) \quad (32b)$$

$$\langle e^{-j[k_{sz}(z' - z''') - k_z(z'' - z''')]} e^{jq_m|z' - z''|} \Phi_{z'/z''} q_m \rangle = \frac{q_m}{2} \left(e^{jw_1} e^{-\sigma_{w_1}^2} - e^{jw_2} e^{-\sigma_{w_2}^2} \right) \quad (32c)$$

$$\langle e^{-j(k_{sz}z' - k_z z'')} e^{jq_m|z' - z''|} \Phi_{z'/z''} q_m \rangle = \frac{q_m}{2} \left(e^{jw_3} e^{-\sigma_{w_3}^2} - e^{jw_4} e^{-\sigma_{w_4}^2} \right) \quad (32d)$$

where

$$\begin{aligned} w_1 &= \omega_1(k_{sz}, k_z, q_m) \\ w_2 &= \omega_1(k_{sz}, k_z, -q_m) \\ w_3 &= \omega_2(k_{sz}, k_z, q_m) \\ w_4 &= \omega_2(k_{sz}, k_z, -q_m) \\ \omega_1(k_{sz}, k_z, q_m) &= -(k_{sz} - q_m)\bar{z}' + (k_z - q_m)\bar{z}'' + (k_{sz} - k_z)\bar{z}''' \\ \omega_2(k_{sz}, k_z, q_m) &= -(k_{sz} - q_m)\bar{z}' + (k_z - q_m)\bar{z}'' \end{aligned} \quad (33)$$

and

$$\begin{aligned} \sigma_{w_1} &= \sigma_{\omega_1}(k_{sz}, k_z, q_m) \\ \sigma_{w_2} &= \sigma_{\omega_1}(k_{sz}, k_z, -q_m) \\ \sigma_{w_3} &= \sigma_{\omega_2}(k_{sz}, k_z, q_m) \\ \sigma_{w_4} &= \sigma_{\omega_2}(k_{sz}, k_z, -q_m) \\ \sigma_{\omega_1}(k_{sz}, k_z, q_m) &= \sigma[k_{sz}^2 + k_z^2 + q_m^2 - (k_{sz} + k_z)q_m - k_z k_{sz} \\ &\quad - (k_{sz} - q_m)(k_z - q_m)\rho(z', z'') \\ &\quad + (k_{sz} - q_m)(k_z - k_{sz})\rho(z', z''') \\ &\quad - (k_z - q_m)(k_z - k_{sz})\rho(z'', z''')] \\ \sigma_{\omega_2}(k_{sz}, k_z, q_m) &= \sigma[k_{sz}^2 + k_z^2 + 2q_m^2 - 2(k_{sz} + k_z)q_m \\ &\quad - 2(k_{sz} - q_m)(k_z - q_m)\rho(z', z'')] / 2 \end{aligned} \quad (34)$$

Putting all these results together and defining new spatial coordinates $\xi = x' - x'''$, $\eta = y' -$

$y''', \zeta' = x'' - x'''$ and $\eta' = y'' - y'''$, we can rewrite (24) as follows

$$\begin{aligned}
 S_{qp}^{dkc} = & \frac{|KE_0|^2}{16\pi^2} \text{Re}\{1/\eta_1\} \sum_{m=1,2} \sum_{r=-1,1} \text{Re}\left\{ \hat{f}_{qp}^* \int_{\mathbb{R}^2} du dv \int d\zeta d\eta d\zeta' d\eta' \right. \\
 & \cdot e^{j[u(\zeta-\zeta') + v(\eta-\eta')] - j[k_{sx}\zeta + k_{sy}\eta]} e^{j[k_x\zeta' + k_y\eta']} \\
 & \cdot D_2(\zeta, \eta, \zeta', \eta'; k_{sz}, k_z, r q_m) \hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, \vec{l}_m^r) \\
 & \cdot e^{-\sigma^2[k_{sz}^2 + k_z^2 + q_m^2 - (k_{sz} + k_z)r q_m - k_{sz}k_z - (k_{sz} - r q_m)(k_z - r q_m)\rho_{12}]} \\
 & \left. \cdot \left(e^{-\sigma^2[(k_{sz} - r q_m)(k_z - k_{sz})\rho_{13} - (k_z - r q_m)(k_z - k_{sz})\rho_{23}] - 1} \right) \right\} \quad (35)
 \end{aligned}$$

with

$$D_2(\zeta, \eta, \zeta', \eta'; k_{sz}, k_z, r q_m) = \int dx''' dy''' e^{-j[(k_{sz} - r q_m)z' - (k_z - r q_m)z'' - (k_{sz} - k_z)z''']} \quad (36)$$

and

$$\begin{aligned}
 z' &= z(x''' + \zeta, y''' + \eta) & \rho_{12} &= \rho(\zeta - \zeta', \eta - \eta') \\
 z'' &= z(x''' + \zeta', y''' + \eta') & \rho_{13} &= \rho(\zeta, \eta) \\
 z''' &= z(x''', y''') & \rho_{23} &= \rho(\zeta', \eta')
 \end{aligned}$$

3.2.3 Complementary Incoherent Power

Finally, the diffuse scattered power for the complementary term is

$$\begin{aligned}
 S_{qp}^{dc} &= \frac{1}{2} \text{Re}\{1/\eta_1\} \left\{ \langle E_{qp}^{sc} E_{qp}^{sc*} \rangle - \langle E_{qp}^{sc} \rangle \langle E_{qp}^{sc*} \rangle \right\} \\
 &= \frac{|KE_0|^2}{2^7 \pi^4} \text{Re}\{1/\eta_1\} \sum_{m,n=1,2} \left\{ \int_{\mathbb{R}^4} du dv du' dv' \int_{S^4} dx' dy' dx'' dy'' dx''' dy''' dx'''' dy'''' \right. \\
 & \cdot e^{j[u(x' - x'') - u'(x''' - x''') + v(y' - y'') - v'(y''' - y''')] - j[k_{sx}(x' - x'') + k_{sy}(y' - y'')]} \\
 & \cdot e^{j[k_x(x'' - x''') + k_y(y'' - y''')]} \left[\langle e^{-jk_{sz}(z' - z''')} e^{jk_z(z'' - z''')} e^{jq_m|z' - z''}| \right. \\
 & \cdot e^{-jq'_n|z'''' - z''''} | \hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, \vec{g}_m) \hat{F}_{qp}^{n*}(\vec{k}^i, \vec{k}^s, \vec{g}_n') \rangle \\
 & \left. - \langle e^{-j(k_{sz}z' + k_z z'')} e^{jq_m|z' - z''}| \hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, \vec{g}_m) \rangle \right. \\
 & \left. \left. \langle e^{j(k_{sz}z''' - k_z z''')} e^{-jq'_n|z'''' - z''''} | \hat{F}_{qp}^{n*}(\vec{k}^i, \vec{k}^s, \vec{g}_n') \rangle \right] \right\} \quad (37)
 \end{aligned}$$

Applying the same arguments used to calculate the averages relevant for the cross term power, we obtain the following relations

$$\begin{aligned}
 & \langle e^{-jk_{sz}(z' - z''')} e^{jk_z(z'' - z''')} e^{jq_m|z' - z''}| e^{-jq'_n|z'''' - z''''} | (\Phi_{z'/z''} q_m)^\alpha (\Phi_{z''/z'''} q'_n)^\beta \rangle \\
 &= \frac{q_m^\alpha q_n^\beta}{4} \left(e^{j\omega_1} e^{-\sigma\omega_1} + (-1)^\alpha e^{j\omega_2} e^{-\sigma\omega_2} + (-1)^\beta e^{j\omega_3} e^{-\sigma\omega_3} + (-1)^{\alpha+\beta} e^{j\omega_4} e^{-\sigma\omega_4} \right) \quad (38)
 \end{aligned}$$

where $\alpha, \beta = 0, 1$ and the other coefficients are compactly given by

$$\begin{aligned}
 \omega_1 &= \pi(k_{sz}, k_z, q_m, q'_n) \\
 \omega_2 &= \pi(k_{sz}, k_z, -q_m, q'_n) \\
 \omega_3 &= \pi(k_{sz}, k_z, q_m, -q'_n) \\
 \omega_4 &= \pi(k_{sz}, k_z, -q_m, -q'_n) \\
 \sigma_{\omega_1} &= \sigma_{\pi}(k_{sz}, k_z, q_m, q'_n) \\
 \sigma_{\omega_2} &= \sigma_{\pi}(k_{sz}, k_z, -q_m, q'_n) \\
 \sigma_{\omega_3} &= \sigma_{\pi}(k_{sz}, k_z, q_m, -q'_n) \\
 \sigma_{\omega_4} &= \sigma_{\pi}(k_{sz}, k_z, -q_m, -q'_n)
 \end{aligned} \tag{39}$$

by including the general functions π and σ_{π} in the form

$$\begin{aligned}
 \pi(k_{sz}, k_z, q_m, q'_n) &= -(k_{sz} - q_m)z' + (k_z - q_m)z'' + (k_{sz} - q'_n)z''' - (k_z - q'_n)z^{IV} \\
 \sigma_{\pi}(k_{sz}, k_z, q_m, q'_n) &= \sigma^2[k_{sz}^2 + k_z^2 + q_m^2 + q_n'^2 - (k_{sz} + k_z)(q_m + q'_n) \\
 &\quad - (k_{sz} - q_m)(k_z - q_m)\rho(z', z'') - (k_{sz} - q_m)(k_{sz} - q'_n)\rho(z', z''') \\
 &\quad + (k_{sz} - q_m)(k_z - q'_n)\rho(z', z^{IV}) + (k_z - q_m)(k_{sz} - q'_n)\rho(z'', z''') \\
 &\quad - (k_z - q_m)(k_z - q'_n)\rho(z'', z^{IV}) - (k_{sz} - q'_n)(k_z - q'_n)\rho(z''', z^{IV})]
 \end{aligned} \tag{40}$$

Upon substituting (38) into (37) we find that

$$\begin{aligned}
 S_{qp}^{dc} &= \frac{|KE_0|^2}{2^9 \pi^4} \text{Re}\{1/\eta_1\} \sum_{m,n=1,2} \sum_{r,r'=-1,1} \left\{ \int_{\mathbb{R}^4} du dv du' dv' \int d\zeta d\eta d\zeta' d\eta' d\tau d\kappa \right. \\
 &\quad e^{j[u(\zeta + \tau - \zeta') - u'\tau + v(\eta + \kappa - \eta') - v'\kappa]} e^{-j(k_{sx}\zeta + k_{sy}\eta)} e^{j(k_x\zeta' + k_y\eta')} \\
 &\quad D_3(\zeta, \eta, \zeta', \eta', \tau, \kappa; k_{sz}, k_z, r q_m, r' q'_n) \\
 &\quad \hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, \vec{l}_m^r) \hat{F}_{qp}^{n*}(\vec{k}^i, \vec{k}^s, \vec{l}_n^{r'}) \\
 &\quad e^{-\sigma^2[k_{sz}^2 + k_z^2 + q_m^2 + q_n'^2 - (k_{sz} + k_z)(r q_m + r' q'_n)]} \\
 &\quad e^{-\sigma^2[(k_{sz} - r q_m)(r q_m - k_z)\rho_{12} + (k_{sz} - r' q'_n)(r' q'_n - k_z)\rho_{34}] } \\
 &\quad \left(e^{-\sigma^2[(k_{sz} - r q_m)(r' q'_n - k_{sz})\rho_{13} + (k_{sz} - r q_m)(k_z - r' q'_n)\rho_{14}] } \right. \\
 &\quad \left. e^{-\sigma^2[(k_z - r q_m)(k_{sz} - r' q'_n)\rho_{23} + (k_z - r q_m)(r' q'_n - k_z)\rho_{24}] } - 1 \right) \left. \right\}
 \end{aligned} \tag{41}$$

where $\zeta = x' - x''', \eta = y' - y''', \zeta' = x'' - x^{IV}, \eta' = y'' - y^{IV}, \tau = x''' - x^{IV}$ and $\kappa = y''' - y^{IV}$, the function D_3

$$\begin{aligned}
 D_3(\zeta, \eta, \zeta', \eta', \tau, \kappa; k_{sz}, k_z, r q_m, r' q'_n) \\
 = \int dx^{IV} dy^{IV} e^{-j[(k_{sz} - q_m)z' - (k_z - q_m)z'' - (k_{sz} - q'_n)z''' + (k_z - q'_n)z^{IV}]}
 \end{aligned} \tag{42}$$

and

$$\begin{aligned}
 z' &= z(x'' + \xi + \tau, y'' + \eta + \kappa) & \rho_{12} &= \rho(\xi + \tau - \xi', \eta + \kappa - \eta') \\
 z'' &= z(x'' + \xi', y'' + \eta') & \rho_{13} &= \rho(\xi, \eta) \\
 z''' &= z(x'' + \tau, y'' + \kappa) & \rho_{14} &= \rho(\xi + \tau, \eta + \kappa) \\
 z'' &= z(x'', y'') & \rho_{23} &= \rho(\xi' - \tau, \eta' - \kappa) \\
 & & \rho_{24} &= \rho(\xi', \eta') \\
 & & \rho_{34} &= \rho(\tau, \kappa)
 \end{aligned}$$

3.3 Bistatic Scattering Coefficient for the Scattered Field

The *radar cross section* of a particle producing isotropic scattering is defined as the ratio between the scattered and incident power densities, S^{scat} and S^{inc} multiplied by the area of the spherical surface centred at the particle and with a radius R equal to the distance between the particle and the observation point

$$\sigma \equiv \frac{4\pi R^2 S^{scat}}{S^{inc}} \quad (43)$$

Next, we define the *radar scattering cross section* of a finite scatterer in a given direction as the cross section of a particle which would scatter isotropically the same power density in any direction, should it be illuminated by the same incident power density.

For the case of a scattering surface, it is adequate to define the *differential scattering coefficient* as the average value of the scattering cross section per unit area, namely,

$$\sigma^o \equiv \frac{4\pi R^2 S^{scat}}{A S^{inc}} \quad (44)$$

where A denotes the area of the surface. Usually, the term “radar scattering cross section” is shortened to “radar cross section”, whereas “differential scattering coefficient” is referred to as “scattering coefficient”.

Both radar cross section and scattering coefficient can be either monostatic or bistatic, when the observation point is located at the site from where the incident field is transmitted or elsewhere, respectively. Thus, the bistatic scattering coefficient associated to the coherent and diffuse fields scattered by a random rough surface are given by

$$(\sigma^o)_{qp}^c = \frac{8\pi R^2}{A \operatorname{Re}\{1/\eta_1\} E_0^2} S_{qp}^c \quad (45a)$$

$$(\sigma^o)_{qp}^d = \frac{8\pi R^2}{A \operatorname{Re}\{1/\eta_1\} E_0^2} (S_{qp}^{dk} + S_{qp}^{dkc} + S_{qp}^{dc}) \quad (45b)$$

where the power densities S_{qp}^c , S_{qp}^{dk} , S_{qp}^{dkc} and S_{qp}^{dc} have been calculated in previous sections.

4. Formulation of the IEM2M Model for Topographical Surfaces

The scattering coefficient in (45) is described in terms of the integrals included in S_{qp}^c , S_{qp}^{dk} , S_{qp}^{dkc} and S_{qp}^{dc} . The coherently scattered power calculated in (3.1) is the final form proposed here. However, the integrals corresponding to the diffuse power can be manipulated further. A distinction is drawn then between surfaces with small or moderate rms height normalized

to wave number, $k\sigma$, and surfaces with larger values for $k\sigma$. Thus, a forward scattering model is defined by Taylor expansion of the exponentials in the corresponding integrands. This is done for each scattering coefficient term in the next subsections.

4.1 Scattering Model for Surfaces with Small or Moderate Heights

When the product of the rms height of the surface by the wave number has a small or moderate value, the argument of the exponential functions in (22), (35) and (41) will also have a small value. It is then useful to write the exponential functions in the form of a Taylor series.

4.1.1 Kirchhoff Term

The exponential function in (22) involving the correlation between the heights of the two scattering centres \vec{r}' and \vec{r}'' can be expanded as

$$e^{p_z^2 \sigma^2 \rho(\xi, \eta)} = \sum_{n=0}^{\infty} \frac{[\sigma^2 p_z^2 \rho(\xi, \eta)]^n}{n!} \quad (46)$$

Consequently, the Kirchhoff term (22) of the scattering coefficient takes on the form

$$(\sigma^0)_{qp}^{dk} = \frac{1}{2} k_1^2 |\hat{f}_{qp}|^2 e^{-\sigma^2 (k_{sz} - k_z)^2} \sum_{n=1}^{\infty} \frac{(\sigma^2 (k_{sz} - k_z)^2)^n}{n!} W_1^{(n)}(k_{sx} - k_x, k_{sy} - k_y) \quad (47)$$

where

$$W_1^{(n)}(k_{sx} - k_x, k_{sy} - k_y) = \frac{1}{2\pi A} \int d\xi d\eta \rho^n(\xi, \eta) e^{-j[(k_{sx} - k_x)\xi + (k_{sy} - k_y)\eta]} D_1(\xi, \eta, k_{sz} - k_z) \quad (48)$$

4.1.2 Cross Term

The exponential functions in (24) can be expanded in the form

$$\begin{aligned} & e^{\sigma^2 [(k_{sz} - r q_m)(k_z - r q_m) \rho(z', z'')]} \left(e^{-\sigma^2 [(k_{sz} - r q_m)(k_z - k_{sz}) \rho(z', z''')] } \right. \\ & \left. \cdot e^{\sigma^2 [(k_z - r q_m)(k_z - k_{sz}) \rho(z'', z''')] } - 1 \right) \\ & = \sum_{i=0}^{\infty} \frac{[\sigma^2 (k_{sz} - r q_m)(k_z - r q_m) \rho(z', z'')]^i}{i!} \\ & \left[\sum_{n=0}^{\infty} \frac{[-\sigma^2 (k_{sz} - r q_m)(k_z - k_{sz}) \rho(z', z''')]^n}{n!} \right. \\ & \left. \sum_{l=0}^{\infty} \frac{[\sigma^2 (k_z - r q_m)(k_z - k_{sz}) \rho(z'', z''')]^l}{l!} - 1 \right] \quad (49) \end{aligned}$$

The interactions of second order can be described as specular reflections and Snell's refractions. Second-order scattering events can occur connecting points within the correlation length or distant from each other. When the interacting point sources are within the correlation length, we will have either $k_{sz} \simeq q_m$, for $r = 1$, or $k_z \simeq -q_m$, for $r = -1$, and the first exponential function in (49) will have a negligible argument, provided that σ is not large. When those points are distant, the correlation function ρ will be very small. Thus, the first

summation in (49) can be approximated by unity for surfaces with small or moderate rms height

$$e^{\sigma^2[(k_{sz}-rq_m)(k_z-rq_m)\rho(z',z'')]} \simeq 1 \tag{50}$$

and hence

$$\begin{aligned} & e^{\sigma^2[(k_{sz}-rq_m)(k_z-rq_m)\rho(z',z'')]} \left(e^{-\sigma^2[(k_{sz}-rq_m)(k_z-k_{sz})\rho(z',z''')] } \right. \\ & \cdot e^{\sigma^2[(k_z-rq_m)(k_z-k_{sz})\rho(z'',z''')] } - 1 \Big) \\ & \simeq \sum_{n=1}^{\infty} \frac{[-\sigma^2(k_{sz}-rq_m)(k_z-k_{sz})\rho(z',z''')]^n}{n!} \\ & + \sum_{l=1}^{\infty} \frac{[\sigma^2(k_z-rq_m)(k_z-k_{sz})\rho(z'',z''')]^l}{l!} \\ & + \sum_{n=1}^{\infty} \frac{[-\sigma^2(k_{sz}-rq_m)(k_z-k_{sz})\rho(z',z''')]^n}{n!} \\ & \cdot \sum_{l=1}^{\infty} \frac{[\sigma^2(k_z-rq_m)(k_z-k_{sz})\rho(z'',z''')]^l}{l!} \tag{51} \end{aligned}$$

This yields

$$\begin{aligned} (\sigma^o)_{qp}^{dkc} &= \frac{k_1^2}{8\pi} \sum_{m=1,2} \sum_{r=-1,1} \text{Re} \left\{ \hat{f}_{qp}^* e^{-\sigma^2[k_{sz}^2+k_z^2-k_{sz}k_z]} \right. \\ & \int_{\mathbb{R}^2} du dv \hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, \vec{l}_m) e^{-\sigma^2[q_m^2-(k_{sz}+k_z)r q_m]} \\ & \cdot \left[\sum_{n=1}^{\infty} \frac{[-\sigma^2(k_{sz}-rq_m)(k_z-k_{sz})]^n}{n!} W_2^{n,0}(\vec{l}_m; \vec{k}^s, \vec{k}^i) \right. \\ & + \sum_{l=1}^{\infty} \frac{[\sigma^2(k_z-rq_m)(k_z-k_{sz})]^l}{l!} W_2^{0,l}(\vec{l}_m; \vec{k}^s, \vec{k}^i) \\ & + \sum_{n=1}^{\infty} \frac{[-\sigma^2(k_{sz}-rq_m)(k_z-k_{sz})]^n}{n!} \\ & \left. \left. \sum_{l=1}^{\infty} \frac{[\sigma^2(k_z-rq_m)(k_z-k_{sz})]^l}{l!} W_2^{n,l}(\vec{l}_m; \vec{k}^s, \vec{k}^i) \right] \right\} \tag{52} \end{aligned}$$

where

$$\begin{aligned} W_2^{(\alpha,\beta)}(u,v,w;\vec{k}^s,\vec{k}^i) &= \\ & \frac{1}{(2\pi)^2 A} \int d\xi d\eta d\xi' d\eta' e^{j[(u-k_{sx})\xi+(v-k_{sy})\eta-(u-k_x)\xi'-(v-k_y)\eta']} \\ & \cdot D_2(\xi,\eta,\xi',\eta',k_{sz},k_z,w)\rho^\alpha(\xi,\eta)\rho^\beta(\xi',\eta') \tag{53} \end{aligned}$$

4.1.3 Complementary Term

The complementary term of the scattering coefficient involves the evaluation of an integral containing the following expression

$$\begin{aligned}
 & e^{-\sigma^2[(k_{sz}-rq_m)(rq_m-k_z)\rho_{12}+(k_{sz}-r'q'_n)(r'q'_n-k_z)\rho_{34}]} \left(e^{-\sigma^2[(k_{sz}-rq_m)(r'q'_n-k_z)\rho_{13}]} \right. \\
 & e^{-\sigma^2[(k_{sz}-rq_m)(k_z-r'q'_n)\rho_{14}+(k_z-rq_m)(k_{sz}-r'q'_n)\rho_{23}+(k_z-rq_m)(r'q'_n-k_z)\rho_{24}] - 1} \Big) \\
 & = \sum_{i=0}^{\infty} \frac{[-\sigma^2(k_{sz}-rq_m)(rq_m-k_z)\rho_{12}]^i}{i!} \sum_{j=0}^{\infty} \frac{[-\sigma^2(k_{sz}-r'q'_n)(r'q'_n-k_z)\rho_{34}]^j}{j!} \\
 & \left[\sum_{h=0}^{\infty} \frac{[-\sigma^2(k_{sz}-rq_m)(r'q'_n-k_z)\rho_{13}]^h}{h!} \sum_{l=0}^{\infty} \frac{[-\sigma^2(k_{sz}-rq_m)(k_z-r'q'_n)\rho_{14}]^l}{l!} \right. \\
 & \left. \sum_{n=0}^{\infty} \frac{[-\sigma^2(k_z-rq_m)(k_{sz}-r'q'_n)\rho_{23}]^n}{n!} \sum_{t=0}^{\infty} \frac{[-\sigma^2(k_z-rq_m)(r'q'_n-k_z)\rho_{24}]^t}{t!} - 1 \right] \quad (54)
 \end{aligned}$$

As explained in the previous subsection, the correlation between points producing effective second-order scattering is negligible. These points are represented in the summation above by the pairs 1 and 2 on the one hand and by 3 and 4 on the other. Thus, the first two summations containing ρ_{12} and ρ_{34} can be approximated by unity. Further, all the products between summations of the form \sum_1^{∞} containing ρ_{13} and ρ_{14} are negligible. This is so because significant correlation between points 1 and both points 3 and 4 would generally imply a significant correlation between 3 and 4. The same reasoning applies to products with ρ_{13} and ρ_{23} , ρ_{23} and ρ_{24} or ρ_{14} and ρ_{24} . Thereby,

$$\begin{aligned}
 & e^{-\sigma^2[(k_{sz}-rq_m)(rq_m-k_z)\rho_{12}+(k_{sz}-r'q'_n)(r'q'_n-k_z)\rho_{34}]} \left(e^{-\sigma^2[(k_{sz}-rq_m)(r'q'_n-k_z)\rho_{13}]} \right. \\
 & e^{-\sigma^2[(k_{sz}-rq_m)(k_z-r'q'_n)\rho_{14}+(k_z-rq_m)(k_{sz}-r'q'_n)\rho_{23}+(k_z-rq_m)(r'q'_n-k_z)\rho_{24}] - 1} \Big) \\
 & \simeq \sum_{h=1}^{\infty} \frac{[-\sigma^2(k_{sz}-rq_m)(r'q'_n-k_z)\rho_{13}]^h}{h!} + \sum_{l=1}^{\infty} \frac{[-\sigma^2(k_{sz}-rq_m)(k_z-r'q'_n)\rho_{14}]^l}{l!} \\
 & + \sum_{n=1}^{\infty} \frac{[-\sigma^2(k_z-rq_m)(k_{sz}-r'q'_n)\rho_{23}]^n}{n!} + \sum_{t=1}^{\infty} \frac{[-\sigma^2(k_z-rq_m)(r'q'_n-k_z)\rho_{24}]^t}{t!} \\
 & + \sum_{h=1}^{\infty} \frac{[-\sigma^2(k_{sz}-rq_m)(r'q'_n-k_z)\rho_{13}]^h}{h!} \sum_{t=1}^{\infty} \frac{[-\sigma^2(k_z-rq_m)(r'q'_n-k_z)\rho_{24}]^t}{t!} \\
 & + \sum_{l=1}^{\infty} \frac{[-\sigma^2(k_{sz}-rq_m)(k_z-r'q'_n)\rho_{14}]^l}{l!} \sum_{n=1}^{\infty} \frac{[-\sigma^2(k_z-rq_m)(k_{sz}-r'q'_n)\rho_{23}]^n}{n!} \quad (55)
 \end{aligned}$$

Introducing this approximation, (41) becomes

$$\begin{aligned}
(\sigma^o)_{qp}^{dc} = & \frac{k_1^2}{2^7 \pi^2} \sum_{m,n=1,2} \sum_{r,r'=-1,1} \left\{ e^{-\sigma^2(k_{sz}^2+k_z^2)} \int_{\mathbb{R}^4} du dv du' dv' \right. \\
& \hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, \vec{l}_m^r) \hat{F}_{qp}^{n*}(\vec{k}^i, \vec{k}^s, \vec{l}_n^{r'}) \\
& e^{-\sigma^2[q_m^2+q_n^2-(k_{sz}+k_z)(r q_m+r' q_n)]} \\
& \left[\sum_{h=1}^{\infty} \frac{[-\sigma^2(k_{sz}-r q_m)(r' q_n-k_{sz})]^h}{h!} W_3^{h,0,0,0}(\vec{l}_m^r, \vec{l}_n^{r'}; \vec{k}^s, \vec{k}^i) \right. \\
& + \sum_{l=1}^{\infty} \frac{[-\sigma^2(k_{sz}-r q_m)(k_z-r' q_n')^l]}{l!} W_3^{0,m,0,0}(\vec{l}_m^r, \vec{l}_n^{r'}; \vec{k}^s, \vec{k}^i) \\
& + \sum_{n=1}^{\infty} \frac{[-\sigma^2(k_z-r q_m)(k_{sz}-r' q_n')]^n}{n!} W_3^{0,0,n,0}(\vec{l}_m^r, \vec{l}_n^{r'}; \vec{k}^s, \vec{k}^i) \\
& + \sum_{t=1}^{\infty} \frac{[-\sigma^2(k_z-r q_m)(r' q_n'-k_z)]^t}{t!} W_3^{0,0,0,t}(\vec{l}_m^r, \vec{l}_n^{r'}; \vec{k}^s, \vec{k}^i) \\
& + \sum_{h=1}^{\infty} \frac{[-\sigma^2(k_{sz}-r q_m)(r' q_n'-k_{sz})]^h}{h!} \\
& \sum_{t=1}^{\infty} \frac{[-\sigma^2(k_z-r q_m)(r' q_n'-k_z)]^t}{t!} W_3^{h,0,0,t}(\vec{l}_m^r, \vec{l}_n^{r'}; \vec{k}^s, \vec{k}^i) \\
& + \sum_{l=1}^{\infty} \frac{[-\sigma^2(k_{sz}-r q_m)(k_z-r' q_n')]^l}{l!} \\
& \left. \left. \sum_{n=1}^{\infty} \frac{[-\sigma^2(k_z-r q_m)(k_{sz}-r' q_n')]^n}{n!} W_3^{0,m,n,0}(\vec{l}_m^r, \vec{l}_n^{r'}; \vec{k}^s, \vec{k}^i) \right] \right\} \quad (56)
\end{aligned}$$

where

$$\begin{aligned}
W_3^{(h,l,n,t)}(u,v,w,u',v',w'; \vec{k}^s, \vec{k}^i) \\
= & \frac{1}{(2\pi)^3 A} \int d\tilde{\xi} d\eta d\tilde{\xi}' d\eta' d\tau d\kappa e^{j[(u-k_{sx})\tilde{\xi}-(u-k_x)\tilde{\xi}'+(v-k_{sy})\eta-(v-k_y)\eta']} \\
& e^{j[(u-u')\tau+(v-v')\kappa]} D_3(\tilde{\xi}, \eta, \tilde{\xi}', \eta', \tau, \kappa; k_{sz}, k_z, w, w') \rho^h(\tilde{\xi}, \eta) \\
& \rho^l(\tilde{\xi} + \tau, \eta + \kappa) \rho^n(\tilde{\xi}' - \tau, \eta' - \kappa) \rho^t(\tilde{\xi}', \eta') \quad (57)
\end{aligned}$$

4.2 Scattering Model for Surfaces with Large Heights

Although a series of the type given in (47) is convergent for any value of the argument, it is only practical to compute it when the argument is not large. Thus, the summations describing the scattering coefficient for the diffuse field in the previous section are not practical for large rms height. Besides, it was assumed that, on the whole, the correlation between points producing second-order scattering was negligible and, as will be shown below, this is not the case for surfaces with large rms height.

4.2.1 Kirchhoff Term

Let us reconsider first the Kirchhoff term in the form given in Subsection 3.2.1

$$S_{qp}^{dk} = \frac{1}{4\pi A} k_1^2 \hat{f}_{qp}^2 \iint d\bar{\xi} d\eta e^{-j[(k_{sx}-k_x)\bar{\xi}+(k_{sy}-k_y)\eta]} (e^{-(k_{sz}-k_z)^2\sigma^2(1-\rho(\bar{\xi},\eta))} - e^{-(k_{sz}-k_z)^2\sigma^2}) D_1(\bar{\xi},\eta;k_{sz}-k_z) \quad (58)$$

Large values for $k_1\sigma$ give rise to very negative arguments in the exponentials of (58). As a matter of fact the coherent term subtracted in this equation is negligible and the additive exponential is significant only when the correlation function is near unity. It is then possible to perform a Taylor expansion of the correlation function about the origin to obtain

$$\begin{aligned} 1 - \rho(\bar{\xi},\eta) &\simeq \frac{1}{2} |\rho_{\bar{\xi}\bar{\xi}}(0)| \bar{\xi}^2 + \frac{1}{2} |\rho_{\eta\eta}(0)| \eta^2 + |\rho_{\bar{\xi}\eta}(0)| \bar{\xi} \eta \\ &\equiv \frac{1}{2} |\rho_{\bar{\xi}\bar{\xi}}^o| \bar{\xi}^2 + \frac{1}{2} |\rho_{\eta\eta}^o| \eta^2 + |\rho_{\bar{\xi}\eta}^o| \bar{\xi} \eta \end{aligned} \quad (59)$$

were the subscripts in ρ denote partial derivatives and the superscript o denotes that the correlation function is evaluated at the origin. Likewise, we expand the function $D_1(\bar{\xi},\eta;k)$ about the origin

$$\begin{aligned} D_1(\bar{\xi},\eta;k) &\simeq D_1(0,0;k) + D_{1,\bar{\xi}}(0,0;k)\bar{\xi} + D_{1,\eta}(0,0;k)\eta \\ &\quad + \frac{1}{2} D_{1,\bar{\xi}\bar{\xi}}(0,0;k)\bar{\xi}^2 + \frac{1}{2} D_{1,\eta\eta}(0,0;k)\eta^2 + D_{1,\eta\bar{\xi}}(0,0;k)\eta\bar{\xi} \\ &\equiv D_1^o(k) + D_{1,\bar{\xi}}^o(k)\bar{\xi} + D_{1,\eta}^o(k)\eta \\ &\quad + \frac{1}{2} D_{1,\bar{\xi}\bar{\xi}}^o(k)\bar{\xi}^2 + \frac{1}{2} D_{1,\eta\eta}^o(k)\eta^2 + D_{1,\eta\bar{\xi}}^o(k)\eta\bar{\xi} \end{aligned} \quad (60)$$

Upon replacing (59) and (60) in (58), we arrive at

$$\begin{aligned} (\sigma^o)_{qp}^{dk} &= \frac{1}{4\pi A} k_1^2 \hat{f}_{qp}^2 \iint d\bar{\xi} d\eta e^{-j[(k_{sx}-k_x)\bar{\xi}+(k_{sy}-k_y)\eta]} \\ &\quad \exp \left[-(k_{sz}-k_z)^2\sigma^2 \left(\frac{1}{2} |\rho_{\bar{\xi}\bar{\xi}}^o| \bar{\xi}^2 + \frac{1}{2} |\rho_{\eta\eta}^o| \eta^2 + |\rho_{\bar{\xi}\eta}^o| \bar{\xi} \eta \right) \right] \\ &\quad [D_1^o(k_{sz}-k_z) + D_{1,\bar{\xi}}^o(k_{sz}-k_z)\bar{\xi} + D_{1,\eta}^o(k_{sz}-k_z)\eta \\ &\quad + \frac{1}{2} D_{1,\bar{\xi}\bar{\xi}}^o(k_{sz}-k_z)\bar{\xi}^2 + \frac{1}{2} D_{1,\eta\eta}^o(k_{sz}-k_z)\eta^2 + D_{1,\eta\bar{\xi}}^o(k_{sz}-k_z)\eta\bar{\xi}] \end{aligned} \quad (61)$$

where the subtraction of the coherent term has been disregarded.

The following integral identity will be used

$$\begin{aligned} \iint_{-\infty}^{\infty} dx dy e^{-(ax^2+by^2+2cxy)} (A + Bx + Cx^2 + Dy + Ey^2 + Fxy) e^{-j(k_x x + k_y y)} \\ = \frac{\pi}{4(ab-c^2)^{(5/2)}} \exp \left\{ -\frac{k_x^2 b - 2ck_x k_y + k_y^2 a}{4(ab-c^2)} \right\} \\ [A\alpha_A(a,b,c) + B\alpha_B(a,b,c,k_x,k_y) + C\alpha_C(a,b,c,k_x,k_y) \\ + D\alpha_D(a,b,c,k_x,k_y) + E\alpha_E(a,b,c,k_x,k_y) + F\alpha_F(a,b,c,k_x,k_y)] \end{aligned} \quad (62)$$

where

$$\begin{aligned}
\alpha_A(a, b, c) &= 4(ab - c^2)^2 \\
\alpha_B(a, b, c, k_x, k_y) &= -2j(ab - c^2)(bk_x - ck_y) \\
\alpha_C(a, b, c, k_x, k_y) &= 2b(ab - c^2) - (bk_x - ck_y)^2 \\
\alpha_D(a, b, c, k_x, k_y) &= -2j(ab - c^2)(ak_y - ck_x) \\
\alpha_E(a, b, c, k_x, k_y) &= 2a(ab - c^2) - (ak_y - ck_x)^2 \\
\alpha_F(a, b, c, k_x, k_y) &= -2c(ab - c^2) + (ck_x - ak_y)(bk_x - ck_y)
\end{aligned} \tag{63}$$

Therefore, (61) results in

$$(\sigma^o)_{qp}^{dk} = \frac{2k_1^2 \hat{f}_{qp}^2 \mathcal{I}^k(\vec{p})}{p_z^{10} \sigma^{10} [|\rho_{\zeta, \xi}^o| |\rho_{\eta, \eta}^o| - |\rho_{\zeta, \eta}^o|^2]^{5/2} A} \exp \left\{ -\frac{p_x^2 |\rho_{\eta, \eta}^o| - 2p_x p_y |\rho_{\zeta, \eta}^o| + p_y^2 |\rho_{\zeta, \xi}^o|}{2p_z^2 \sigma^2 (|\rho_{\zeta, \xi}^o| |\rho_{\eta, \eta}^o| - |\rho_{\zeta, \eta}^o|^2)} \right\} \tag{64}$$

where

$$\begin{aligned}
\mathcal{I}^k(\vec{p}) &= D_1^o(p_z) \tilde{\alpha}_A + D_{1, \xi}^o(p_z) \tilde{\alpha}_B + \frac{1}{2} D_{1, \xi, \zeta}^o(p_z) \tilde{\alpha}_C \\
&\quad + D_{1, \eta}^o(p_z) \tilde{\alpha}_D + \frac{1}{2} D_{1, \eta, \eta}^o(p_z) \tilde{\alpha}_E + D_{1, \eta, \zeta}^o(p_z) \tilde{\alpha}_F
\end{aligned} \tag{65}$$

with $\vec{p} = \vec{k}^s - \vec{k}^i$, and

$$\begin{aligned}
\tilde{\alpha}_A &= \alpha_A(\kappa_1 |\rho_{\zeta, \xi}^o|, \kappa_1 |\rho_{\eta, \eta}^o|, \kappa_1 |\rho_{\zeta, \eta}^o|) \\
\tilde{\alpha}_\zeta &= \alpha_\zeta(\kappa_1 |\rho_{\zeta, \xi}^o|, \kappa_1 |\rho_{\eta, \eta}^o|, \kappa_1 |\rho_{\zeta, \eta}^o|, p_x, p_y) \quad \zeta = B, C, D, E, F \\
\kappa_1 &= p_z^2 \sigma^2 / 2
\end{aligned} \tag{66}$$

The expression obtained in (61) is the result obtained from classic geometric optics, multiplied by a factor of correction due to the deterministic component of the surface.

4.2.2 Cross Term

From Subsection 3.2.2 we get

$$\begin{aligned}
(\sigma^o)_{qp}^{dkc} &= \frac{k_1^2}{2^5 \pi^3 A} \sum_{m=1,2} \sum_{r=-1,1} \operatorname{Re} \left\{ \hat{f}_{qp}^* \int_{\mathbb{R}^2} du dv \int d\zeta d\eta d\zeta' d\eta' \right. \\
&\quad \cdot e^{j[u(\zeta - \zeta') + v(\eta - \eta')]} e^{-j[k_{sx}\zeta + k_{sy}\eta]} e^{j[k_x\zeta' + k_y\eta']} \\
&\quad \cdot D_2(\zeta, \eta, \zeta', \eta'; k_{sz}, k_z, r q_m) \hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, \vec{l}_m) \\
&\quad \cdot e^{-\sigma^2[(k_{sz} - r q_m)(k_z - r q_m)(1 - \rho_{12})]} \left[e^{-\sigma^2[(k_{sz} - r q_m)(k_{sz} - k_z)(1 - \rho_{13})]} \right. \\
&\quad \left. \cdot e^{-\sigma^2[(r q_m - k_z)(k_{sz} - k_z)(1 - \rho_{23})]} - e^{-\sigma^2(k_{sz} - k_z)^2} \right] \left. \right\} \tag{67}
\end{aligned}$$

Some simplifications are applicable but, before introducing them, some remarks are in order. As in Paragraph 4.1.2, the approach is seeing the interactions of second order as specular reflections and Snell's refractions. Also, surface integration is taken over two regions for each

correlation function, the region where points are close in terms of the correlation length and the region where points are distant from each other. The correlation function ρ_{12} links the points that are connected by second-order scattering events, and the functions ρ_{13} and ρ_{23} relate a point acting as a secondary wave source of second order and a point with a first-order scattering role. The second type of functions are present due to the fact that the cross term described by (67) is an interference between first and second-order scattering in the calculation of the scattered power.

The situation is more complicated now than in (58) as the sign of the arguments in the exponential functions depends on the value of $k_{sz} - r q_m$ and $k_z - r q_m$. We observe the following

1. The coherent component in (67) can be written as

$$e^{-\sigma^2[(k_{sz}-r q_m)(k_z-r q_m)(1-\rho_{12})]} e^{-\sigma^2(k_{sz}-k_z)^2} \\ = e^{-\sigma^2[k_{sz}^2+k_z^2+q_m^2-(k_{sz}+k_z)r q_m-k_{sz}k_z-(k_{sz}-r q_m)(k_z-r q_m)\rho_{12}]} \quad (68)$$

The second exponential at the l.s. of (68) has a large negative argument for large $k\sigma$ values. Therefore, the coherent term will be very small except, perhaps, when the argument of the first exponential at the l.s. of (68) has a positive argument. For this to happen, we need either $k_{sz} - q_m > 0$ when $r = 1$ or $k_z + q_m < 0$ when $r = -1$. In both cases, according to the argument of the exponential at the r.s. of (68), the product of the two exponential functions with different signs in their argument is negligible. Thereby, the coherent component subtracted in (67) is not significant, as we should expect from a surface with large rms height.

2. For the incoherent term, and according to the aforementioned distinction between the two areas of integration for each correlation function, we note that

- (a) If the three correlation functions ρ_{12} , ρ_{13} and ρ_{23} are all very small, the exponential functions yield

$$e^{-\sigma^2(k_{sz}-r q_m)(k_z-r q_m)} e^{-\sigma^2(k_{sz}-r q_m)(k_{sz}-k_z)} e^{-\sigma^2(r q_m-k_z)(k_{sz}-k_z)} \\ = e^{-\sigma^2(k_{sz}-r q_m)^2} e^{-\sigma^2(r q_m-k_z)(k_{sz}-k_z)} \\ = e^{-\sigma^2(k_z-r q_m)^2} e^{-\sigma^2(k_{sz}-r q_m)(k_{sz}-k_z)} \\ = e^{-\sigma^2(k_{sz}-k_z)^2} e^{-\sigma^2(k_{sz}-r q_m)(k_z-r q_m)} \quad (69)$$

From (69), it is clear that the product of the three exponential functions is negligible no matter the sign of $k_{sz} - r q_m$ and $k_z - r q_m$.

- (b) If two correlation functions are very small and the other one is close to unity, then we obtain similar identities to (69). For instance, provided that $\rho_{12} \simeq 1$, the product of exponentials is written as

$$e^{-\sigma^2(k_{sz}-k_z)^2} e^{-\sigma^2(k_{sz}-r q_m)(k_z-r q_m)(1-\rho_{12})} \simeq e^{-\sigma^2(k_{sz}-k_z)^2} \quad (70)$$

and can be neglected. The same holds for either of the other two correlation functions.

- (c) The region of the integration domain where two correlation functions are close to unity and the other is negligible can be regarded as having measure zero. For example, if $\rho_{12} \simeq 1$ and $\rho_{13} \simeq 1$, we expect $\rho_{23} \simeq 1$, that is, if the pair of points (1,2) and (1,3) are highly correlated, then the pair (2,3) is generally expected to be highly correlated, too.
- (d) When the three correlation functions are all close to unity, the exponentials can have moderate or small arguments and therefore they do contribute to the integral. Thereupon, the most significant region of the integration domain corresponds to small values of ξ , η , ξ' and η' and ρ_{12} , ρ_{13} and ρ_{23} can be Taylor expanded about the origin. To see the order of approximation to be taken for each correlation function we investigate their physical meaning. The exponential function containing ρ_{12} represents the interference between the sources located at points 1 and 2, which are the secondary wave sources involved in a second-order scattering event. On the other hand, ρ_{13} and ρ_{23} represent the interference between one of those second-order sources on the surface and the source located at point 3, which is a first-order - or Kirchhoff - secondary wave source. As the Kirchhoff field is expected to be of a higher magnitude than the complementary field, we expand ρ_{13} and ρ_{23} about the origin up to second order and ρ_{12} only up to first order.

According to these remarks, the product of exponential functions in (67) can be replaced by

$$\begin{aligned}
 & e^{-\sigma^2[(k_{sz}-rq_m)(k_z-rq_m)(1-\rho_{12})]} \\
 & \left[e^{-\sigma^2[(k_{sz}-rq_m)(k_{sz}-k_z)(1-\rho_{13})]} e^{-\sigma^2[(rq_m-k_z)(k_{sz}-k_z)(1-\rho_{23})]} - e^{-\sigma^2(k_{sz}-k_z)^2} \right] \\
 & \simeq e^{-\frac{1}{2}\sigma^2(k_{sz}-rq_m)(k_{sz}-k_z)[|\rho_{\xi\xi}^0|\xi^2+|\rho_{\eta\eta}^0|\eta^2+2|\rho_{\xi\eta}^0|\xi\eta]} \\
 & e^{-\frac{1}{2}\sigma^2(rq_m-k_z)(k_{sz}-k_z)[|\rho_{\xi\xi}^0|\xi'^2+|\rho_{\eta\eta}^0|\eta'^2+2|\rho_{\xi\eta}^0|\xi'\eta']}
 \end{aligned} \tag{71}$$

However, it is important to note that this replacement is only possible when the arguments of the exponential functions at the r.s. of (71) are negative. Therefore, if $(k_{sz}-rq_m)$ or (rq_m-k_z) are not positive, the substitution is not possible and $\exp\{-\sigma^2[(k_{sz}-rq_m)(k_z-rq_m)(1-\rho_{12})]\}$ cannot be discarded. The assumption here is to consider that the reflections and refractions involved in second-order scattering are unlikely to produce first deviations where the modulus of the z-component of the wave vector increases, such that $rq_m < k_z$, or second deviations where it decreases, such that $k_{sz} < rq_m$. Thus the integration domain in u and v , Γ_r , will be constrained to the following conditions

$$\Gamma_r : \begin{cases} q_m < |k_z| & \text{if } r = -1 \\ q_m < k_{sz} & \text{if } r = 1 \end{cases} \tag{72}$$

We expand also D_2 in (36) about the origin

$$\begin{aligned}
 D_2(\xi, \eta, \xi', \eta'; k, k', k'') &= D_2^0(k, k', k'') + \sum_{\beta=\xi, \eta, \xi', \eta'} D_{2,\beta}^0(k, k', k'') \beta \\
 &+ \frac{1}{2} \sum_{\beta, \gamma=\xi, \eta, \xi', \eta'} D_{2,\beta,\gamma}^0(k, k', k'') \beta \gamma
 \end{aligned} \tag{73}$$

where the subscripts denote partial derivatives and the superscript o in D_2 means that this function or its derivatives have been evaluated at the origin.

By making use of (62), we obtain that (67) can be written as

$$\begin{aligned}
 (\sigma^o)_{qp}^{dkc} &= \frac{2k_1^2}{\sigma^{20}\pi A} \sum_{m=1,2} \sum_{r=-1,1} \operatorname{Re} \left\{ \hat{f}_{qp}^* \int_{\Gamma_r} du dv \right. \\
 &\quad \frac{1}{p_{sz}^{(r)5} p_{iz}^{(r)5} p_z^{10} (|\rho_{\xi,\xi}^o| |\rho_{\eta,\eta}^o| - |\rho_{\xi,\eta}^o|^2)^5} \mathcal{I}^{kc}(\vec{k}^i, \vec{k}^s, \vec{l}^r_m) \\
 &\quad \hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, \vec{l}^r_m) \\
 &\quad \cdot \exp \left\{ -\frac{p_{sx}^2 |\rho_{\eta,\eta}^o| - 2p_{sx} p_{sy} |\rho_{\xi,\eta}^o| + p_{sy}^2 |\rho_{\xi,\xi}^o|}{2\sigma^2 p_{sz}^{(r)} p_z (|\rho_{\xi,\xi}^o| |\rho_{\eta,\eta}^o| - |\rho_{\xi,\eta}^o|^2)} \right\} \\
 &\quad \cdot \exp \left\{ -\frac{p_{ix}^2 |\rho_{\eta,\eta}^o| - 2p_{ix} p_{iy} |\rho_{\xi,\eta}^o| + p_{iy}^2 |\rho_{\xi,\xi}^o|}{2\sigma^2 p_{iz}^{(r)} p_z (|\rho_{\xi,\xi}^o| |\rho_{\eta,\eta}^o| - |\rho_{\xi,\eta}^o|^2)} \right\} \left. \right\}
 \end{aligned} \tag{74}$$

where

$$\mathcal{I}^{kc}(\vec{k}^s, \vec{k}^i, u, v, r q_m) = \hat{\alpha}^t D_2 \hat{\alpha}' \tag{75}$$

with

$$\begin{aligned}
 \hat{\alpha} &= \begin{bmatrix} \hat{\alpha}_A \\ \hat{\alpha}_B \\ \hat{\alpha}_C \\ \hat{\alpha}_D \\ \hat{\alpha}_E \\ \hat{\alpha}_F \end{bmatrix} & \hat{\alpha}' &= \begin{bmatrix} \hat{\alpha}'_A \\ \hat{\alpha}'_B \\ \hat{\alpha}'_C \\ \hat{\alpha}'_D \\ \hat{\alpha}'_E \\ \hat{\alpha}'_F \end{bmatrix} \\
 D &= \begin{bmatrix} D_2^o & D_{2,\xi'}^o & D_{2,\xi',\xi'}^o/2 & D_{2,\eta'}^o & D_{2,\eta',\eta'}^o/2 & D_{2,\xi',\eta'}^o \\ D_{2,\xi}^o & D_{2,\xi,\xi'}^o & 0 & D_{2,\xi,\eta'}^o & 0 & 0 \\ D_{2,\xi,\xi}^o/2 & 0 & 0 & 0 & 0 & 0 \\ D_{2,\eta}^o & D_{2,\xi',\eta}^o & 0 & D_{2,\eta,\eta'}^o & 0 & 0 \\ D_{2,\eta,\eta}^o/2 & 0 & 0 & 0 & 0 & 0 \\ D_{2,\xi,\eta}^o & 0 & 0 & 0 & 0 & 0 \end{bmatrix}
 \end{aligned} \tag{76}$$

and

$$\begin{aligned}
 \hat{\alpha}_A &= \alpha_A (\kappa_2^{(r)} |\rho_{\xi,\xi}^o|, \kappa_2^{(r)} |\rho_{\eta,\eta}^o|, \kappa_2^{(r)} |\rho_{\xi,\eta}^o|) \\
 \hat{\alpha}'_A &= \alpha_A (\kappa_3^{(r)} |\rho_{\xi,\xi}^o|, \kappa_3^{(r)} |\rho_{\eta,\eta}^o|, \kappa_3^{(r)} |\rho_{\xi,\eta}^o|) \\
 \hat{\alpha}_\zeta &= \alpha_\zeta (\kappa_2^{(r)} |\rho_{\xi,\xi}^o|, \kappa_2^{(r)} |\rho_{\eta,\eta}^o|, \kappa_2^{(r)} |\rho_{\xi,\eta}^o|, p_{sx}, p_{sy}) \\
 \hat{\alpha}'_\zeta &= \alpha_\zeta (\kappa_3^{(r)} |\rho_{\xi,\xi}^o|, \kappa_3^{(r)} |\rho_{\eta,\eta}^o|, \kappa_3^{(r)} |\rho_{\xi,\eta}^o|, p_{ix}, p_{iy}) \\
 \zeta &= B, C, D, E, F \\
 \kappa_2^{(r)} &= p_{sz}^{(r)} p_z \sigma^2 / 2 \\
 \kappa_3^{(r)} &= p_{iz}^{(r)} p_z \sigma^2 / 2
 \end{aligned} \tag{77}$$

$$\begin{aligned}
p_{sx} &= k_{sx} - u & p_{sy} &= k_{sy} - v & p_{sz}^{(r)} &= k_{sz} - r q_m \\
p_{ix} &= u - k_x & p_{iy} &= v - k_y & p_{iz}^{(r)} &= r q_m - k_z \\
\vec{p} &= \vec{k}^s - \vec{k}^i & & & &
\end{aligned} \tag{78}$$

The notation has been simplified for the matrix D , where all the elements are evaluated at $(k_{sz}, k_z, r q_m)$. The modulation due to the topography of the surface is contained in the function $\mathcal{I}^{kc}(\vec{k}^s, \vec{k}^i, u, v, r q_m)$.

4.2.3 Complementary Term

Recalling the results of Subsection 3.2.3 for the complementary term of the diffuse scattered power, we can write

$$\begin{aligned}
S_{qp}^{dc} &= \frac{k_1^2}{2^{10} \pi^3 A} \sum_{m,n=1,2} \sum_{r,r'=-1,1} \left\{ \int_{\mathbb{R}^4} du dv du' dv' \int d\xi d\eta d\xi' d\eta' d\tau d\kappa \right. \\
&\quad e^{j[u(\xi+\tau-\xi')-u'\tau+v(\eta+\kappa-\eta')-v'\kappa]} e^{-j(k_{sx}\xi+k_{sy}\eta)} e^{j(k_x\xi'+k_y\eta')} \\
&\quad D_3(\xi, \eta, \xi', \eta', \tau, \kappa; k_{sz}, k_z, r q_m, r' q'_n) \\
&\quad \hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, \vec{l}_m) \hat{F}_{qp}^{n*}(\vec{k}^i, \vec{k}^s, \vec{l}_n^{r'}) \\
&\quad e^{-\sigma^2[(k_{sz}-r q_m)(k_z-r q_m)(1-\rho_{12})+(k_{sz}-r' q'_n)(k_z-r' q'_n)(1-\rho_{34})]} \\
&\quad \left[e^{-\sigma^2[(k_{sz}-r q_m)(k_{sz}-r' q'_n)(1-\rho_{13})+(k_{sz}-r q_m)(r' q'_n-k_z)(1-\rho_{14})]} \right. \\
&\quad \left. e^{-\sigma^2[(k_z-r q_m)(r' q'_n-k_{sz})(1-\rho_{23})+(k_z-r q_m)(k_z-r' q'_n)(1-\rho_{24})]} \right. \\
&\quad \left. - e^{-\sigma^2(k_{sz}^2+k_z^2)} \right] \left. \right\} \tag{79}
\end{aligned}$$

Although the higher dimensionality in (79) makes this integral more complex than (67) the same principles used to simplify the exponential functions apply in both cases. Yet, instead of repeating the same reasoning as in the previous paragraph, we will try to make use of the physics already found there. In (67) we had the interference between the first-order component and the second-order components of the incoherently scattered field. We found that the most meaningful contribution comes from the interference between the first-order secondary sources located near the second-order secondary sources, which are in turn close to one another. This means that the waves transmitted by these secondary sources interfere more constructively when the sources are near each other, as we might expect from a rough surface with high rms height and small or moderate correlation length. Furthermore, the coherently scattered power for such a surface is negligible. Assuming that this is also the case for the complementary term of the scattered power, where the interference occurs between second-order secondary waves only, we will get significant contribution for the integral over small values of $\xi, \eta, \xi', \eta', \tau$ and κ . As we did for the cross term, the order of the Taylor series for the correlation functions is different for each function. We assume that the most significant interferences occur between the secondary sources which do not belong to the same second-order scattering event. Thus, ρ_{14} , ρ_{23} and ρ_{24} are approximated at second order, whereas ρ_{12} and ρ_{34} are approximated at first order. The correlation function ρ_{13} describes the interference between the secondary sources of the outgoing field and are also approximated only at first

order. Then, we obtain the following approximation

$$\begin{aligned}
 & e^{-\sigma^2[(k_{sz}-r q_m)(k_z-r q_m)(1-\rho_{12})+(k_{sz}-r' q'_n)(k_z-r' q'_n)(1-\rho_{34})]} \\
 & \left[e^{-\sigma^2[(k_{sz}-r q_m)(k_{sz}-r' q'_n)(1-\rho_{13})+(k_{sz}-r q_m)(r' q'_n-k_z)(1-\rho_{14})]} \right. \\
 & e^{-\sigma^2[(k_z-r q_m)(r' q'_n-k_{sz})(1-\rho_{23})+(k_z-r q_m)(k_z-r' q'_n)(1-\rho_{24})]} \\
 & \left. - e^{-\sigma^2(k_{sz}^2+k_z^2)} \right] \tag{80} \\
 & \simeq e^{-\frac{1}{2}\sigma^2(k_{sz}-r q_m)(r' q'_n-k_z)[|\rho_{\xi\xi}^0|(\xi+\tau)^2+|\rho_{\eta\eta}^0|(\eta+\kappa)^2+2|\rho_{\xi\eta}^0|(\xi+\tau)(\eta+\kappa)]} \\
 & e^{-\frac{1}{2}\sigma^2(k_z-r q_m)(r' q'_n-k_{sz})[|\rho_{\xi\xi}^0|(\xi'-\tau)^2+|\rho_{\eta\eta}^0|(\eta'-\kappa)^2+2|\rho_{\xi\eta}^0|(\xi'-\tau)(\eta'-\kappa)]} \\
 & e^{-\frac{1}{2}\sigma^2(k_z-r q_m)(k_z-r' q'_n)[|\rho_{\xi\xi}^0|\xi'^2+|\rho_{\eta\eta}^0|\eta'^2+2|\rho_{\xi\eta}^0|\xi'\eta']}
 \end{aligned}$$

Similar comments to those made after (71) are in order. Thus, (80) is to be used under the constrains of $(k_{sz} - r q_m) > 0$, $(r q_m - k_z) > 0$, $(k_{sz} - r' q'_n) > 0$, and $(r' q'_n - k_z) > 0$. The substitution (80) is then introduced into (79) with the domain of integration for (u, v, u', v') restricted to $\Gamma_r \times \Gamma'_{r'}$

$$\Gamma_r : \begin{cases} q_m < |k_z| & \text{if } r = -1 \\ q_m < k_{sz} & \text{if } r = 1 \end{cases} \quad \Gamma'_{r'} : \begin{cases} q'_n < |k_z| & \text{if } r = -1 \\ q'_n < k_{sz} & \text{if } r = 1 \end{cases} \tag{81}$$

It is now convenient to redefine the integration coordinates as follows

$$\begin{aligned}
 \xi'' &= \xi + \tau & \eta'' &= \eta + \kappa \\
 \xi''' &= \xi' - \tau & \eta''' &= \eta' - \kappa
 \end{aligned} \tag{82}$$

Accordingly, the modulation function D_3 is reformulated as \hat{D}_3

$$\begin{aligned}
 & \hat{D}_3(\xi', \eta', \xi'', \eta'', \xi''', \eta'''; k_{sz}, k_z, r q_m, r' q'_n) \\
 & \equiv D_3(\xi'' + \xi''' - \xi', \eta'' + \eta''' - \eta', \xi', \eta', \xi' - \xi''', \eta' - \eta'''; k_{sz}, k_z, r q_m, r' q'_n) \tag{83}
 \end{aligned}$$

and then the following Taylor series is carried out as

$$\begin{aligned}
 \hat{D}_3(\xi', \eta', \xi'', \eta'', \xi''', \eta'''; k_{sz}, k_z, r q_m, r' q'_n) &= \hat{D}_3^0(k, k', k'', k''') \\
 &+ \sum_{\beta=\xi', \eta', \xi'', \eta'', \xi''', \eta'''} \hat{D}_{3,\beta}^0(k, k', k'', k''') \beta \\
 &+ \frac{1}{2} \sum_{\beta, \gamma=\xi', \eta', \xi'', \eta'', \xi''', \eta'''} \hat{D}_{3,\beta, \gamma}^0(k, k', k'', k''') \beta \gamma
 \end{aligned} \tag{84}$$

The spatial coordinates can be integrated in (79) with the help of (62) to produce

$$\begin{aligned}
 (\sigma^o)_{qp}^{dc} = & \frac{k_1^2}{2\sigma^{30}A} \sum_{m,n=1,2} \sum_{r,r'=-1,1} \left\{ \int_{\mathbb{R}^4} du dv du' dv' \right. \\
 & \frac{1}{p_{sz}^{(r)5} p_{sz}'^{(r)5} p_{iz}^{(r)10} p_{iz}'^{(r)10} (|\rho_{\xi,\xi}^o| |\rho_{\eta,\eta}^o| - |\rho_{\xi,\eta}^o|^2)^{15/2}} \\
 & \mathcal{I}^c(\vec{k}^s, \vec{k}^i, \vec{l}_m^r, \vec{l}_n^{r'}) \\
 & \hat{F}_{qp}^m(\vec{k}^i, \vec{k}^s, \vec{l}_m^r) \hat{F}_{qp}^{n*}(\vec{k}^i, \vec{k}^s, \vec{l}_n^{r'}) \\
 & \cdot \exp \left\{ - \frac{(p'_{ix} - p_{sx})^2 |\rho_{\eta,\eta}^o| - 2(p'_{ix} - p_{sx})(p'_{iy} - p_{sy}) |\rho_{\xi,\eta}^o| + (p'_{iy} - p_{sy})^2 |\rho_{\xi,\xi}^o|}{2\sigma^2 p_{iz}^{(r)} p_{iz}'^{(r)} (|\rho_{\xi,\xi}^o| |\rho_{\eta,\eta}^o| - |\rho_{\xi,\eta}^o|^2)} \right\} \\
 & \cdot \exp \left\{ - \frac{p_{sx}^2 |\rho_{\eta,\eta}^o| - 2p_{sx} p_{sy} |\rho_{\xi,\eta}^o| + p_{sy}^2 |\rho_{\xi,\xi}^o|}{2\sigma^2 p_{sz}^{(r)} p_{sz}'^{(r)} (|\rho_{\xi,\xi}^o| |\rho_{\eta,\eta}^o| - |\rho_{\xi,\eta}^o|^2)} \right\} \\
 & \cdot \exp \left\{ - \frac{p_{sx}'^2 |\rho_{\eta,\eta}^o| - 2p_{sx}' p_{sy}' |\rho_{\xi,\eta}^o| + p_{sy}'^2 |\rho_{\xi,\xi}^o|}{2\sigma^2 p_{iz}^{(r)} p_{iz}'^{(r)} (|\rho_{\xi,\xi}^o| |\rho_{\eta,\eta}^o| - |\rho_{\xi,\eta}^o|^2)} \right\} \left. \right\} \quad (85)
 \end{aligned}$$

where

$$\begin{aligned}
 p_{sx} &= k_{sx} - u & p_{sy} &= k_{sy} - v & p_{sz}^{(r)} &= k_{sz} - r q_m \\
 p_{ix} &= u - k_x & p_{iy} &= v - k_y & p_{iz}^{(r)} &= r q_m - k_z \\
 p'_{sx} &= k_{sx} - u' & p'_{sy} &= k_{sy} - v' & p'_{sz}^{(r)} &= k_{sz} - r' q'_n \\
 p'_{ix} &= u' - k_x & p'_{iy} &= v' - k_y & p'_{iz}^{(r)} &= r' q'_n - k_z \quad (86)
 \end{aligned}$$

and \mathcal{I}^c is the modulation factor due to the topography. It is given in terms of a tensorial product

$$\mathcal{I}^c(\vec{k}^s, \vec{k}^i, u, v, r q_m, u', v', r' q'_n) = \sum_{i,j,k=1}^6 \hat{D}_3^{i,j,k} \check{\alpha}_i \check{\alpha}'_j \check{\alpha}''_k \quad (87)$$

where the tensor $\hat{D}_3^{i,j,k}$ is defined by the partial derivatives of D_3 on $\xi', \eta', \xi'', \eta'', \xi''', \eta'''$ in the following manner: a) the first superscript denotes the degree of derivation on the pair of variables (ξ', η') ; thus, 1 refers to no derivation, 2 and 3 refer to the first and second derivative on ξ' , 4 and 5 to the first and second derivative on η' , and 6 to the cross derivative on ξ' and η' ; b) the second and third superscripts have equivalent meanings for the pairs of variables (ξ'', η'') and (ξ''', η''') , respectively; c) all the tensor elements corresponding to derivatives of order higher than 2 are set to zero. The vectors $\check{\alpha}$, $\check{\alpha}'$ and $\check{\alpha}''$ are given by

$$\check{\alpha} = \begin{bmatrix} \check{\alpha}_A \\ \check{\alpha}_B \\ \check{\alpha}_C \\ \check{\alpha}_D \\ \check{\alpha}_E \\ \check{\alpha}_F \end{bmatrix} \quad \check{\alpha}' = \begin{bmatrix} \check{\alpha}'_A \\ \check{\alpha}'_B \\ \check{\alpha}'_C \\ \check{\alpha}'_D \\ \check{\alpha}'_E \\ \check{\alpha}'_F \end{bmatrix} \quad \check{\alpha}'' = \begin{bmatrix} \check{\alpha}''_A \\ \check{\alpha}''_B \\ \check{\alpha}''_C \\ \check{\alpha}''_D \\ \check{\alpha}''_E \\ \check{\alpha}''_F \end{bmatrix}$$

$$\begin{aligned}
\check{\alpha}_A &= \alpha_A (\kappa_4^{(r)} |\rho_{\zeta, \zeta}^o|, \kappa_4^{(r)} |\rho_{\eta, \eta}^o|, \kappa_4^{(r)} |\rho_{\zeta, \eta}^o|) \\
\check{\alpha}'_A &= \alpha_A (\kappa_5^{(r)} |\rho_{\zeta, \zeta}^o|, \kappa_5^{(r)} |\rho_{\eta, \eta}^o|, \kappa_5^{(r)} |\rho_{\zeta, \eta}^o|) \\
\check{\alpha}''_A &= \alpha_A (\kappa_6^{(r)} |\rho_{\zeta, \zeta}^o|, \kappa_6^{(r)} |\rho_{\eta, \eta}^o|, \kappa_6^{(r)} |\rho_{\zeta, \eta}^o|) \\
\check{\alpha}'_{\zeta} &= \alpha_{\zeta} (\kappa_4^{(r)} |\rho_{\zeta, \zeta}^o|, \kappa_4^{(r)} |\rho_{\eta, \eta}^o|, \kappa_4^{(r)} |\rho_{\zeta, \eta}^o|, p'_{ix} - p_{sx}, p'_{iy} - p_{sy}) \\
\check{\alpha}'_{\zeta} &= \alpha_{\zeta} (\kappa_5^{(r)} |\rho_{\zeta, \zeta}^o|, \kappa_5^{(r)} |\rho_{\eta, \eta}^o|, \kappa_5^{(r)} |\rho_{\zeta, \eta}^o|, p_{sx}, p_{sy}) \\
\check{\alpha}''_{\zeta} &= \alpha_{\zeta} (\kappa_6^{(r)} |\rho_{\zeta, \zeta}^o|, \kappa_6^{(r)} |\rho_{\eta, \eta}^o|, \kappa_6^{(r)} |\rho_{\zeta, \eta}^o|, p'_{sx}, p'_{sy}) \\
\zeta &= B, C, D, E, F \\
\kappa_4^{(r)} &= p_{iz}^{(r)} p'_{iz}{}^{(r)} \sigma^2 / 2 \\
\kappa_5^{(r)} &= p_{sz}^{(r)} p'_{iz}{}^{(r)} \sigma^2 / 2 \\
\kappa_6^{(r)} &= p_{iz}^{(r)} p'_{sz}{}^{(r)} \sigma^2 / 2
\end{aligned} \tag{88}$$

Effect of Geometrical Shadowing in Random Rough Surfaces The derivation of the far-zone scattered field with IEM2M is a second-order approach based on the Kirchhoff surface fields. As already mentioned, this causes the field components of the model to be approximations to the exact first and second-order scattered field components. One of the corrections that can be made to improve these approximations is to include the shadowing effects that are not considered in the Kirchhoff surface fields, on which the whole derivation is based. The Kirchhoff approximation fails to take account of the different states of illumination under the incident field. These states range from full illumination to complete shadowing by other parts of the surface, as well as regimes of semishadowing caused by diffraction. The replacement of semishadowed regions by sharply edged illuminated and shadowed regions is made by assuming ray paths instead of waves. This approximation is referred to as geometrical shadowing and is the type of shadowing that will be considered here.

The first well known attempt to include geometrical shadowing effects in rough surfaces was made by Beckmann Beckmann (1965). However, Brockelman and Hagfors' results Brockelman & Hagfors (1966) obtained by Monte-Carlo simulation proved to be in great disagreement with Beckmann's predictions. Two different shadowing functions were introduced by Wagner Wagner (1966) and shortly afterwards by Smith Smith (1967). Hardin Hardin (1971) extended the theory to allow the source to be at a finite height above the surface, making a special case of Wagner's theory when the source is at an infinite height. Bass and Fuks also investigated rough surface shadowing in Bass & Fuks (1979). We will follow the recent study by Kapp and Brown Kapp & Brown (1994), based on Ricciardi and Sato's work on first passage time problems for Gaussian processes Ricciardi & Sato (1983; 1986), which shows how Wagner's shadowing function can be obtained in a more rigorous way than in Wagner (1966). The theory is extended here to include topographical surfaces. Bass and Fuks' considerations Bass & Fuks (1979) regarding the shadowing phenomenon in the framework of perturbation series are also considered.

5. Shadowing: Formulation for First Crossing Problems

The random rough surface $z = \zeta(x, y)$ which serves as the target of our scattering experiment is assumed to be represented by a Gaussian distribution. Thus, for any set of n pairs

$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, the joint pdf f_{z_1, z_2, \dots, z_n} of the points $z_i = z(x_i, y_i)$ on the surface is given by

$$f_{z_1, z_2, \dots, z_n}(z_1, z_2, \dots, z_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n D^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2D\sigma^2} \sum_{i,k=1}^n D_{ik}(z_i - \bar{z}_i)(z_k - \bar{z}_k) \right\} \quad (89)$$

where

$$D = \det(\rho_{ik}), \quad \rho_{ik} = \sigma^{-2} E \{z_i z_k\}$$

and D_{ik} is the cofactor of the element ρ_{ik} in the determinant D . Each point on the surface is assumed to have a distinct mean value but all are described by a single variance σ^2 .

Let S be a ray impinging upon the surface at point $\vec{r}_0 = (x_0, y_0, z_0)$ given by

$$\begin{cases} z_S = z_0 + a(x - x_0) \\ y_S = y_0 \end{cases} \quad (90)$$

with an angle $\theta = \text{arccota}$ over the normal, where the ray is chosen to lie in the $y = y_0$ plane for convenience. The z coordinate of the ray will be written as $z_S(\vec{r}_0, x)$ in what follows. We define $g(\zeta, S, x; x_0, y_0 | z_0) dx dy_0$ as the probability that ζ will cross the incoming ray S in the interval $(x, x + dx) \times (y_0, y_0 + dy_0)$ but not in the segment $(x_0, x) \times (y_0, y_0 + dy_0)$, with $x > x_0$, given that the height at (x_0, y_0) is z_0 . The function $g(\zeta, S, x; x_0, y_0 | \zeta(x_0, y_0) = z_0)$ can be written as

$$\begin{aligned} g(\zeta, S, x; x_0, y_0 | \zeta(x_0, y_0) = z_0) dx dy_0 &= \Pr[\zeta \text{ crosses } S \text{ from below in} \\ &\quad (x, x + dx) \times (y_0, y_0 + dy_0) | \zeta(x_0, y_0) = z_0] \\ &\quad - \int_{x_0}^x dx_1 \Pr[\zeta \text{ crosses } S \text{ from below in} \\ &\quad (x, x + dx) \times (y_0, y_0 + dy_0) \text{ and in} \\ &\quad (x_1, x_1 + dx_1) \times (y_0, y_0 + dy_0)] \\ &\quad \text{but not in } x' : x' \in (x, x_1) | \zeta(x_0, y_0) = z_0] \end{aligned} \quad (91)$$

Thereby, $g(\zeta, S, x | \zeta(x_0, y_0) = z_0)$ can be found by iterating (91) to obtain the following infinite series

$$\begin{aligned} g(\zeta, S, x; x_0, y_0 | z_0) &= w_1(x; x_0, y_0 | \zeta(x_0, y_0) = z_0) \\ &\quad - \int_{x_0}^x dx_1 w_2(x, x_1; x_0, y_0 | \zeta(x_0, y_0) = z_0) \\ &\quad + \int_{x_0}^x dx_1 \int_{x_0}^{x_1} dx_2 w_3(x, x_1, x_2; x_0, y_0 | \zeta(x_0, y_0) = z_0) - \dots \\ &\quad + (-1)^n \int_{x_0}^x dx_1 \int_{x_0}^{x_1} dx_2 \dots \int_{x_0}^{x_{n-1}} dx_n \\ &\quad \quad w_{n+1}(x, x_1, \dots, x_n; x_0, y_0 | \zeta(x_0, y_0) = z_0) \\ &\quad + \dots \end{aligned} \quad (92)$$

where $w_i(x, x_1, \dots, x_{i-1}; x_0, y_0 | \zeta(x_0, y_0) = z_0) dx dx_1 \dots dx_{i-1} dy_0$ is the joint probability that the ray ζ crosses S i times from below ("up-crossing"), specifically in the intervals $(x, x + dx) \times$

$(y_0, y_0 + dy_0), (x_1, x_1 + dx_1) \times (y_0, y_0 + dy_0), \dots, (x_{i-1}, x_{i-1} + dx_{i-1}) \times (y_0, y_0 + dy_0)$, given $\zeta(x_0, y_0) = z_0$. These pdf's can be written as

$$w_i(x_1, \dots, x_i; x_0, y_0 | \zeta(x_0, y_0) = z_0) = \int_a^\infty dz'_1 \cdots \int_a^\infty dz'_i \prod_{j=1}^i (z'_j - a) \tag{93}$$

$$f_{i,i}[z_S(\vec{r}_0, x_1), \dots, z_S(\vec{r}_0, x_i); z'_1, \dots, z'_i | \zeta(x_0, y_0) = z_0]$$

with $f_{i,i}[z_S(\vec{r}_0, x_1), z_S(\vec{r}_0, x_2), \dots, z_S(\vec{r}_0, x_i); z'_1, z'_2, \dots, z'_i | \zeta(x_0, y_0) = z_0]$ being the joint pdf of $\zeta(x_k, y_0) = z_S(\vec{r}_0, x_k)$ for $k = 1, \dots, i$, conditional upon $\zeta(x_0, y_0) = z_0$. Ricciardi and Sato obtained a similar infinite series for $g(\zeta, S, x; x_0, y_0 | \zeta(x_0, y_0) = z_0)$ in Ricciardi & Sato (1983; 1986).

6. Shadowing Function

We will introduce two assumptions:

- i. the heights and slopes at the shadowing points are uncorrelated with the height of the shadowed points, and
- ii. the shadowing points are uncorrelated with each other.

Under these approximations, the joint pdf in (93) satisfies

$$f_{i,i}[z_S(\vec{r}_0, x_1), z_S(\vec{r}_0, x_2), \dots, z_S(\vec{r}_0, x_i); z'_1, z'_2, \dots, z'_i | \zeta(x_0, y_0) = z_0]$$

$$= \left(\frac{1}{2\pi\sigma\sigma'} \right)^i \prod_{k=1}^i e^{-\frac{(z_S(\vec{r}_0, x_k) - z(x_k, y_0))^2}{2\sigma^2}} e^{-\frac{(z'_k - z'(x_k, y_0))^2}{2\sigma'^2}}$$

$$= f_{1,1}[z_S(\vec{r}_0, x_1); z'_1] f_{1,1}[z_S(\vec{r}_0, x_2); z'_2] \cdots f_{1,1}[z_S(\vec{r}_0, x_i); z'_i] \tag{94}$$

The probability density function that a point on the surface at (x_0, y_0) will not be shadowed when the surface is illuminated by a plane wave of propagation vector \hat{k} is

$$W(\hat{k}; x_0, y_0) = \int_{-\infty}^\infty dz_0 W(\hat{k}; x_0, y_0 | \zeta(x_0, y_0) = z_0) p(\zeta(x_0, y_0) = z_0) \tag{95}$$

where

$$W(\hat{k}; x_0, y_0 | \zeta(x_0, y_0) = z_0) = 1 - \int_{x_0}^\infty dx g(\zeta, S, x; x_0, y_0 | \zeta(x_0, y_0) = z_0) \tag{96}$$

Combining (92), (93) and (94) with (96), we obtain

$$W(\hat{k}; x_0, y_0 | \zeta(x_0, y_0) = z_0) = e^{-G_\infty(x_0, y_0; z_0; \hat{k})} \tag{97}$$

with

$$G_\infty(x_0, y_0; z_0; \hat{k}) = \frac{1}{2\sqrt{2}\sqrt{\pi}\sigma} \int_0^\infty dx e^{-\frac{(z_S(z_0, x) - \bar{z}(x, y_0))^2}{2\sigma^2}}$$

$$\left[\sigma' \sqrt{\frac{2}{\pi}} e^{-\frac{(a - z'(x, y_0))^2}{2\sigma'^2}} - (a - z'(x, y_0)) \operatorname{erfc}\left(\frac{a - z'(x, y_0)}{\sqrt{2}\sigma'}\right) \right] \tag{98}$$

where $\bar{z}(x, y)$ represents the mean height at (x, y) , $\bar{z}(x, y)$ is the mean slope at this point and σ'^2 is the variance of the slope. Therefore,

$$W(\hat{k}; x_0, y_0) = \int_{-\infty}^\infty dz_0 p(\zeta(x_0, y_0) = z_0) e^{-G_\infty(x_0, y_0; z_0; \hat{k})} = \langle e^{-G_\infty(x_0, y_0; z_0; \hat{k})} \rangle_{z_0} \tag{99}$$

where $\langle \rangle_{z_0}$ denotes an average over z_0 values. Obtaining a 3-D shadowing function from (99) is immediate. Thus,

$$\mathcal{S}(\hat{k}^i) = W(\hat{k}) \equiv \langle e^{-G_\infty(x_0, y_0; z_0; \hat{k})} \rangle_{(x_0, y_0; z_0)} \tag{100}$$

where $\langle \rangle_{(x_0, y_0; z_0)}$ denotes an average over z_0 as well as x_0 and y_0 . The pdf for the variables x_0 and y_0 is a uniform distribution over a finite, topographical surface. Yet, it is important to note that we have assumed a surface with infinite dimensions in (96) and hence also in obtaining (97). However, it is possible to assume that the border effects due to a finite surface are negligible so (100) can be derived from (99) in order to compute the shadowing.

7. Bistatic Shadowing

The problem of shadowing is present both in the directions of incidence and scattering. Expressions for the relevant shadowing functions in first and second-order scattering events are derived in this section.

7.1 Single Scattering

To introduce a bistatic shadowing function for first-order scattering, let us first consider an incident ray S_i and a reflected or scattered ray S_s crossing a point (x_0, y_0, z_0) on the surface with angles θ_i and θ_s over the normal and slopes $a_i = \cot \theta_i$ and $a_s = \cot \theta_s$. Likewise, k_i and k_s represent the propagation vectors of the plane waves along the incident and reflected ray directions. The probability $W(A, B)$ that the surface will not cross either S_i (event A) or S_s (event B) anywhere equals the product of the probability that it will not cross S_i , W_A , and the conditional probability that it will not cross S_s given that it does not cross S_i , $W(B|A)$. Within a solid angle “pencil” or neighbourhood around the ray S_1 and up to some distance or radius from (x_0, y_0, z_0) , the event B is correlated to event A and $W(B|A) \neq W(B)$ in general.

Both this radius and the width of the pencil are proportional to the correlation length of the surface. In the surfaces we are considering there are two correlation lengths, namely, the one corresponding to the deterministic component which shapes the surface as topographical and the one corresponding to the random component. The former is larger than the latter. The correlation between the statistical events A and B is only due to the random component of the correlation. Therefore, the scope of the statistical interference of A and B is small at the scale of the whole surface. Hence, we can approximate $W(B|A) = W(B)$ for cases other than backscattering and write

$$\begin{aligned} W(\hat{k}^i, \hat{k}^s; x_0, y_0 | \zeta(x_0, y_0) = z_0) \\ = W(\hat{k}^i; x_0, y_0 | \zeta(x_0, y_0) = z_0) W(\hat{k}^s; x_0, y_0 | \zeta(x_0, y_0) = z_0) \end{aligned} \tag{101}$$

where $W(\hat{k}^i, \hat{k}^s; x_0, y_0 | \zeta(x_0, y_0) = z_0)$ is a more rigorous notation for $W(A, B)$. Hence, the following bistatic shadowing function is found

$$\mathcal{S}(\hat{k}^i, \hat{k}^s) = \langle e^{-[G_\infty(x_0, y_0; z_0; \hat{k}^i) + G_\infty(x_0, y_0; z_0; \hat{k}^s)]} \rangle_{(z_0; x_0, y_0)} \tag{102}$$

For the case of backscattering, $W(B|A) = 1$ and

$$\mathcal{S}(\hat{k}^i, \hat{k}^s) = \mathcal{S}(\hat{k}^i) \tag{103}$$

7.2 Second Order Scattering

For the case of second-order scattering we will apply, in a reiterative fashion, the result given in (102) for bistatic shadowing. However, there are some remarks to be made. First, we differentiate between those intermediate plane waves propagating through the medium below the surface, $|\vec{l}_2\rangle$, and those propagating through the incidence medium, $|\vec{l}_1\rangle$. Then it is necessary to consider that the intermediate plane waves travel both upwards and downwards. We present shadowing functions for all these four combined cases.

Let us consider the scattering event of a second-order deflection where the intermediate plane wave propagates upwards within the incidence medium. The bistatic shadowing function $\mathcal{S}(\hat{k}^i, \hat{l}_1^+)$ defined in (102) represents the fraction of the surface which scatters the incident power outwards. Therefore, $1 - \mathcal{S}(\hat{k}^i, \hat{l}_1^+)$ is the fraction of the scattered power that is once more intercepted by the surface. In the same fashion, only the fraction $\mathcal{S}(\hat{l}_1^+, \hat{k}^s)$ of the surface rescatters the power into the \hat{k}^s direction. Hence, the second order shadowing function for “reflected” intermediate waves can be written as

$$\mathcal{S}_1(\hat{k}^i, \hat{l}_1^+, \hat{k}^s) = [1 - \mathcal{S}(\hat{k}^i, \hat{l}_1^+)] \mathcal{S}(\hat{l}_1^+, \hat{k}^s) \quad (104)$$

Likewise, we obtain

$$\mathcal{S}_1(\hat{k}^i, \hat{l}_1^-, \hat{k}^s) = \mathcal{S}(\hat{k}^i, \hat{l}_1^-) \mathcal{S}(\hat{l}_1^-, \hat{k}^s) \quad (105)$$

as $\mathcal{S}(\hat{k}^i, \hat{l}_1^-)$ is the fraction of the first-order scattered power that will impinge again upon the surface.

When the intermediate wave planes propagate through the medium below the surface, the same principles as above apply. The only difference is that the computation of the bistatic shadowing functions have to be made with the surface equation $z = \zeta(x, y)$ replaced by $z = -\zeta(x, y)$. If we denote such shadowing functions as $\mathcal{S}'(\hat{k}^i, \hat{l}_2^+)$ and $\mathcal{S}'(\hat{l}_2^+, \hat{k}^s)$, the second-order shadowing for “refracted” intermediate waves is given by

$$\mathcal{S}_2(\hat{k}^i, \hat{l}_2^+, \hat{k}^s) = \mathcal{S}'(\hat{k}^i, \hat{l}_2^+) \mathcal{S}'(\hat{l}_2^+, \hat{k}^s) \quad (106)$$

$$\mathcal{S}_2(\hat{k}^i, \hat{l}_2^-, \hat{k}^s) = [1 - \mathcal{S}'(\hat{k}^i, \hat{l}_2^-)] \mathcal{S}'(\hat{l}_2^-, \hat{k}^s) \quad (107)$$

Acknowledgements

This chapter is dedicated to the memory of Tanos (Tony) Elfouhaily, an extraordinary man whose early departure left our scientific community utterly desolated.

Appendix A. The C Coefficients.

As stated before, all the expressions for f_{qp} and F_{qp} are given in Alvarez-Perez (2001), namely, in its equations (A1) to (A5) for f_{qp} and (B3) to (B10) for F_{qp} . Equations (A5) were criticized as a result of a misinterpretation: they do not imply any relationship for the fields but only for the scattering coefficient as effective reflection coefficients, yet they do guarantee reciprocity. However, the C coefficients are written there in a very general manner which requires a great deal of work by the implementer. Here these coefficients are worked out and incorporate the

first power of the $\pm(k_{sz} + k_z)$, $(k_{sz} \mp k_z^{(2)})$ and $(\pm k_{sz}^{(2)} - k_z)$ included in the i_n 's.

$$C_1(k_x, k_y, -k_z) = -C_1(k_{sx}, k_{sy}, -k_{sz}) = k_1 \cos \phi_s (\cos \theta_s - \cos \theta)$$

$$C_1(k_x, k_y, \pm k_z^{(2)}) = \cos \phi_s (k_1 \cos \theta_s \mp \sqrt{k_2^2 - k_1^2 \sin^2 \theta})$$

$$C_1(k_{sx}, k_{sy}, \pm k_{sz}^{(2)}) = \cos \phi_s (k_1 \cos \theta \pm \sqrt{k_2^2 - k_1^2 \sin^2 \theta_s})$$

$$C_2(k_x, k_y, -k_z) = k_1^2 \cos \theta (\cos \phi_s - \cos \phi_s \cos \theta \cos \theta_s - \sin \theta \sin \theta_s)$$

$$C_2(k_{sx}, k_{sy}, -k_{sz}) = k_1^2 \cos \theta_s (\cos \phi_s - \cos \phi_s \cos \theta \cos \theta_s - \sin \theta \sin \theta_s)$$

$$C_2(k_x, k_y, \pm k_z^{(2)}) = \cos \theta [\cos \phi_s (k_2^2 \mp k_1 \cos \theta_s \sqrt{k_2^2 - k_1^2 \sin^2 \theta}) - k_1^2 \sin \theta \sin \theta_s]$$

$$C_2(k_{sx}, k_{sy}, \pm k_{sz}^{(2)}) = \pm k_1 \sin \theta \sin \theta_s \sqrt{k_2^2 - k_1^2 \sin^2 \theta_s} \\ - \cos \phi_s [\cos \theta (k_2^2 - k_1^2 \sin^2 \theta_s) \pm k_1 \sqrt{k_2^2 - k_1^2 \sin^2 \theta_s}]$$

$$C_3(k_x, k_y, -k_z) = -k_1^2 \sin \theta (\cos \phi_s \cos \theta_s \sin \theta - \cos \theta \sin \theta_s)$$

$$C_3(k_{sx}, k_{sy}, -k_{sz}) = -k_1^2 \sin \theta_s (\cos \phi_s \cos \theta \sin \theta_s - \cos \theta_s \sin \theta)$$

$$C_3(k_x, k_y, \pm k_z^{(2)}) = -k_1 \sin \theta (k_1 \cos \phi_s \cos \theta_s \sin \theta \mp \sqrt{k_2^2 - k_1^2 \sin^2 \theta} \sin \theta_s)$$

$$C_3(k_{sx}, k_{sy}, \pm k_{sz}^{(2)}) = -k_1 \sin \theta_s (k_1 \cos \phi_s \cos \theta \sin \theta_s \pm \sqrt{k_2^2 - k_1^2 \sin^2 \theta_s} \sin \theta)$$

$$C_4(k_x, k_y, -k_z) = -k_1 \cos \theta [\cos \phi_s (\cos \theta \cos \theta_s - 1) + \sin \theta \sin \theta_s]$$

$$C_4(k_{sx}, k_{sy}, -k_{sz}) = -k_1 \cos \theta_s [\cos \phi_s (\cos \theta \cos \theta_s - 1) + \sin \theta \sin \theta_s]$$

$$C_4(k_x, k_y, \pm k_z^{(2)}) = \cos \theta [\cos \phi_s \cos \theta_s (k_1 \cos \theta_s \mp \sqrt{k_2^2 - k_1^2 \sin^2 \theta}) \\ - k_1 \sin \theta_s (\sin \theta - \cos \phi_s \sin \theta_s)]$$

$$C_4(k_{sx}, k_{sy}, \pm k_{sz}^{(2)}) = \cos \theta_s [\cos \phi_s (k_1 \pm \cos \theta \sqrt{k_2^2 - k_1^2 \sin^2 \theta_s}) - k_1 \sin \theta \sin \theta_s]$$

$$C_5(k_x, k_y, -k_z) = C_2(k_x, k_y, -k_z)$$

$$C_5(k_{sx}, k_{sy}, -k_{sz}) = C_2(k_{sx}, k_{sy}, -k_{sz})$$

$$C_5(k_x, k_y, \pm k_z^{(2)}) = \pm \sqrt{k_2^2 - k_1^2 \sin^2 \theta} [\cos \phi_s \cos \theta_s (k_1 \cos \theta_s \mp \sqrt{k_2^2 - k_1^2 \sin^2 \theta}) \\ - k_1 \sin \theta_s (\sin \theta - \cos \phi_s \sin \theta_s)]$$

$$C_5(k_{sx}, k_{sy}, \pm k_{sz}^{(2)}) = \cos \theta_s [\cos \phi_s (k_2^2 \pm k_1 \cos \theta \sqrt{k_2^2 - k_1^2 \sin^2 \theta_s}) - k_1^2 \sin \theta \sin \theta_s]$$

$$C_6(k_x, k_y, -k_z) = C_6(k_{sx}, k_{sy}, -k_{sz}) = C_6(k_x, k_y, \pm k_z^{(2)}) = C_6(k_{sx}, k_{sy}, \pm k_{sz}^{(2)}) = 0$$

$$\begin{aligned}
C_7(k_x, k_y, -k_z) &= -k_1 \sin \phi_s (\cos \theta \cos \theta_s - 1) \\
C_7(k_{sx}, k_{sy}, -k_{sz}) &= k_1 \cos \theta_s \sin \phi_s (\cos \theta - \cos \theta_s) \\
C_7(k_x, k_y, \pm k_z^{(2)}) &= \sin \phi_s [\cos \theta_s (k_1 \cos \theta_s \mp \sqrt{k_2^2 - k_1^2 \sin^2 \theta}) + k_1 \sin^2 \theta_s] \\
C_7(k_{sx}, k_{sy}, \pm k_{sz}^{(2)}) &= \cos \theta_s \sin \phi_s (k_1 \cos \theta \pm \sqrt{k_2^2 - k_1^2 \sin^2 \theta_s}) \\
\\
C_8(k_x, k_y, -k_z) &= k_1^2 \cos \theta \sin \phi_s (\cos \theta_s - \cos \theta) \\
C_8(k_{sx}, k_{sy}, -k_{sz}) &= k_1^2 \cos \theta_s \sin \phi_s (\cos \theta_s - \cos \theta) \\
C_8(k_x, k_y, \pm k_z^{(2)}) &= \cos \theta \sin \phi_s (k_2^2 \cos \theta_s \mp k_1 \sqrt{k_2^2 - k_1^2 \sin^2 \theta}) \\
C_8(k_{sx}, k_{sy}, \pm k_{sz}^{(2)}) &= -\cos \theta_s \sin \phi_s (k_2^2 \cos \theta \pm k_1 \sqrt{k_2^2 - k_1^2 \sin^2 \theta_s}) \\
\\
C_9(k_x, k_y, -k_z) &= -k_1^2 \sin \phi_s \sin^2 \theta (\cos^2 \theta_s + \sin \theta \sin \theta_s) \\
C_9(k_{sx}, k_{sy}, -k_{sz}) &= 0 \\
C_9(k_x, k_y, \pm k_z^{(2)}) &= C_9(k_x, k_y, -k_z) \\
C_9(k_{sx}, k_{sy}, \pm k_{sz}^{(2)}) &= 0 \\
\\
C_{10}(k_x, k_y, -k_z) &= k_1 \cos \theta \sin \phi_s (\cos \theta - \cos \theta_s) \\
C_{10}(k_{sx}, k_{sy}, -k_{sz}) &= k_1 \sin \phi_s (\cos \theta \cos \theta_s - 1) \\
C_{10}(k_x, k_y, \pm k_z^{(2)}) &= -\cos \theta \sin \phi_s (k_1 \cos \theta_s \mp \sqrt{k_2^2 - k_1^2 \sin^2 \theta}) \\
C_{10}(k_{sx}, k_{sy}, \pm k_{sz}^{(2)}) &= -\sin \phi_s (k_1 \pm \cos \theta \sqrt{k_2^2 - k_1^2 \sin^2 \theta_s}) \\
\\
C_{11}(k_x, k_y, -k_z) &= -C_8(k_x, k_y, -k_z) \\
C_{11}(k_{sx}, k_{sy}, -k_{sz}) &= -C_8(k_{sx}, k_{sy}, -k_{sz}) \\
C_{11}(k_x, k_y, \pm k_z^{(2)}) &= \mp \sqrt{k_2^2 - k_1^2 \sin^2 \theta} \sin \phi_s (k_1 \cos \theta_s \mp \sqrt{k_2^2 - k_1^2 \sin^2 \theta}) \\
C_{11}(k_{sx}, k_{sy}, \pm k_{sz}^{(2)}) &= \mp \sqrt{k_2^2 - k_1^2 \sin^2 \theta_s} \sin \phi_s (k_1 \cos \theta \pm \sqrt{k_2^2 - k_1^2 \sin^2 \theta_s}) \\
\\
C_{12}(k_x, k_y, -k_z) &= C_{12}(k_x, k_y, \pm k_z^{(2)}) = 0 \\
C_{12}(k_{sx}, k_{sy}, -k_{sz}) &= C_{12}(k_{sx}, k_{sy}, \pm k_{sz}^{(2)}) = -k_1^2 \sin \phi_s \sin^2 \theta_s
\end{aligned}$$

With these expressions it is straightforward to prove that the SPM limit for the most general case of bistatic scattering is reached when we take (1) to first order in σ^2 . Probably the formal character of the C 's as given in Alvarez-Perez (2001) has precluded other authors to properly implement the model, as it is the case in Fung et al. (2002) or Du (2008), where incorrect IEM2M results were provided. A Mathematica version of the code is available from the author upon request.

8. References

- Alvarez-Perez, J. L. (2001). An extension of the IEM/IEMM surface scattering model, *Waves Random Media* **11**: 307–29.
- Bass, F. G. & Fuks, I. M. (1979). *Wave Scattering from Statistically Rough Surfaces*, Pergamon Press.
- Beckmann, P. (1965). Shadowing of rough surfaces, *IEEE Transactions on Antennas and Propagation* **13**: 384–388.
- Brockelman, R. A. & Hagfors, T. (1966). Note on the effect of shadowing on the backscattering of waves from random rough surfaces, *IEEE Transactions on Antennas Propagation* **14**(5): 621–629.
- Chen, K., Wu, T., Tsang, L., Li, Q., Shi, J. & Fung, A. (2003). Emission of rough surfaces calculated by the integral equation method with comparison to three-dimensional moment method simulations, *IEEE Transactions on Geoscience and Remote Sensing* **41**(1): 90–101.
- Chen, K., Wu, T., Tsay, M. & Fung, A. (2000). A note on the multiple scattering in an IEM model, *IEEE Transactions on Geoscience and Remote Sensing* **38**(1): 249–256.
- Du, Y. (2008). A new bistatic model for electromagnetic scattering from randomly rough surfaces, *Waves in Random and Complex Media* **18**(1): 109–128.
- Elfouhaily, T. M. & Guerin, C.-A. (2004). A critical survey of approximate scattering wave theories from random rough surfaces, *Waves in Random Media* **14**(4): R1–R40.
- Fung, A. & Chen, K. (2004). An update on the IEM surface backscattering model, *IEEE Geoscience and Remote Sensing Letters* **1**(2): 75–77.
- Fung, A. K. (1994). *Microwave scattering and emission models and their applications*, Artech House.
- Fung, A., Liu, W., Chen, K. & Tsay, M. (2002). An improved IEM model for bistatic scattering from rough surfaces, *Journal of Electromagnetic Waves and Applications* **16**(5): 689–70.
- Fung, A. & Pan, G. W. (1986). An integral equation model for rough surface scattering, *Proceedings of the International Symposium on Multiple Scattering of Wave in Random Media and Random Surface* (Pennsylvania State University Press) pp. 701–714.
- Hardin, J. C. (1971). Theoretical analysis of rough surface shadowing from point source radiation, *Journal of the Acoustical Society of America* **52**: 227–233.
- Hsieh, C.-Y., Fung, A. K., Nesti, G., Sieber, A. J. & Coppo, P. (1997). A further study of the IEM surface scattering model, *IEEE Transactions on Geoscience and Remote Sensing* **35**(4): 901–909.
- Kapp, D. A. & Brown, G. S. (1994). Effect of correlation between shadowing and shadowed points in rough surface scattering, *IEEE Transactions on Antennas and Propagation* **42**(8): 11554–1160.
- Poggio, A. J. & Miller, E. K. (1973). *Integral Equation Solution of three dimensional scattering problems (in Computer Techniques for Electromagnetics, ed. R. Mittra)*, Pergamon.
- Ricciardi, L. M. & Sato, S. (1983). A note on first passage time problems for Gaussian processes and varying boundaries, *IEEE Transactions on Information Theory* **29**(3): 454–457.
- Ricciardi, L. M. & Sato, S. (1986). On the evaluation of first passage time densities for Gaussian processes, *Signal Processing* **11**: 339–357.
- Smith, B. G. (1967). Geometrical shadowing of a random rough surface, *IEEE Transactions on Antennas and Propagation* **15**(5): 668–671.
- Wagner, R. J. (1966). Shadowing of randomly surfaces, *Journal of the Acoustical Society of America* **41**(1): 138–147.

- Wu, T. & Chen, K. (2004). A reappraisal of the validity of the IEM model for backscattering from rough surfaces, *IEEE Transactions on Geoscience and Remote Sensing* **42**(4): 743–53.
- Wu, T., Chen, K., Shi, J., Lee, H.-W. & Fung, A. (2008). A study of an AIEM model for bistatic scattering from randomly rough surfaces, *IEEE Transactions on Geoscience and Remote Sensing* **46**(9): 2584–98.

Microwave Remote Sensing of Soil Moisture in Semi-arid Environment

A. K. M. Azad Hossain and Greg Easson
*The University of Mississippi
United States of America*

1. Introduction

Soil moisture, defined as the water content in the upper layer of soil, is the hydrologic variable that controls the interactions (and feedbacks) between land surface and atmospheric processes (Hossain and Anagnostou, 2005). Soil moisture is important in the distribution of precipitation between runoff and infiltration (Baghdadi et al., 2006). Soil moisture monitoring and characterization of the spatial and temporal variability of soil moisture at scales from small catchments to large river basins is important in the understanding of subsurface - land surface - atmospheric interactions as well as, drought analysis, crop yield forecasting, irrigation planning, flood protection, and forest fire prevention (Georgakakos and Baumer, 1996; Robock et al., 2003). Surface soil moisture distribution information is a critical forcing variable in many Soil Vegetation Atmosphere Transfer (SVAT) models to estimate profile soil moisture at daily time steps. Soil moisture distribution also plays a key role in the prediction of erosion and sediment loads in watershed streams and ponds. In arid and semi-arid watersheds soil moisture content has been used as a surrogate indicator of general plant health (Moran et al., 2004).

Research in the extraction of soil surface moisture information from remotely sensed imagery has been an important research topic in the last decade. Optical remote sensing data have been used successfully for mapping and monitoring relative variations in soil moisture when reflective data are combined with thermal data from the same sensors (e.g., Carlson et al., 1995; Lambin and Ehrlich, 1996; Gillies et al., 1997; Hossain and Easson, 2008 & 2006). Microwave remote sensing techniques provide a direct measurement of the surface soil moisture for a range of vegetation cover conditions within reasonable error bounds (Jackson, 2002). Passive microwave remote sensing uses radiometers that detect and measure the natural thermal microwave emissions of a particular frequency, within a narrow band. This measurement provides the brightness temperature, which includes contributions from the atmosphere, reflected sky radiation, and the land surface. The Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E) sensor on NASA's Aqua satellite and the Soil Moisture and Ocean Salinity (SMOS) mission of European Space Agency (ESA) are the two major currently operating passive microwave sensors dedicated to soil moisture mapping. Active microwave sensors produce energy and

measure the amount of energy returned from the target to yield a variable called the backscattering coefficient (σ° or β°). The backscattering coefficient is related to the surface reflectivity, which is used to determine surface soil moisture (Ulaby et al., 1986). Radarsat 2 Synthetic Aperture Radar (SAR), ENVISAT Advance Synthetic Aperture Radar (ASAR) and Advanced Land Observing Satellite (ALOS) Phased Array Type L-band Synthetic Aperture Radar (PALSAR) are the most widely used currently operating spaceborne active microwave sensors.

This chapter provides a brief overview of active microwave soil moisture remote sensing and presents a case study of soil moisture mapping in a semi-arid environment. The overview of active microwave soil moisture remote sensing includes a discussion of the basic principles of microwave remote sensing emphasizing soil surface moisture information extraction and different methods for soil moisture estimation using SAR data with emphasis on the existing algorithms for SAR based soil moisture mapping in semi-arid environment. The case study presents research in southeastern New Mexico that explored the linear and the non-linear relationships between radar reflectivity (backscatter) and soil moisture and determined the impact of vegetation in soil moisture estimation.

1.1 Basic Principles of Synthetic Aperture Radar (SAR)

RADAR (RADio Detection And Ranging) sensors operate in the microwave portion of the electromagnetic spectrum beyond the visible and thermal infrared regions. Imaging radars are generally considered to include wavelengths from 1 mm to 1 m. RADAR is an active sensor, transmitting a signal of electromagnetic energy, illuminating the terrain, and recording or measuring the response returned from the target or surface. Thus, the term “active microwave” is often synonymous with radar (Henderson and Lewis, 1998). As an active sensor, radars are independent of the sun and sun conditions and can operate day or night. Radar can, in effect, collect data on a 24 hour basis. Unlike optical sensors, imaging radars are not affected by cloud or haze and operate generally independent of weather conditions.

SAR

Traditionally (before 1978) radar imaging was conducted using Real Aperture Radar (RAR) systems. RAR transmits a narrow angle beam of pulse radio wave in the range direction at right angles to the flight direction (called the azimuth direction) and receives the backscatter from the targets, which is transformed into a radar image from the received signal. Aperture is the opening used to collect the reflected energy used to form an image. In the case of radar imaging the aperture is the antenna and for RAR systems, only the amplitude of each echo return is measured and processed. The spatial resolution of a RAR system is mainly determined by the size of the antenna. For any given wavelength, the larger the antenna the better the spatial resolution. It is difficult to attach large antenna to aircraft or spaceborne sensor systems. For example a 1 km diameter antenna is needed in order to obtain 25 m resolution with L band ($\lambda=25$ cm) at a distance of 100 km from a target.

In order to overcome this limitation, radar systems with a synthetic aperture have been developed, which simulate an artificial (or virtual) antenna. Synthetic Aperture Radar (SAR) takes advantage of the Doppler history of the radar echoes generated by the forward motion

of the spacecraft to synthesize a large antenna, enabling high azimuthal resolution in the resulting image using a physically small antenna and longer radar wavelength.

Active Microwave Region in EMS

The microwave portion of the Electromagnetic Spectrum (EMS) is large, relative to the visible, and there are several wavelength ranges or bands used in radar imaging. Imaging radar systems operate at specific wavelengths or frequencies in the EMS. The active microwave regions include X, C, L and K band, which refers to specific segments of the microwave portion of the EMS. For example, an X band system would be radar that operates at a single wavelength within this band (e.g., 3.2 cm) (Henderson and Lewis, 1998). Most of the spaceborne radar systems operate in C (5.7 cm) and L (24 cm) bands.

Radar Imaging Geometry

Interpretation of SAR imagery for information extraction requires an understanding of radar imaging geometry, the nature of interaction between radar energy and surface features, and the parameters used to characterize the performance of different SAR systems. Figure 1 is a schematic diagram that illustrates the geometry of radar imaging and related radar terminology. In radar imaging systems, the platform (a) travels forward in the flight direction (b) with the nadir point (c) directly beneath the platform. The microwave beam (k) is transmitted obliquely at right angles to the direction of flight, illuminating a swath (usually the width of the imaging area) (f), which is offset from nadir. Range (e) refers to the across-track dimension perpendicular to the flight direction (b), while azimuth (d) refers to the along-track dimension parallel to the flight direction (b). Near range (h) is the portion of the imaging swath closest to the nadir track and far range (g) is the portion of the imaging swath farthest from the nadir track. Depression angle (α) is the angle between the horizontal and a radar ray path. Slant range distance (i) is the radial line of sight distance between the radar and each target on the surface. Ground range distance is the true horizontal distance along the ground corresponding to each point measured in slant range. Incidence angle (θ) is the angle between the radar beam and the perpendicular to the ground surface. Look angle (β) is the angle at which the radar looks at the surface, or the angle between vertical and the ray path.

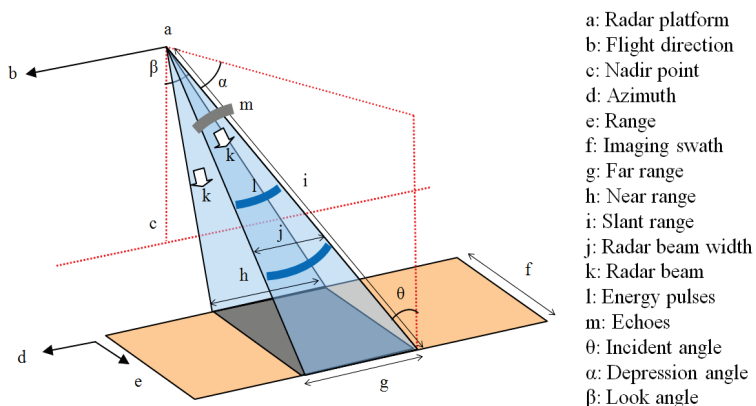


Fig. 1. Radar imaging geometry (adopted from Henderson and Lewis, 1998).

Target Interaction

In active microwave system, the energy received by a radar derives from one or more individual reflections. In a typical scene, there are many scatterers contributing to the energy received from a given region, and the locations of the individual scatterers are random (not necessarily maintaining any pattern). Usually the net signal for the scene is described by using an average over the region, leading to a distributed or diffuse scatterer model (Raney, 1998). In active systems, the brightness or darkness of the image is dependent on the amount of the transmitted energy that is returned back to the radar from targets on the surface as a measure of reflectivity (backscatter). Bright areas are produced by strong radar responses and darker areas are from weaker radar responses. Sigma-naught (σ^0) is commonly used to describe the average reflectivity or scattering co-efficient of a radar scene. Beta-naught (β^0) interprets the brightness estimates of mean reflectivity. Beta-naught (β^0) separates the radiometric response and reflectivity dependent on the terrain properties, such as local slope. According to Glen and Carr (2004):

$$\sigma^0 = \beta^0 + 10 \log_{10}(\sin l) \quad (1)$$

$$\beta^0 = 10 \log_{10}(DN^2 + A_3 / A_2) \quad (2)$$

Where, DN = Digital number, A_3 = fixed offset from the radiometric data record, A_2 = look-up table (LUT) value, and l = local incidence angle.

Beta-naught (β^0) is suggested as standard terminology for the output product of most imaging radars (Raney, 1998).

Response to radar energy (for a given wavelength and polarization) by the target is primarily dependent on three factors: surface roughness, local incident angle and the dielectric constant of the target materials (Raney, 1998). Surface Roughness is measured as the average height variation in the surface cover (measured in cm). Rayleigh (1945) proposed that the criterion for roughness depended on three parameters; incident angle, wavelength and surface irregularities. A surface is considered smooth if the height variations are smaller than the radar wavelength. Specular reflection is caused by a smooth surface where the incident energy is reflected and not backscattered. This results in smooth surfaces appearing as darker toned areas on an image. Diffuse reflection is caused by a rough surface, which scatters the energy equally in all directions. A significant portion of the energy will be backscattered to the radar, such that a rough surface will appear lighter in tone on an image.

Corner reflection occurs when the target object reflects most of the energy directly back to the antenna resulting in a very bright appearance to the object. This occurs where there are buildings, metallic structures (urban environments) and cliff faces, folded rock (natural environments).

Image geometry and radiometry are influenced by the angle of the incident illumination with respect to the local slope of the scene towards the radar. The image brightness per pixel is sensitive to the local incident angle, however, variations in the aspect angle due to the component of slope in the azimuth direction has negligible impact on image brightness (Guindon, 1990). Maxwell's formulation considered the propagation of electromagnetic

waves through media characterized by certain physical constants, one of which is the dielectric constant. The dielectric constant (or the complex permittivity) is the principal description of the medium's response to the presence of an electrical field (Raney, 1998). The dielectric constant is measured as the ratio of the permittivity of a substance to the permittivity of free space. The material's dielectric constant depends weakly on frequency, but its loss tangent depends strongly on frequency (Raney, 1998).

System Parameters

Along with wavelength (or frequency), polarization and resolution are the two most important system parameters that are used to interpret the performance of an imaging radar system. Polarization of the radar signal is the orientation of the electromagnetic field and is a significant factor by which the radar signal interacts with objects on the ground, reflecting back the resulting energy. Most radar imaging sensors are designed to transmit microwave radiation either horizontally polarized (H) or vertically polarized (V), and receive either the horizontally or vertically polarized backscattered energy. Polarizing radar has four possible combinations of both transmit and receive polarizations: HH - for horizontal transmit and horizontal receive, VV - for vertical transmit and vertical receive, HV - for horizontal transmit and vertical receive, (cross-polarized), VH - for vertical transmit and horizontal receive (cross-polarized). The spatial resolution of a radar system is the ability to distinguish between different objects, and it is dependent on the properties of the microwave radiation and geometric effects. There are two types of spatial or ground resolution: range resolution (across-track resolution) and azimuth resolution (along-track resolution). Range resolution requires that the objects be separated by more than half the pulse length. Azimuth resolution is dependent on the angular width of the radiated microwave beam and the slant range distance. The azimuth resolution becomes more coarse with increasing distance from the sensor. More detail discussion on the imaging radar's system parameters can be found in Henderson and Lewis (1998). Table 1. shows the system parameters of commonly used radar imaging systems.

System parameters	Radarsat 1/2 SAR	ERS SAR	ENVISAT ASAR	ALOS PALSAR
Incidence angle (°)	20-50	23	15-45	10-51
SAR Band	C	C	C	L
Wavelength (cm)	5.7	5.7	5.7	23
Polarization	HH	VV	HH, VV, VH, HV	HH, VV, VH, HV
Resolution (m)	3-100	30	10-100	10-100
Repeat pass	24	35	35	46

Table 1 System parameters of different SAR platforms

1.2 Estimation of Soil Moisture Using SAR

Measuring and mapping soil moisture has been investigated using scatterometers, satellites, space shuttles, and airborne synthetic aperture radars (SAR) for many years. SAR data are well suited for estimating soil moisture due to the relationship of the dielectric constant and soil moisture. According to Ulaby et al. (1987), for a given soil condition (roughness or texture) radar backscatter is linearly dependent on volumetric moisture (m_v) in the upper 2 to 5 cm of soil with a correlation $R^2 \sim 0.8$ to 0.9 .

In bare soil conditions, radar scattering is determined by surface roughness (geometry of the air soil boundary) and the microwave dielectric properties of the soil medium. The geometric factors affect the shape of the scattering pattern for an incident wave while the dielectric properties control the magnitude of reflection, absorption and transmission (Dobson and Ulaby, 1998). The average dielectric properties of the soil medium are dependent on the engineering properties of the soil including moisture content, density, texture, mineralogy and fluid chemistry. The dielectric constant of a material consists of two parts: real (ϵ') and imaginary (ϵ''). In case of a perfectly dry soil, the relative dielectric constant is independent of soil type (Dobson et al., 1985). Dielectric constant in liquid water shows strong dependence on the microwave frequency and weak sensitivity to physical temperature. When compared to dry soil the real part of the relative dielectric constant is 30 times greater and the imaginary part is about two orders of magnitude larger (Dobson and Ulaby, 1998). Due to the relationship between the dielectric constant in liquid water and microwave frequency the addition of water in liquid form to soil changes the dielectric constant of the mixture markedly.

The penetration depth into a soil by microwave is proportional to radar wavelength (Dobson and Ulaby, 1998). Significant penetration can occur in low loss materials, such as arid soils. Maximum penetration can occur in dry soil condition, whereas, the least penetration will occur in wet soils.

Studies, particularly in the past decade, have resulted in many methods, algorithms, and models relating satellite-based imagery from radar backscatter to surface soil moisture, however, no operational algorithm exists using radar data acquired by existing spaceborne sensors (Borgeaud and Saich, 1999). According to Moran et al. (2004) the promising approaches using SAR sensors for soil moisture estimation include: semi-empirical approaches, change detection, data fusion, and SAR with microwave scattering models.

Semi-empirical algorithms generally use SAR imagery of single wavelength, incident angle and polarization. Multiple passes and/or ancillary information are required for better accuracy. This approach is often scene or site dependent (Moran et al., 2000; Quesney et al., 2000). The change detection algorithms require multiple passes of SAR data and this approach has potential as an operational application (Engman, 1994). The reason for the operational suitability of this approach is that the algorithm is based on the assumption that the temporal variability of surface roughness and vegetation is at longer time scale than that of soil moisture content, and therefore, the change in radar backscatters between repeat passes results from the change in soil moisture. (Lu and Meyer, 2002). Research has been conducted into data fusion using passive and active microwave data, and microwave and optical data. Passive-active microwave data fusion algorithms use active backscattering coefficient to determine fine resolution vegetation biomass and surface roughness, and passive brightness temperature to estimate soil moisture content (Bindlish and Barros, 2002). The microwave-optical data fusion algorithms simplify the inverse problem for soil moisture content estimation on the basis of complementarily or interchangeability of optical and SAR data (Changey et al., 1995). Empirical, semi-empirical and theoretical microwave scattering models are available for use with the SAR data for soil moisture estimation. The Water Cloud Model (WCM) (Attema and Ulaby, 1978) and the Integral Equation Model

(IEM) (Fung et al., 1992; Colpitts, 1998) are common examples of this kind of models. The models are inverted to estimate soil moisture content from radar backscattering coefficient. Soil moisture estimation accuracy is higher by this approach, however, model parameterization is complex. These SAR based soil moisture estimation algorithms are described and explained in detail in Henderson and Lewis (1998) and in Moran et al. (2004).

1.3 Microwave Remote Sensing of Soil Moisture in Semi-arid Environment

The presence of vegetative cover introduces complexity into soil moisture mapping due to the interaction of the microwave energy with the vegetation and soil. Depending on the amount of vegetation present, its dielectric properties, height and geometry (size, shape and orientation of its component parts), the sensitivity of microwave backscatter to volumetric soil moisture may be significantly reduced. Previous studies indicate that in a semi-arid environment the influence of sparse vegetation on the linear relationship between radar backscatter and soil moisture is negligible or can be ignored (Thomas et al., 2004; Lin and Wood, 1993; and Dubois et al., 1995).

The research results presented in this chapter attempted to verify this concept in the semi-arid environment of southeastern New Mexico. This research also explores the non-linear relationship between soil moisture and radar backscatter, and investigates the impact of vegetation in soil moisture estimation using microwave imagery in this area.

2. Study Site

Nash Draw, located in part of the northeastern Chihuahuan desert in southeastern New Mexico (Figure 2), is a karst valley that developed in response to subsurface dissolution of evaporites and subsidence of the overlying strata (Holt et al., 2005). It is a complex example of the localized effects of evaporite karst on surface topography, near-surface geology, and hydrology (Powers et al., 2006). Although this area is in a semi-arid environment, the vegetation pattern is not uniform. Nash Draw covers an area of 400 sq. km and a subset area of 225 sq. km was selected as the study site.

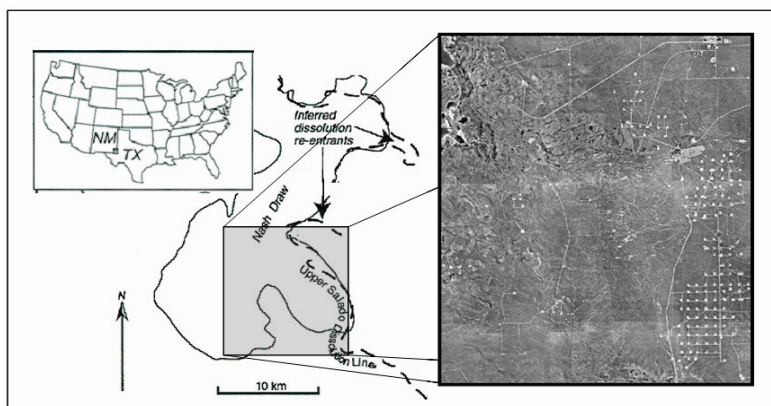


Fig. 2. Location of study site (Modified from Holt et al., 2005)

3. Data Used

Rainfall in Nash Draw is unreliable and erratic. Hydrologic data shows that August is commonly the wettest part of the season and October marks the end of the rainy season. It was assumed that imagery acquired during the months of August through November would record the maximum variation of soil moisture in the study site.

3.1 Synthetic Aperture Radar (SAR) Imagery

Due to its high spatial resolution, Radarsat 1 SAR Fine imagery was considered to be the best imagery for estimating soil moisture in the study site. The Alaska Satellite Facility (ASF) in Fairbanks, Alaska served as the Radarsat 1 data node for the United States. The ASF acquired 5 scenes of Radarsat 1 SAR Fine imagery covering the Nash Draw area for this research. The imagery was acquired at 10 m spatial resolution with 50 X 50 km swath coverage. The scenes were acquired in descending mode at 37 ° incidence angle. The image acquisition dates include August 02 and 26, September 19, October 13 and November 06 of 2006. Figure 3 shows the acquired SAR imagery.

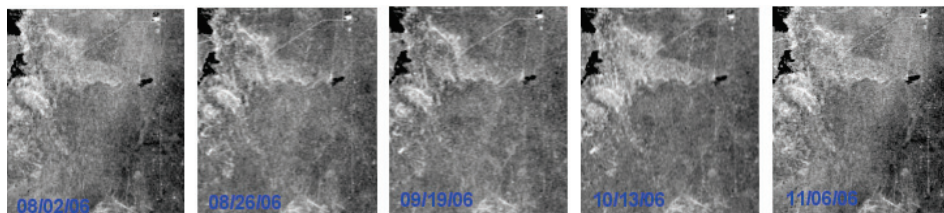


Fig. 3. Acquired SAR imagery

SAR Pre-processing

Initially all acquired SAR imagery were received as Level 0 products and then converted to level 1 products. These Level 1 SAR data have undergone several pre-processing phases including terrain correction, calibration and filtering to be ready for soil moisture estimation.

Data Calibration

An initial experiment was conducted using one radar scene to determine the suitability of σ° or β° for soil moisture estimation in our study site. It was found that backscatter values as β° had a better correlation with field measured soil moisture than backscatter values as σ° in our study site (Hossain and Easson, 2007). We believe that due to the low topographic relief in the study site variations in radar backscatter expressed as β° produced better results. Based on these observations we decided to calibrate all the radar scenes as β° for this project.

Removing Speckles

Coherent imaging systems produce images with a granular appearance, with a multitude of bright and dark spots caused by random constructive and destructive interference of the wavelets returning from the various scatterers within the resolution cell of the system (Goodman, 1975). From the mathematical point of view, the effect of this interference process can be regarded as a multiplicative noise, called speckle. In Synthetic Aperture

Radar (SAR) imagery, the presence of speckle affects the procedures for texture class discrimination. Speckle noise needs to be reduced to preserve edges and image texture (D'Elia et al., 2004). The most well known model used as the basis for the development of many of the existing speckle filters is the multiplicative model with an exponential probability density function (Touzi, 2002). To remove the speckles in the radar scenes we applied different types of filtering techniques with different window sizes. We found 5x5 Lee filtering provided better results and this method was used to de-speckle the acquired radar scenes.

Geometric Correction

The processed (calibrated and filtered) SAR data and the GIS coverage of the sample locations were georectified to the same projection system. All imagery and GIS data were georectified using recent aerial photograph and Universal Transverse Mercator (UTM) projection system.

3.2 Elevation Data

The knowledge of the local incident angle is essential for the quantitative estimation of soil moisture and roughness from SAR data. In the absence of topographic relief, the local incident angle equals the radar look angle. This is not true for terrain with larger topographic variations where the local incident angle becomes a function of the radar look angle and the local terrain slope. This makes the straightforward surface parameter estimation difficult (Hajnsek and Pottier, 2000). It is necessary to terrain correct the SAR data to allow geometric overlays of remotely sensed data from different sensors and/or geometries. The average elevation of the study site varies from 900 m to 1100 m, which is considered low relief for processing SAR imagery. However, despite the minimal topographic influence, a terrain correction was performed using USGS 30 m digital elevation model (DEM) to achieve greater geometric accuracy of the data.

3.3 Insitu Soil Moisture Data

Near-real time soil moisture data is needed to quantitatively map soil moisture with reasonable accuracy from SAR data. Soil samples were collected in selected parts of Nash Draw and analyzed to calculate volumetric soil moisture for calibration of the SAR imagery acquired on August 02, 2006 and November 06, 2006. Eighty soil samples were collected within a site covering 225 sq. km in Nash Draw.

Sampling Technique

A stratified soil sampling technique (Dane and Topp, 2002) was used in the acquisition of the soil samples. Using this method, the study site was divided into several grids and a simple random sampling technique was used in each grid with prior definition of sample size. The study site was divided into 4 equal parts and random sample points were selected in each part using a 500 m grid spacing, with 20 samples collected in each quadrant. Accessibility and variation in soil types was given preference for selecting sampling sites. Samples were collected for measuring volumetric soil moisture. Figure 4 shows the distribution of the collected soil sample locations in the study site.

RADARSAT 1 acquires imagery in HH polarization and in C band with 5.7 cm wavelength and 5.3 GHz frequency (RADARSAT International, 1995). The RADARSAT 1 beam should be able to penetrate the ground at least up to 3 cm in dry conditions. Therefore, sample collection was limited to within 3 cm below the land surface. A cylinder with 3 cm length and 5.7 cm diameter was used to collect soil cores. Figure 4 shows the procedure of soil sample collection.

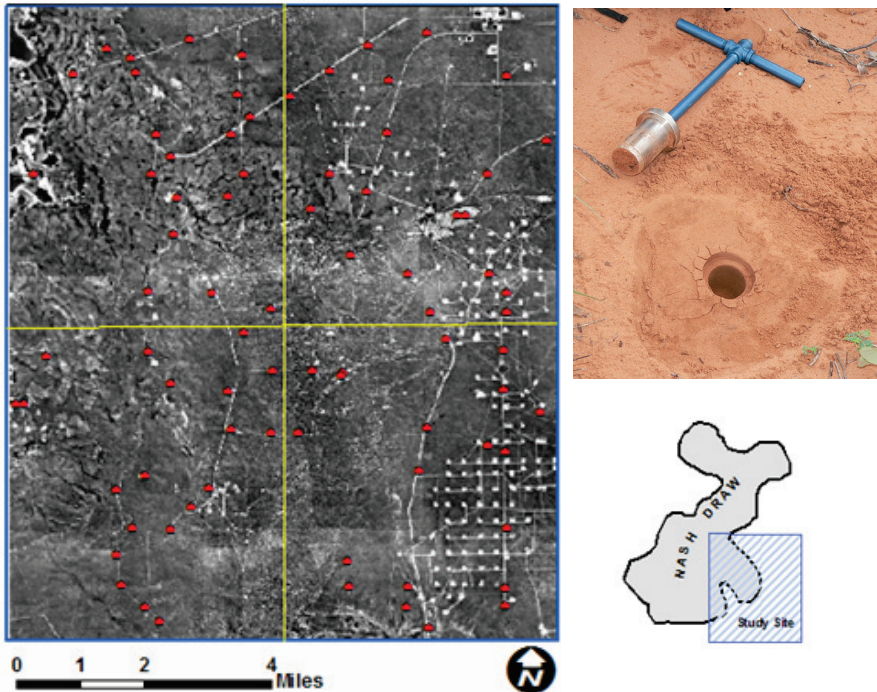


Fig. 4. Location and distribution of soil samples collected to measure the in situ soil moisture content (Left), and demonstration of soil sample collection (Right)

Volumetric Soil Moisture Measurement

Gravimetric soil moisture was measured and values were converted to volumetric soil moisture using sample volume. The ASTM D 2216-98 standard procedure (with modification) was used to calculate the gravimetric soil moisture and Equations (3) and (4) were used to calculate the volumetric moisture content.

$$w_v = \frac{V_w}{V} \times 100 \quad (3)$$

$$w_v = \frac{W_{ms} - W_{ds}}{76.55} \times 100 \quad (4)$$

Where, w_v = volumetric soil moisture (%), V_w = volume of moisture content (cc), V = volume of sample = 76.55 (cc), W_{ms} = weight of moist soil (gm), W_{ds} = weight of dry soil (gm), $V_w = W_{ms} - W_{ds}$ (volume of 1 gm water = 1 cc).

Soil moisture measurements for sample # 15 and sample # 115 were excluded from all analysis due to the proximity of these samples to a lake in the study area (Figure 5). Samples # 15 and # 115 had soil moisture measurements of 24.6% and 22.48%, respectively. Soil moisture measurements for samples # 31 and # 32 (from August data set) and for samples # 131 and # 132 (from November data set) were also excluded from the analysis due to the distortion of SAR backscatter at these sample locations. These samples are located in dune sand and the backscatter values in the SAR imagery for that area are distorted apparently due to total reflection of the radar signal (Figure 6). Table 2. shows the statistics of the in situ soil moisture measurements.

Statistics	Measurement Dates	
	August 01-03, 2006	November, 04-06, 2006
Mean	6.48	2.80
Minimum	2.55	0.26
Maximum	14.53	10.16
St. Deviation	2.56	2.13

Table 2. Statistics of in situ soil moisture measurements

3.4 Vegetation Maps

Vegetative cover can strongly influence soil moisture mapping with SAR imagery due to the interaction of the microwaves with the vegetation and soil. The amount of vegetation, its dielectric properties and distribution pattern can significantly impact the sensitivity of microwave backscatter to volumetric soil moisture. A vegetation distribution map is needed for mapping soil moisture using SAR with reasonable accuracy.

A vegetation map for the study site was not available at the same resolution as the SAR imagery. The acquired SAR imagery was used to produce a time series of vegetation maps for the study site.

General Vegetation Distribution Pattern Map

We combined the acquired five radar scenes to produce a temporal data set [Figure 7(a)]. The temporal data set was then used to perform multi-temporal analysis to create a vegetation map of the study site. Vegetation signatures were obtained from the multi-temporal data for areas with little or no vegetation, sparse vegetation and dense vegetation. The obtained signatures were applied to the radar imagery to produce the vegetation map [Figure 7(b) and 7(c)]. Finally we simplified the vegetation map by dividing the study site into two zones: Zone 1 represents the area characterized by sparse, thin or no vegetation and Zone 2 represents areas characterized by comparatively denser vegetation [Figure 7(d)].

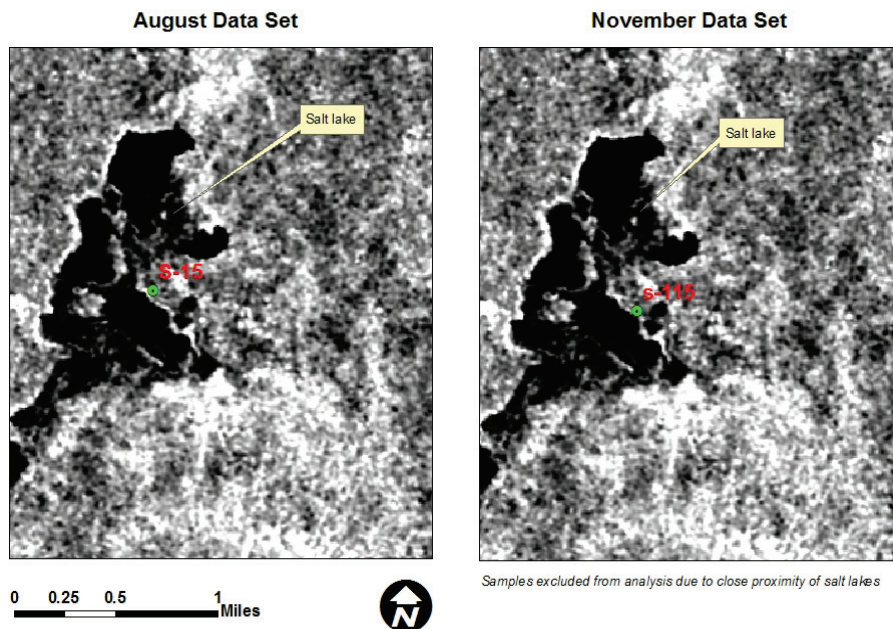


Fig. 5. Exclusion of samples # S-15 and S-115

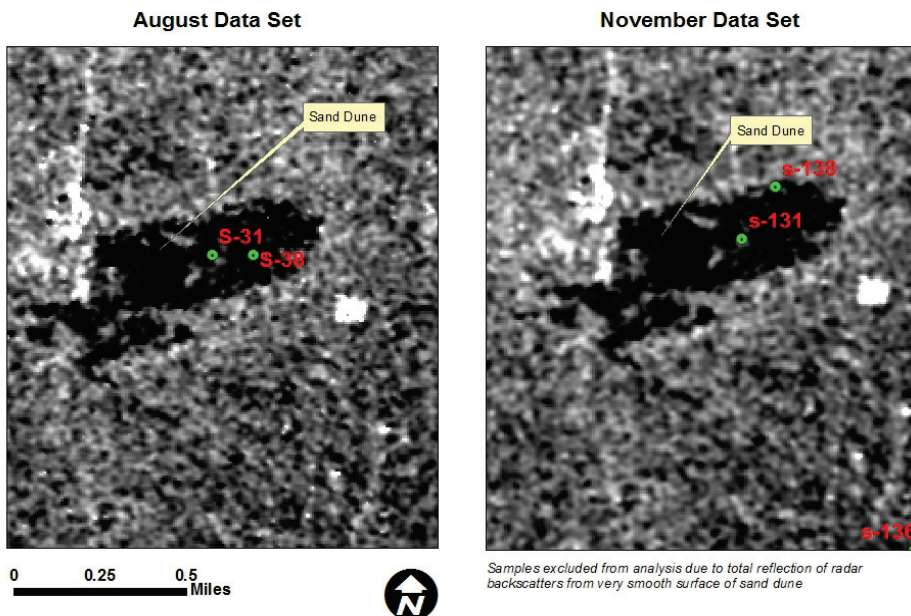


Fig. 6. Exclusion of samples # S-31, S-38, S-131 and S-138

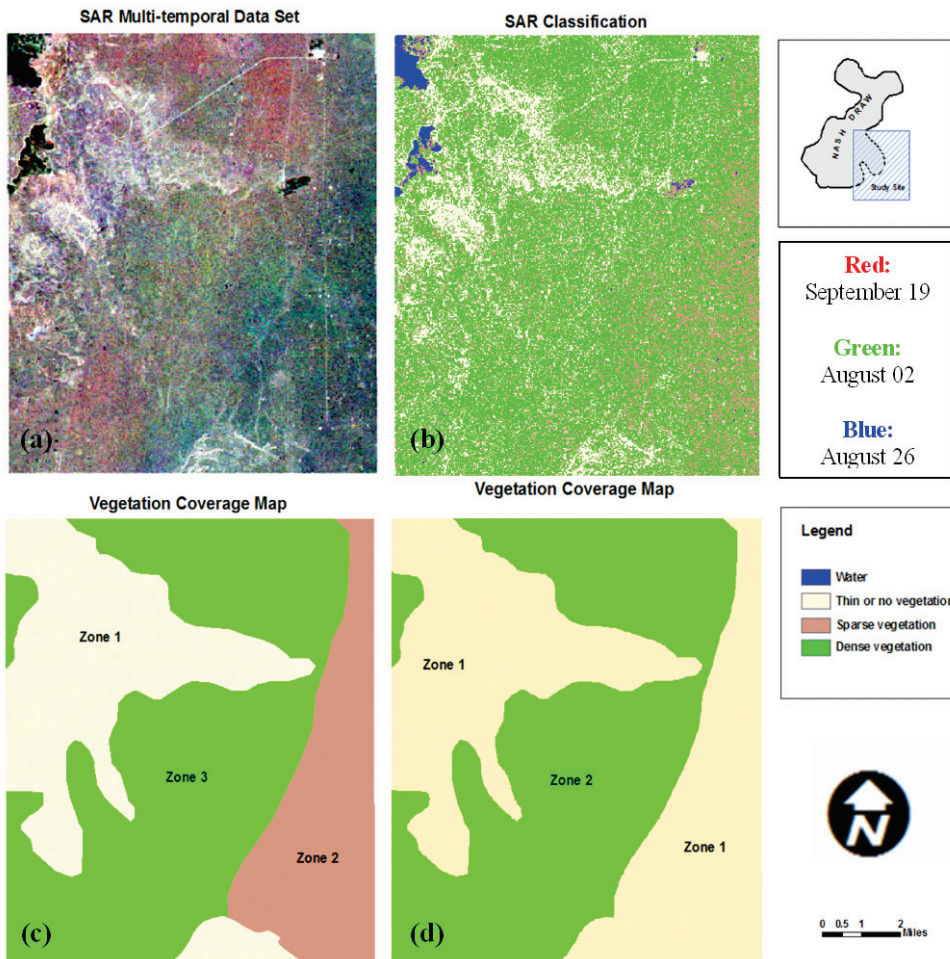


Fig. 7. Vegetation mapping

4. Methods

The algorithms developed by a semi-empirical approach are the most common and widely used (Moran et al., 2000). In this study simple linear regression was performed between soil moisture obtained from the field data and radar backscatter to study the linear relationship between radar reflectivity and soil moisture. An Artificial neural network (ANN) was used to study the non-linear relationship between radar backscatter (reflectivity) and soil moisture. Correlation coefficient (R^2) values were used to evaluate the suitability of the numerical models to map soil moisture in southeastern New Mexico

4.1 Simple Linear Regression

According to Ulaby et al. (1996), the radar backscatter from a surface with vegetation consists of three components: (1) product of the backscatter contribution of bare soil surface (σ°) and the two way attenuation of the vegetated layer (τ^2), (2) the direct backscatter contribution of the vegetation layer (σ_{dv}°), and (3) multiple scattering involving the vegetation elements and the ground surface (σ_{int}°).

$$\sigma^\circ = \tau^2 \sigma_s^\circ + \sigma_{dv}^\circ + \sigma_{int}^\circ \quad (5)$$

In case of densely vegetated surface, $\tau^2 \sim 0$ and σ° is determined largely by volumetric scattering from the vegetation canopy (Moran et al., 2004). For sparsely vegetated surfaces, $\tau^2 \sim 1$ and the second and third terms in the Equation (5) are negligible, and in this situation σ° is determined by the soil roughness and moisture content (Moran et al., 2004). Therefore, for bare soil, σ_s° has a functional relation with soil moisture, m_s and surface roughness, R (Engman and Chauhan, 1995), and it can be expressed as follows:

$$\sigma_s^\circ = f(R, m_s) \quad (6)$$

This indicates that for a target with uniform R , m_s can be estimated using the following expression:

$$m_s = a + b\sigma^\circ \quad (7)$$

Where a and b are regression coefficients, which are usually determined from field experiments encompassing the target-invariant R and the scene-invariants SAR wavelength (λ), incidence angle (θ_i), polarization, and calibration. However, Equation (7) is only valid for a given sensor, land use, and soil type, and for targets when τ^2 , σ_{dv}° and σ_{int}° are known or negligible (Moran et al., 2004).

Quesney et al. (2000) resolved Equations (5)-(7) to derive soil moisture information from ERS SAR measurements over an agricultural watershed in France on the basis of a priori vegetation classification of the site and in situ soil moisture measurements. This study separated the areas with low vegetation biomass for soil moisture estimation.

Similarly, for a semiarid watershed in Arizona, Moran et al. (2000) utilized the difference between dry- and wet-season SAR σ° ($\Delta\sigma^\circ$) to normalize the effects of surface roughness and topography on ERS SAR measurements. Thoma et al. (2004) improved on this approach to minimize empiricism and used a quantitative form of $\Delta\sigma^\circ$ to map soil moisture for an entire watershed with RADARSAT for three dates in 2003. In these studies, the effects of sparse vegetation were found to be negligible and could be ignored. These observations were also supported with similar findings by Lin and Wood (1993), Chanzy et al. (1997), Demircan et al. (1993), Dobson et al. (1992), and Dubois et al. (1995).

Our study site, Nash Draw is characterized by a semi-arid environment and generally sparse vegetation cover. According to the previous researchers, the effects of sparse vegetation in the radar backscattering should be insignificant and can be ignored for soil

moisture estimation in this type of study site. It is important to note for our study site we modified the Equation (7) by using β° instead of σ° as the input of SAR reflectivity values.

$$m_s = a + b\beta^\circ \quad (8)$$

Radar β° backscatter values were extracted for the soil sample locations where in situ soil moisture measurements were made and compared with volumetric soil moisture values for the sample locations. As discussed above, we used simple linear regression for the comparison and to calculate the coefficients a and b in the numerical model. We developed the model for the entire study site and also for different zones in the study site, based on vegetation density, to determine how well the model works for different land cover types, including vegetated areas and bare lands.

4.2 Non-Linear Regressions

Many SAR-based soil moisture estimation models assume that soil moisture distribution is linearly related to the radar reflectivity of the soil surface (e.g., Ulaby et al., 1996; Moran et al., 2000; Dobson et al., 1992; Dubois et al., 1995). Limited studies were conducted in the past to explore the non-linear relationship between soil moisture and radar backscatter (reflectivity). In this study we developed neural networks based numerical models to estimate soil moisture using SAR data in Nash Draw and to explore the non-linear relationship between soil moisture and SAR backscatter.

Artificial Neural Networks

Artificial neural networks, are a branch of artificial intelligence (Gardner and Dorling, 1998) in which the solution to a problem is learned from a set of examples (Bishop, 1994). A neural network can be regarded as a nonlinear mathematical function, which transforms a set of input variables into a set of output variables. The use of neural networks has been shown to be effective alternatives to more traditional statistical techniques (Schalkoff, 1992). Neural networks can be trained to approximate any smooth, measurable function (Hornik, et al., 1989), can model highly non-linear functions and can be trained to be accurately generalized when presented with unseen data (Gardner and Dorling, 1998). In a typical neural network model, a single neuron forms a weighted sum of the inputs x_1, \dots, x_d given by $a = \sum w_i x_i$, and then transforms this sum using a non-linear activation function $g(\)$ to give a final output $z = g(a)$ (Figure 8).

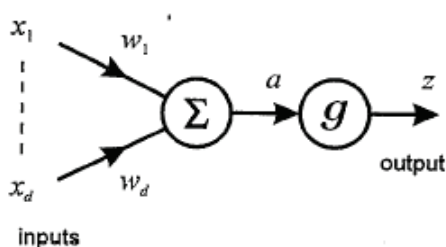


Fig. 8. A single processing unit in neural networks

A feed forward neural network can be regarded as a nonlinear mathematical function, which transforms a set of input variables into a set of output variables. The multilayer perceptron is the most widely used feed forward neural networks. Figure 8 shows a single processing unit of neural networks. If we consider a set of m such units, all with common inputs, then we arrive at a neural network having a single layer of adaptive parameters (weights). The output variables are denoted by z_j and are given by Equation (9).

$$z_j = g\left(\sum_{i=0}^d w_{ji}x_i\right) \quad (9)$$

Where w_{ji} is the weight for input i to j , and $g(\)$ is an activation function as discussed previously.

The neural network model developed to estimate soil moisture in Nash Draw includes only one input, the radar backscatter values. This model demonstrates the nature of non-linear relationship between radar backscatter and soil moisture.

5. Results

5.1 Simple Linear regressions

The simple linear regression models for the entire study site are shown in Figure 9. Figure 10 shows the regression models for different vegetation zones for August, 2006 data set, and Figure 11 shows the regression models for different vegetation zones for November, 2006 data set. Equation (10) and Equation (11) represent the numerical models for the entire study site for August, 2006 data set and November 2006 data set respectively. Equation (12) and Equation (13) represent the numerical models for thinly vegetated areas (Zone 1) for August, 2006 data set and November 2006 data set respectively. Equation (14) and Equation (15) represent the numerical models for densely vegetated areas (Zone 2) for August, 2006 data set and November 2006 data set respectively. Table 3 shows the results of the regression models for August, 2006 and November, 2006 respectively.

$$m_{ss_aug} = 18.47 + 0.82\beta^\circ \quad (10)$$

$$m_{ss_nov} = 6.13 + 0.23\beta^\circ \quad (11)$$

$$m_{z1_aug} = 27.4 + 1.34\beta^\circ \quad (12)$$

$$m_{z1_nov} = 13.33 + 0.67\beta^\circ \quad (13)$$

$$m_{z2_aug} = 4.11 - 0.11\beta^\circ \quad (14)$$

$$m_{z2_nov} = 3.52 + 0.14\beta^\circ \quad (15)$$

Domain	Correlation Coefficient (R^2)	
	August 2006	November 2006
Entire study site	0.24	0.05
Thinly vegetated areas (Zone 1)	0.61	0.51
Densely vegetated areas (Zone 2)	0.01	0.04

Table 3. Regression results

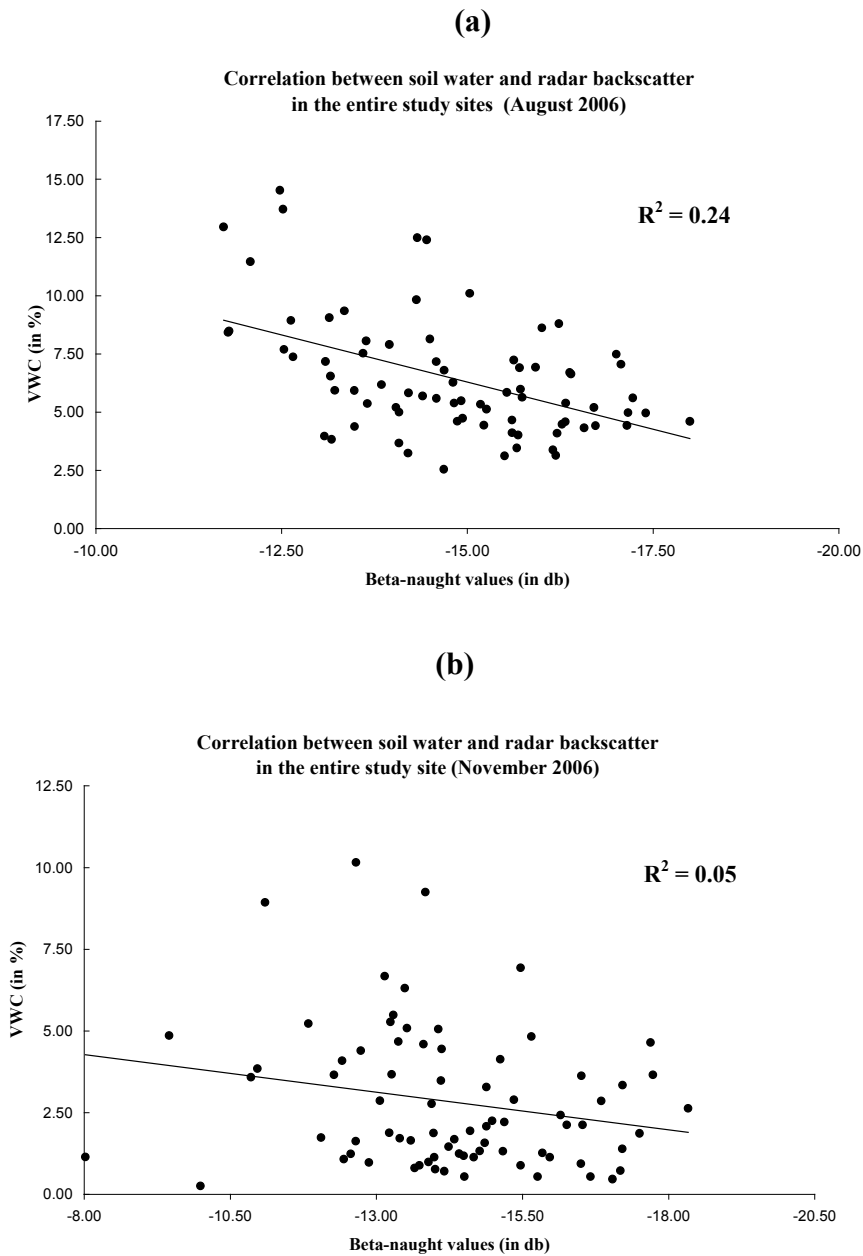


Fig. 9. Regression chart for entire study sites (a) August, 2006 and (b) November, 2006.

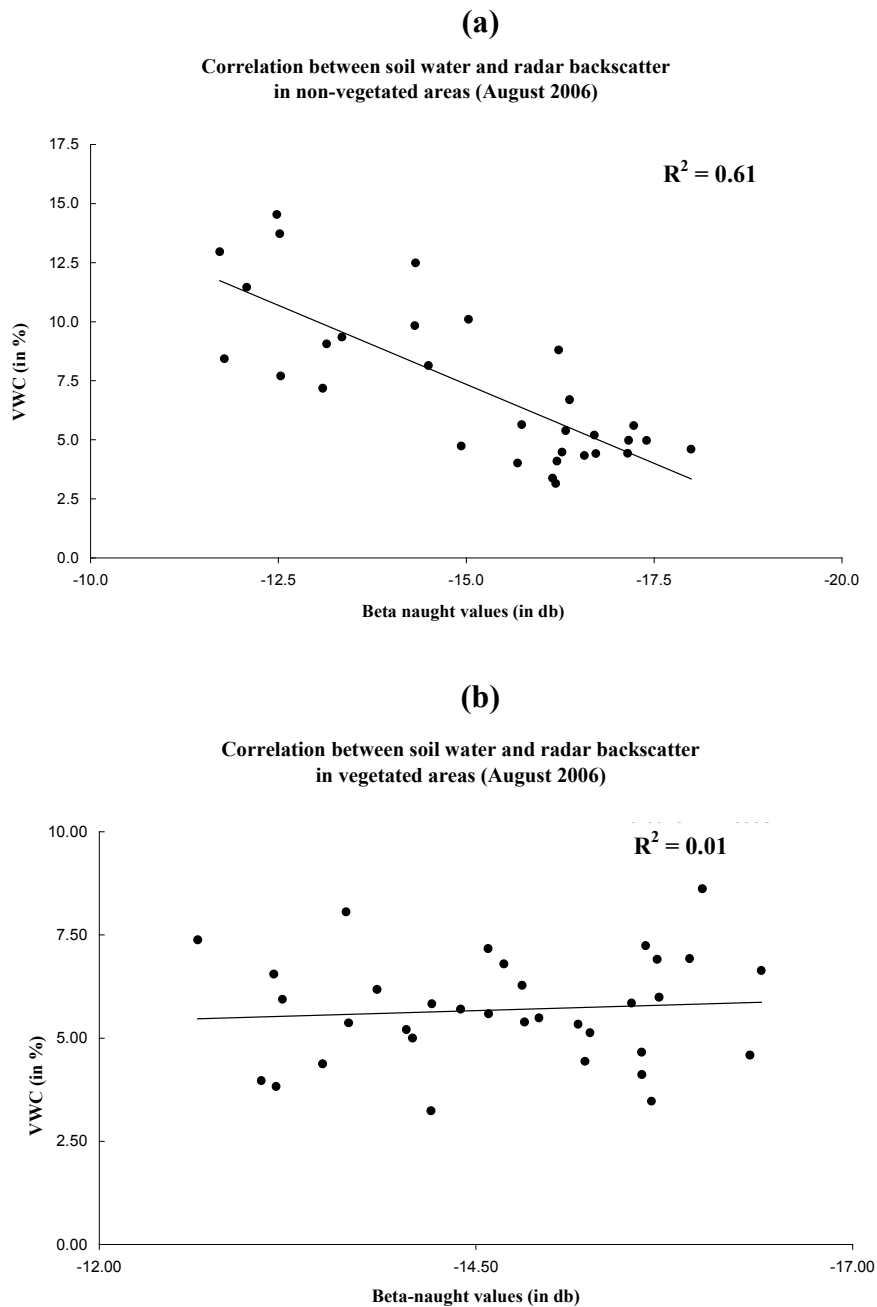
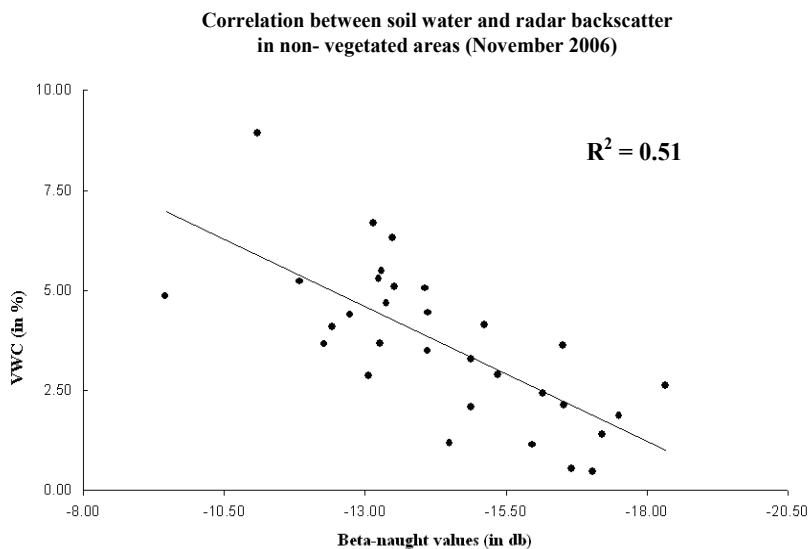


Fig. 10. Regression chart for different zones for August, 2006 data (a) Zone 1& (b) Zone 2 (31 and 34 data values used in the analysis for zone 1 and Zone 2 respectively).

(a)



(b)

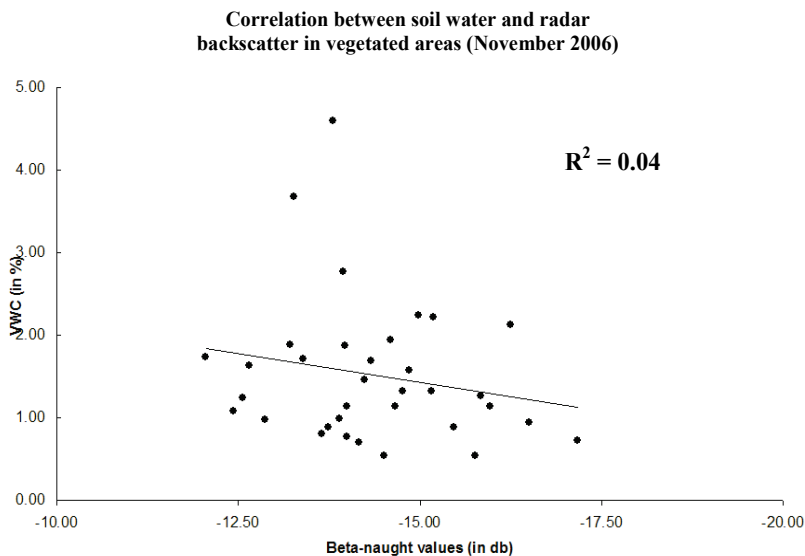


Fig. 11. Regression chart for different zones for November, 2006 data (a) Zone 1 & (b) Zone 2 (31 and 34 data values used in the analysis for zone 1 and Zone 2 respectively).

The R^2 values of the numerical models developed for the entire study site indicate that the linear relationship between the radar backscatter values obtained for the entire study site and soil moisture is not well defined. We observed this for both the August ($R^2 = 0.24$) and November ($R^2 = 0.05$) data sets. The R^2 values of the numerical models developed for vegetated areas (Zone2) for both August and November data set are very low (0.01 and 0.04 respectively). We found higher R^2 values for the numerical models developed for non-vegetated or sparsely vegetated areas (Zone 1) for both August and November data set (0.61 and 0.51 respectively). These findings indicate that the relationship between radar backscatter and soil moisture in densely vegetated areas is not linear.

5.2 Non-linear Regressions (Neural Networks)

We developed neural network based non-linear numerical models for soil moisture estimation for the entire study site using the August data set. We developed models using radar backscatter values as the only input to investigate the non-linear relationship between the radar backscatter (reflectivity) and soil moisture. We used JMP 6.0 statistical software to perform the neural networks based analysis. The model correlation coefficient (R^2) and cross validation correlation coefficient (CV R^2) were used to evaluate the model performance for soil moisture prediction. A neural network with 3 hidden nodes resulted in correlation coefficient (R^2) and the cross validation correlation coefficient (CV R^2) of 0.24 and 0.11 respectively. This result indicates that the non-linear relationship between radar backscatter and soil moisture is also not well defined for the entire study site.

5.3 Model Evaluation

We evaluated the correlation coefficients (R^2) and cross validation correlation coefficients (CV R^2) (for the neural networks based models) developed by both linear and non-linear regressions for soil moisture estimation in Nash Draw, NM and made the following observations.

- Simple linear regression between radar backscatter values and in situ soil moisture measurements can be used to develop SAR based soil moisture estimation model with model R^2 values of 0.51 to 0.61, but the model application should be restricted to non-vegetated to thinly vegetated areas.
- Neural network based non-linear regressions using radar backscatter values and in situ soil moisture measurements can be used to develop soil moisture estimation model for the entire study site with a model R^2 value of 0.24 and CV R^2 of 0.11.

6. Conclusions and Discussion

Earlier researchers reported that in a semi-arid environment with sparse vegetation, there is a linear relationship between soil moisture and radar backscatter. Our research shows that in semi-arid environment the influence of vegetation can influence the accuracy of the soil moisture estimation using the linear relationship between the radar backscatter and soil moisture. This observation is supported by the lower R^2 values (0.24 – August data set and 0.05 – November data set) obtained for the numerical models developed for the entire study site, the higher R^2 values (0.61 – August data set and 0.51 – November data set) obtained for

the numerical models developed for the parts of the study site identified as very thin or sparsely vegetated areas, and very low R^2 values (0.01 - August data set and 0.04 - November data set) obtained for the numerical models developed for the parts of the study site identified as more densely vegetated areas.

The non-linear relationship between radar reflectivity and soil moisture was investigated using a numerical model developed by feed forward neural networks with radar backscatter values and near real time in situ soil moisture measurements. The model was developed for the entire study site and did not show a strong non-linear relationship between radar backscatter (reflectivity) and soil moisture. The correlation coefficient (R^2) (0.24) did not improve from that obtained by simple linear regression (0.24) between radar backscatter and soil moisture.

This research indicates that in semi-arid environment vegetation coverage can significantly reduced the accuracy of soil moisture estimation and mapping using numerical models based on simple linear and non-linear relationships between radar backscatter values derived from high resolution SAR imagery and near real time in situ soil moisture measurements. This research also shows that numerical models based on only radar backscatter and near real time in situ soil moisture measurements can only be used in thinly vegetated to bare soil conditions in a semi-arid environment to estimate and map soil moisture with improved accuracy ($R^2 = 0.51$ to 0.61).

We recommend to include soil type, soil salinity and surface elevation information (in addition to vegetation coverage and in situ soil moisture measurements) in both linear and non-linear numerical models to improve the accuracy of SAR based soil moisture estimation in semi-arid environment without separating the vegetated and non-vegetated zones.

7. Acknowledgements

Thanks are due to the NASA Applied Sciences Program and the University of Mississippi Geoinformatics Center (UMGC) for funding the project through a Rapid Prototyping Capability (RPC) for Earth-Sun Systems Sciences project at The University of Mississippi. Thanks are also due to the Alaska Satellite Facility (ASF) for providing the SAR imagery; Dr. Dennis Powers, Glen Garrett and Dirk O'Daniel for their assistance in the field to acquire soil samples for soil moisture analysis; and Patrick Yamnik for assisting in image processing and GIS analysis.

8. References

- ASTM (Designation: D 2216 - 98). (1999). *Standard test method for laboratory determination of water (moisture) content of soil and rock by mass* (reprinted from the annual book of ASTM standards), American Society for Testing and Materials, 100 Barr Harbor Dr., West Conshohocken, PA 19428, 5p.
- Attema, E. & Ulaby, F. (1978). Vegetation modeled as a water cloud, *Radio Science*, vol. 13, pp. 357-364, ISSN: 0048-6604.

- Baghdadi, N.; Holah, N. & Zribi, M. (2006). Soil moisture estimation using multi-incidence and multi-polarization ASAR data, *International Journal of Remote Sensing*, vol. 27, no. 9-10, pp. 1907-1920, ISSN: 1366-5901.
- Bishop, C. M. (1994). Neural networks and their applications, *Review of Scientific Instruments*, vol. 65, no. 6, pp. 1803-1833, ISSN: 0034-6748.
- Bindlish, R. & Barros, A.P. (2001). Parameterization of vegetation backscatter in radar-based soil moisture estimation, *Remote Sensing of Environment*, vol. 76, no. 1, pp. 130-137, ISSN: 0034-4257.
- Borgeaud, M. & Saich, P. (1999). Status of the retrieval of bio- and geophysical parameters from SAR data for land applications, *Proceedings of the International Geoscience and Remote Sensing Symposium, IGARSS '1999, 28 June - 02 July, 1999, Hamburg, Germany*, pp. 1901-1903.
- Carlson, T.N.; Gillies, R.R. & Schmugge, T.J. (1995). An interpretation of NDVI and radiant surface temperature as measures of surface soil water content and fractional vegetation cover, *Agricultural and Forest Meteorology*, vol. 77, no. 3-4, pp. 191-205, ISSN: 0168-1923.
- Chanzy, A.; Bruckler, L. & Perrier, A. (1995). Soil evaporation monitoring: a possible synergism of microwave and infrared remote sensing, *Journal of Hydrology*, vol. 165, no. 1-4, pp. 235-259, ISSN: 0022-1694.
- Chanzy, A.; Kerr, Y.; Wigneron, J.P. & Calvet, J.C. (1997). Soil moisture estimation under sparse vegetation using microwave radiometry at C-band, *Proceedings of the International Geoscience and Remote Sensing Symposium, IGARSS '1997, 3-8 August 1997, Singapore*, pp. 1090-1092.
- Colpitts, B.G. (1998). The integral equation model and surface roughness signatures in soil moisture and tillage type determination, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, pp. 833-837, ISSN: 0196-2892.
- Dane, J. H. & Topp, G. C. (2002). *Methods of soil analysis, Part 4, Physical Methods, SSSA Book Series, No. 5*, 1692p, ISBN 0-89118-841-X.
- D'Elia C.; Ferraiuolo, G.; Pascazio, V. & Schirnzi, G. (2004). An MRF based technique for speckle reduction in SAR images, *Proceedings of the International Geoscience and Remote Sensing Symposium, IGARSS, 20-24 September 2004, Anchorage, AK*, vol. 6, pp. 4211- 4214.
- Demircan, A.; Rombach, M. & Mauser, W. (1993). Extraction of soil moisture from multitemporal ERS-1 SLC data of the Freiburg test site, *Proceedings of the International Geoscience and Remote Sensing Symposium, IGARSS '1993, 18-21 August 1993, Tokyo, Japan*, vol. 4, pp. 1794-1796.
- Dobson, M. C. & Ulaby, F. T. (1998). Mapping soil moisture distribution with imaging radar. In: *Principles & Applications of Imaging Radar, Manual of Remote Sensing, 3rd Edition, Volume 2*, Henderson, F. M. & Lewis, A. J., pp. 407-433, John Wiley & Sons, Inc., ISBN: 0-471-29406-3, New York.
- Dobson, M. C.; Ulaby, F. T.; El-Rayes M. & Hallikainen. (1985). Microwave dielectric behavior of wet soil, part II: four component dielectric mixing model, *IEEE Transaction on Geoscience and Remote Sensing*, vol. GE-24, no. 1, pp. 517-526, ISSN: 0196-2892.

- Dobson, M.C. ; Pierce, L.; Arabandi, K.; Ulaby, F.T. & Sharik, T. (1992). Preliminary analysis of ERS-1 SAR for forest ecosystem studies, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, pp. 203–21, ISSN: 0196-2892.
- Dubois, P.C.; van Zyl, J. & Engman, E.T. (1995). Measuring soil moisture with imaging radars, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, pp. 915–926, ISSN: 0196-2892.
- Engman, E.T. (1994). The potential of SAR in hydrology, *Proceedings of the International Geoscience and Remote Sensing Symposium, IGARSS '94*, 8–12 August 1994, Pasadena, Calif. IEEE, Piscataway, N.J. pp. 283–285.
- Engman, E.T. & Chauhan, N. (1995). Status of microwave soil moisture measurements with remote sensing, *Remote Sensing of Environment*, vol. 51, pp. 189-198, ISSN: 0034-4257.
- Farnsworth R.K.; Barret E.C. & Dhanju M.S. (1984). *Application of remote sensing to hydrology including ground water*, IHP-II Project A. 1.5, UNESCO, Paris, France.
- Fung, A.K.; Li, Z. & Chen, K.S. (1992). Backscattering from a randomly rough dielectric surface, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, pp. 356–369, ISSN: 0196-2892.
- Gardner, M.W. & Dorling, S.R. (1998). Artificial Neural Networks (The Multilayer Perceptron) - a review of applications in the atmospheric sciences, *Atmospheric Environment*, vol. 32, no. 14/15, pp. 2627-2636, ISSN: 1352-2310.
- Georgakakos, K. P. & Baumer, O. W. (1996). Measurement and utilization of on-site soil moisture data, *Journal of Hydrology*, vol. 184, pp. 131–152, ISSN: 0022-1694.
- Gillies, R. R.; Carlson, T. N.; Cui, J.; Kustas, W. P. & Humes, K. S. (1997). A verification of the 'triangle' method for obtaining surface soil water content and energy fluxes from remote measurements of the Normalized Difference Vegetation Index and surface radiant temperature, *International Journal of Remote Sensing*, vol. 18, pp. 3145–3166, ISSN: 0143-1161.
- Glenn, N. F. & Carr, J. R. (2004). Establishing a relationship between soil moisture and RADARSAT-1 SAR data obtained over the Great Basin, Nevada, U.S.A., *Canadian Journal of Remote Sensing*, vol. 30, no. 2, pp.176–181, ISSN: 0143-1161.
- Goodman, J. W. (1975). *Statistical properties of laser speckle patterns, in laser speckle and related phenomena*, Edited by J. C. Dainty, Springer, 1975, pp. 9-75.
- Guindon, B. (1990). Development of a shape-from-shading technique for the extraction of topographic models from individual spaceborne SAR images, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28 no. 4, pp. 654–661, ISSN: 0196-2892.
- Hajnsek, I. & Pottier E. (2000). Terrain correction for quantitative moisture and roughness retrieval using polarimetric SAR data, *Proceedings of the International Geoscience and Remote Sensing Symposium IGARSS 2000*, vol. 3, pp. 1307-1309.
- Henderson, F. M. & Lewis, A. J. (1998). *Principles & Applications of Imaging Radar, Manual of Remote Sensing, 3rd Edition, Volume. 2*, pp. 01-07, John Wiley & Sons, Inc., ISBN: 0-471-29406-3, New York.
- Holt, R.M.; Beauheim, R.L.& Powers, D.W. (2005). Predicting fractured zones in the Culebra Dolomite, In: Faybishenko, B., Witherspoon, P.A., and Gale, J., eds., *Dynamics of Fluids and Transport in Fractured Rock: AGU Geophysical Monograph Series*, vol. 162, p. 103-116, ISSN 0065-8448.
- Hornik, K.; Stinchcombe, M. & White, H. (1989). Multilayer feed forward networks are universal approximators, *Neural Networks*, vol. 2, pp. 259-366, ISSN: 0893-6080.

- Hossain, F. & Anagnostou, E.N. (2005). Numerical investigation of the impact of uncertainties in satellite rainfall and land surface parameters on simulation of soil moisture, *Advances in Water Resources*, vol. 28, no. 12, pp. 1336-1350, ISSN: 0309-1708.
- Hossain, A. & Easson, G. (2008). Evaluating the potential of VI-LST Triangle Model for quantitative estimation of soil moisture using optical imagery, *Proceedings of the International Geoscience and Remote Sensing Symposium IGARSS 2008*. IEEE International, vol. 3, issue, 7-11 July 2008, pp. 879 - 882.
- Hossain, A. & Easson, G. (2006). Mapping spatial variation in surface moisture using reflective and thermal ASTER imagery for Southern Africa, *ASPRS 2006 Annual Conference*, Reno, Nevada, May 1-5, 2006.
- Hossain, A.; Easson, G.; Powers, D. W. & Holt, R. M. (2007). Impact of variation in reflectivity of microwave data for soil moisture estimation in semi-arid environment, *Sigma Xi Poster Presentation*, The University of Mississippi, University, MS.
- Jackson, T. J. (2002). Remote sensing of soil moisture: implications for groundwater recharge, *Hydrogeology Journal*, vol. 10, pp. 40-51, ISSN: 1431-2174.
- Kelly R. E. J.; Davie T. J. A. & Atkinson P. M. (2003). Explaining temporal and spatial variation in soil moisture in a bare field using SAR imagery, *International Journal of Remote Sensing*, vol. 24, no. 15, pp. 3059-3074, ISSN: 0143-1161.
- Lambin, E. F. & Ehrlich, D. (1996). The surface temperature-vegetation index space for land cover and land-cover change analysis, *International Journal of Remote Sensing*, vol. 17, pp.1087-1105, ISSN: 0143-1161.
- Lin, D.S. & Wood, E.F. (1993). Behavior of AirSAR signals during MACEurope '91, *Proceedings of the International Geoscience and Remote Sensing Symposium, IGARSS '1993*, 18-21 August 1993, Tokyo, Japan, vol. 4, pp. 1800-1802.
- Lu, Z., & Meyer, D.J. (2002). Study of high SAR backscattering caused by an increase of soil moisture over a sparsely vegetated area: implications for characteristics of backscatter, *International Journal of Remote Sensing*, vol. 23, pp. 1063-1074, ISSN: 0143-1161.
- Meijerink, A. M. J.; Schultz G. A. & Engman, E. T. (eds.). (2000). *Remote sensing in hydrology and water management*, Springer, Berlin Heidelberg New York, pp 305-325.
- Moran, S. M.; Clarke, T. R.; Inoue, Y. & Vidal, A. (1994). Estimating crop water deficit using the relationship between surface-air temperature and spectral vegetation index, *Remote Sensing of Environment*, vol. 49, pp. 246-263, ISSN: 0034-4257.
- Moran, M.S.; Hymer, D.C.; Qi, J. & Sano, E.E. (2000). Soil moisture evaluation using multi-temporal Synthetic Aperture Radar (SAR) in semiarid rangeland, *Agricultural and Forest Meteorology*, vol. 105, pp. 69-80, ISSN: 0168-1923.
- Moran, S. M.; Christa D.; Peters, L.; Watts J. M. & McElroy, S. (2004). Estimating soil moisture at the watershed scale with satellite-based radar and land surface models, *Canadian Journal of Remote Sensing*, vol. 30, no. 5, pp. 805-826, ISSN: 1712-7971.
- Nemani, R. & Running, S., 1993, Developing satellite derived estimates of surface moisture status, *Journal of Applied Meteorology*, vol. 32, pp. 548-557, ISSN: 1558-8424.

- Powers, D.W.; Beauheim, R.L.; Holt, R.M. & Hughes, D.L. (2006). Evaporite karst features and processes at Nash Draw, Eddy County, New Mexico, In: Caves & Karst of Southeastern New Mexico, Land, L. and others, eds., *NM Geological Society Fifty-seventh Annual Field Conference Guidebook*, pp. 253-266, ISBN: 1-58546-092-3 (ISSN: 0077-8567).
- Quesney, A.; Le Hégarat-Masclé, S.; Taconet, O.; Vidal-Madjar, D; Wigneron, J.P.; Loumagne, C. & Normand, M. (2000). Estimation of watershed soil moisture index from ERS/SAR data, *Remote Sensing of Environment*, vol. 72, pp. 290-303, ISSN: 0034-4257.
- RADARSAT International, 1995, RADARSAT Illuminated: your guide to products and services, Unpublished manual, 60p.
- Rayleigh, J.W. S. & Robert B. L. (1945). *The theory of sound*, vol. 1, ISBN: 0-486-60292-3, New York.
- Raney, R. K. (1998). Radar fundamentals: Technical perspective. In: *Principles & Applications of Imaging Radar, 3rd Edition, Volume 2*, Henderson, F. M. & Lewis, A. J., pp. 407-433, John Wiley & Sons, Inc., ISBN: 0-471-29406-3, New York.
- Robock; Alan; Lifeng, L.; Eric, F. W.; Fenghua W.; Kenneth, E.; Mitchell; Paul, R. H.; John C. S.; Dag L.; Brian C.; Justin S.; Qingyun D.; Wayne H.; Rachel T. P.; Dan T. J.; Jeffrey B. B. & Kenneth C. C. (2003). Evaluation of the North American Land Data Assimilation System over the Southern Great Plains during the warm season, *Journal of Geophysical Research*, vol. 108 (D22), no. 8846, ISSN: 0148-0227.
- Schalkoff, R. (1992). Pattern recognition: Statistical, structural and neural approaches, John Wiley and Sons, New York.
- Schneider, K., & Oppelt, N. (1998). The determination of mesoscale soil moisture patterns with ERS data, In *Proceedings of the International Geoscience and Remote Sensing Symposium, IGARSS '98*, 6-10 July 1998, Seattle, Wash. IEEE, New York. pp. 1831-1833.
- Thoma, D. P.; Moran, M. S.; Bryant, R.; Rahman, M.; Holifield-Collins, C. D.; Skirvin, S.; Sano, E. E. & Slocum, K. (2006). Comparison of four models to determine surface soil moisture from C-band radar imagery in a sparsely vegetated semiarid landscape, *Water Resources Research*, vol. 42, pp. 1-12, ISSN: 0043-1397.
- Touzi R. (2002). A review of speckle filtering in the context of estimation, theory, *IEEE Transactions on Geoscience And Remote Sensing*, vol. 40, pp. 2392-2404, ISSN: 0196-2892.
- Ulaby, F.T.; Moore, M. K. & Fung, A.K. (1986). *Microwave remote sensing, active and passive, from theory to application, Volume 3*, Artech House, ISBN: 0-89006-192-0. Norwood, MA
- Ulaby, F.T.; Held, D.; Donson, M.C. & McDonald, K.C. (1987). Relating polarization phase difference of SAR signals to scene properties, *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-25, pp. 83-92, ISSN: 0196-2892.
- Ulaby, F.T.; Dubois, P.C. & Van Zyl, J. (1996). Radar mapping of surface soil moisture, *Journal of Hydrology*, vol. 184, pp. 57-84, ISSN: 0022-1694.

Ensemble of retrieval algorithms and electromagnetic models for soil and vegetation water content estimation from SAR images

Claudia Notarnicola
*EURAC-Institute for Applied Remote Sensing,
Bolzano, Italy*

1. Introduction

In the last two decades, the interest for the estimation of Earth surface parameters from remotely sensed data has increased in the scientific community. Within this field, one of the most challenging and attractive problems is represented by the estimation of soil moisture (SM) and vegetation water content (VWC) as they are fundamental in many disciplines.

The prediction of SM variations is equally important at mesoscale and smaller scales. Mesoscale atmospheric models have demonstrated sensitivity to spatial gradients while at field level, SM can be considered storage of water between rainfalls and evaporation thus acting as a regulator to fundamental hydrologic processes such as infiltration and runoff (Delworth, 1988).

Surface SM information is also a critical forcing variable in many Soil Vegetation Atmosphere Transfer (SVAT) models which are able to estimate SM values at daily time steps.

Vegetation is a fundamental component of every ecosystems and VWC is one of the most important biochemical components with 35-95% of the vegetation body. VWC yields information about the physiological conditions of the plants. Furthermore, estimation of VWC from local to global scales is central to the understanding of biomass burning processes, water stress and drought condition. The prediction of this variable can be important for irrigation strategies and for yield forecasting (Pennuelas et al., 1993).

Spaceborne and airborne microwave sensors are best suited for the detection of water content (Ulaby et al., 1986). The retrieval of biophysical parameters from remotely sensed data falls within the category of inverse problems where, from a vector of measured values, m , one wishes to infer the set of ground parameters, x , that gave rise to them. The inverse problem is typically ill-posed due to its non-linearity between remote sensing measurements and ground parameters. Furthermore, many aspects of the natural surfaces, such as surface roughness and the amount and type of vegetation, alter the radar backscatter.

Many approaches have been developed in order to provide possible solution to these inverse problems, spanning from empirical and semi-empirical approaches to sophisticated machine learning techniques.

The development of empirical models has been studied both as a first approach to study the relationship between remotely sensed signals and surface parameters and to obtain a simple inversion model in itself. The frequently used linear approach is based on regression coefficients generated by the observations over a specific site (Prevot et al, 1993, Dubois et al 1994). One of the first empirical models was proposed by Oh *et al.*, 1992 on bare soils, where the co-polarized and cross-polarized ratios of the backscattering coefficients are expressed in terms of the surface parameters. The Oh model, which is developed from multi-polarization radar data, was revealed to be poorly effective when tested on synthetic aperture radar (SAR) data. Subsequently, Dubois et al. 1995 developed an empirical inversion model from scatterometer data and applied it to SAR data in the case of bare soils. The Dubois inversion model was found to be applicable to the different forms of measured data and tends to be quite accurate with a root-mean-square error (rmse) of 4.2% on SM values. Although the Dubois model performed well, it is site specific and is only valid under the conditions in which the measurements were taken. As a result of the way empirical models are developed and their relative inversion procedure, they have a limited range of applicability. The complexity and nonlinearity of the problems cannot be taken into account in empirical formulations, thus leading to the necessity of considering theoretical backscattering models. Many theoretical models have been developed in order to describe the interaction between the electromagnetic radiation and natural surfaces. They can represent a great variety of situations and still have the possibility to consider cases that have not been taken into account by the empirical models. On the other side, theoretical models are developed under several hypotheses that may not be completely verified in field conditions. One main limitation of a theoretical model is considered the description of the surface morphology. One of the most widely used descriptions is based on two parameters: 1) the standard deviation (SD) of heights s and 2) the correlation length l . The SD of heights is an estimate of the variance of the vertical dimension of the soil surface profile, whereas its correlation function relates the statistical correlation between any two points on a given surface. The surface correlation length l is usually defined as the displacement for which the correlation function is equal to $1/e$ (Ulaby et al, 1986). This parameterization is often considered critical because they do not completely describe the variability of natural surfaces (Mattia & Le Toan, 1999). The SD of heights can have an accuracy of only about 10%, the correlation length measurements vary as much as an order of magnitude (Dubois et al, 1995, Notarnicola et al, 2003). Although they have the capacity to generalize and treat a great variety of situations, theoretical forward-scattering algorithms are of a certain complexity and are sometimes difficult to invert due to the requirements of several parameters in the computations.

To overcome this difficulty, typical inversion techniques are iterative methods and statistical approaches. Bindlish and Barros (Bindlish & Barros, 2000) used the integral equation model (IEM) with the Jacobian method—an iterative scheme—to perform the inversion on multifrequency multipolarization SAR data from Washita '94. In this case, the retrieval can be performed on all the surface parameters, as they are included in the IEM. This algorithm, which is tested only with one data set in a single sensor configuration, produces SM estimates with an average error of 3.4%. Statistically based inversion methods, such as the Bayesian approach, have been in existence for a long time and are based on probabilities that a given set of measurements comes from certain surface parameter values. The probability density functions (pdfs) are estimated by training, where samples of sensor and

surface measurements are presented in the algorithm. The practical use of Bayes' theorem is to turn probabilities that can be estimated from a training set into those that are required for the estimation of the unknown surface parameters (Marchant & Onyongo, 2003). A useful property of a Bayesian method is that it is optimal in the sense that it minimizes the expected error. Another important aspect is that, to derive these general pdfs, as performed with the Bayesian methodology, a large amount of experimental data is needed. The experimental data should cover a wide spectrum of real situations to obtain reliable statistical functions, but the inversion technique itself does not represent "the solution." In fact, the inversion procedure has the same limitation as the forward model as it relies on limited surface parameter conditions. As an example, Haddad and Dubois (Haddad & Dubois, 1994), starting from the forward model proposed by Oh *et al.* (Oh *et al.*, 1992) used a Bayesian approach to determine the inverse model. As the model was based on a data set with a low correlation length, it failed to be applicable to the data sets without this condition.

A suitable method for this kind of multidimensional retrieval is the neural network (NN). It can be trained to extract surface parameters from remotely sensed data, and in this way, it can perform the same function as a statistical inversion method. The training data for the NNs can be obtained from theoretical forward-scattering models, thus allowing the control of the range of parameters with which the network is trained. Artificial NNs (ANNs) have a number of advantages and disadvantages compared to conventional statistical algorithms. One advantage of an NN is that it can identify subtle and nonlinear patterns, which is not always the case with traditional statistical methods (Beale & Jackson, 1992). In addition, NNs do not require normally distributed continuous data and may be used to integrate data from different sources with poorly defined or unknown distributions. Another advantage is that NNs are able to take a specific set of input data and generalize a solution set, which may give the correct answer for unknown input patterns that are similar, but not identical, to the input data. One of the problems is the difficulty in adequately configuring and training a network. There are no given rules for the configuration of the network (in terms of the number of hidden nodes, hidden layers, etc.). The training process has to be carefully controlled due to the risk of overtraining the network. Overtraining is a phenomenon whereby the network learns a training data set to an excellent level but cannot accurately predict the correct answer with independent test data. Furthermore, overtraining frequently happens when the number of training data is limited as often are the remotely sensed data sets (Notarnicola *et al.*, 2008)

The main drawback of an NN is that the inverse empirical mapping established between remotely sensed data and surface parameters cannot be explicitly written down, and the user can generally only act on some configuration parameters but not on the analytical expression that leads to the results.

New approaches are emerging in the last years for the estimation of biophysical parameters; one of the most used is the Support Vector Regression (SVR).

SVR, initially developed for classification purposes, is now being applied also to the estimation of biophysical parameters. SVR is based on a geometrical rather than a statistical approach, because it bases the estimation on both the geometrical distances between samples and the maximization of the geometrical margin instead of on the estimation of the posterior probability distribution over the samples. For this reason, there are two main advantages with respect to NN and statistical approach. The SVR method is less sensitive to

the limited availability of training samples with respect to other machine learning techniques and to the overfitting of the datasets, thus leading to high generalization capabilities (Camps-Valls et al., 2006). Till now this approach has not yet applied for the SM estimation.

Another way to overcome the difficulties of the single approach is to use the concept of ensemble. Ensembles are widely used in machine learning techniques and the main idea of ensemble learning is to employ multiple learners and combine their predictions.

The last decade has seen many works related to ensemble learning systems. These systems are groups of machine learning approaches where each learner provides an estimate of a target variables that after are combined in different ways in order to reduce the generalization error if compared to the single learner (Brown et al., 2005).

The different estimates are usually combined through a combination function, commonly a majority vote for classification and a linear combination for regression. It is a good improvement in the combined estimates if the individual estimators should exhibit different patterns of generalization

As an example some works on ensemble of neural networks are reported. Cho and Kim (1995) combined the results from multiple neural networks using fuzzy logic which resulted in more accurate classification. Bishop (1995) affirms that if L networks produce errors which have zero mean and are uncorrelated, then the sum-of-squares error can be reduced by a factor of L simply by averaging the predictions of the L networks. Liu and Yao (1999) proposed the Negative Correlation Learning (NCL) algorithm wherein a penalty term is added to the error function which helps in making the individual predictors as different from each other as possible while encouraging the accuracy of individual predictors. This enables the mapping function learnt by the ensemble to generalize better when an unseen input is to be processed.

In this context, this chapter will address assessed remote sensing procedures, such as empirical models, Bayesian methods for the estimation of SM and VWC from multi-frequency and multi-polarization SAR images in synergy with optical sensors and electromagnetic models. Initially, the procedures are used as separate inversion methods. In this case, limitations and potentialities are illustrated. Subsequently, each method is considered as an element of an ensemble from which then the best estimates are drawn.

The basic concept behind this ensemble method is that each single methodology has its advantage and disadvantage and it is able to detect some features with high accuracy and other features with low accuracy. The idea of ensemble learning is to employ multiple learners and combine their predictions. Numerous works applied in different context have demonstrated that the ensemble estimate accuracy is quite often much higher than the accuracy of the single predictor (Ueda & Nakano, 1996).

This work presents an innovative approach for the ensemble of regression algorithms by considering both different regression techniques applied to different sensor configurations thus exploiting the capability of different frequencies/polarization combination to estimate soil and vegetation features.

The chapter is organized as follows. Section 2 is devoted to the description of analyzed experimental data sets. Section 3 illustrates the most used electromagnetic models whose simulations will be used in the inversion procedure. These procedures are outlined in section 4. The results of the different procedures are discussed in section 5. Section 6 introduces the concept of *ensemble estimates* and discusses the results of this technique with

respect to the results obtained from the different procedure. Conclusions and future applications are drawn in section 7.

SM	Soil moisture
GSM	Gravimetric soil moisture
VWC	Vegetation water content
VSM	Volumetric soil moisture
σ^0	Backscattering coefficient
τ^2	Two-way attenuation of the vegetation layer
ϵ	Dielectric constant
SD	Standard deviation
s	Standard deviation of height
l	Correlation length
IEM	Integral Equation Model
WCM	Water Cloud Model

Table 1. Summary of scientific notation and most used acronyms.

2. Data set description

SMEX'02 is a remote sensing experiment that was carried out in Iowa in 2002 (http://nsidc.org/data/amsr_validation/soil_moisture/smex02/), mainly focused on modelling and algorithm validation over a range of SM conditions with moderate to high vegetation biomass conditions. The main site, chosen for intensive sampling SM, vegetation and surface roughness, was the Walnut Creek watershed (Figure 1), where 32 fields, 10 soybean and 21 corn fields, were sampled intensively. The field and sensor data acquired during this experiment are particularly suitable to our analysis because of:

- The number of fields that were considered in the experiment with different level of soil and vegetation moisture;
- The acquisition of both radar and optical data and the extensive ground measurements carried out within each field.

2.1 Soil moisture measurements

SM sampling in the Watershed sites was carried out to provide a reliable estimate of the mean and variance of the volumetric SM of the surface SM for fields that are approximately 800 m by 800 m. These measurements are used primarily to support the aircraft based microwave investigations, which were conducted between 0900 and 1200 local time. At four standard locations in each site the gravimetric soil moisture (GSM) was sampled on each day of sampling with a 0-6 cm scoop tool. This GSM sample was then split into 0-1 cm and 1-6 cm samples providing a rough estimate of the site average 0-1 cm GSM. GSM is converted to volumetric soil moisture (VSM) by multiplying GSM and bulk density of the soil. Bulk density was sampled one time at each of these four locations using an extraction technique. VSM values are calculated by using GSM and bulk density that are the parameters directly measured in the fields. The soil texture data for the SMEX'02 study area were obtained from CONUS-SOIL dataset (Miller & White, 1998). Soil texture is of the utmost importance in physical models for estimation of soil dielectric properties. In fact the Hallikainen empirical formula derives the soil dielectric constant from SM and soil texture values (Hallikainen, 1995). The values of the real part of the dielectric constant along with

the roughness parameters are the inputs to the theoretical models used in this inversion approach. This part is described in the following sections.

2.2 Vegetation water content measurements

VWC (kg/m²) was measured several times in 32 fields with four rounds. VWC in plant stems and leaves were computed as

$$VWC' = B_g' - B_d' \text{ (g/plant)} \quad (1)$$

where B_g' is the green biomass + tare weight and B_d' is the dry biomass + tare weight. This assumes that water loss from the tares (paper bags) was negligible in comparison with that from the plant samples. In row crops, areal stand density (ASD; plants/m²) was estimated from the row plant density (RD; plants/m) by using

$$ASD = RD/RS \quad (2)$$

where RS is the row spacing. VWC (kg/m²) was then computed as

$$VWC = 10^{-3} * VWC' * ASD. \quad (3)$$

However not every field was sampled during each round. This implies that a measured VWC value is not available for all days of acquisitions. For each field-date combination, three locations in the field were visually selected from airborne digital imagery to represent average, minimum and maximum canopy conditions. Above ground biomass was removed and wet and dry weights were used to compute VWC. For this investigation, all samples within a field on a given date were averaged and this single value was used.

Other ground truth measurements used in this work include surface roughness in terms of standard deviation of heights and correlation length.

2.3 Remotely sensed data

The AirSAR images (resolution: 8 -12 m ground range) were acquired on 1, 5, 7, 8, 9 July 2002. The LANDSAT (resolution: 30 m) images were acquired contemporary to SAR on 1, 8 July 2002.

The five L- and C-band images were processed by the AirSAR operational processor providing calibrated data sets. The absolute and relative calibration accuracy obtained for each sensor, as reported in the literature (van Zyl, 1992), are listed in table 2.

ABSOLUTE/RELATIVE	C-BAND	L-BAND
AIRSAR	±1.0 dB / ±0.4 dB	±1.2 dB / ±0.5 dB

Table 2. AIRSAR calibration accuracy.

From sensitivity studies (Dubois et al., 1995), in order to avoid errors in the SM estimation larger than 4.2%, the relative calibration error should be less than 0.5 dB and the absolute calibration error should be less than 2.0 dB, because the inversion is also more sensitive to relative than absolute calibration errors.

During the campaign, two Landsat Thematic Mapper (TM) scenes from Landsat 5 and three Landsat Enhanced Thematic Mapper plus (ETM+) from Landsat 7 were acquired during the primary study period. They were mainly used to calculate the brightness temperature and the indices, the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Water Index (NDWI). These two indices are also very important factors in estimating VWC which is needed for SM estimation using microwave methods. The images were atmospherically and radiometrically corrected to produce the at-ground reflectance and then the NDVI and NDWI indices (Gao et al., 1996).

In this work, the data acquired on 1st July and some data taken randomly from the other dates were considered as training samples. The fact to not consider exclusively the data coming from one single day allows the results to be independent from the specific soil and weather conditions of a single date.

3. Electromagnetic models

As the proposed approaches, both the empirical and the statistical methods, consider in different ways simulated data, theoretical models for bare and vegetated soils are briefly described. For bare soil, the SAR response has been simulated by means of the Integral Equation Model (IEM), (Fung, 1994). This model, with respect to other electromagnetic models, has the advantage of being applicable to a wide range of roughness scale. For the model, the input parameters are the real part of the dielectric constant, the standard deviation of height and the correlation length. The dielectric constant is linked directly to VSM and soil texture through some well known and validated experimental relationships (Hallikainen, 1985).



Fig. 1. Distribution of the Walnut Creek fields on the LANDSAT image. The size of the watershed is 18 km wide and 36 km long. The coordinate of the image centre are 449205.0E/4645240.0N (UTM Zone 15, NAD 83). The large gray areas are towns.

In the IEM formulation, the like polarized backscattering coefficients for surfaces with small or medium roughness are given by:

$$\sigma_{pp}^0 = \frac{k^2}{2} \exp(-2k_z^2 s^2) \sum_{n=1}^{\infty} |I_{pp}^n| \frac{W^{(n)}(-2k_x, 0)}{n!}, \quad (4)$$

where k is the wave number, θ is the incidence angle, $k_z = k \cos \theta$, $k_x = k \sin \theta$ and pp refers to the horizontal (HH) or vertical (VV) polarization state and s is the standard deviation of terrain heights. The term I_{pp}^n depends on k , s and on R_H , R_V , the Fresnel reflection coefficients in horizontal and vertical polarizations. The Fresnel coefficients depend directly on the dielectric constant. The symbol $W(-2k_x, 0)$ is the Fourier transform of the n^{th} power of the surface correlation coefficient. In this context, an exponential correlation function has been adopted that seems to better describe the properties of natural surfaces (Fung, 1994). For vegetated soils, the simple approach, based on the so-called water-cloud model (WCM), was developed by Attema and Ulaby (1978), who proposed to represent, in a radiative transfer model, the vegetation canopy as a uniform cloud whose spherical droplets are held in place structurally by dry matter. The WCM represents the power backscattered by the whole canopy σ^0 as the incoherent sum of the contribution of the vegetation σ_{veg}^0 and the contribution of the underlying soil σ_{soil}^0 , which is attenuated by the vegetation layer through the vegetation transmissivity τ^2 . For a given incidence angle the backscatter coefficient is represented by the general form:

$$\sigma^0 = \sigma_{\text{veg}}^0 + \tau^2 \sigma_{\text{soil}}^0. \quad (5)$$

Particularly, this expression can be written in more detailed way:

$$\sigma^0 = A \text{VWC} \cos \theta (1 - \tau^2) + \tau^2 \sigma_{\text{soil}}^0, \quad (6)$$

where VWC is the vegetation water content (kg/m^2), θ the incidence angle, σ_{soil}^0 represents the backscattering coefficient of bare soil that in this case calculated by using the IEM model, τ^2 is the two-way vegetation transmissivity with $\tau^2 = \exp(-2B \text{VWC} / \cos \theta)$. The parameters A and B depend on the canopy type and require an initial calibration phase where they have to be found in dependence of the canopy type.

In this work the model simulation enters differently in the inversion procedure. For the Bayesian approach, the simulated data are generated in order to compare them to the measured data and to create the noise probability density function (pdf) as detailed in the section devoted to this approach. The formulation of the WCM has been used in the derivation of the empirical models for the consideration of all the scattering components that have to be taken into account in the interaction between vegetation-soil and the SAR signal.

4. Description of inversion methodologies

4.1 Empirical methods

The empirical approach has been developed in two separate versions, one for the VWC estimates and the other one for the SM estimates.

For VWC, the linear relationship has been modeled as follows:

$$VWC = A\sigma_1 + B\sigma_2 + C \tag{7}$$

where σ_1 , σ_2 are the backscattering coefficients with the following configurations:

- σ_1 and σ_2 are respectively σ_{HH} , σ_{VV} for C band
- σ_1 and σ_2 are respectively σ_{HH} , σ_{VV} for L band
- σ_1 and σ_2 are respectively σ_{HH} for C band and σ_{HH} for L band

Within 32 fields, some of them have been chosen randomly and considered as training fields. The training data belong to the acquisitions carried out on 1st July, thus assuring that the comparison with the Bayesian approach results is performed under identical training conditions. For the test, the data acquired on 8th July were used. The choice for the training and test data were dictated from the availability of Landsat image in contemporary acquisitions with SAR data. The correlation coefficients and the F-values for the empirical correlations in the training data are listed in table 3.

Empirical model	R ²	F	P
σ_{HH} , σ_{VV} for C band	0.68	26.2	< 0.05*
σ_{HH} , σ_{VV} for L band	0.64	21.8	< 0.05*
σ_{HH} C band/ L band	0.68	26.4	< 0.05*

Table 3. Correlation coefficients (R²), F test values (F) and level of confidence (P) for the empirical models to retrieve VWC (*indicates significance at the 0.05 probability level).

For SM, a different kind of empirical relationship has been supposed because a simple linear relationship similar to the one for VWC did not produce acceptable results. An approach was proposed by Notarnicola et al. 2006, following an approach developed by Chen et al. 2003 and based on a previous work by Dubois et al. 1995. This empirical approach was derived and tested on a subset of the SMEX'02 data, producing acceptable results. However, when applied to the whole data, the results were not satisfactory. Then, in order to take into account the different components in the interaction of the SAR signal with the soil and vegetation, the empirical model has been inspired to the vegetation theoretical model described in section 3.

The SM has been supposed to be a function of backscattering coefficients, of VWC, of the roughness parameter s and of a combination of the roughness parameter multiplied by an attenuation factor expressed as $\exp(-VWC)$:

$$SM = A\sigma_1 + B\sigma_2 + C VWC + D s + E s \exp(-VWC) + F. \tag{8}$$

This relationship should take into account the following contributions due to the interaction among the signal, the canopy and the soil (Attema & Ulaby, 1978)

- the relationship to the backscattering coefficients is considered as a kind of mean values of the overall responses of soil, vegetation and their interaction;
- the relationship to VWC is fundamental as VWC plays a key role in these densely vegetated fields on the retrieval of SM as already demonstrated in Notarnicola et al. 2007. It quantifies the contribution of VWC to the detected signal.

- the contribute of the soil is divided in two terms, one is SM which in this case is the parameter to be estimated and the other is the roughness parameters s . As showed in previous studies (Notarnicola et al 2007, Du et al, 2008), this last parameter plays an important role also for densely vegetated fields.
- the relationship to $s \cdot \exp(-VWC)$ takes into consideration double bouncing effect which may appear especially for tall plants such as corn plants in case of shorter wavelength (C band). The contribution of the soil is represented by the s parameter multiplied by $\exp(-VWC)$ which represent the attenuation of the signal from soil due to the presence of vegetation.

The correlation coefficients and the F-values for the empirical correlations in the training data are listed in table 4.

Empirical model	R ²	F	P
$\sigma_{HH}^0, \sigma_{VV}^0$ for C band	0.32	2.30	> 0.05***
$\sigma_{HH}^0, \sigma_{VV}^0$ for L band	0.42	3.41	< 0.10**
σ_{HH}^0 C band/ σ_{HH}^0 L band	0.48	4.20	< 0.05*

Table 4. Correlation coefficients (R²), F test values (F) and level of confidence (P) for the empirical models to retrieve SM (*indicates significance at the 0.05 probability level; **indicates significance at the 0.10 probability level;***indicates that based on F test the relationship is considered not reliable).

The data in table 3 and 4 illustrate the difficulty to infer information about SM especially in the case of the C band. If the data are further divided in two groups, soybean and corn fields, the correlation improves notably for corn (R²=0.63) but the correlation is not considered reliable for the F test. For the soybean fields, the correlation does not change considerably with respect to the values shown in table 4.

The training data were used to evaluate the multiple regressions. The obtained relationships are then applied to the test data in order to verify their generalization capabilities and robustness. This analysis is illustrated in the section dedicated to the results comparison.

4.2 Bayesian methodology

The main aim is to infer the soil parameter values, S_i , that for vegetated soils are the soil dielectric constant ϵ , the standard deviations of heights, s , and the correlation length, l , and the vegetation water content VWC by measuring features f_1, f_2, \dots , in this case represented by backscattering coefficients, $\sigma_{1m}, \sigma_{2m}, \dots$. The procedure is divided into training and test phase.

In the training phase, the conditional probability $P(\sigma_{1m}, \sigma_{2m}, \dots | S_i)$ can be estimated by using the Bayes' theorem from a part of the data. This is the probability of finding that particular vector of features σ_i , given specific values of S_i .

By using IEM, theoretical values of the sensors responses, in correspondence to ground truth, are obtained. The latter are compared to the experimental values introducing random variables, N_i , not depending on ϵ, s and l and representing a function that takes into account some noise factors such as the sensor noise, the error introduced by IEM and the contribute of vegetation (Notarnicola et al., 2006). The problem consists in finding an estimate of the

$P(\sigma_{1m}, \sigma_{2m}, \dots, | S_i)$ by taking into account the presence of this noise factor N_i and setting the relationship between measured and simulated data as follows:

$$\sigma_{im} = N_i \sigma_{ith} \quad (9)$$

where σ_{im} and σ_{ith} are respectively the measured and theoretical values of sensor responses. Once calculated the function $P(\sigma_{1m}, \sigma_{2m}, \dots, | S_i)$, the Bayes' theorem allows for the calculation of the posterior probability from the above conditional probability and the prior probability:

$$P(S_i | \sigma_{1m}, \sigma_{2m}, \dots) = \frac{P(S_i) P(\sigma_{1m}, \sigma_{2m}, \dots | S_i)}{\int P(S_i) P(\sigma_{1m}, \sigma_{2m}, \dots | S_i) dS_i} \quad (10)$$

In the case of bare fields, the theoretical values calculated by the IEM model should be as close as possible to the measured ones and then the pdf mean should be close to the value of 1 with a standard deviation that represents the field variability as well as the sensor error. For vegetated areas, the resulting pdf means should quantify the different behavior of radar signal for bare and vegetated fields. Thus pdfs should contain information on some vegetation parameters that influence the radar signal. Particularly, a good correlation has been found between pdf means and VWC. Instead of correlating pdf means directly to measured VWC, the estimates of this parameter, obtained from a LANDSAT image, have been considered. The purpose is to verify whether the pdf mean variations can be predicted using VWC derived from other remotely sensed data. The methodology for the calculation of VWC from LANDSAT images has been derived and tested in Jackson et al. 2004. The pdf means have been correlated to these LANDSAT derived VWC. A linear relationship has been presumed among pdf means and VWC, initially in the following form:

$$Pdf_1 = a_1 VWC + b_1 \quad (11)$$

$$Pdf_2 = a_2 VWC + b_2 \quad (12)$$

The general trend indicates that pdf means decrease as VWC increases. However the trend is not constant, a group of data belonging to corn fields has a particular behavior and also if the VWC is relatively high (around 4 kg/m²) the corresponding pdf means is high as well. This is in contrast to what established before. This group is made up of pdf values that indicate a relative small difference between the measured and the theoretical backscattering coefficients. This may be ascribed to the presence of a rough surface whose contribute to theoretical backscattering coefficients is higher with respect to a smooth surface (Ulaby et al.,1986).

Within each vegetated group, soybean fields are characterized by low values of s , around 0.6 cm, which determine low values of theoretical backscattering coefficients. Then the ratio between measured and theoretical values is below 1 even if the vegetation is not very dense. On the other side, the roughness in the corn fields is characterized by higher values of s . The rougher surface contributes with high theoretical backscattering coefficients and determines values of the ratio not as low as expected in the case of this dense corn vegetation.

The correlation between pdf means, VWC and *s* has been also considered in the inversion procedures as a multiple fit:

$$Pdf_1 = a_1 VWC + b_1 s + c_1 \tag{13}$$

$$Pdf_2 = a_2 VWC + b_2 s + c_2 . \tag{14}$$

Table 5 reports the correlation coefficients (R^2) for the considered remotely sensed data configuration for the linear relationships (11) and (12) between pdf means and VWC values and the linear relationships (13) and (14) among pdf means, VWC values and the roughness parameter *s*.

Polarization/frequency	Only VWC R^2	VWC + roughness R^2
C_{HH}+C_{Vv}	0.23	0.50
L_{HH}+L_{Vv}	0.61	0.85
C_{HH}+L_{HH}	0.37	0.52

Table 5. Correlation coefficients (R^2) for a linear relationship between pdf means and the VWC values (column 2) and among pdf means, VWC values the roughness parameter *s* (standard deviation of heights) (column 3).

The aim of the training phase is to evaluate the pdf $P(S_i | \sigma_{1m}, \sigma_{2m}, \dots)$ while in the test phase the expression (10) is applied on the second half of the acquired data in order to verify the prediction capability of this methodology.

The dependence of the pdf means on the amount of VWC introduces a new variable in the inversion problem (Notarnicola et al, 2007) that can be used to extract VWC values themselves from the radar signal. With the introduction of the VWC as a new variable *k*, the posterior pdf expressed in (10) can be written as follows:

$$P(\epsilon, s, l, k | \sigma_{1m}, \sigma_{2m}, \dots) = \frac{P_{\text{prior}}(\epsilon, s, l, k) P(\sigma_{1m}, \sigma_{2m}, \dots | \epsilon, s, l, k)}{\iiint_{\epsilon, s, l, k} P(\epsilon, s, l, k) P(\sigma_{1m}, \sigma_{2m}, \dots | \epsilon, s, l, k) d\epsilon ds dl dk} \tag{15}$$

As the main interest was to extract dielectric constant values from which SM can be calculate (Hallikainen et al., 1985), a first integration over the pdf $P(\epsilon, s, l, k | \sigma_{1m}, \sigma_{2m}, \dots)$ is performed with respect to the roughness parameters, *s* and *l*, and *k* over their range of values in order to obtain a marginal distribution:

$$P(\epsilon | \sigma_{1m}, \sigma_{2m}, \dots) = \frac{\iiint_{s, l, k} P(\epsilon, s, l, k) P(\sigma_{1m}, \sigma_{2m}, \dots | \epsilon, s, l, k) ds dl dk}{\iiint_{\epsilon, s, l, k} P(\epsilon, s, l, k) P(\sigma_{1m}, \sigma_{2m}, \dots | \epsilon, s, l, k) d\epsilon ds dl dk} . \tag{16}$$

This distribution represents the probability of the different dielectric constant values for the possible combination of measured backscattering coefficients $\sigma_{1m}, \sigma_{2m}, \dots$, (Notarnicola & Posa., 2004).

Analogous calculation can be performed for the variable k which represents VWC. The pdf $P(\varepsilon, s, l, k | \sigma_{1m}, \sigma_{2m}, \dots)$ has to be integrated over the whole range of dielectric constant values and roughness parameters in order to obtain a pdf that retains exclusively information on the VWC:

$$P(k | \sigma_{1m}, \sigma_{2m}, \dots) = \frac{\iiint_{\varepsilon, s, l} P(\varepsilon, s, l, k) P(\sigma_{1m}, \sigma_{2m}, \dots | \varepsilon, s, l, k) d\varepsilon ds dl}{\iiint_{\varepsilon, s, l, k} P(\varepsilon, s, l, k) P(\sigma_{1m}, \sigma_{2m}, \dots | \varepsilon, s, l, k) d\varepsilon ds dl dk} \quad (17)$$

From this distribution the mean value and the variance of the estimator can be extracted (Gelman, 1995) as follows:

$$\bar{\varepsilon} = \int_{\varepsilon} \varepsilon \cdot P(\varepsilon | \sigma_{1m}, \sigma_{2m}, \dots) d\varepsilon \quad \bar{k} = \int_k k \cdot P(k | \sigma_{1m}, \sigma_{2m}, \dots) dk \quad (18)$$

$$\sigma^2(\varepsilon) = \int_{\varepsilon} (\varepsilon - \bar{\varepsilon})^2 \cdot P(\varepsilon | \sigma_{1m}, \sigma_{2m}, \dots) d\varepsilon \quad \sigma^2(k) = \int_k (k - \bar{k})^2 \cdot P(k | \sigma_{1m}, \sigma_{2m}, \dots) dk \quad (19)$$

In all these calculations, the prior pdf for the parameters, over which integration is performed, has to be specified. In the integration for the calculation of the marginal distribution, the prior pdf has been considered uniform across the whole possible range of values. This means that no supplementary information about these parameters was considered apart from their range of values. The dielectric constant has been integrated in the range from 2 to 20 and the VWC in the range 0.1 to 8 kg/m². The integration window for s is [0.1 cm, 3.0 cm] and for l is [0.1 cm, 21.0 cm], they cover most of the surface measurements. The purpose was to verify the capability to extract dielectric constant and VWC values independently from roughness levels. This procedure has been applied to backscattering coefficients $\sigma_{1m}, \sigma_{2m}, \dots$ in the following configurations:

- C band, HH and VV polarizations;
- L band, HH and VV polarizations;
- C and L band, HH polarization.

5. Results of the single methodologies and relative comparison

As illustrated in previous paragraphs, the inversion methodologies have been applied to different sensors configurations, trying to exploit if the combination of different polarizations and/or bands may help to extract the soil features. In fact, due to the different way C band or L band signals interact with soil and the above canopy layer, they are sensitive to different surface characteristics. Then their use is important to the concept of the ensemble that will be described in the next section.

In this paragraph, the results of the empirical and Bayesian methodologies are illustrated and evaluated in terms of:

- Correlation coefficients, R^2 , between the estimates and the ground truth values
- Root Mean Square Error, RMSE, between the estimates and the ground truth values.

This analysis is carried out on the test data. Tables 6 and 7 list the performance characteristics of the single procedure for each sensor configuration respectively for VWC and SM estimates. The best performances are done by the C and combination of C and L band data for the Bayesian approach, while for the empirical approach only the L band retain the good performances obtained during the training phase.

Methods	R ²	RMSE (kg/m ²)
Empirical C band	0.20	2.44
Empirical L band	0.56	1.29
Empirical C - L band	0.25	2.27
Bayesian C band	0.64	1.30
Bayesian L band	0.46	1.46
Bayesian C - L band	0.55	1.29

Table 6. Correlation coefficients (R²), RMSE for the comparison between the different estimates and the ground truth values for VWC values.

Methods	R ²	RMSE (cm ³ /cm ³)
Empirical C band	0.20	0.11
Empirical L band	0.0006	0.09
Empirical C - L band	0.05	0.12
Bayesian C band	0.14	0.11
Bayesian L band	0.17	0.08
Bayesian C - L band	0.47	0.05

Table 7. Correlation coefficients (R²), RMSE for the comparison between the different estimates and the ground truth values for SM values.

As expected the estimation of SM is quite difficult, thus determining values of R² not higher than 0.47 and high RMSE up to 0.12 cm³/cm³. The performance of the empirical and Bayesian approach improves if the extreme values of SM are excluded from the error computation. In this case, the results are illustrated in table 8 where values of SM higher than 0.27 cm³/cm³ and lower than 0.10 cm³/cm³ have been excluded.

Methods	R ²	RMSE (cm ³ /cm ³)
Empirical C band	0.11	0.06
Empirical L band	0.40	0.05
Empirical C - L band	0.42	0.08
Bayesian C band	0.22	0.10
Bayesian L band	0.45	0.05
Bayesian C - L band	0.65	0.02

Table 8. Correlation coefficients (R²), RMSE for the comparison between the different estimates and ground truth values for SM values, excluding extreme values.

This indicates that both algorithms are not able to predict the extreme values of the SM range. For low values, it depends on the fact that the signal for soil is weak and difficult to be disentangled from the vegetation signal. For high values, the signal from soil is strong but in the case of C band the effect of absorption from 'narrow leaf' plants, such as soybean, determines a lower signal reaching the sensor (Macelloni et al., 2001). The L band estimates are the only one able to predict highest values of SM.

For the Bayesian methodology, similar analyses were also found in Notarnicola et al. 2006. In that case, the methodologies were applied only to few fields of the same data sets. With respect to the accuracy reported in Notarnicola et al., 2006, a worsening in the performance is found. In particular the data set includes all the fields in the WC basin and the fields located in the eastern part which exhibits anomalous values of SM, some very high values around $0.35 \text{ cm}^3/\text{cm}^3$ and some values lower than $0.05 \text{ cm}^3/\text{cm}^3$.

If the watershed is divided in two parts, the western and the eastern part, the performances of the algorithm for SM retrieval differ significantly. The correlation coefficients R^2 are equal to 0.33 and 0.70, not significantly different from those found in Notarnicola et al. 2006.

Furthermore, the performances notably change if in the data set the soybean and corn fields are considered separately. This happens only for the Bayesian approach while the results for the empirical approach remain the same. The results for the Bayesian approach are reported in table 8.

Similar characteristics are also found in (Lakhankar et al., 2009), where it is proved that the RMSE is dependent on the level of vegetation of the different fields. Furthermore, in the case of C band, the signal coming from the VWC dominates over the signal coming from soil. In fact, when the vegetation has low value of VWC, such as in the case of soybean fields, the C band is able to provide acceptable estimates for SM. In the case of corn fields, the best results are obtained with the combination of C and L band, one sensitive to VWC and the other to the surface contribution. In this case, the discrepancies may be ascribed to the fact that in the Bayesian formulation the double bouncing between soil and corn trunk effect is not taken into account. This effect in such kind of plants with broad leaves could dominate (Macelloni et al., 2001).

Methods	R^2	RMSE (cm^3/cm^3)
Corn fields		
Bayesian C band	0.13	0.13
Bayesian L band	0.17	0.09
<i>Bayesian C - L band</i>	<i>0.47</i>	<i>0.06</i>
Soybean fields		
<i>Bayesian C band</i>	<i>0.69</i>	<i>0.03</i>
Bayesian L band	0.18	0.07
<i>Bayesian C - L band</i>	<i>0.67</i>	<i>0.04</i>

Table 9. Correlation coefficients (R^2), RMSE for the comparison between the different estimates and ground truth values for SM values and for the Bayesian approach. With respect to table 7, in this case, the soybean and corn fields are considered separately. In italics, the values significantly different from the ones found in whole data sets are indicated.

6. Ensemble estimates and relative results

The idea to use the ensemble concepts emerges from the previous analysis on the results of the single inversion techniques. Different wavelengths (C - L band) or their combination can be used to extract information according to different types of vegetation, different level of SM and VWC. This information stemming from the previous analysis can be inserted in an ensemble approach. The problem can be formalized in the following way.

The knowledge about the function f , which performs the inversion from the signal domain to the feature domain, is represented by a learning sample of n independent observations:

$$L = \{(S_i, \sigma_i); i = 1, \dots, n\}. \quad (20)$$

An algorithm a is used to fit a model $a(\cdot | L)$ to the data L . Based on this model certain objects of interest $a(S_i | L)$ which describe the distribution of S_i given σ_i can be computed. In this analysis, $a(S_i | L)$ are the linear regression and the Bayesian approach and the objects of interest may be the predicted values

$$S_{ip} = a(S_i | L). \quad (21)$$

At least for regression ($S_i \in \mathbb{R}$) and binary classification problems ($S_i \in \{-1, 1\}$), an ensemble a_E of K basis models can be written as a linear combination of K predictions derived from model a_k , which was fitted using a special learning sample L^k :

$$a_E(S_i | L) = \sum_k \beta_k a_k(S_i | L^k). \quad (22)$$

For real valued responses in regression problems, the prediction of the ensemble is a weighted sum of the predictions of the basis models. The next part of the section is dedicated in finding the best solution in order to create weighted estimates starting from the estimates of the single learners.

In this case, the single learners are represented by the empirical and Bayesian approaches applied to different sensor configurations. As indicated in the section dedicated to the results analysis, the information given by the different approaches and the different configurations are in many cases complementary with respect to the type of vegetation, and of SM values. Then they can be considered as the members of an ensemble and the main aim is to find the best combination of members which then will lead to find the best estimates for the inversion problem.

In this case, one of the main differences with respect to the traditional ensemble techniques is that the single learner is trained separately and then the estimates are considered as part of an ensemble.

The inversion approaches have been applied to the training data in order to calculate the RMSE considering the following configurations:

- For SM, 5 different levels: 0.0-0.10 - 0.10-0.15 - 0.15-0.20 - 0.20-0.25- higher than 0.25 cm^3/cm^3 ;
- Within each of these groups this is the further distinction between corn and soybean.
- For VWC, three main groups have been considered, 0.0- 1.0, 1.0-3.0, higher than 3.0 kg/m^2 .

For each of these groups, the RMSE errors have been calculated in order to verify for which of the six inversion procedures adopted it is possible to find the lowest value of RMSE. The output of this procedure is illustrated in the following tables:

Methods/ranges	0.0-1.0	1.0-3.0	> 3.0
Empirical C band			
Empirical L band		x	
Empirical C - L band			
Bayesian C band	x		x
Bayesian L band			
Bayesian C - L band			

Table 10. Approaches which exhibit the lowest RMSE in three different VWC ranges

Methods/ranges	0.0-0.10	0.10-0.15	0.15-0.20	0.20-0.25	> 0.25
Empirical C band- corn			x		
Empirical C band- soybean					
Empirical L band -corn					
Empirical L band- soybean			x	x	
Empirical C - L band - corn	x	x			
Empirical C -L band- soybean					
Bayesian C band - corn					
Bayesian C band - soybean	x				
Bayesian L band - corn				x	x
Bayesian L band - soybean		x			x
Bayesian C - L band- corn					
Bayesian C - L band - soybean					

Table 11. Approaches which exhibit the lowest RMSE in five different SM ranges, further divided in corn and soybean groups.

This analysis is the base for the application on the test data sets. The six approaches have been applied to the test data and the best estimates have been calculated by using the following three steps:

- Step 1 Calculation of the estimates average, considering all the values if they fall in the same range and excluding one or two values if they disagree with the other ones. If there is a conflict between an empirical estimate and a Bayesian one, the last has been chosen as it is most reliable in many cases. This first step is useful to individuate the range of the estimates and then adopt the best estimator. For VWC the range of the parameters has been also compared with the estimates deriving from the LANDSAT images by using the approach of Jackson et al. 2004.
- Step 2. Considering each estimate, a RMSE coming from the training data have been associated and a new mean has been calculated by considering only the first three values which have the lowest RMSE.
- Step 3. To the two mean values calculated at point 1 and 2, a correction factor is applied which gives more weight to the mean value with the lowest variance. Furthermore, in case of presence of high values of SM, the results from the Bayesian approach in L band has been used as it is the only approach which is able to detect high values of SM.

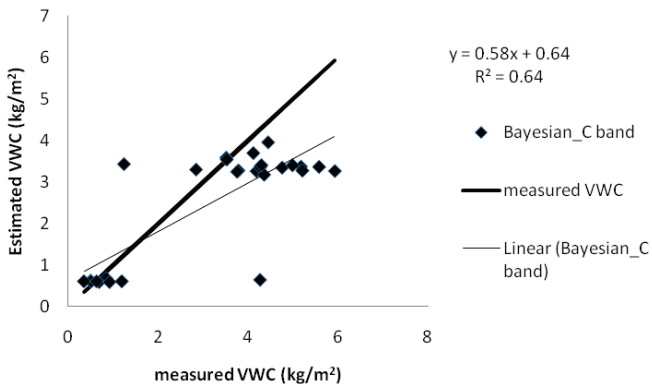
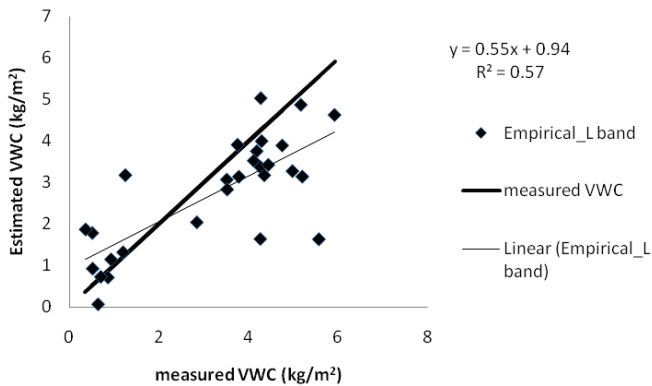
The idea of the procedures originates from the ability of the different procedure and configurations to detect some specific soil and vegetation characteristics.

The output of these procedures has been reported in table 12.

Ensemble	R ²	RMSE
VWC	0.66	1.20 (kg/m ²)
SM	0.83	0.03 (cm ³ /cm ³)

Table 12. Results from the ensemble approach applied to VWC and SM estimates.

The results reported in table 12 indicate a notable improvement in the estimation of both SM and VWC considering the R² between measured and estimated values. For the RMSE, the improvement is evident especially for SM, while for VWC the ensemble RMSE is similar to the one found for the Bayesian approach considered as a single learner. Anyhow, the VWC values were already quite well estimated from the single approaches, and then the ensemble approach is not expected to improve much more the estimation as revealed comparing both R² and RMSE (Brown et al, 2005) On the other side, it is interesting to highlight the information for SM that has been extracted from the single learners and that contribute to determine the better estimates. The results of the ensemble technique are illustrated in figure 2 for VWC and in figure 3 for SM. In each figure, there are four graphs where the results from the three approaches with the highest accuracy and the ensemble results are reported.



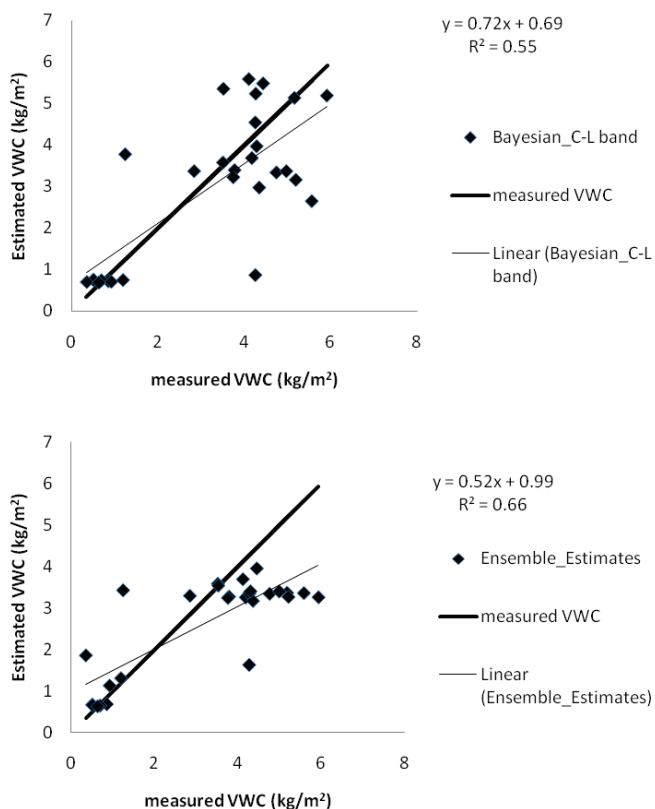
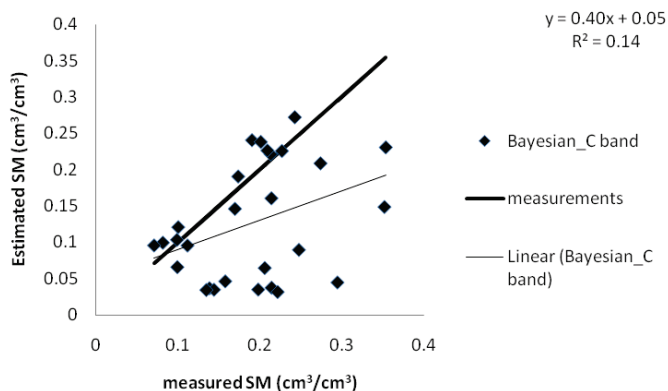


Fig. 2. Results from the single approaches (the three ones with the highest accuracy) and the ensemble approach applied to VWC estimates. The mean rmse of ground data is around 20%. Each graph reports the correlation coefficient R^2 and the linear fit between measured and estimated VWC.



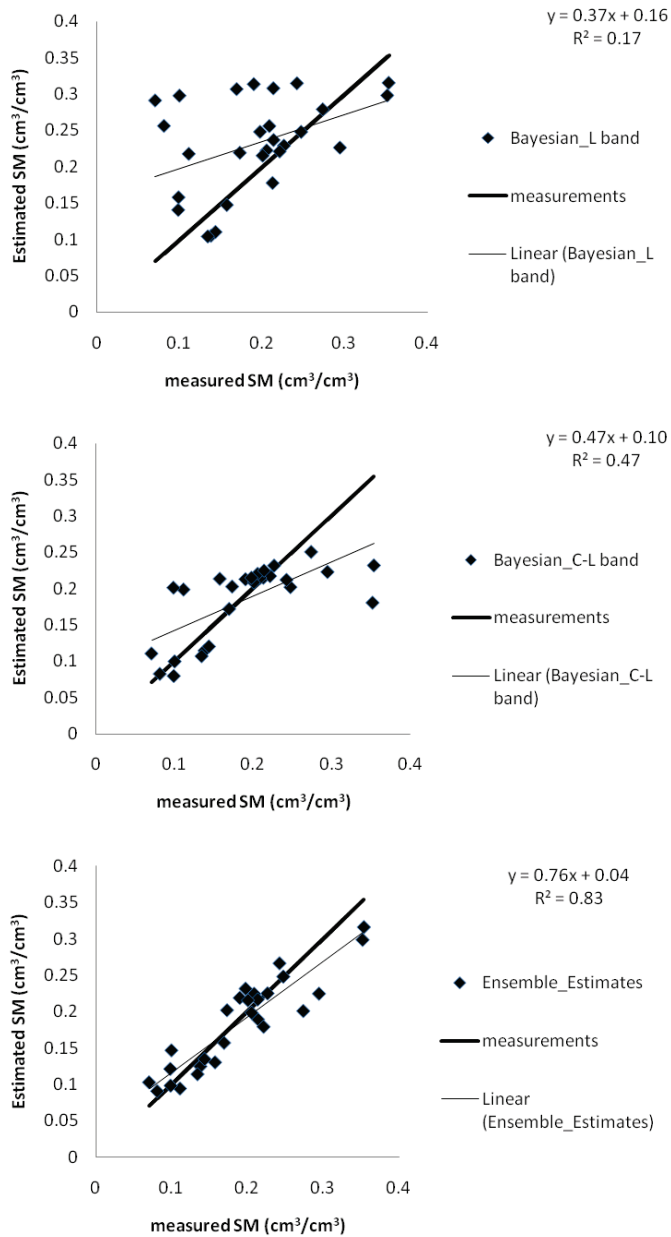


Fig. 3. Results from the single approaches (the three ones with the highest accuracy) and the ensemble approach applied to SM estimates. The confidence interval for measured SM values is $\pm 0.05 \text{ cm}^3/\text{cm}^3$. Each graph reports the correlation coefficient R^2 and the linear fit between measured and estimated SM.

7. Conclusions and future applications

The main aim of this chapter is to illustrate the application of some standard inversion procedures, an empirical and a Bayesian approach for the estimation of VWC and SM from radar images in cases of densely vegetated fields. In this analysis, the presence of vegetation determines a strong disturb to the evaluation of SM. Both methodologies make use or are related to the formulation of theoretical electromagnetic models such as IEM for bare soils and WCM for vegetated fields. The approaches have been applied considering one frequency channel, C or L band and their combination. In all the case, both co-polarized channels, HH and VV, have been used. Subsequently these single learners have been considered as members of an ensemble and a procedure mainly based on the variance minimization has been applied to derive the best estimates.

Results from the single learners indicate that for VWC:

- The algorithms are able to detect three main ranges: from 0.0 to 1.0 Kg/m², from 1.0 to 3.0 Kg/m² and values higher than 3 Kg/m².
- The Bayesian approach determines the best estimates especially in terms of RMSE.
- In the case of Bayesian approach both C and L band can provide reliable estimates with high correlation coefficients and low RMSE values.

While for SM:

- The empirical approach works better if the extreme value of SM are excluding from the computation of R² and RMSE. This demonstrates the ability of the approach to determine an average SM status, but in case of extreme situations such as very low or high values of SM, the algorithm is not enough sensitive to these values and able to disentangle the vegetation effect from the radar signal.
- Also the Bayesian approach is sensitive to this problem even in a minor way. In fact the estimates improve if some anomalous SM values are eliminated. These SM high values are not correlated to high values of backscattering coefficients or VWC.
- In the Bayesian approach, the different use of C and L band emerges if soybean and corn fields are analyzed separately. In this case, for the corn fields, only the combination of C and L band can provide estimates with acceptable R² and RMSE. For soybean fields, good results are determined by both C band and the combination of C and L band.

These analyses are the starting point from which the ensemble part derives. It is clear that there is not a unique method which can provide reliable estimates for all types of soil condition in terms of vegetation and SM status. These are due to the limitation of the method itself, for example generally the empirical approaches are quite site specific, but in some cases, each method or sensor configuration is able to detect some specific characteristics and is insensitive to some others.

The ensemble approach used in this work considers the single estimates and determines the best estimates based on an approach which aims at minimizing the variance in an iterative way. Results from the ensemble learner indicate:

- For VWC the improvement is not so evident, even because the single estimates were already good enough.

- The net improvement is evident for SM, where diverse capability of each single learner to detect specific SM condition (e.g. Bayesian approach L band was the only one able to predict high values of SM) emerges.

For further validation, this new procedure will be applied to other data sets and enriched with other inversion techniques.

8. References

- Attema, E.P.W., & Ulaby, F.T. (1978). Vegetation modeled as a water cloud. *Radio Science*, 13, 357 - 364.
- Beale, R. & Jackson, T. *Neural Computing: An Introduction*. Bristol, U.K.: Adam Hilger, 1991.
- Bindlish, R. & Barros, A. P. (2000). Multifrequency soil moisture inversion from SAR measurements with the use of IEM, *Remote Sens. Environ.*, vol. 71, no. 1, pp. 67-88.
- Bishop, C. M., (1995). *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press.
- Brown, G., Wyatt, J.L., Tino, P. (2006) Managing Diversity in Regression ensembles, *Journal of machine learning research* 6, 1621-1650, 2006.
- Camps-Valls, G., Bruzzone L, Rojo-Alvarez, J., Melgani, F. (2006). Robust Support Vector Regression for Biophysical variable estimation from remotely sensed Images, *IEEE Trans. Geosci. Remote Sens.Letters*, vol. 3, no. 3,pp. 339-343.
- Chen, D., Jackson, T. J., Li, F., Cosh, M. H., Walthall, C. & Anderson M. (2003). Estimation of vegetation water content for corn and soybeans with a normalized difference water index (NDWI) using Landsat Thematic Mapper data, *Proc. IGARSS*, Toulouse, France, pp.2853-2856.
- Cho, S. & Kim, J.H. (1995). Multiple Network Fusion Using Fuzzy Logic. *IEEE Transactions on Neural Networks*, 6(2), 497-501.
- Delworth, T.L., & Manabe, S. (1988). The influence of potential evaporation on the variabilities of the simulated soil wetness and climate. *J. Climate*, 1(5), 523-547.
- Du, Y., Luo, Y.L., Yan, W.Z.,(2008) An electromagnetic scattering model for soybean canopy, *Progress in electromagnetic Research*, PIER 79, 209-223.
- Dubois P. C., van Zyl J., & Engman T. (1995). Measuring soil moisture with imaging radars," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 4, pp. 915-926.
- Fung, A. K. (1994) *Microwave Scattering and Emission Models and their Application*. Artech House, Boston.
- Gao, B.C. (1996). NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space, *Remote Sensing of Environment*, Vol. 58, 257-266
- Gelman, R., Carlin, J. B., Stern, H. S. & Rubin, D. B. (1995). *Bayesian Data Analysis*. London, U.K.: Chapman & Hall, 1995.
- Haddad, Z. S. & Dubois, P. (1994). Bayesian estimation of soil parameters from remote sensing data," in *Proc. IGARSS*, vol. 3, pp. 1421-1423.
- Hallikainen, M. T., Ulaby, F. T., Dobson, M. C., El-Rayes, M. A., & Wu, L.. (1985). Microwave dielectric behavior of wet soil—Part I: Empirical models and experimental observations," *IEEE Trans. Geosci. Remote Sens.*, vol. GRS-23, no. 1, pp. 25-34, Jan. 1985.
- Jackson, T. J., Chen, D., Cosh, M., Li, F., Anderson, M., Walthall, C., Doriaswamy, P. & Hunt, E. R. (2004). Vegetation water content mapping using Landsat data derived

- normalized difference water index for corn and soybeans. *Rem. Sen. Environment*, 92: 475 – 482.
- Lakhankar, T., Ghedira, H., Temimi, M., Sengupta, M., Khanbilvar, R., Blake, R., (2009). Non-parametric methods for soil moisture retrieval from satellite remote sensing data, *Remote Sensing*, 1, 3-21. doi: 10.3390/rs1010003.
- Lin D. S., Wood E. F., Bevan K., & Saatchi S. (1994). Soil moisture estimation over grass-covered areas using AIRSAR, *Int. J. Remote Sens.*, vol. 15, no. 11, pp. 2323-2343.
- Liu, Y., & Yao, X. (1999). Ensemble Learning via Negative Correlation. *Neural Networks*, 12(10), 1399-1404.
- Macelloni, G., Paloscia, S., Pampaloni, P., Marliani, F., Gai, M., (2001). The relationship between the backscattering coefficient and the biomass of narrow and broad leaf crops, *IEEE Transaction on Geoscience and Remote Sensing*, vol. 39, no. 4, pp. 873-884.
- Marchant, J. A. & Onyango, C. M. (2003). Comparison of a Bayesian classifier with a multilayer feed-forward neural network using the example of plant/weed/soil discrimination, *Comput. Electron. Agric.*, vol. 39, no. 1, pp. 3-22.
- Mattia F. & Le Toan T. (1999). Backscattering properties of multi-scale rough surfaces, *J. Electromagn. Waves Appl.*, vol. 13, no. 4, pp. 493-527.
- Miller, E.A., & White, R.A., (1998) A contaminous United States Multilayer soil characteristics data set for regional climate and hydrology modeling, *Earth Interact.*, vol. 2, n.2, pp.1-26.
- Notarnicola C., D'Alessio A. C., Casarano D., Posa F., & Sabatelli V. (2003). Use of a C-band ground-based scatterometer to monitor surface roughness and soil moisture changes, *Subsurf. Sens. Technol. Appl.: Int. J.*, vol. 4, no. 2, pp. 187-206.
- Notarnicola, C., & Posa, F. (2004). Bayesian Algorithm for the Estimation of the Dielectric Constant from Active and Passive Remotely Sensed Data. *IEEE Geoscience and Remote Sensing Letters*, 1(3), 179-203.
- Notarnicola, C., Angiulli M. & Posa, F. (2006). Use of radar and optical remote sensing data for soil moisture retrieval over vegetated areas. *IEEE Trans. Geoscience Remote Sensing*, 44(4).
- Notarnicola, C. & Posa, F. (2007). Inferring vegetation water content from C and L band images, *IEEE Transactions on Geoscience and Remote Sensing*, vol.45, no.10, p.3165-3171.
- Notarnicola C., Angiulli M. & Posa F. (2008). Soil Moisture Retrieval From Remotely Sensed Data: Neural Networks Approach Versus Bayesian Method, *IEEE Transaction on Geoscience and Remote Sensing*, vol. 42, no. 2, pp. 547-557, February 2008.
- Oh Y., Sarabandi K., & Ulaby F. T. (1992). An empirical model and an inversion technique for radar scattering from bare soil surfaces, *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 2, pp. 370-381, Mar. 1992
- Pennuelas J., Filella L., Biel C., Serrano L. & Save R.. (1993). "The reflectance at the 950– 970 μm region as an indicator of plant water status, *International Journal of Remote Sensing*, 14, 1887-1905.
- Prevot L., Dechambre M., et al., (1993). Estimating the characteristics of vegetation canopies with airborne radar measurements, *Int. J. Remote Sens.*, vol. 14, no. 15, pp. 2803–2818.
- Ueda, N. & Nakano, R. (1996). Generalization error of ensemble estimators. *Proceedings of International Conference on Neural Networks*, pages 90–95, 1996.

- Ulaby, F. T. , Moore R. K. & Fung A. K., (1986). *Microwave Remote Sensing: Active and Passive*. Norwood, MA: Artech House, 1986, vol. 2.
- van Zyl, J., Carande. R., Lou. Y., Miller T., & Wheeler, K. (1992). The NASA/JPL three-frequency polarimetric AIRSAR system. *IEEE IGARSS Dig.*, 1, 649-651.

Methodology for investigation of the factors for georadar signals influencing the directional pattern of synthetic aperture radar

Zolotarev I.D.
Omsk State University (OMSU)
Russia
Miller Ya.E.
ACADEMY MBF
Russia

1. Introduction

The SAR pattern width is the most significant characteristics of modern georadars providing remote sensing of the Earth within the specified radar swath. A simplified consideration of formation of the SAR directional pattern is given for an idealized case of signal pickup from the equidistant points along the vehicle trajectory at its constant travel speed. Nevertheless, the signal parameters influencing the character of the directional pattern of the synthetic aperture radar are not constant at implementation of the georadar with SAR. This results in swinging in time of the SAR directional pattern in addition to its widening; therefore, the characteristics of the detected extended object on the Earth may differ from the real ones.

The given chapter contains a new methodology for research and optimization of the directional pattern for the interferometric SAR and the calculated examples of the above-mentioned methodology. A peculiar method of transients determination in the selective filters entering the SAR path at the radar signal passing through them lies in the basis of the calculation procedure for determination of swinging of the radar antenna directional pattern. There is taken into account influence on the SAR characteristics of non-equidistance of the readings along the vehicle trajectory. The Doppler effect influence on formation of the antenna directional pattern is also under consideration. There is given the method of taking into account the out-of-parallelism of the beams for each point of the sensed surface at formation of the SAR directional pattern.

It is worth mentioning that even a small deviation of the SAR directional pattern caused (in particular) by dynamic mode of the georadar path operation may result in a considerable inaccuracy of information acquisition at remote sensing of the Earth. For example, at the vehicle altitude of $h = 500$ km and the antenna direction error of 1° , the error in determination of the coordinates of each detected point of the Earth surface is around 10 km (which is not permissible for information acquisition at remote sensing of the Earth).

The given chapter deals with consideration of a combined influence of the above-mentioned factors on the SAR characteristic. There are given recommendations on minimization of a dynamic error of the directional pattern of synthetic aperture radar.

2. Influence of transients on the interferometric SAR characteristics for the sensing pulse with rectangular envelope curve

Numerous works on the SAR equipment are devoted to formation of the directional pattern of the required type for detecting and ranging the objects with provision of the required angular resolution. The use of an interferometric approach at SAR designing permits to increase the angle resolution. A high range resolution is achieved by the maximum possible shortening of a pseudorandom sequence discrete. Formation of the antenna directional pattern (ADP) with synthetic aperture requires sampling at the specified points of the radar carrier trajectory of the amplitude and the phase of the received signal that is significantly lower than the noise level before the correlator. The task of the signal correlation processing is obtaining the required signal-to-noise ratio with the equivalent gain equal to 60-80 dB (Boerner, 2000; Boerner, 2004; Antipov et al., 1988; Filippov et al., 1994).

A significant limiting factor is the transients that inevitably take place in various sections of the signal processing path that have the frequency selective properties; these sections include a physical antenna, the phase-shifting circuits, the summers and the multipliers. Despite the fact that the problems of a steady-state mode for SAR are represented by a broad scope of research, operation of the given systems in the dynamic mode has hardly been described (Vendic & Parnes, 2002). Most probably, this may be justified by a high level of laboriousness of the oscillatory systems research with the accuracy up to a signal phase. Potential possibilities to increase the range resolution are defined by a minimal realizable duration of the sequence discrete. In this case, phase overshooting at each discrete occurring due to the transients limits the informational possibilities of the radars with the PSK and FSK signals.

The impact of transients on the SAR directional pattern is revealed in the work. The given result was acquired on the base of the "fast" inverse Laplace transform (FILT) method developed earlier by one of the authors (the method permits to get a selective system response with the accuracy up to a signal phase and provide solution to the amplitude-phase-frequency problem in radio electronics on the FILT base) (Zolotarev, 1969; Zolotarev, 1996; Zolotarev, 1999; Zolotarev et al., 2004; Zolotarev et al., 2005).

There are used the analog frequency converters for the ultra wideband signals with duration of about 1 ns after the antenna path in the SAR. In this case, the subsequent selective filters determine the resulting channel bandpass. That is why the given work deals with analysis of impact of the transients occurring in these filters on the SAR directional pattern.

There will be analyzed 2 identical unilateral selective elements as a bandpass filter (BF). The BF transfer characteristic can be written down in the form of a fractional rational function

$$K(s) = K_0 \left[\frac{s + b}{s^2 + 2\alpha s + \omega_r^2} \right]^2 = K_0 \left[\frac{s + b}{(s + \alpha)^2 + \omega_0^2} \right]^2$$

Here the damping constant α equals to a half of the bandpass of a separate selective section, ω_r - the resonance frequency, $\omega_0 = (\omega_r^2 - \alpha^2)^{1/2}$ - the filter free frequency; let's assume $b = 2\alpha$.

The image of the radio pulse of intermediate frequency ω_{imd} with τ duration

$$f_{in}(s) = A_0 \left[\frac{s \cdot \sin \psi + \omega_{imd} \cos \psi}{s^2 + \omega_{imd}^2} - \frac{s \cdot \sin \psi_\tau + \omega_{imd} \cos \psi_\tau}{s^2 + \omega_{imd}^2} e^{-s\tau} \right], \quad \psi_\tau = \psi + \omega_{imd}\tau.$$

For the signal image at the BF output we have $f_{out}(s) = f_{in}(s)K(s)$.

Thus, according to FILT (Zolotarev et al., 2005), transition into the space of the originals gives a complex representation of the signal at the filter output

$$\begin{aligned} \dot{f}_{out}(t) = & A_0 \dot{K}(j\omega_{imd}) e^{j(\omega_{imd}t + \psi)} [1(t) - 1(t - \tau)] + 2jK_0 \sum_{l=0}^{n-1} \dot{B}_l t^{n-l-1} e^{(-\alpha + j\omega_0)t} 1(t) - \\ & - 2jK_0 \sum_{l=0}^{n-1} \dot{B}_{l,\tau} (t - \tau)^{n-l-1} e^{(-\alpha + j\omega_0)(t-\tau)} 1(t - \tau). \end{aligned} \quad (1)$$

where the complex constants \dot{B}_l may be found from the expression

$$\dot{B}_l = \sum_{h=0}^s \frac{(-1)^h C_{n+h-1}^h}{(n-l-1)!(l-h)!} \cdot \frac{1}{(2j\omega_0)^{n+h}} \frac{d^{l-h}}{ds^{l-h}} \left[\frac{(s \sin \psi + \omega_{imd} \cos \psi)(s+b)^n}{s^2 + \omega_{imd}^2} \right]_{s=-\alpha + j\omega_0}.$$

Let's look for the real signal as $f_{out}(t) = \text{Im} \{ \dot{f}_{out}(t) \}$.

Let's represent the complex output signal as

$$\dot{f}_{out}(t) = \dot{f}_{norm}(t) \dot{N}(t), \quad (2)$$

where the BF response to a Monoharmonic signal is assumed as a normalizing function represented as $\dot{f}_{norm}(t) = A_0 \dot{K}(j\omega_{imd}) e^{j(\omega_{imd}t + \psi)}$. $\dot{N}(t)$ - the normalized complex envelope curve of the signal at the output of the BF under investigation, module $N(t)$ characterizes the behavior of the signal envelope curve at the BF output, and function $\delta(t) = \arg \{ \dot{N}(t) \}$ determines the current behavior of its phase.

For a plane wave front, the phase difference (caused by the wave arrival under the θ angle) for the base between the adjacent readings a , is determined with the expression

$$\Delta\varphi = \frac{2\pi a}{\lambda} \sin \theta. \quad (3)$$

For the side-lobe suppression, the law of the amplitude distribution along the aperture L is

chosen in the following way: $I(z) = 1 + \Delta \cos(2\pi z/L)$, $z \leq L/2$, where Δ is assumed to be equal to 0.4 (Sazonov, 1988). SAR directional pattern $F(\theta)$ for 100 readings along the spacecraft trajectory is represented in Figure 1, where θ is given in radians.

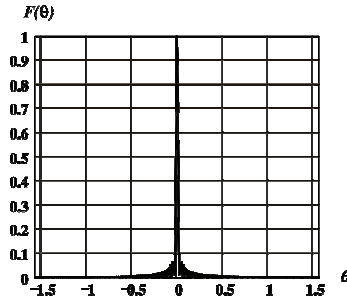


Fig. 1. SAR directional pattern

In fact, $\Delta\varphi$ will have the increment $\delta(t)$ caused by the transient. It will result in dependence of real θ on the time, i.e.

$$\theta(t) = \theta + \Delta\theta(t), \quad \Delta\theta(t) = \delta(t) \frac{\lambda}{2\pi a \cdot \cos \theta}. \tag{4}$$

Figure 2 shows the calculated charts in nondimensional time αt for the transient and the corresponding positions of the SAR directional pattern for various transient time points. The number of the readings along the spacecraft trajectory chosen for calculation equals to 100.

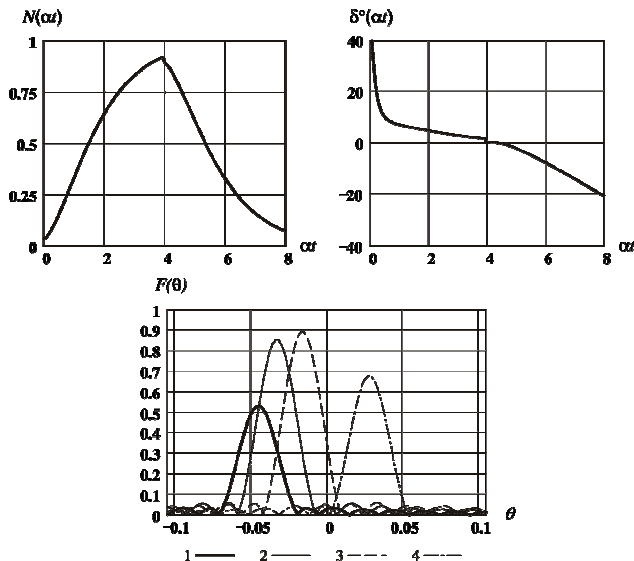


Fig. 2. Design parameters: Q-factor $Q = \omega_r / 2\alpha = 2$, pulse duration $\alpha\tau = 4$; $1 - t = 0.5\tau$, $2 - t = 0.8\tau$, $3 - t = \tau$, $4 - t = 1.4\tau$.

The above-mentioned charts show that “swinging” of the SAR directional pattern in respect to the one calculated for the steady-state mode increases along with the increase of the signal bandwidth and the bandpass filter. Thus, there is limited the accuracy of the direction finding according to the angular coordinates of the detected object. As modern SARs use ultra wideband signals, one should take into account the directional pattern time shift that constitutes the values comparable with the SAR beam width.

Proceeding from the actual dynamic operation mode of the system with SAR, the obtained results make it possible to estimate the limit capabilities of building the synthetic antennae that apply the interference principle.

3. Influence of transients on the interferometric SAR characteristics for sensing pulse with bell-shaped envelope curve

As it is shown in the work (Zolotarev et al., 2005), transients in the elements of the SAR formation circuit provide a significant impact on the direction of the directional pattern major maximum. In this case, when implementing the SAR radars, it is necessary to pay serious attention to the actual SAR characteristic obtained in the result of the corresponding signal conversion. Ignoring of this factor may result in rough errors at determination of the detected surface parameters. A lot of works contain a supposition that smoothing of the envelope curve shape will decrease the influence of transients on the SAR characteristic. Due to this, it seems to be important to consider the SAR formation at the use of a bell-shaped sensing signal with the Gaussian envelope curve. The radio pulses with a sinus-quadratic envelope curve are the characteristics similar to the given signal. Let’s consider the SAR formation for the given signal type at various Q-factors of the antenna filters and signal duration.

There will be analyzed 3 identical unilateral selective elements as a band pass filter. The BF transfer characteristic will be written down as a fractional rational function

$$K(s) = K_0 \left[\frac{s + b}{s^2 + 2\alpha s + \omega_r^2} \right]^3, \text{ Q-factor of the filter } Q = \frac{\omega_r}{2\alpha}.$$

Here, damping constant α equals to a half of the bandpass of a separate selective section, ω_r - the resonance frequency, let’s assume $b = 2\alpha$.

The sensing signal with a sinus-quadratic envelope curve is written down as

$$f_{in}(t) = A_0 \sin^2(2\Omega t) \sin(\omega_c t + \psi) [1(t) - 1(t - \tau)]. \quad (5)$$

Let’s transform the last expression into the form of

$$f_{in}(t) = A_0 \left\{ \frac{1}{2} \sin(\omega_c t + \psi) - \frac{1}{4} \sin[(\omega_c - 2\Omega)t + \psi] - \frac{1}{4} \sin[(\omega_c + 2\Omega)t + \psi] \right\} [1(t) - 1(t - \tau)]$$

The image of a radio pulse with the ω_c frequency, τ duration and the bell-shaped envelope curve

$$f_{in}(s) = \frac{A_0}{2} \left[\frac{s \cdot \sin \psi + \omega_c \cos \psi}{s^2 + \omega_c^2} - \frac{s \cdot \sin \psi_\tau + \omega_c \cos \psi_\tau}{s^2 + \omega_c^2} e^{-s\tau} \right] -$$

$$- \frac{A_0}{4} \left[\frac{s \cdot \sin \psi + (\omega_c + 2\Omega) \cos \psi}{s^2 + (\omega_c + 2\Omega)^2} - \frac{s \cdot \sin \psi_\tau + (\omega_c + 2\Omega) \cos \psi_\tau}{s^2 + (\omega_c + 2\Omega)^2} e^{-s\tau} \right] -$$

$$- \frac{A_0}{4} \left[\frac{s \cdot \sin \psi + (\omega_c - 2\Omega) \cos \psi}{s^2 + (\omega_c - 2\Omega)^2} - \frac{s \cdot \sin \psi_\tau + (\omega_c - 2\Omega) \cos \psi_\tau}{s^2 + (\omega_c - 2\Omega)^2} e^{-s\tau} \right],$$

$$\psi_\tau = \psi + \omega_c \tau .$$

We have $f_{out}(s) = f_{in}(s)K(s)$ for the signal image at the BF output.

Then, according to the FILT (Zolotarev, 1969; Zolotarev et al., 2004; Zolotarev et al., 2005), the transition into space of the originals gives a complex representation of the signal at the filter output $\dot{f}_{out}(t)$, the real signal will be found as $f_{out}(t) = \text{Im}\{\dot{f}_{out}(t)\}$.

Let's represent the complex output signal as

$$\dot{f}_{out}(t) = \dot{f}_{norm}(t)\dot{N}(t),$$

where the BF response to a Monoharmonic signal is assumed as a normalizing function represented as $\dot{f}_{norm}(t) = A_0 \dot{K}(j\omega_c) e^{j(\omega_c t + \psi)}$. $\dot{N}(t)$ - the normalized complex envelope curve of the signal at the output of the BF under investigation, module $N(t)$ characterizes behavior of the signal envelope curve at the BF output and function $\delta(t) = \arg\{\dot{N}(t)\}$ determines the current behavior of its phase.

Concerning the plane wave front, the phase difference (caused by a wave arrival under the θ angle) for the base between the adjacent readings a , is determined with the expression

$$\Delta\varphi = \frac{2\pi a}{\lambda} \sin \theta$$

In fact, $\Delta\varphi$ will have increment $\delta(t)$ caused by the transient. It will result in dependence of real θ on the time, i.e.

$$\theta(t) = \theta + \Delta\theta(t), \quad \Delta\theta(t) = \delta(t) \frac{\lambda}{2\pi a \cdot \cos \theta}.$$

Figure 3, 4 shows the calculated charts in nondimensional time at for the transient and the corresponding positions of the SAR directional pattern for various time points of transients. The number of readings along the spacecraft trajectory chosen for calculation at the SAR formation equals to 1000, $a = \lambda/2$, $\lambda = 0.1$ m.

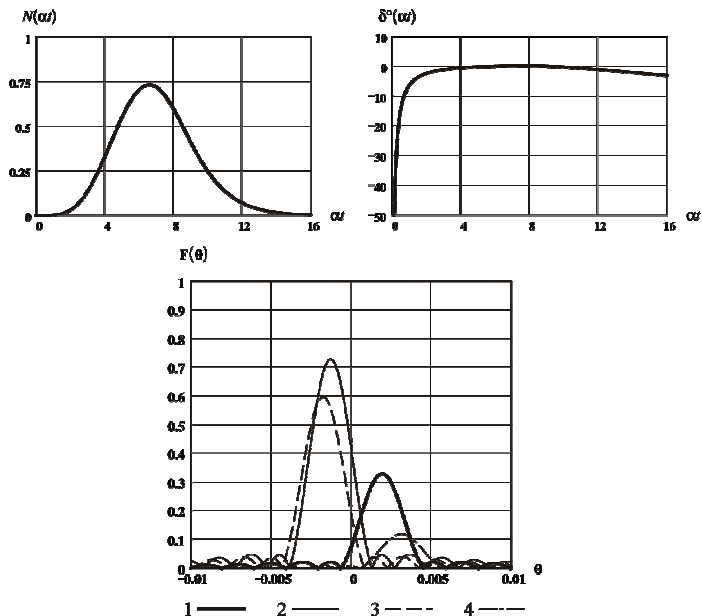


Fig. 3. Design parameters: Q-factor $Q = \omega_r / 2\alpha = 25$, pulse duration $\alpha\tau = 8$; 1 - $t = 0.5\tau$, 2 - $t = 0.8\tau$, 3 - $t = \tau$, 4 - $t = 1.4\tau$.

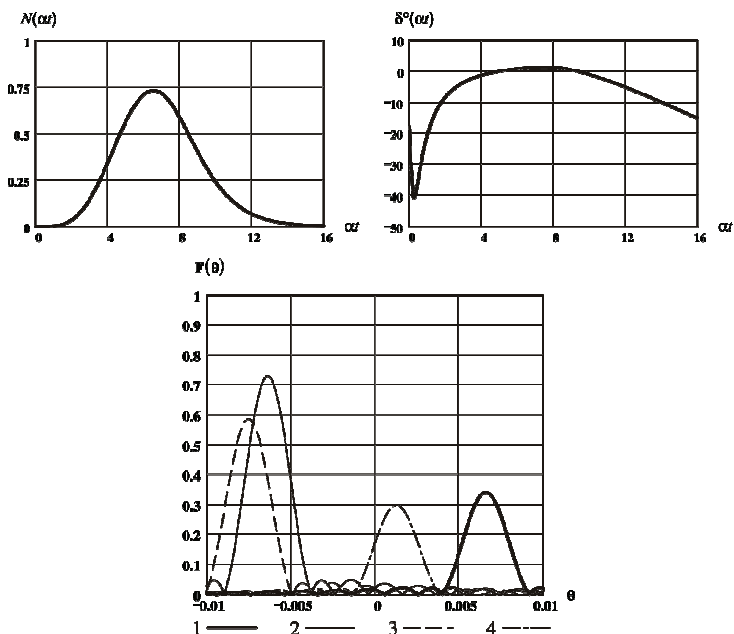


Fig. 4. Design parameters: Q-factor $Q = 5$, pulse duration $\alpha\tau = 8$; 1 - $t = 0.5\tau$, 2 - $t = 0.8\tau$, 3 - $t = \tau$, 4 - $t = 1.4\tau$.

As it proceeds from the calculated charts (Figure 3, 4), the maximum of the SAR directional pattern turns out to be shifted for the bell-shaped (sinus-quadratic) pulse regarding the case of the transients' absence. This shift depends on the filters Q-factor value and rises together with the increase of the signal bandwidth and also depends on the current time of the transient. With the flight altitude being $h = 5000$ m, this shift in the horizontal plane for the object that is being detected reaches considerable values of about several hundreds meters. That is why when designing radars with the SAR, it is necessary to pay serious attention to minimization of the error caused by the transients in the antenna circuit.

4. Research of the effect produced by transients on the correlation properties of the signals with pseudorandom phase shift keying in the systems of the radar remote sensing of the Earth

Significance of modern radar methods of sensing the Earth caused rapid development of the given scientific and engineering areas and their practical application in various research fields of the Earth geostructure. The most important parameters determining quality of these systems are the lock range in the plane that is perpendicular to the carrier path, narrowing of the directional pattern owing to the antenna aperture synthesis as well as the duration of a sensing radio pulse signal providing sequential scanning of the Earth surface along the narrow directional pattern of the synthetic antenna. Nowadays the range resolution of about ten centimeters (at sequential scanning) is treated as the upper reachable limit for the systems of the Earth Remote Sensing (ERS) (Zolotarev et al., 2006). In this case, the value of High Frequency (HF) filling of the radar signal is usually about 3-10 GHz. One of the most important requirements for the given systems is a high level of coherence of HF filling of the sensing signals that is required to form the narrow directional pattern of the antenna with a synthetic aperture. The second requirement proceeds from the necessity to ensure the signal level high increase over the interference signal when making the decision concerning the properties of the Earth surface sensed area. The above-mentioned requirement justifies formation of a pseudorandom sequence of the sensing radar signal with phase shift keying. In this case, the extraction of a low-level signal from the noise is carried out by the correlation device (Varakin, 1985) which is the "heart" of the ERS system. As the sensing signals are distinguished by high frequencies of HF filling, it is necessary to convert frequency (for their primary processing) with use of the intermediate frequency (IF) filters. In this case, the minimal filling frequency constitutes the value of around 1 GHz. With the current level of the processor equipment development, the given condition requires usage of the analogous IF filters at primary processing of the received signal. The unavoidable transients appearing in this case lead to distortion of the phase and envelope curve of each sequence element, and in the end they may result in a significant deterioration of the ERS system correlation device operation (Zolotarev et al., 2004; Zolotarev et al., 2005). However, laboriousness and inconvenience of obtaining the accurate solutions for analyzing the correlator operation with the filters resulted in almost complete lack of the research conducted in the given direction. This prevents from obtaining reliable recommendations when building the ERS system correlator and makes one decide in favor of the idealized model of its operation.

The new results obtained in the given work on the basis of the fast inverse Laplace transform method (Zolotarev, 1969; Zolotarev, 2004) ensuring description of the transient

with the accuracy of up to a signal phase, make it possible to get reliable recommendations when building the high-precision ERS equipment that applies correlation processing of the signals. Figure 5 shows an example of a pseudorandom sequence (PRS) at the output of the correlation device for the ERS system.

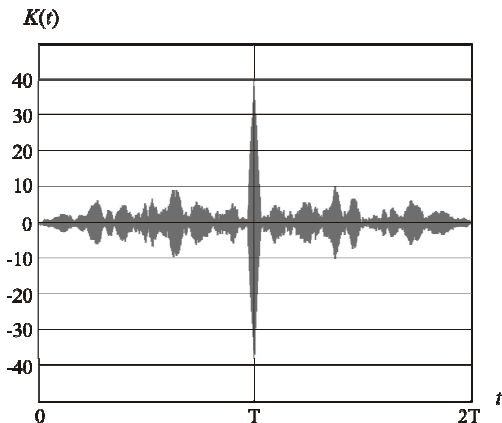


Fig. 5. Calculated parameters: element duration $\tau = 10$ ns, IF filter frequency 1 GHz, sequence length $N = 31$, interference-to-signal ratio $P_n / P_s = 10$.

In the system under implementation the PRS duration constitutes 1023 elements which allows a significant signal level increase above the noise. This permits (owing to the use of the polarity effects and a thin phase structure of the central peak of the correlation function) to obtain important additional information on the results of scanning the Earth.

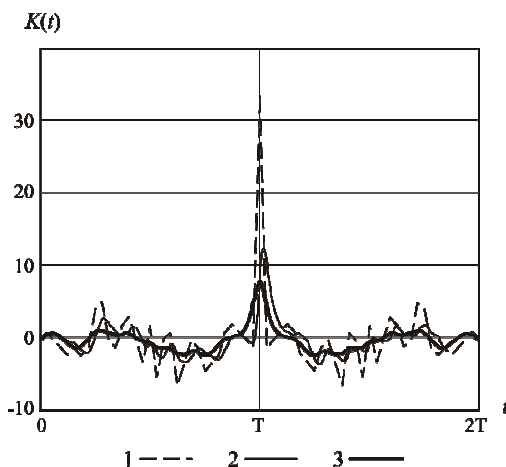


Fig. 6. 1 - ACF for the signal undistorted by the transient; 2 - CCF for the signals from the main path and the reference one; 3 - CF, the reference signal coincides (in its form) with the input signals that have passed through the filters.

One of the ways of building the correlation function for the correlator with filters is that a high frequency component is filtered after the signal correlation processing by means of frequency conversion. Figure 6 shows the output signal for the given case. Curves *b* and *c* correspond to the PRS passing through the detuned filter (the value of detuning equals to a half of the bandpass filter).

The proper operation of the system may be ensured only with the transients taken into account, and in particular, when applying normalization of the levels subject to combination of the parameters of the filters and the signal.

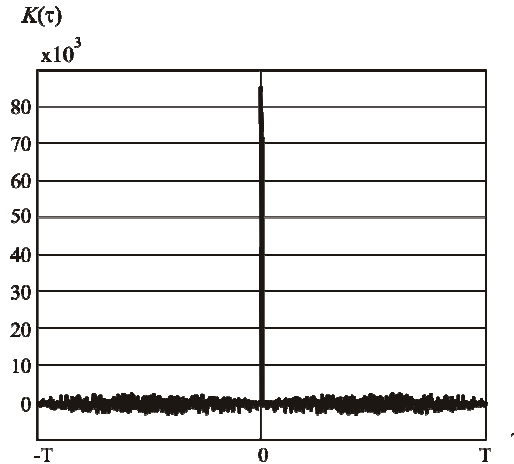


Fig. 7. Correlation function for the sequence with a 1023-element length.

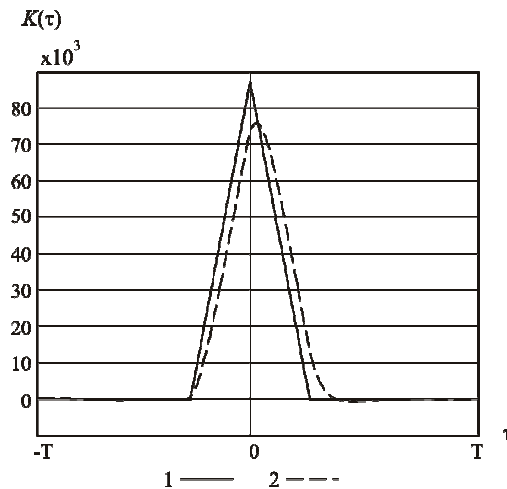


Fig. 8. The correlation peak at the enlarged scale: 1 - ACF for the signal undistorted by transient; 2 - CCF for the signals from the main path and the reference one.

As it proceeds from Figure 6, the transients provide a significant impact on the form of the signal at the correlator output in case of a relatively small number of the elements of

sequence. Thus the range resolution of the ERS system becomes worse. If there is a considerable increase in the sequence length (in this case it is 1023 elements), the influence of the transients on the range resolution reduces significantly. However, as the research conducted revealed, the transients provide a considerable impact on the dynamic shift of the synthetic antenna directional pattern and therefore, there is a decrease in the accuracy of the object location on the Earth surface (Zolotarev et al., 2005; Zolotarev et al., 2006).

5. Research of the influence of transients, non-equidistance of the taken readings, divergence of beams on the interferometric SAR characteristics

There is under consideration a combined influence of the transients in the filters of the radar system selective circuits, non-equidistance of the taken readings and divergence of the beams at the distance up to the Earth surface reflecting elements that is comparable with a synthetic antenna aperture value. There is taken into account influence of the above-mentioned factors on the resolution capability of the radar system for the Earth remote sensing. The transients lead to swinging of the SAR antenna pattern; the other indicated factors result in widening of the synthetic antenna pattern. There are given the corresponding relationships and diagrams that make it possible to take into account the influence of the above-mentioned factors and determine the ways for reduction of the destructive factors influence on the synthetic antenna pattern.

The results of the work are original as a combined influence of the factors has not been under consideration before. According to the calculation results, the factors provide a rather considerable influence on the form of the antenna directional pattern that may result in serious errors when determining the characteristics of the extended object lying within the radar swath.

The research conducted in the given work has revealed that it is impossible to develop the radar system with application of the interferometric SAR without an obligatory consideration of the combined influence of the indicated factors on the SAR ADP.

1. Influence of the transients in the selective filters and the antenna-feeder section of the system path forming the SAR. In this case, there is under consideration the case of application of the identical filters in the selective path that is extremely complicated for analysis. To conduct research of the transients influence, there was applied the method developed in (Zolotarev, 1969; Zolotarev, 2004), providing a fast inverse Laplace transform at conducting research of the dynamic modes of oscillatory systems. As the systems of interferometric SAR formation are the phase ones, it is highly necessary to apply the given method, as it allows obtaining of the exact analytical expressions with the accuracy of up to a phase for the response of the system selective path to the radiofrequency pulse excitation. The band filter represented by 4 identical unilateral selective elements will be under consideration as a selective path. The transfer characteristic of the BF will be written down as a fractional rational function

$$K(s) = K_0 \left[\frac{s + b}{s^2 + 2\alpha s + \omega_r^2} \right]^4, \text{ Q-factor of filter } Q = \frac{\omega_r}{2\alpha}.$$

Here damping constant α equals to a half of the bandpass of a separate selective section, ω_r - resonance frequency, let's assume $b = 2\alpha$.

The sensing signal with a rectangular envelope is written down as

$$f_{in}(t) = A_0 \sin(\omega_c t + \psi) [1(t) - 1(t - \tau)]$$

The image of a radio pulse with the ω_c frequency and τ duration

$$f_{in}(s) = A_0 \left[\frac{s \cdot \sin \psi + \omega_c \cos \psi}{s^2 + \omega_c^2} - \frac{s \cdot \sin \psi_\tau + \omega_c \cos \psi_\tau}{s^2 + \omega_c^2} e^{-s\tau} \right],$$

$$\psi_\tau = \psi + \omega_c \tau$$

We have $f_{out}(s) = f_{in}(s)K(s)$ for the signal image at the BF output.

In this case, according to the FILT (Zolotarev, 1969; Zolotarev, 2004), transition into the space of the originals gives a complex representation of the signal at the filter output

$\dot{f}_{out}(t)$, the real signal may be found as $f_{out}(t) = \text{Im}\{\dot{f}_{out}(t)\}$.

Let's represent the complex output signal as

$$\dot{f}_{out}(t) = \dot{f}_{norm}(t)\dot{N}(t),$$

where the BF response to a Monoharmonic signal is assumed as a normalizing function

represented as $\dot{f}_{norm}(t) = A_0 \dot{K}(j\omega_c) e^{j(\omega_c t + \psi)}$. $\dot{N}(t)$ - the normalized complex envelope curve of the signal at the output of the BF under investigation, module $N(t)$ characterizes

behavior of the signal envelope curve at the BF output and function $\delta(t) = \arg\{\dot{N}(t)\}$

determines the current behavior of its phase.

Concerning the plane wave front, the phase difference (caused by a wave arrival under the θ angle) for the base between the adjacent readings a , is determined with the expression

$$\Delta\varphi = \frac{2\pi a}{\lambda} \sin \theta$$

In fact, $\Delta\varphi$ will have the increment $\delta(t)$ caused by the transient (Zolotarev et al., 2006). It will result in dependence of the real θ on the time, i.e.

$$\theta(t) = \theta + \Delta\theta(t), \quad \Delta\theta(t) = \delta(t) \frac{\lambda}{2\pi a \cdot \cos \theta}$$

Figure 9 shows the calculated charts of the corresponding positions of the SAR directional pattern for various time points of the transients. The number of the readings along the vehicle trajectory chosen for calculation at the SAR formation $N = 500$, $a = \lambda/2$, $\lambda = 0.1$ m.

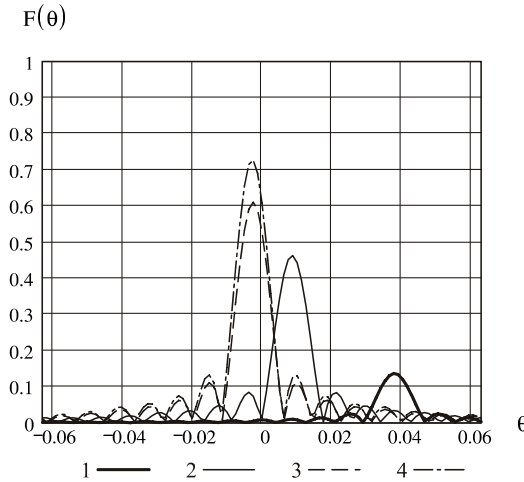


Fig. 9. Calculated parameters: Q-factor of selective system $Q = \omega_r / 2\alpha = 25$, pulse duration $\alpha\tau = 4$; 1 – $t = 0.2\tau$; 2 – $t = 0.4\tau$; 3 – $t = 0.6\tau$; 4 – $t = 0.8\tau$

As it proceeds from the calculated charts (Figure 9), the maximum of the SAR directional pattern turns out to be shifted regarding the case of the transients' absence. This shift depends on the filters Q-factor value and rises together with the increase of the signal bandwidth and depends on the current time of the transient. When the flight altitude $h = 4,000$ m, this shift in a horizontal plane for the detected object reaches considerable values of about several hundred meters. That is why when designing radars with the SAR, it is necessary to pay special attention to minimization of the error caused by the transients in the antenna circuits.

2. Non-equidistance of the taken readings along the vehicle trajectory is an important factor that shall be taken into account at the SAR ADP formation.

Now, unlike the previous point, we will consider the base between the adjacent readings as a random quantity corresponding to the Gaussian law. Let $\{a_i\}$ be a sequence of the distances between the adjacent readings of the reflected signal along the vehicle trajectory with the mean value equal to a and the dispersion σ . So, the difference of the phases between the adjacent readings may be determined with the following expression:

$$\Delta\varphi_i(\theta) = \frac{2\pi \cdot \sin(\theta) \cdot a_i}{\lambda}, \quad i = 1..N. \tag{6}$$

Then there is used an interferometric approach for building the SAR ADP, the number of the readings taken along the vehicle trajectory is assumed equal to 500 (Figure 10). For a sidelobe suppression (Figure 11) the amplitude distribution law along the aperture L is chosen in the following way $I(z) = 1 + \Delta\cos(2\pi z/L)$, $z \leq L/2$, where Δ assume equal to 0.4 (Sazonov, 1988).

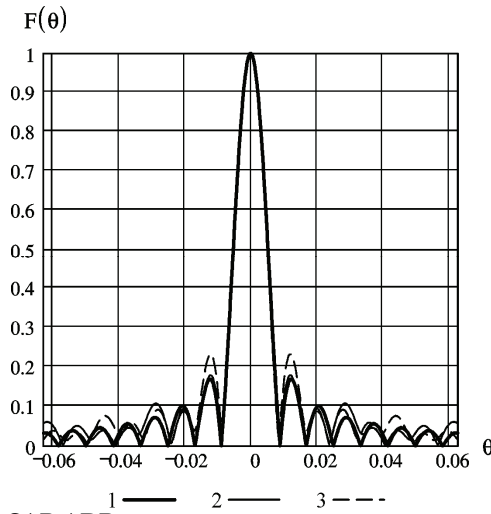


Fig. 10. Interferometric SAR ADP.

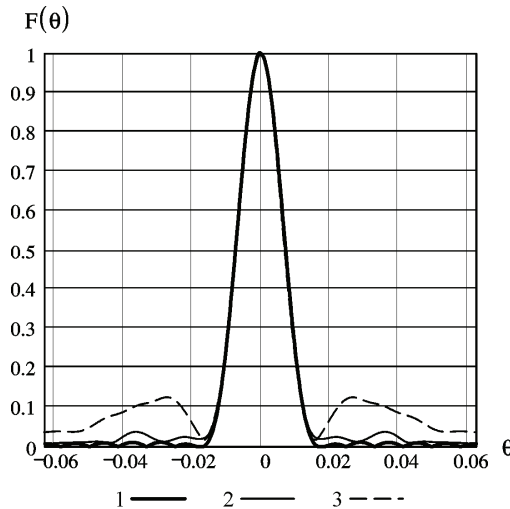


Fig. 11. Interferometric SAR ADP with sidelobe suppression.

The obtained charts show that the increase in dispersion for the distance between the adjacent readings results in a significant increase in the value of the SAR ADP sidelobes. As the research shows, to minimize the ADP sidelobe level, it is necessary to decrease the dispersion value. Introduction of the cosine amplitude distribution of the readings due to aperture (Figure 11) also contributes to it.

3. A significant deterioration of the angle selectivity at the SAR directional pattern formation is determined by out-of-parallelism of the beams for each point of the sensed surface (Figure 12). The vehicle altitude increase above the surface is also a means for reducing out-of-

parallelism of the beams. As a rule, the vehicles with a greater altitude have a greater velocity (for example, the low-altitude vehicles - up to 5 km, jet planes - about 10 km, medium-altitude satellites - about 1,500 km). Correspondingly, there is increase in the number of the readings taken within the same time interval that contributes to narrowing the SAR ADP.

In case of taking into consideration out-of-parallelism of the beams, the difference of the phases between the signals of the adjacent readings is written down in the following way:

$$\Delta\varphi(\theta) = \frac{2\pi}{\lambda} h \left(\sqrt{1 + \left(\operatorname{tg}\theta + \frac{a}{h}\right)^2} - \frac{1}{\cos\theta} \right). \quad (7)$$

Building of the SAR directional pattern shall be carried out as in the previous cases.

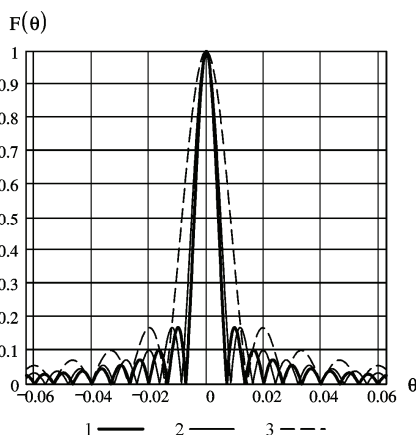


Fig. 12. SAR ADP at $N = 500$; beam parallelism case - 1, out-of-parallelism case: 2 - vehicle's altitude $h = 4,000$ m; 3 - $h = 3,000$ m

One shall keep in mind that increase in the velocity altitude for maintaining the radio links energetics requires increase in the radiant power of the transmitter installed on the vehicle.

The research conducted revealed that all the above-indicated factors provide a significant impact on the SAR ADP characteristics. It is worth mentioning that when designing the corresponding systems (for example, radar remote sensing of the Earth), it is necessary to pay special attention to minimization of the locating angle dynamic error by means of time shifting of the directional pattern. In this sense one should refer to the methods of compensation of the transients in the selective filters of the path.

Another significant factor providing an impact on the SAR ADP is out-of-parallelism of the beams. Here one can come across some contradictions as at present a special attention is paid to the large-scale ground maps that require application of the flight vehicles at relatively low altitudes.

Non-equidistance of the readings also provides an impact on the quality of the formed SAR ADP. That is why it is important to ensure the equipment building with a rigid timing of the reflected signal readings.

6. Influence of a combination of factors on the character of synthetic aperture radar directional pattern: transients, non-equidistance of the readings and out-of-parallelism of the beams at the Earth remote sensing

There is under consideration a combined influence of the transients in the filters of the radar system selective circuits, non-equidistance of the taken readings and divergence of the beams at the distance up to the Earth surface reflecting elements that is comparable with the synthetic antenna aperture value. The transients lead to swinging of the SAR antenna pattern; the other indicated factors result in widening of the synthetic antenna pattern and a sidelobe increase.

A combined influence of the factors was not under consideration before, though they provide a rather considerable influence on the form of the antenna directional pattern which may result in serious errors when determining the characteristics of the extended object lying within the radar swath (Zolotarev et al., 2007).

There is under investigation influence of the transients in the selective filters of the system path forming the SAR. In this case, there is under consideration application of the identical filters in the selective path that is extremely complicated for analysis. To conduct research of the transients influence, there was applied the method developed in (Zolotarev, 1969; Zolotarev, 2004), providing a fast inverse Laplace transform at conducting research of the dynamic modes of oscillatory systems. As there are interferometric SARs under investigation, it is necessary to apply the given method, as it allows obtaining of the analytical expressions with the accuracy of up to a phase for the response of the system selective path to the radiofrequency pulse excitation.

The band filter represented by 4 identical unilateral selective elements will be under consideration as a selective path. The transfer characteristic of the BF will be written down as a fractional rational function

$$K(s) = K_0 \left[\frac{s + b}{s^2 + 2\alpha s + \omega_r^2} \right]^4, \text{ Q-factor of filter } Q = \frac{\omega_r}{2\alpha}, \text{ assume } b = 2\alpha.$$

The sensing signal with a rectangular envelope curve will be written down as

$$f_{in}(t) = A_0 \sin(\omega_c t + \psi) [1(t) - 1(t - \tau)]$$

The image of a radio pulse with the ω_c frequency and τ duration

$$f_{in}(s) = A_0 \left[\frac{s \cdot \sin \psi + \omega_c \cos \psi}{s^2 + \omega_c^2} - \frac{s \cdot \sin \psi_\tau + \omega_c \cos \psi_\tau}{s^2 + \omega_c^2} e^{-s\tau} \right],$$

$$\psi_\tau = \psi + \omega_c \tau$$

We will have $f_{out}(s) = f_{in}(s)K(s)$ for the signal image at the BF output. According to the FILT (Zolotarev, 1969; Zolotarev, 2004), transition into the space of the originals gives a complex representation of the signal at the filter output $\dot{f}_{out}(t)$, the real signal can be found as $f_{out}(t) = \text{Im}\{\dot{f}_{out}(t)\}$.

Let's represent the complex output signal as

$$\dot{f}_{out}(t) = \dot{f}_{norm}(t)\dot{N}(t),$$

where the BF response to a Monoharmonic signal is assumed as a normalizing function represented as $\dot{f}_{norm}(t) = A_0 \dot{K}(j\omega_c) e^{j(\omega_c t + \psi)}$, module of multiplicative function $N(t)$ characterizes behavior of the signal envelope curve at the BF output and function $\delta(t) = \arg\{\dot{N}(t)\}$ determines the current behavior of its phase.

Concerning the plane wave front, the phase difference (caused by the wave arrival under the θ angle) for the base between the adjacent readings a , is determined with the expression

$$\Delta\varphi = \frac{2\pi a}{\lambda} \sin\theta$$

In fact, $\Delta\varphi$ will have the increment $\delta(t)$ caused by the transient (Zolotarev et al., 2006). It will result in dependence of the real θ on the time, i.e.

$$\theta(t) = \theta + \Delta\theta(t), \quad \Delta\theta(t) = \delta(t) \frac{\lambda}{2\pi a \cdot \cos\theta}.$$

Non-equidistance of the taken readings along the vehicle trajectory is an important factor that shall also be taken into account at the SAR ADP formation.

Let's consider the base between the adjacent readings as a random quantity corresponding to the Gaussian law. Let $\{a_i\}$ be sequence of the distances between the adjacent readings of the reflected signal along the vehicle trajectory with the mean value equal to a and the dispersion σ . So, the difference of the phases between the adjacent readings may be determined with the following expression:

$$\Delta\varphi_i(\theta) = \frac{2\pi \cdot \sin(\theta) \cdot a_i}{\lambda}, \quad i = \overline{1..n}.$$

A significant deterioration of the angle selectivity at the SAR directional pattern formation is conditioned by out-of-parallelism of the beams for each point of the sensed surface. The vehicle altitude increase above the surface is also a means for reducing out-of-parallelism of the beams. Correspondingly, there is increase in the number of readings taken for the same time interval that contributes to narrowing the SAR ADP.

In case of taking into consideration out-of-parallelism of the beams, the difference of the phases between the signals of the adjacent readings is written down in the following way:

$$\Delta\varphi(\theta) = \frac{2\pi}{\lambda} h \left(\sqrt{1 + \left(\operatorname{tg}\theta + \frac{a}{h} \right)^2} - \frac{1}{\cos\theta} \right).$$

For a sidelobe suppression (Figure 13) the amplitude distribution law along the aperture L is chosen in the following way $I(z) = 1 + \Delta \cos(2\pi z/L)$, $z \leq L/2$, where Δ is assumed equal to 0.4 (Sazonov, 1988).

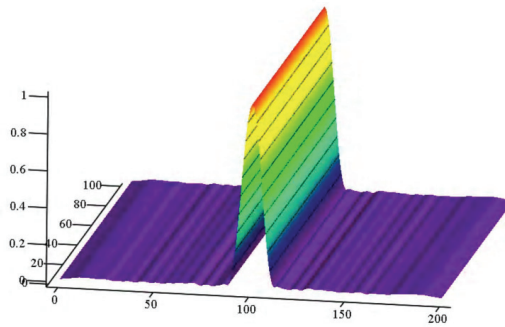


Fig. 13. Synthetic antenna directional pattern without taking into account the destructive factors (transients, non-equidistance of the taken readings, out-of-parallelism of the beams). $n = 500$, $a = \lambda / 2$, $\lambda = 0.1$ m.

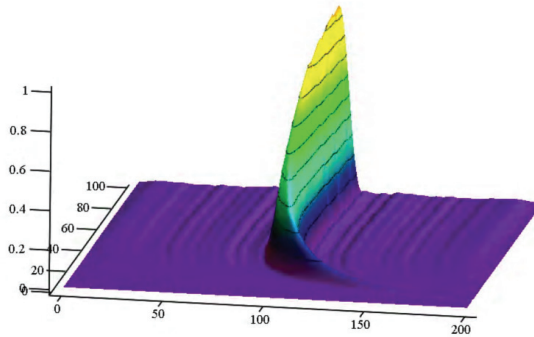


Fig. 14. Dynamic of the SAR directional pattern behavior when taking into account influence of the transients. The calculated parameters are chosen the same as the ones in Figure 13, but there is taken into account availability of 4 identical filters in the system selective path, Q-factor of the selective system $Q = \omega_r / 2\alpha = 25$, pulse duration $\alpha\tau = 4$.

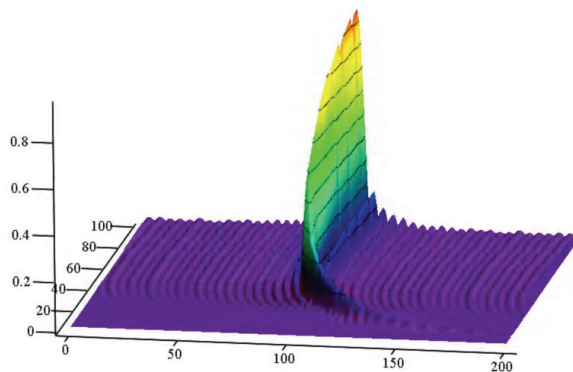


Fig. 15. In this figure there is taken into account a combined influence of the 3 indicated factors determining deformation of the SAR directional pattern: $\sigma = 0,25$, vehicle altitude $h = 4,000$ m.

7. Conclusion

The conducted research revealed that transients provide the most critical influence on the SAR directional pattern. It is difficult to eliminate the dynamic error of the SAR ADP, and at a high flight altitude of modern vehicles even a small angle deviation results in a wrong estimation of the location of a surface-reflecting element. It is worth mentioning that when designing the corresponding systems (for example, a radar remote sensing of the Earth), special attention shall be paid to minimization of the dynamic error of the locating angle due to the directional pattern time shifting.

Increase in the distance dispersion between the adjacent readings results in a significant increase in the SAR ADP sidelobes. That is why it is important to ensure equipment building with a rigid timing of the reflected signal readings. As research shows, to minimize the ADP sidelobe level, it is necessary to decrease the dispersion value.

Another significant factor providing an impact on the SAR ADP is out-of-parallelism of the beams. At present a special attention is paid to the large-scale ground maps that require application of the flight vehicles at relatively low altitudes. In this case, the factor of out-of-parallelism of the beams demonstrates itself more vividly.

In general, when designing the SAR implementation systems, it is necessary to take into consideration a combined influence of all the discussed factors.

8. Acknowledges

The authors would like to express our sincere gratitude to T.O. Pozharsky, an Omsu post-graduate student for the calculations he made and for his active participation in the debates of the results. Mr. Pozharsky derived the formula (7) that takes into account the influence of divergence of the beams.

9. References

- Antipov, V.N.; Goryainov, V.T. & Kulin, A.N. (1988). *Radar Stations with Digital Synthesizing of Antenna Aperture*, under the editorship of V.T. Goryainov, M.: Radio and Communication, 1988. – p. 304.
- Boerner, W.-M. (2000). *Invited Lecture for the Fifteens Anniversary of Radio Engineering Faculty of Tomsk State University of Control Systems and Electronics, Tomsk, Russia, 2000 October 12.*
- Boerner, W.-M. (2004). *From Airborne Via Drones To Space-Borne Polarimetric-Interferometric SAR Environmental Stress-Change Monitoring – A Comparative Assessment Of Its Applications*, Proceedings on EUSAR2004 5th European Conference on Synthetic Aperture Radar, Ulm, Germany, 2004, May 25-27.
- Filippov, V.S.; Ponomarev, L.I. & Grinev, A.Y. (1994). *Antennae and Microwave Devices. Designing of the Phased Array Antennae /* under the editorship of D.I. Voskresensky, M.: Radio and Communication, 1994. – p. 592.
- Sazonov, D.M. (1988). *Antennae and Microwave Devices*, Book for the University Radio-Technical Specialties. - M.: Higher School, 1988. – 432 p.
- Varakin, L.E. (1985). *Communication Systems with Noise-Like Signals*, M.: Radio and Communication, 1985.

- Vendic, O.G. & Parnes, M.D. (2002). *Antennae with Electric Scanning*, M.: Science-press, 2002, - p. 232.
- Zolotarev, I.D. (1969). *Transient Processes in Resonance Amplifiers of the Pulse-Position Measuring Systems*, Novosibirsk: - Science. - Siberian Division of USSR Academy of Sciences. - 1969.
- Zolotarev, I.D. (1996). *The new Approach in Determination of the Problem "Amplitude, Phase, Frequency" in the Theory of Signals and Systems*, Abstracts of the XXV General Assembly URSI. - Lille. - France. - 1996.
- Zolotarev, I.D. (1999). *Complex Signal and the Problem of the Amplitude-Phase-Frequency for Ultra Wideband Processes*. Collection of the papers of the V international scientific and technical conference "Radiolocation, Navigation, Communication". - Voronezh: VNIIS, 1999. - v. 1. - p. 348-356.
- Zolotarev, I.D. (2004). *The Method Simplifying Inverse Laplace Transformation At Oscillatory Processes Researches. The "Amplitude, Phase, Frequency" Problem In Radioelectronics And Its Solution*, Tutorial. - Omsk: OmSU Publishing, 2004. - 132 p.
- Zolotarev, I.D.; Miller, Ya.E. & Pozharsky, T.O. (2004). *Research of Passing of Ultrawideband Phase-Shift Keyed (PSK) Radar Signals Through Selective Filter at Various Forms of Enveloping Curves of Discrete*, Abstracts RADAR2004 International Conference on Radar Systems, Toulouse, France, 2004, October 18-22.
- Zolotarev, I.D.; Miller, Ya.E. & Pozharsky, T.O. (2005). *Research of Deformation of Fine Phase Structure of Ultra Wideband Radar Signals when Passing Through System of Identical Selective Filters*, IGARSS 2005 Proceedings, 0-7803-9051-2/05 2005 IEEE, COEX, Seoul, Korea, 2005, July 25-29.
- Zolotarev, I.D.; Miller, Ya.E. & Pozharsky, T.O. (2005). *Influence of transients on the interferometric SAR characteristics*, "CEOS SAR Workshop 2005" Materials. - Adelaide. - Australia. - 2005.
- Zolotarev, I.D.; Miller, Ya.E. & Pozharsky, T.O. (2006). *Influence of transients on the interferometric SAR characteristics*, 6th European Conference on SAR "EUSAR 2006" Materials, Dresden, Germany, 16-18 May, 2006.
- Zolotarev, I.D.; Miller, Ya.E. & Pozharsky, T.O. (2007). *Research of influence of transients, non-equidistance of the taken readings, divergence of beams on characteristics of interferometric SAR*, Proceedings of the IEEE International Geoscience And Remote Sensing Symposium "IGARSS 2007", Barcelona, Spain, 23-27 July, 2007.