PhD Dissertation

International Doctorate School in
Information and Communication Technologies

DISI-University of Trento

ADVANCED METHODS FOR THE RETRIEVAL OF GEO-/BIO-PHYSICAL VARIABLES FROM REMOTE SENSING IMAGERY

Luca Pasolli

Advisor:
Prof. Lorenzo Bruzzone
University of Trento

Co-Advisor:
Dr. Claudia Notarnicola
Eurac Research

March 2012

A Stefania

Non cercare di diventare un uomo di successo,
ma piuttosto un uomo di valore.

Albert Einstein

Abstract

The retrieval of geo-/bio-physical variables from remote sensing imagery is a challenging and important research field. On the one hand, advances in electronics, engineering and space sciences are offering to the users community new sensors capable to acquire information on the Earth surface with higher accuracy and improved features with respect to the past. On the other hand, the need of large-scale, accurate and up-to-date mapping and monitoring of natural targets and physical processes is becoming fundamental for many application domains. This calls for the development of accurate, robust and effective retrieval methodologies.

The main goal of this thesis is to investigate and develop advanced methods and systems for the retrieval of geo-/bio-physical variables from satellite remote sensing imagery being able to exploit the potential of new and upcoming satellite systems and support real application domains. Special attention has been devoted to the definition of methods and to the analysis of data acquired in the challenging mountain environment.

The activity carried out and presented in this dissertation is oriented to investigate the main limitations of the existing methodologies for addressing the estimation problem and to develop novel and improved systems that can overcome the drawbacks identified. In particular, the following main novel contributions are proposed in this thesis:

a) A theoretical and empirical comparative analysis of non-linear machine learning regression methods, namely the Multi-Layer Perceptron Neural Network and the Support Vector Regression, for soil moisture retrieval in different operational scenarios.

b) A novel multi-objective model-selection strategy for tuning the free parameters of non-linear regression methods taking into account different quality metrics that are jointly optimized.

c) A novel hybrid approach to the retrieval of geo-/bio-physical variables from remote sensing data integrating both theoretical electromagnetic models and field reference measurements.

d) A sensitivity analysis and a retrieval system for soil moisture content estimation from new generation SAR imagery in an Alpine catchment.

e) *An empirical study on the effectiveness of fully-polarimetric SAR signals for soil moisture estimation in mountain areas.*
f) *An improved algorithm for mapping and monitoring Green Area Index (GAI) in Alpine pastures and meadows from satellite MODIS imagery.*

Qualitative and quantitative experimental results obtained on real remotely sensed data confirm the effectiveness of the proposed solutions.

Keywords

Geo-/Bio-Physical Variables, Soil moisture content, Green Area Index (GAI), Retrieval, Support Vector Regression (SVR), Theoretical model inversion, Parameter optimization, Synthetic Aperture Radar (SAR), Polarimetry, Multispectral image, Alpine environment, Image processing, Remote sensing.

Acknowledgments

This thesis is the result of more than three years of work. During this chapter of my life I had the opportunity to learn and improve myself from the professional, scientific and personal viewpoints. I had also the chance to meet and know new people, to travel, to present my work in international conferences and workshops and to face new challenges. All these experiences will profoundly an positively affect the rest of my life and for this reason I think it is worh spending few words to say thank you.

Thank you to Claudia and Lorenzo. They believed in me and supported my work since the beginning. Their suggestions, scientific guidance and experience have been foundamental in the past years for developing the research presented in this document. I really appreciated the helpfulness and kindness they always demonstrated to me and I really hope to have the pleasure of continuing this collaboration in the next years.

I would like to thank the European Accademy (EURAC), and in particular the Institute for Applied Remote Sensing under the guidance of Marc, which supported my PhD and provided me this geat opportunity.

Data are a foundamental ingredient of my PhD research. Without data, none of the analysis and developments presented hereafter would have been possible. Thank you to European Space Agency, for providing the RADARSAT2 images in the framework of the SOFIA project, and to the MODIS science team, for developing the MODIS products and making them available to the scientific community.Thank you to Giacomo and the Institute for Alpine Environment at EURAC, for the nice trips on the steep slopes in Matcher valley during our field campaigns; to Georg Wohlfahrt and the Biomet group at the University of Innsbruck, for providing us with the GAI field measurements; to all the people who acquired, processed and made available the *Scatterometer*, *ActPass* and *SMEX* data set used in this work.

A mention goes to all my collegues and friends at RSlab and EURAC who shared with me these last three years. They were always king and helpful whenever I had a doubt or a favour to ask and with them I had the opportunity to enjoy little things, like a coffe or a break, in my everyday life.

A particular and sincere thank to my parents Annamaria and Paolo and to my sister Veronica. They always encouraged and helped me during these (many) years of study. Their unconditioned and total support has been a reference point whatever difficulty and uncertainty I had to face.

Finally, I would like to express my profound gratefulness to Stefania. During the last six years that we spent together, she always believed in me and supported my choices despite the uncertainty that comes along with the PhD carrier. Her love, her smile, her right word in the right moment (she is always right!) have been foundamental for reaching this important achievement of my life.

Contents

Introduction and Thesis Overview **1**

Brief Introduction to Remote Sensing 1
Information Extraction from Remote Sensing Imagery: the Geo-/Bio-Physical variable retrieval problem 3
The Retrieval Process: a Pattern Recognition Perspective 5
Retrieval Methods 6
Motivations, Objectives and Novel Contributions of the Thesis 9
Structure of the Thesis 13

Comparative Analysis of Non-Linear Machine Learning Regression Methods for the Retrieval of Soil Moisture in Different Operational Scenarios **15**

2.1 Introduction and Motivation 15
2.2 The ε-insensitive Support Vector Regression Technique 18
2.3 Data Set Description and Design of the Experiments 20
2.3.1 Field Measurements 20
2.3.2 Simulated Samples 21
2.3.3 Design of Experiments 22
2.4 Experimental Results 24
2.4.1 Results Scenario 1 25
2.4.2 Results Scenario 2 28
2.5 Discussion and Conclusion 31

A Novel Multi-Objective Strategy for Tuning the Free Parameters of Non-Linear Regression Methods **33**

3.1 Introduction and Motivation 33
3.2 Proposed Multi-Objective Model-Selection Strategy 36
3.2.1 Problem Formulation 36
3.2.2 Proposed Multi-Objective Strategy 39
3.3 Experimental Analysis 42

3.3.1 Data Set Description 42
3.3.2 Design of Experiments 44
3.3.3 Experiment 1: Model-Selection Using All the Criterion Functions 45
3.3.4 Experiment 2: Model-Selection Using Two Criterion Functions 48
3.4 Discussion and Conclusion 50

ANovel Hybrid Approach to the Retrieval of Geo-/Bio-Physical Variables from Remote Sensing Data 53

4.1 Introduction and Motivation 53
4.2 Proposed Hybrid Retrieval Approach 56
4.2.1 Global Deviation Bias (GDB) Strategy 57
4.2.2 Local Deviation Bias (LDB) Strategy 57
4.3 Experimental Analysis 60
4.3.1 Data Set Description 60
4.3.2 Design of Experiments 61
4.3.3 Experiment 1: Correction with the Global Deviation Bias (GDB) Strategy 62
4.3.4 Experiment 2: Correction with the Local Deviation Bias (LDB) Strategy 64
4.4 Conclusion 67

Retrieval of Soil Moisture Content from New Generation RADARSAT2 SAR Imagery in an Alpine Catchment 69

5.1 Introduction and Motivation 69
5.2 Study Area and Data Set Description 71
5.2.1 Study Area 71
5.2.2 Satellite Images 72
5.2.3 Field Measurements 74
5.2.4 Ancillary Data 76
5.3 Sensitivity Analysis 76
5.3.1 Effect of Topography 78
5.3.2 Effect of Vegetation/Land-Cover Heterogeneity 79
5.4 Proposed Retrieval System 81
5.5 Retrieval Accuracy Assessment 83
5.5.1 Quantitative Assessment with Point Measurements 83
5.5.2 Qualitative Assessment with Distributed Maps 84
5.6 Conclusion 85

Effectiveness of Fully-Polarimetric SAR Data for Soil Moisture Retrieval in Mountain Areas **87**

6.1 Introduction and Motivation 87
6.2 Study Area and Data Set 89
6.3 Design of the Experiments 90
6.3.1 Feature Extraction Strategies 90
6.3.2 Experimental Setup 91
6.4 Experimental Results 92
6.4.1 Experiments 1: Intensity & Phase Processing 95
6.4.2 Experiments 2: α/A/H Polarimetric Decomposition 95
6.4.3 Experiments 3: Independent Component Analysis 96
6.4.4 Qualitative Assessment with Soil Moisture Content Maps 96
6.5 Conclusion 98

An Improved Algorithm for the Retrieval of Green Area Index from MODIS Imagery in Mountain Grasslands in the Alps **101**

7.1 Introduction and Motivation 101
7.2 Study Site and Data 105
7.2.1 Satellite Data 106
7.2.2 Ground Measurements 107
7.2.3 Ancillary Data 107
7.3 Methods 108
7.3.1 Reference Spectral Library Generation 109
7.3.2 Inversion Algorithm 114
7.3.3 Data Processing and GAI Map Generation 115
7.3.4 Validation 117
7.4 Results 118
7.4.1 Visual Analysis of Estimated GAI Maps 118
7.4.2 Analysis of Estimated GAI Dynamics 122
7.4.2.1 Temporal Dynamics 122
7.4.2.2 Range Dynamics 122
7.4.3 GAI Accuracy Assessment 125
7.5 Discussion 126
7.6 Conclusions 130

Conclusions **133**

8.1 Summary and Discussion 133
8.2 Final Remarks and Future Developments 137

List of Publications **139**

Bibliography **143**

Chapter 1

Introduction and Thesis Overview

In this chapter an overview of the contents of this PhD thesis is given. After a brief introduction on remote sensing systems, the main research topic addressed in this dissertation, i.e. the retrieval of geo-/bio-physical variables, is introduced and the related works in this research field are presented. Then the objectives and motivations of the thesis work as well as the main novel contributions proposed are highlighted. Finally, the structure of the document is illustrated.

Brief Introduction to Remote Sensing

Remote Sensing (RS) is the science and technology of acquiring and interpreting information regarding a scene of interest without being directly in contact with the item (or items) under investigation (*Jia and Richards, 2004*). In remote sensing, a sensor is devoted to the measurement of the energy emitted and/or scattered/reflected by an object and propagated as electromagnetic radiation through the space. Then, the measured energy is converted into a signal (e.g., an electrical current) and passed to a processing unit to be stored and/or analyzed. A large variety of systems fall within this broad definition, ranging from the human visual system to planetary observation systems. In this thesis, we will focus the attention on imaging systems onboard of satellite platforms for the observation and monitoring of the Earth surface, often referred as Earth Observation (EO) remote sensing systems. Since the 50s, when the first artificial satellites were launched, EO systems were used for military and civil operations. Thanks to the possibility to acquire information over large areas with a reduced cost and on a regular basis, satellite remote sensing growth progressively to become nowadays a key technology in most of human activities.

The research in many fields of electronic, informatics and signal processing made it possible the development of many kinds of EO systems, which can be categorized according to different criteria. Depending on the source of energy involved in the image acquisition, one can distinguish between passive and active systems. Passive systems rely on the presence of

an external illumination source, such as the sun or the target itself. The electromagnetic energy measured by the sensor is divided into a certain number of spectral bands, so that the sensor is called multi-spectral (or hyper-spectral if the number of spectral bands is higher than some tens). The width of each spectral band defines the spectral resolution of the system. Examples are the optical/thermal scanners and the microwave radiometers. Active systems, instead, exploit an artificial source of energy, which is controlled by the sensor. The sensor intrinsically measures: i) the time delay between the emission and return of the energy pulse, to locate the target in space and height; and ii) the power of the received electromagnetic signal, which provides information on the characteristics of the investigated object. Examples of are Radio Detection and Ranging (RADAR) and Light Detection and Ranging (LIDAR) systems, which operate in the microwave and optical domain, respectively. Despite the definition of spectral bands is improper for active sensors, new generation active systems allow the acquisition of multiple information of the scene of interest, e.g., by controlling the polarization or the frequency of the emitted and received electromagnetic signals. With regard to the spatial resolution, defined as the smallest distance between two objects that can be distinguished by the sensor, remote sensing systems can be divided into: very high resolution systems, with resolution in the order or less than one meter; high resolution systems, with resolution of some up to teens of meters; medium resolutions systems, with resolution in the order of hundreds of meters; and low resolutions systems, from several hundreds to some kilometers. Finally, remote sensing systems can be classified according to the revisit frequency, i.e., the minimum time interval that occurs between two consequent acquisitions over the same target area, into: very high revisit frequency systems, with daily acquisition capability; high revisit frequency systems, with acquisition every few days; medium revisit frequency systems, with temporal intervals between consecutive acquisitions of tens of days.

The large variety of remote sensing sensors available to the users community made it possible in the last decades to deepen the knowledge on the Earth system and its processes from many different perspectives, pushing at the same time the technology through the development of even more advanced systems in a virtuous circle. Satellite remote sensing, however, is intrinsically confronted to a trade-off between spatial resolution, revisit frequency and band richness. This issue is partially overcome by the latest and upcoming EO missions, which rely on constellations of satellite systems to increase the revisit frequency maintaining appealing characteristics from the spatial and spectral viewpoints. Examples are the four satellites COSMO SkyMed constellation of the Italian Space Agency (ASI) (*WWW1*) and the family of Sentinel missions, developed in the framework of the Global Monitoring for Environment and Security (GMES) program by the European Space Agency (ESA) (*WWW2*). The latter is foreseen to be launched in 2013, when the first satellites of each constellation will be inserted in orbit. Each 2-satellites constellation will have a different payload (including Synthetic Aperture Radar (SAR) and multi-spectral high-resolution scanner) thus

providing a valuable and unique database of high-resolution and high revisit frequency (1 to 3 days) EO imagery.

Information Extraction from Remote Sensing Imagery: the Geo-/Bio-Physical Variables Retrieval Problem

The analysis of remote sensing imagery is usually dichotomized into *image*-centered and *data*-centered approaches (*Showengerdt, 2007*). The former approach refers to the interpretation of the scene on the basis of the spatial relationships among features on the ground. The information can be extracted from the scene either by an experienced user through photointerpretation or by a computer aided system in a quantitative and objective manner (*Richards and Xia, 2006*). Typical examples are the identification of land-cover/land-use classes on the ground and the recognition of changes among them. Data-centered approaches perform the analysis of the acquired scene driven by the data, i.e., the physical measurements of emitted and/or scattered/reflected electromagnetic energy. To this type of analysis usually belong intrinsically quantitative tasks, such as the measure of spectral absorption of a target, the estimation of fractional abundances of surface materials and the retrieval of geo-/bio-physical variables. In this dissertation we focus the attention on the last issue.

Geo-/bio-physical variables are continuous attributes that quantify physical and/or structural properties of natural targets. Typical examples are the temperature and the moisture percentage of soil superficial layers, the depth and density of snow packs (which product gives rise to the so called snow water equivalent (SWE)), the concentration of biological particles and chemical pollution in costal sea waters, the amount of leaf covered area per ground unit (the so called leaf area index (LAI)), biomass and leaf bio-chemical constituents of a vegetated target, and many others. Geo-/bio-physical variables are of fundamental importance in several application and research domains. Information on soil moisture content and snow pack parameters (especially wetness and density) revealed to be a crucial support to the forecast of natural risks, such as drought, flooding, landslides and avalanches (*Sandholt et al., 2002; Mätzler, 1987; Perry, 2000; Engman and Gurney, 1991; Hino et al., 1988*). Snow water equivalent provides useful indications for the estimation of water availability for irrigation and hydropower purposes in mountain areas (*Baghdadi et al., 1997; Goodison and Louie, 1986*). Precision farming and vegetation stress monitoring activities often relies on the precise knowledge on plant bio-chemical and structural properties together with water availability and moisture concentration indicators (*Bolten et al., 2010; Stimson et al., 2005; Penuelas et al., 1994; Bastiaanssen and Bos, 1999; Heathman et al., 2003*). In meteorology an accurate representation of soil and vegetation geo-/bio-physical variables is fundamental to effectively quantify exchanges of energy and greenhouse gasses between the land surface and the atmosphere (*Sellers et al., 1997; Turner et al., 2004*). Many other domains could be cited,

such as natural resources management (*Bastiaanssen et al., 2005; Hall and Martinec, 1985*) and climate changes monitoring (*Sellers, 1996; Derksen et al., 1998; Rodriguez-Iturbe et al., 1999*). With this regard, several geo-/bio-physical variables have been identified by the Global Climate Observing System (GCOS) as Essential Climate Variables (*WWW3*) All the aforementioned applications and research domains require spatially and often also temporally distributed measurements of the variables of interest (*Schulz et al., 2006*). Satellite remote sensing, thanks to its synoptic and regular imaging of the Earth surface, implicitly fulfills these requirements.

From a physical viewpoint, changes in the chemical, physical and structural characteristics of a target (either natural or man-made) determine variations of its electromagnetic response in terms of absorption, emission, transmission and reflection (*Schanda, 1986; Slater, 1980; Ulaby et al., 1986a*). The possibility to quantitatively infer the geo-/bio-physical variable of interest from the measurements performed by a remote sensing sensor is based on this behavior. However, this task is not straightforward for many reasons:

- *The complexity and non-linearity that characterize usually the relationship between remote sensing measurements and target variables* (*Twomey, 1977*). On the one hand, geo-/bio-physical variables may affect to different extents the electromagnetic properties of a target along their range of variability (e.g., saturation and other non-linear effects may occur, (*Haboudane et al., 2004*). On the other hand, electromagnetic radiation shows a different sensitivity to the different physical phenomena depending on the wavelength of the signal (*Ulaby et al., 1986b; Jaquemoud and Baret, 1990; Verhoef, 1984*).

- *The ill-posed nature of the retrieval problem*. The total electromagnetic response of a target is typically the result of multiple contributions, each one determined by a different structural, chemical or physical characteristic (*Jackson and Schmugge, 1991*). This aspect determines to the so-called variable equifinality issue, i.e., the phenomenon whereby similar electromagnetic responses can be associated with different geo-/bio-physical variable configurations (*Beven & Freer, 2001; Beven, 2006*).

- *The image formation process at sensor level*. Remote sensing sensors provide a quantized representation of the investigated scene in the spatial domain. The electromagnetic energy measured within an elementary resolution cell is the result of the presence of multiple objects on the ground with slightly (or sometimes strongly) different characteristics. This behavior is the origin of a mixed contribution at sensor level. Even by increasing the spatial resolution this mixing phenomenon cannot be completely cancelled, since it remains in pixels representing the boundaries between objects (*Showengerdt, 2007*). Moreover, the response corresponding to a pixel can be also affected by radiation components coming from the surrounding of the area investigated (*Borengasser et al., 2008*).

- *The influence of external disturbing factors*. The remote sensing acquisition systems is not ideal, but affected by disturbing factors such as the noise and non-linearity at sensor level and the presence of the atmosphere. Even if these issues can be determined and corrected to some extent with the help of calibration and atmospheric correction procedures, they may still corrupt the signal measured at sensor level and thus introduce further ambiguity and complexity in the retrieval process (*Chen, 1997; Hadjimitsis et al. 2004*).

This list, which is not meant to be exhaustive since many other issues can be encountered when dealing with the geo-/bio-physical variable retrieval in specific application contexts (e.g., the influence of topography in mountain areas), points out the general complexity of the retrieval process. These considerations call for the definition and use of proper methods for processing satellite remote sensing data and retrieve the desired geo-/bio-physical variable.

The Retrieval Process: a Pattern Recognition Perspective

From a methodological viewpoint, the retrieval of geo-/bio-physical variables from remote sensing data has been mainly treated as a more general pattern recognition problem. Pattern Recognition (PR) has been defined as “the discipline which learns some theories and methods to design machines that can recognize patterns in noisy data or complex environment” (*Srihari, 1993*). Ripley (*1996*) outlined pattern recognition as follows: “Given some examples of complex signals and the correct decisions for them, it makes decisions automatically for a stream of future examples”. These definitions point out the key elements of a PR system: i) it recognizes patterns in noisy data or complex environment; and ii) it performs the decision for future examples in an automatic manner. A general pattern recognition system is the ensemble of several processing steps as synthetized in Figure 1.1. From a conceptual viewpoint, these steps can be grouped into i) data pre-processing and ii) data analysis. The former consists of all the steps required for reducing possible errors, noise and other disturbing factors in the data associated with the acquisition phase and for extracting the most relevant information (i.e., the features) for the addressed problem. Data pre-processing has thus a crucial role for the definition of an effective pattern recognition system. The latter deals with the recognition process *per se*. Both data pre-processing and data analysis might rely on the availability of prior knowledge on the addressed problem.

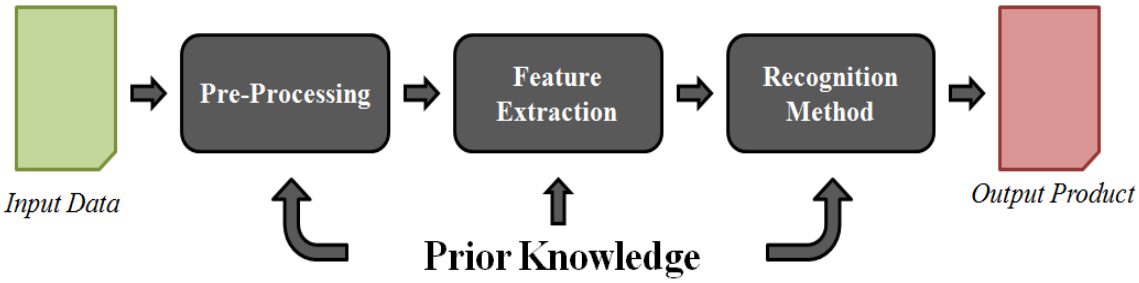


Figure 1.1. Block scheme of a general pattern recognition system.

A brief overview of the state-of-the-art methods for data analysis with specific regard to the geo-/bio-physical variables retrieval problem (i.e., the retrieval methods) is provided in the following. With regard to data pre-processing, a comprehensive overview of such techniques is outside the main objective of this dissertation. Indeed, they are typically tailored more to the characteristics of the acquisition sensor considered rather than to the specific pattern recognition task addressed (i.e., the retrieval problem). We refer the reader to (*Oliver and Quegan, 2004; Girard and Girard, 2003*) and the introduction of the next chapters of this thesis for more details on this issue.

Retrieval Methods

The retrieval method is the core of a retrieval system. It assumes that the addressed retrieval problem can be expressed in terms of a mapping between a set $\boldsymbol{x} = (x^1, x^2, \dots, x^m)$ of m features extracted from the signals acquired using remote sensors and the desired continuous variable y. From an analytical viewpoint, this concept can be expressed as:

$$y = f(\boldsymbol{x}) + e \qquad (1.1)$$

where f denotes the desired and unknown mapping and e is a Gaussian random variable with zero mean and unit variance taking into account all the noise contributions affecting the retrieval problem. From the methodological perspective, the retrieval of y corresponds to the problem of determining a mapping f' as close as possible to the true mapping f.

In the geo-/bio-physical variable retrieval literature, this task has been addressed mainly following two approaches: i) the derivation of empirical data-driven relationships; and ii) the inversion of physical based analytical models.

The first approach relies on the availability of a set of reference samples, i.e., couples of *in-situ* measurements of the desired target variable associated with the corresponding measurements of the remote sensor. These samples are exploited for deriving an empirical mapping, e.g., by means of statistical regression techniques in combination with parametric (linear, logarithmic or polynomial) functions. Then the identified relationship is extended to

the whole satellite image. Examples can be found in studies for the retrieval of vegetation characteristics from optical remote sensing data (*Colombo et al. 2003; Heiskanen, 2006*) and suspended chemical and biological particles in coastal waters (*Teodoro et al., 2007*). Analytically more sophisticated parametric functions have been defined when the complexity of the retrieval problem increases. This is the case of the operational SeaWiFS chlorophyll concentration algorithm (*O'Relley et al., 1998*), were ratios between spectral bands and log-transformations were used to take into consideration the non-linear behavior of the investigated mapping. Empirical relationships are appealing since they are typically fast to derive and quite accurate. Moreover, they abstract complex physical phenomena to a higher level, which can be easily addressed by non-experts without a specific background in the field. The main drawback is the need of a set of possibly well representative reference samples. The collection of ground measurements requires the human intervention and is usually a time-consuming and expensive task. Moreover, errors may occur for various reasons during the measurement process. This aspect affects the quality and quantity of reference samples available. Another important issue is the fact that empirical relationships are typically site and sensor dependent, since derived from samples collected under specific operational conditions. This limits the possibility to extend their use to different areas and different remote sensing systems, since they remain valid only under the conditions in which reference samples have been collected (*Colombo et al., 2003; Meroni et al., 2004*).

The second approach demands the definition of the desired mapping function to analytical electromagnetic models. Such models are based on a solid physical based description of the mechanisms involving the interaction of the electromagnetic radiation and the target object of interest. In the direct operational way, they simulate the response of a target object as function of: 1) the target characteristics (i.e., structural, chemical and biophysical variables); and ii) the signal characteristics (i.e., wavelength, incidence/reflection angle, etc.). Thus in the inverse operational way they can be used to represent the mapping between the measurements at the remote sensor and the variable of interest. A wide variety of analytical electromagnetic models has been proposed in the literature, with different levels of complexity and generality. When dealing with microwave emission and scattering, probably the most widely used model is the Integral Equation Model by Fung et al. (*1992*), which has been further improved to take into account multiple scattering effects (the Advanced IEM, *Fung et al., 2000*) and often coupled with layered structured or three dimensional models to handle complex target such as vegetated areas and snow packs (*Guoquing and Rason, 1995; Karam et al., 1992*). In the field of vegetation variable retrieval from optical signals, the PROSAIL model has been used in a wide variety of remote sensing studies (Jaquemoud and Baret, 1990; Jaquemoud et al., 2000; Verhoef, 1984). Many other examples can be found in the literature (*Turner et al., 2004; Schlerf and Atzberger, 2006*). Thanks to the solid physical foundation and the wide range of applicability (in terms of both target properties and system characteristics), electromagnetic models can operate in more general scenarios that are difficult to represent through the

collection of *in-situ* measurements. For this reason, they are particularly appealing to address the estimation of geo-/bio-physical variables from remote sensing data. A major concern is related to the fact that they rely on hypothesis and assumptions that simplify their representation of real phenomena. This issue is intrinsic in the modelization process and can be reduced (but not completely eliminated) by increasing the complexity of the model, at the price of a reduced generalization ability (*Darvishzadeh et al., 2008*). Another drawback of electromagnetic models is their high complexity and dependence on a huge number of input parameters. This makes the inversion process often analytically not tractable. To face this problem, many different inversion strategies have been proposed in the literature. The most common ones are: i) iterative search algorithms such as the Nelder-Mead and the Netwon-Rapson methods (*Meroni et al., 2004; Paloscia et al., 2008*), which iteratively try different model parameter configurations to minimize a dissimilarity measure between simulated and measured electromagnetic response of a target object; ii) look-up table matching, which searches among a set of pre-computed simulated spectra the most similar to the remote measurement (*Darvishzadeh et al., 2008*); and iii) regression methods, which exploit a set of simulated samples (i.e., couples of target geo-/bio-physical variables and simulated electromagnetic responses) to infer the inverse theoretical mapping (*Song et al., 2009*).

Regardless to the approach considered, either empirical or physical based, the high complexity and non-linearity of retrieval problems required the development and use of more advanced methods. A class of highly powerful regression methods, which has been successfully introduced in the field of geo-/bio-physical variable estimation since two decades generating a rising interest in the remote sensing community, is represented by non-linear machine learning techniques. Due to advanced learning strategies, such techniques can learn and approximate even complex non-linear mappings exploiting the information contained in a set of reference samples. Another advantage is the fact that no assumptions have to be made about the data distribution (for this reason, non-linear machine learning methods are often referred as distribution free). Due to this property, the retrieval process can integrate data coming from different sources with poorly defined (or unknown) probability density functions and relating well with the target variable. Among machine learning techniques, one can recall Artificial Neural Network (ANN) (*Reale and Jackson, 1991*) and Support Vector Regression (SVR) (*Vapnik, 1995*) methods. ANN is probably the most largely used technique in the field of geo-/bio-physical variable retrieval and has been widely investigated in many application domains. Tsang et al. (*1992*) addressed the estimation of snow variables from the Special Sensor Microwave/Imager (SSM/I) data by means of a neural network trained with dense medium multiple scattering model. The effectiveness of neural network model inversion for estimating soil moisture in comparison with well-known inversion strategies, namely the Bayesian method and the simplex algorithm, is investigated in Paloscia et al. (*2008*), and Notarnicola et al. (*2008*). Final evaluations point out that ANNs are a good trade-off in terms of accuracy, stability and computational speed with respect to the other strategies

investigated. Other interesting examples can be found in the field of vegetation parameters retrieval (*Del Frate et al., 2003*). Also SVR is becoming popular in the field of geo-/bio-physical variables retrieval, especially in the last few years, despite its use up to now is still limited. Few papers investigated the effectiveness of this method for the retrieval of vegetation characteristics, open water chemical and biological particles concentration and land and sea surface tempearture (*Durba et al., 2007, Bruzzone and Melgani, 2005; Moser and Serpico, 2009*). Achieved results point out in general the promising features of this method, such as the good intrinsic generalization ability and the robustness to noise and limited availability in the reference samples. However, the full potential of the method has not been fully exploited yet.

Objectives, Motivations and Novel Contributions of the Thesis

Despite the strong research effort already demonstrated by the scientific community, the retrieval of geo-/bio-physical variables from remote sensing imagery is a challenging and important research field. On the one hand, advances in electronics, engineering and space sciences are offering to the users community new sensors able to acquire information from the Earth surface with higher accuracy and improved features with respect to the past. On the other hand, the need of large-scale, accurate and up-to-date mapping and monitoring of natural targets and physical processes is growing significantly. This calls for a continuous research and investigations for the development of even more accurate and robust methodologies.

The objective of the thesis is to investigate and develop advanced methods and systems for improving the accuracy and robustness of the retrieval process from satellite remote sensing imagery. The study will be carried out with particular regard to the retrieval of geo-bio-physical variables in the challenging mountain environment. The motivations for the interest in this specific context are mainly the following:

- Mountain regions present extreme operational conditions, i.e., very limited availability of representative reference samples due to the difficulty in the field measurement collection and high heterogeneity of the local conditions at the ground. State-of-the-art methodologies require further investigations and improvements to tackle the complexity of the retrieval in this environment.
- The new and upcoming satellite systems (e.g., the Sentinel satellites family) will provide large amounts of data with improved spatial and temporal resolutions. This opens the path for better studying and monitoring heterogeneous and complex environments such as the mountain regions, provided that proper methodologies are available to fully exploit the potential of these new data.

- At the best or our knowledge, the retrieval of geo-/bio-physical variables from satellite remote sensing imagery in mountain areas has been only marginally investigated up to now. Very few studies can be found in the remote sensing literature on this topic (*Paloscia et al., 2010; Vescovo and Gianelle, 2008; Heiskanen, 2006*). Nonetheless, accurate mapping of geo-/bio-physical properties of the natural targets present in this environment is of utmost importance for many application domains (e.g., risk forecasting, natural resources monitoring) but almost impossible to realize without the synoptic view of remote sensing.

The activity carried out and presented in this dissertation is mainly oriented to investigate some of the limitations of the existing methods for the addressed estimation problem and to develop novel and improved systems that can overcome the identified drawbacks. In particular, the following main novel contributions are proposed in this thesis:

A. A comparative analysis of non-linear machine learning regression methods, namely the Multi-Layer Perceptron Neural Network (MLP-NN) and the Support Vector Regression (SVR), for soil moisture retrieval in different operational scenarios.

Soil moisture estimation is one of the most challenging problems in the context of geo-/bio- physical variables retrieval. Several studies attempted to address this issue, with a clear growing interest in the last years in the use of non-linear machine learning techniques. In this thesis we introduce the use of the of the ε-insensitive Support Vector Regression technique for the retrieval of soil moisture content from microwave remotely sensed data at field/basin scale. At the time of this work, no studies existed in the literature that investigated the effectiveness of this technique in this specific context. SVR has attractive properties, such as ease of use, good intrinsic generalization capability and robustness to the noise in the reference data, which make it a valid candidate as an alternative to more traditional neural network based techniques, e.g., the MLP-NN, considered as benchmark methods in this specific field. The effectiveness of the SVR methods is investigated in different operational scenarios, namely: i) the inversion of a physical based electromagnetic model, which guarantees the availability of numerous and well representative labeled reference samples; and ii) the inference of an empirical based model from a limited number of labeled reference samples measured in the field. Moreover, in each scenario different combinations of input features, namely microwave signal measurements for different sensor frequency, polarization and acquisition angles, are tested. The performance of the investigated methods is discussed in terms of estimation accuracy, generalization ability, computational complexity and ease of use. This analysis is important as preparatory step for the development of accurate and robust retrieval systems and provides useful indications for building soil moisture estimation processors for upcoming satellite missions or near real time applications as well.

B. *A novel multi-objective strategy for tuning the free parameters of non-linear regression methods.*

Non-linear machine learning techniques depend on a set of free-parameters which control the learning process of the method and thus have a direct impact on the accuracy and generalization ability of the final estimation system. The tuning of these parameters is thus a major step in the definition of accurate and robust estimation systems. In this thesis, we propose to address this task in the framework of the multi-objective optimization. Two or more metrics that quantify from different and competing perspectives the goodness of a given parameter configuration are taken into account and jointly optimized according to the concept of Pareto optimality (*Feldman, 1980*). This allows preserving the meaning of each metric involved in the process and deriving multiple optimal solutions to the tuning problem, each one leading to a different optimal trade-off among the metrics. The proposed strategy leads to two main advantages with respect to mono-objective strategies: i) the intrinsic improved robustness and efficiency, since multiple metrics are jointly exploited; and ii) the possibility to select the parameter configuration that shows the desired trade-off among different criteria and thus best meets both the application constraints and the requirements of the specific estimation problem. These features are fundamental for the efficient and effective definition of robust estimation systems.

C. *A novel hybrid approach to the retrieval of geo-/bio-physical variables from remote sensing data integrating both theoretical electromagnetic models and field reference measurements.*

As pointed out in the literature review reported the previous section, the retrieval of geo-/bio-physical variables can be addressed according to two approaches: the inversion of physical based theoretical electromagnetic models or the inference of an empirical relationship from field reference samples. Both approaches present strengths, but also limitations that may strongly affect the accuracy and the robustness of the estimation process. In order to convey the strengths of both approaches into a unique estimation method, in this thesis we develope a novel hybrid approach based on the integrated use of a set of (few) field reference samples and a physical based theoretical model. The estimation process is modeled by two terms: the first one expresses the relationship between the input features and the target variable according the model based on the physics of the considered problem; the second one corrects the deviation between theoretical model estimates and true target values according to an empirical data-driven relationship inferred from the (few) field reference samples. On the one hand, the proposed methodology allows to preserve the robustness and generality of theoretical model based estimation, which stems from the rigorous theoretical foundation. On the other hand, it reduces the bias and imprecisions due to simplifications in the analytical

formulations of the model. This approach is particularly useful for those application domains, such as the soil moisture retrieval, where the limited availability of field reference samples and the spatial and temporal variability of the target geo-/bio-physical variable represent a limitation for the development and applicability of robust and general empirical models.

D. *A sensitivity analysis and a retrieval system for soil moisture content estimation from new generation SAR imagery in an Alpine catchment.*

Satellite SAR imagery has been widely investigated and exploited for the retrieval of soil biophysical characteristics, in particular soil moisture content, with promising results. However, very limited attention has been devoted to the analysis of mountain areas. The analysis of the SAR signal in this challenging environment is complicated by the effects of vegetation, topography and land-cover heterogeneity. In this thesis, the potential, but also the limitations, of the new generation satellite SAR sensors for the retrieval of soil moisture content in the Alpine area is investigated. The study consists first in a sensitivity analysis aimed at understanding both the sensitivity of the SAR backscattering signal to soil moisture content and the influence on it of the local characteristics of the investigated area in terms of topography, land-cover and status of the vegetation coverage. Then an advanced retrieval system to address the soil moisture retrieval problem tailored to this environment is presented. The proposed system takes advantage from the available ancillary data to disentangle the effects of topography and land-cover heterogeneity on the SAR signal and exploits the methodologies anticipated in the previous points to cope with the complexity of the retrieval process.

E. *An empirical study on the effectiveness of fully-polarimetric SAR signals for soil moisture estimation in mountain areas.*

In order to deal with the disturbing effect of vegetation and roughness with regard to the estimation of soil moisture from SAR signals, in the literature multi-frequency and multi-angle approaches have been proposed. The rationale is that the SAR signal is sensitive to the target properties in a different way and to a different extent depending on the system frequency and incidence angle. The combination of the signals acquired with different system configurations leads to a better characterization of these disturbing factors and thus to a more accurate estimation of the variable of interest. However, multi-frequency and multi-angle scenes over the same area of interest are very difficult to acquire from satellite platforms. More efficient is the acquisition of SAR imagery using multiple polarization configurations. This modality is currently operational for many satellite SAR sensors. In this thesis the effectiveness of polarimetric features for the challenging task of soil moisture retrieval in mountain areas is investigated. The polarimetric information is processed according to standard intensity and phase

extraction techniques as well as advanced polarimetric decompositions and data driven transformation algorithms. Different combinations of the extracted features are then investigated by providing them as input to the retrieval system and evaluating the resulting soil moisture estimates in terms of quantitative accuracy metrics and spatial patterns consistency. The proposed analysis provides valuable insights about the exploitation of polarimetric SAR signals in mountain areas useful for future studies and for the definition of operational products in similar environmental conditions, as well as for the development of future satellite SAR missions.

F. An improved algorithm for mapping and monitoring Green Area Index (GAI) in Alpine pastures and meadows from satellite MODIS imagery.

Leaf Area Index is one of the most important structural variables of vegetation canopies. The corresponding parameter for fully photo-synthetically active vegetation canopies, such as grasslands, is often referred as Green Area Index (GAI). The synoptic monitoring of this parameter at regional scale is crucial for the understanding of hydrological processes, energy fluxes between the ground surface and the atmosphere, climatic trends and the impact of human activities on them. The retrieval of these variables has been largely addressed with the use of medium resolution satellite data, which have the advantage of being acquired globally, with high repetition rate and on a regular basis. While focusing on the global consistency of the estimates, current products present typically limited accuracy and reliability for specific biomes, such as the Alpine grasslands. To address this issue, this thesis presents an improved algorithm for the retrieval of GAI from moderate resolution satellite MODIS imagery specifically customized for Alpine meadows and pastures. In particular, the main novel aspects of the proposed algorithm, which is based on the inversion of a radiative transfer model, are: i) the improved spatial resolution with respect to the existing operational products; ii) the tuning of the transfer model on the characteristics of Alpine pastures and meadows; and iii) the accounting for the local topographic characteristics. These features open the path for the exploitation of moderate resolution satellite imagery for novel and more accurate monitoring analyses in the challenging mountain environment at regional scale.

Structure of the Thesis

The dissertation is organized into eight chapters.

In this chapter we briefly introduced the context and the state-of-the-art on the main topic addressed in this thesis, namely the retrieval of geo-/bio-physical variables from remote sensing data. Moreover, the objectives, motivations and main novel contributions of this work have been highlighted.

Chapters 2 and 3 focus on the use of non-linear machine learning methods for the retrieval of geo-/bio-physical variables from remote sensing data. In particular, Chapter 2 introduces the Support Vector Regression technique in the context of the retrieval of an important environmental variable, i.e., the soil moisture content, from microwave remote sensing data. The chapter discusses its effectiveness in two different operational scenarios in terms of accuracy, generalization ability and ease of use in comparison with the Multi-Layer Perceptron Neural Network, which is considered as a benchmark in this specific field. Chapter 3 deals with the problem of tuning the free parameters of non-linear machine learning methods, which is a crucial task for the achievement of good accuracy and generalization ability from the considered method. The problem is addressed in the framework of the multi-objective optimization according to the concept of Pareto optimality, which allows one deriving multiple optimal trade-off solutions among which selecting the one that best meets the application constraints and requirements.

Chapter 4 deals with the problem of overcoming the dichotomy between physical based and empirical based approaches to the retrieval of geo-/bio-physical variables. To this aim a novel hybrid approach is proposed, which tries to integrate the robustness and generalization ability of the physical based model inversion approach with the information extracted from a set of (few) field reference samples.

The remaining chapters focus the attention on operational estimation scenarios. Chapters 5 and 6 deal indeed with the topic of retrieving soil moisture content from new generation satellite SAR imagery in the complex Alpine environment. Chapter 5 presents a sensitivity analysis of the SAR signal to both soil moisture content and the local characteristics (e.g., topography) of the investigated area. Then it introduces an advanced retrieval system, which benefits from the methodological innovations introduced in the previous chapters and from the availability of ancillary data to deal with the complexity of the retrieval problem. Chapter 6 goes a step further in the analysis by investigating the fully polarimetric capability of the SAR imagery considered. This is done in order to understand the potentialities and the limitations of this advanced acquisition modality of new generation SAR systems with regard to the analysis and mapping of soil variables in the complex mountain environment.

Chapter 7 focuses on the mapping and monitoring of another crucial geo-/bio-physical environmental variable, i.e., the green area index (GAI) of Alpine grasslands. In this context an improved retrieval algorithm is presented, which tries to overcome the main limitations of existing products by increasing the spatial resolution, tuning the retrieval on the spectral characteristics of the target area and accounting in the retrieval process for the local topography of the scene.

Finally, Chapter 8 draws the conclusion on the presented activities and on the novel contributions to the state-of-the-art in the field of geo-/bio-physical variables retrieval introduced in this dissertation. Moreover, it discusses possible future developments of this PhD work.

Chapter 2

Comparative Analysis of Non-Linear Machine Learning Regression Methods for the Retrieval of Soil Moisture Content in Different Operational Scenarios

This chapter introduces an advanced state-of-the-art non-linear regression method, the Support Vector Regression technique, in the context of the retrieval of an important environmental variable, namely the soil moisture content, from microwave remote sensing data. Its effectiveness is investigated in two different operational scenarios: i) the inversion of an electromagnetic physical model, which guarantees the availability of numerous and well representative reference samples for the training of the method; and ii) the inference of empirical relationships from a limited number of field reference samples. The performance in terms of accuracy, generalization ability and ease of use are discussed in comparison with the Multi-Layer Perceptron Neural Network, which is considered as a benchmark in this specific field. Useful guidelines for building soil moisture estimation processors for upcoming satellite missions or operational near-real-time applications as well are drawn at the end of this analysis.

2.1 Introduction and Motivation

Soil moisture content (SMC) is a key variable in many hydrological and meteorological processes. It controls the infiltration rate during precipitation events, the evapotranspiration and the micrometeorology of soils (*Rodriguez-Iturbe et al., 1999*). Thus it influences both global water and energy balances. As a consequence, the information about the spatial distribution and concentration of soil moisture is of great importance in both hydrological applications, such as floods predictions in case of extreme rainfall events, watershed management during dry periods, irrigation scheduling, precision farming, and earth sciences,

like climate change analysis and meteorology (*Bastianseens and Bos, 1999; Sandholt et al., 2002; Heathmann et al., 2003*).

Several techniques have been proposed for *in-situ* measurements of soil moisture content (*Walker et al., 2004*). However, field measurements are typically expensive and time demanding. Moreover, they provide point-like information which is difficult to spatialize due to the spatial variability of the soil characteristics. For this reason, the retrieval of soil moisture from remote sensing signals represents an efficient and effective solution to support field measurements. Microwave signals are widely used for this purpose, thanks to their well-established sensitivity to the amount of water in the soil surface (*Wang, 1980; Ulaby et al., 1986b; Ulaby, 1974; Jackson, 1997*). However, the complexity and non-linearity of the retrieval problem as well as the ambiguities that may affect the microwave signal, such as surface roughness and vegetation, make the soil moisture estimation problem particularly challenging (*Ulaby et al., 1978; Ulaby et al., 1979; Dobson and Ulaby, 1981; Lakhankar et al., 2009; Mattia et al., 1997*).

For the analysis and characterization of the soil surface, empirical relationships derived from experimental field reference samples were first exploited as proposes by Lin et al. (*1994*). However, simple parametric function resulted to be inadequate for modelling the complexity of the soil moisture content retrieval. Analytically more sophisticated functions can be instead used, as proposed by Dubois et al. (*1995*), with encouraging accuracies for the investigated area. Apart from some attempts, however, very limited effort has been put in this kind of approach for addressing the retrieval problem.

Analytical electromagnetic models, on the contrary, have been in use for a long time in this challenging field. In particular the Integral Equation Model (IEM) by Fung et al. (*1992*) and the emissivity model by Wang and Choudhury (*1995*) are widely used in the remote sensing community, thanks to their wide range of applicability (in terms of both target properties and system characteristics) and validated behaviour (*Fung, 1994*). These properties and the solid theoretical foundation make the use of such models particularly appealing in soil moisture estimation from remote sensing data, especially when no (or very limited) field reference samples are available. This condition is quite common in near real time operational conditions. However, electromagnetic may provide a simplified representation of the reality. One of the most critical issues in this regard is the description of the surface morphology. Commonly used parameterizations assume wide-sense stationary statistics across the horizontal plane, which limit in some cases the possibility to completely describe the variability of natural surfaces (*Mattia and Le Toan, 1999*). Another major issue when using theoretical models, as pointed out in Chapter 1, is their inversion, thus implying the use of advanced (and often computationally demanding) inversion strategies (*Bindlish and Barros, 2000; Tabatabaeenejad and Moghaddam, 2009; Oh, 2006*).

In the last decade non-linear machine learning (ML) techniques, in particular Artificial Neural Networks (ANNs), have been successfully introduced in the field of soil variables

retrieval. In particular, such techniques were exploited as an effective solution to address the inversion of electromagnetic models (*Del Frate et al., 2003; Notarnicola et al., 2008; Paloscia et al., 2008*). A set of simulated reference samples (i.e., simulated backscattering coefficients associated with the corresponding soil characteristic variables) is generated exploiting the theoretical model in the forward operational way. These samples are then used to train the ANN method by providing as input the simulated backscattering coefficients and as output the corresponding target variable (e.g., the soil moisture content). A second subset of simulated reference samples is usually exploited for model-selection purposes, i.e., to tune the free-parameters and define the architecture of the network. Once trained, the technique should reproduce the desired mapping, thus can be exploited for estimating the desired target parameter from remote sensing imagery. In the prediction phase, ANNs are typically fast. This feature is particularly appealing in near real time operational conditions or when large amounts of data, such as large satellite scenes, have to be processed. However, the training/tuning process is usually a complex task. Indeed no clear rules are available for the definition of the network architecture (number of hidden layers and nodes). This process has to be carefully controlled, since the method may incur in the so-called over-fitting, the phenomenon whereby the network learns too precisely the training samples losing the ability to correctly predict independent unknown samples. This issue becomes as stronger as smaller is the set of reference samples used in the training phase.

Another promising ML technique is the ε-insensitive Support Vector Regression (SVR) (*Vapnik, 1995*). Among others, key features of this method are: i) the exploitation of the margin principle; ii) the possibility of solving the learning problem using a convex cost function; and iii) the sparseness of the solution. These characteristics result in a good intrinsic generalization capability and a limited complexity in handling the training phase. Some previous studies have investigated the effectiveness of SVR for the estimation of geo-/bio-physical variables from remotely sensed data, indicating good, competitive results on various application domains with respect to other state-of-the-art methods (*Bruzzone and Melgani, 2005; Camp-Valls et al., 2006*). However, almost no effort has been devoted to the investigation and the use of this promising method for the retrieval of soil moisture content, with the exception of the study presented by Ahmad et al. (*2010*), which however investigates the use of SVR method for soil moisture content retrieval for a coarse scale mapping from low resolution satellite data, thus implicitly neglecting the influence of local scale surface roughness.

In this chapter, the effectiveness of the SVR technique to address the problem of soil moisture content estimation at field/basin scale is investigated. At this scale local phenomena, such as surface roughness, affect the microwave signal thus increasing the complexity and non-linearity of the estimation process. Up to now, no investigations on the use of the SVR technique in this challenging and interesting operational condition have been proposed. In greater detail, the main novel and interesting aspects of this analysis are:

1. The assessment of the SVR method in two distinct operational scenarios, namely: i) the inversion of a physical based electromagnetic model, the common operational scenario in this application field, which guarantees the availability of numerous and well representative labelled reference samples; and ii) the inference of an empirical based model from a limited number of labeled reference samples measured in the field. This second scenario is interesting for assessing the generalization capability and robustness of the method in case of limited availability of reference samples
2. The test of different combination of input features, namely microwave signal measurements for different sensor frequency, polarization and acquisition angles.

For comparison purposes, a Multi-Layer Perceptron Neural Network, referred as a benchmark in this field, is considered. For both methods, the achieved performance in terms of estimation accuracy, generalization ability, computational complexity and ease of use is thus discussed. The aim is to derive indications about the effectiveness of the investigated methods useful for developing robust and accurate retrieval systems and building soil moisture content processors for the upcoming satellite missions or near real time applications as well.

The rest of the chapter is organized as follows. The next section will introduce the basic principles and the analytical formulation of the ε-insensitive SVR technique, which will be exploited in this and other analysis performed later on in this dissertation. Section 2.3 will describe the data set exploited for the analysis and provide the details of the experimental analysis. Section 2.4 is devoted to the presentation and discussion of the achieved experimental results. Finally, the conclusion of the study will be drawn in Section 2.5.

2.2 The ε-insensitive Support Vector Regression Technique

Given a general retrieval problem (see equation 1.1 in Chapter 1), the goal of the ε-insensitive SVR technique is to find a function *f'*as smooth as possible that approximates *f* while keep at most a deviation ε from the target *y* (*Vapnik, 1995*). For this purpose, the original *m*-dimensional input domain is mapped into a space with higher dimensionality, where the function underlying the data is supposed to have an increased flatness. Thus *f'* is approximated in a linear way:

$$f'(\boldsymbol{x}) = \boldsymbol{w} \cdot \Phi(\boldsymbol{x}) + b \qquad (2.1)$$

where $\boldsymbol{w}$ represents the vector of weights of the linear function, $\Phi(\cdot)$ is the mapping that projects the samples from the original into the higher dimensional feature space and *b* is the bias.

The optimal linear function in the transformed space is selected by minimizing the structural risk, which is the combination of the training error (empirical risk) and the model complexity (confidence term) evaluated on a set of N_{tr} reference training samples. The first term is calculated according to a ε-insensitive loss function (see Figure 2.1), where ε

quantifies the tolerance to errors, i.e., it allows defining an insensitive tube surrounding the function *f'* (thus increasing the robustness of the technique to small errors and to the noise in the training set). Equivalently, the penalty can be expressed by means of non-negative slack variables ξ , ξ^* which measure the distance (in the target space) of the training samples lying outside the ε-insensitive tube from the tube itself. The second term of the cost function is expressed through the Euclidean norm of the weight vector $\boldsymbol{w}$, which can be inversely related to the geometrical margin of the corresponding solution and thus (under a geometrical interpretation) to the complexity of the model. The cost function to minimize becomes:

$$\psi(\boldsymbol{w}, \xi) = C \sum_{i=1}^{N} (\xi_i + \xi_i^*) + \frac{1}{2} \|\boldsymbol{w}\|^2 \qquad (2.2)$$

and is subject to the following constraints:

$$\begin{cases} y_i - [\boldsymbol{w} \cdot \Phi(\boldsymbol{x}_i) + b] \leq \varepsilon + \xi_i \\ [\boldsymbol{w} \cdot \Phi(\boldsymbol{x}_i) + b] - y_i \leq \varepsilon + \xi_i^* \quad i = 1,2, \dots, N_{tr} \\ \xi_i, \xi_i^* \geq 0 \end{cases} \qquad (2.3)$$

C is a regularization parameter that allows one to tune the trade-off between the complexity (or flatness) of the function *f'* and the tolerance to empirical errors.

The constrained optimization problem in (2.2) can be reformulated through a Lagrange functional, which leads in the dual formulation to a convex (easy to handle) quadratic problem (QP) and thus to a unique solution (the global minimum of the cost function). The final prediction function, in terms of the samples in the original input domain, becomes:

$$f'(\boldsymbol{x}) = \sum_{i \in N} (\alpha_i - \alpha_i^*) K(\boldsymbol{x}_i, \boldsymbol{x}) + b \qquad (2.4)$$

where α_i and α_i^* represent the Lagrange multipliers of the QP and $K(\cdot,\cdot)$ is a kernel function. The latter must satisfy the Mercer's theorem, so that it can be associated with some type of inner product in the high-dimensional feature space (i.e., $K(\boldsymbol{x}_i, \boldsymbol{x}) = \Phi(\boldsymbol{x}_i) \cdot \Phi(\boldsymbol{x})$). Thus, the kernel function allows evaluating the similarity between a pair of samples in the transformed feature space as a function of the samples in the input space, i.e., without the explicit definition of the mapping function $\Phi(\cdot)$. This strongly reduces the analytical complexity related to the mapping issue. Commonly adopted kernels are polynomial and Gaussian radial basis functions (RBF) kernels (*Bruzzone and Melgani, 2005; Moser and Serpico, 2009*). Lagrange multipliers weight each training sample according to its importance in determining the solution function *f'*. Samples associated with a nonzero Lagrange multiplier are called support vectors. The other samples have no weight in the definition of the result since they fall within the ε-tube. Consequently, to increase ε means to reduce the number of support vectors. This will increase the sparseness of the final representation of the data at the price of lower approximation accuracy on training samples. In this sense, ε quantifies the trade-off between data sparseness and approximation accuracy of the model.

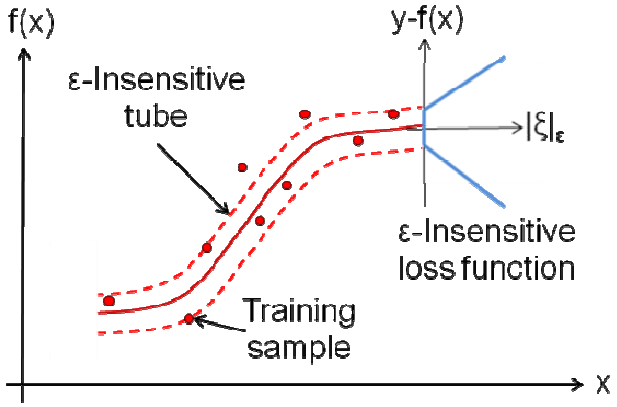


Figure 2.1. Example of ε-insensitive tube defined by a linear loss-function for a one-dimensional learning problem.

Some attractive features of the SVR result from the analytical formulation presented above:

1. *Good intrinsic generalization ability*, thanks to the use of the ε-insensitive cost function (which introduces sparseness into the model) and the optimization of both empirical error and model complexity to drive the learning process.
2. *Good capability to handle high-dimensional input spaces*. This property stems again from the way in which generalization is controlled during the learning (minimization of both empirical risk and confidence term).
3. *Limited complexity and high stability of the learning process*, due to the convexity of the optimization problem and the use of the kernel trick.
4. *Ease of use*, since relatively few free parameters (i.e., the regularization coefficient C, the width of the insensitive tube ε and the kernel type and parameters) have to be tuned.

2.3 Data Set Description and Design of the Experiments

2.3.1 Field Measurements

The field measurements considered in this work belong to two data sets associated with two different acquisition campaigns on bare soils. The first one (from this point referred as *ActPass* data set) consists of 27 samples acquired by the University of Bern's researchers in three bare and essentially smooth fields (*Wegmueller et al., 1994*). Roughness properties in terms of standard deviation of height s and correlation length l are reported in Table 2.1. The remote acquisitions were performed with a truck-mounted fully polarimetric radiometer/scatterometer operating at different frequencies and incidence angles.

Table 2.1. Main characteristics of the field soil measurements considered in the analysis.

Data Set	# of samples	Field Type	Roughness characteristics (*s* and *l*)	Soil moisture characteristics (*SMC*)
ActPass	27	Bare or lightly vegetated	$0.4\text{cm} \leq s \leq 1.7\text{cm}$ $2.0\text{cm} \leq l \leq 5.0\text{cm}$	$6\% \leq \text{SMC} \leq 38\%$
Scatterometer	17	Bare	$1.3\text{cm} \leq s \leq 2.1\text{cm}$ $5.0\text{cm} \leq l \leq 13.0\text{cm}$	$8\% \leq SMC \leq 32\%$

The use of the same antenna for both active and passive measurements ensured that the two instruments observed the target under almost identical spatial conditions. The ground truth measurements of the volumetric soil moisture content (*SMC*) (i.e., the average water content of the top 4 cm of the soil) were performed taking 5 soil samples of a known volume and drying them at 105°C according to the guidelines given by Cihlar et al. (*1987*). The percentage error of the measurements is about 10%. For reasons of homogeneity with the simulated data (described in the next subsection), *SMC* values were converted into relative dielectric constant (from this point indicated as *diel*) through a well-known relationship (*Hallikainen et al., 1985*).

The second data set (indicated as *Scatterometer* data set) is made up of 17 samples acquired in two bare fields at different times, from 1998 to 2004 (*Mattia et al., 2003*). For this data set, soil roughness conditions were extremely variable even within the same field, as reported in Table 2.1, due to plowing practice. Remote measurement acquisitions were accomplished with a C-band scatterometer using HH and VV polarizations at 23° and 40° incidence angles, while ground truth measurements of soil moisture content were collected with the same procedure described above (*Cihlar et al., 1987*). Then soil moisture measurements were converted into relative dielectric constant values.

2.3.2 Simulated Samples

Simulated data are generated according to two theoretical models extensively validated and analyzed in the literature: the Integral Equation Model (IEM) (*Fung et al., 1992*) for backscattering coefficients and the Wang model (*Wang and Choudhury, 1995*) for emissivity values. Both models have been extensively used in the literature and validated with field measurements, including those presented above (*Ruf and Zhang, 2003; Wang, 1994*).

The key equation in the IEM formulation, which allows one deriving the simulated backscattering coefficient given as input the soil characteristics (relative dielectric constant and roughness) and the sensor configuration (frequency, polarization and incidence angle), is the following:

$$\sigma_{pp}^{0} = \frac{k^2}{2} e^{-2k_z^2 s^2} \sum_{k=1}^{\infty} \left| I_{pp}^{n} \right| \frac{W^{(n)}(-2k_x, 0)}{n!} \qquad (2.5)$$

where k is the wavenumber, θ is the incidence angle, $k_x = k \cdot \sin(\theta)$, $k_z = k \cdot cos(\theta)$, pp refers to the HH or VV polarization states and s is the terrain standard deviation of height. I_{pp}^{n} depends on k, s and the Fresnel coefficients in H and V polarizations. $W^{(n)}(-2k_x, 0)$ is the Fourier transform of the n-th power of a predefined surface correlation function, which in this work has been derived according to an exponential correlation function. The exponential function is the most widely used, despite it does not completely describe the variability of natural surfaces (*Mattia and Le Toan, 1999*)

Concerning the radiometer responses, assuming to consider a simple homogeneous isothermal soil medium with plain air-soil boundary, the brightness temperature of the soil surface can be expressed as:

$$T_B(\theta, p) = e^{sp}(\theta, p) T_{s_{pp}}^{0} \qquad (2.6)$$

where T is the soil temperature, θ the incidence angle, p the polarization and $e^{sp}(\theta, p)$ the soil emissivity, which assumes the following expression according to Wang and Choudhury (*1995*):

$$e^{sp} = 1 - [(1 - a) r_H + a r_V] e^{-4k^2 s^2 cos^2 \theta} \qquad (2.7)$$

where r_H and r_V are the surface reflectivities for the H and V polarization, respectively; a is a mixing polarization parameter depending on the operating frequency and the terrain standard deviation of height s. As for the IEM model, by using this formulation it is possible to derive emissivity values given as input a configuration of the soil and sensor characteristics.

The generation process is performed by considering ranges of soil characteristics (i.e., dielectric constant *diel*, standard deviation of height *s* and correlation length *l*) wider to some extent than those of the field measurements, with the aim to obtain a set of simulated samples capable of representing more general situations. For both *ActPass* and *Scatterometer* data sets, more than 1000 samples are generated. A Gaussian noise contribution (estimated on the basis of the real data) is then added to these samples (*Notarnicola et al., 2008*). Further details on the ranges and settings considered during the model simulations are available in Table 2.2.

2.3.3 Design of Experiments

To realize the analysis presented in this work, several experiments are carried out, all of them requiring two disjoint set of labeled samples: i) the reference set, which is used to train/tune the regression technique; and ii) the test set, which allows evaluating the estimation performance in case of unseen samples. By exploiting in different ways field measurements and simulated samples, two scenarios are defined and investigated:

Table 2.2. Ranges of soil variables considered for model simulations.

Data Set	Standard deviation of height (*s*)	Correlation length (*l*)	Relative dielectric constant (*diel*)
ActPass	0.3cm ≤ *s* ≤ 1.9cm step 0.1 cm	2.0cm ≤ *l* ≤ 5.0cm step 1.0 cm	2 ≤ *diel* ≤ 20 step 1
Scatterometer	1.3cm ≤ *s* ≤ 2.1cm step 0.1 cm	5.0cm ≤ *l* ≤15.0cm step 1 cm	2 ≤ *diel* ≤ 20 step 1

1. *Scenario 1: inversion of a physical based electromagnetic model*. The regression method is exploited for approximating the inverse theoretical mapping between electromagnetic measurements and soil variables. To this aim, the following procedure is adopted. The samples simulated according to the IEM and Wang models are randomly divided into two equally sized subsets, i.e., training and validation sets. The first one is used for the training of the estimator, while the second one is used for model-selection purposes. Different configurations of the free parameters, defined according to a pre-defined grid, are tested and the one which provides the best performance on the validation samples is selected. Finally, the regressor is assessed using the field measurements.
2. *Scenario 2: inference of an empirical based mapping*. In this operational condition, both training/tuning and performance assessment of the method are done using field measurements. Due to the limited number of field measurements available, a cross-validation scheme has been adopted (*Cherkassky and Mulier, 1998*). The rationale behind this approach is to split the available samples into *t* subsets; *(t-1)* are used for the training/tuning the regression technique, which accuracy is assessed on the remaining subset. The procedure is iterated *t* times, each time changing the configuration of reference and test sets. At the end, all samples are considered as test.

In both scenarios, various combinations of active and/or passive microwave measurements with the following sensor and acquisition geometry configurations are considered:

- *ActPass data set:*
    a) Backscattering coefficient polarization HH and emissivity polarization H both at 4.6 GHz and 20° incidence angle (indicated as *ActPass HH4.6-H4.6*).
    b) Backscattering coefficient polarization HH at 4.6 GHz and emissivity polarization H at 2.5 GHz both at 20° incidence angle (indicated as *ActPass HH4.6-H2.5*).
    c) Backscattering coefficient polarizations HH and VV and emissivity polarizations H and V both at 4.6 GHz and 20° incidence angle (indicated as *ActPass HHVV4.6-HV4.6*).

- *Scatterometer data set*:
    a) Backscattering coefficient polarizations HH VV at 23° incidence angle (indicated as *Scatterometer HHVV 23°*).
    b) Backscattering coefficient polarizations HH VV at 23° and 40° incidence angles (indicated as *Scatterometer HHVV 23°/40°*).

The aim is to investigate whether and to what extend different input features affect the estimation accuracy of the considered estimation technique and if the choice of these features depends on the operational scenario. These configurations are motivated by: i) similarities to the acquisition modes of operational microwave remote sensing systems; and ii) continuity with previous research published in the literature (*Notarnicola et al., 2008*).

Concerning the SVR technique set-up, the LIBSVM software (*Chang and Lin, 2011*) is used for all the experiments. A Gaussian RBF kernel is adopted thanks to the limited computational overhead and the good performance achieved in previous analyses (*Bruzzone and Melgani, 2005*). The ranges of the free parameters are $[10^{-3}; 10^{3}]$ for γ (the RBF kernel width) and C, respectively; $[10^{-4}; 10]$ for ε. With regard to the MLP Neural Network, the Matlab "Neural Network Toolbox" is used. We refer to that Notarnicola et al. (*2008*) for details on the MLP-NN implementation and tuning.

2.4 Experimental Results

The estimation accuracy of the investigated SVR technique is evaluated on the set of test samples according to two figures of merit, namely the mean squared error (*MSE*) and the mean relative error (*MRE*), defined as follows:

$$MSE = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} \left(diel_i - \widehat{diel}_i \right)^2 \qquad (2.8)$$

$$MRE = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} \frac{\left| diel_i - \widehat{diel}_i \right|}{diel_i} \qquad (2.9)$$

where N_{test} denotes the number of test samples, $diel_i$ and $\widehat{diel}_i$ the measured and predicted dielectric constant values for the i-th sample, respectively. These figures of merit are compared with those obtained by the MLP NN on the same set of samples. Also scatter plots of estimated versus measured dielectric constant values have been considered for performance evaluation, in combination with linear trend line and squared correlation coefficients (R^2).

2.4.1 Results: Scenario 1

The results achieved in terms of mean estimation errors in the first scenario investigated are reported in Table 2.3. Considering the whole range of variability of the target parameter, SVR achieves better accuracy than MLP for both data sets and for all the input feature configurations, with the exception of the *MRE* value in the case of the *ActPass HHVV4.6-HV4.6* configuration. These results suggest that the SVR technique has a better generalization ability than the MLP NN does, i.e., a better capability to learn a mapping that provides higher accuracy in the prediction of unknown real samples. This capability stems from the use of the ε-insensitive cost function and the optimization of both empirical and structural risks during the training phase.

Analyzing in more detail the results achieved by varying the configurations of sensor type, acquisition geometry and mode, a general tendency to improve the estimation accuracy by increasing and diversifying the information given as input to the SVR technique is observed. This effect can be seen also with the MLP NN. Both methods achieve the highest accuracy (in terms of *MSE*) when two polarizations (*HHVV4.6-HV4.6*) and two acquisition angles (*HHVV 23°/40°*) are provided in input for the *ActPass* and *Scatterometer* data sets, respectively. *MRE* values confirm this trend in the case of the MLP NN, while are almost constant in the case of the SVR. The improvement in estimation accuracy is mainly for values of dielectric constant lower than 10. This effect can be explained by the fact that differences in the signal acquired by the remote sensor among polarizations or acquisition angles are stronger for dry soils than for wet soils. The use of additional features helps the estimation technique to disentangle the ambiguity (greater for dry soils than for wet soils) introduced by soil roughness into microwave signals. However, while MLP shows a degrade in the estimation accuracy for high soil moisture content values by increasing the number of features provided in input, SVR maintains similar performance in this operating range for all the different configurations considered. This observation points out that the SVR method has a better capability to exploit the information provided by the additional input features and has higher robustness to the noise associated with additional input channels/acquisition angles.

By analyzing the scatter plots of estimated versus measured dielectric constant values (see the examples reported in Figure 2.2) it is possible to observe the tendency of the MLP regressor to associate high dielectric constant values with outlier or noisy samples, thus explaining the worse performance in terms of estimation accuracy (especially *MSE*) and the slight overestimation of the linear tendency line. In contrast, the estimates provided by SVR present a quite flat trend (as indicated by the slope coefficients in the graphs), slightly over-estimating and under-estimating low and high values of dielectric constant, respectively, but providing on average a small error even in presence of outliers. The difference between the behaviours of SVR and MLP estimates is enhanced in the case of the *Scatterometer* data set, probably due to its higher ambiguity with respect to the *ActPass* data set (i.e., a larger dynamic range of roughness parameter values and the absence of passive microwave

measurements). This suggests that SVR, especially in the presence of noisy and ambiguous input samples, is prone to learn a mapping with a flatter behaviour, which provides estimates with a reduced error, at the price of a reduced dynamic range of the output values at the extremes. In contrast, MLP shows the capability to cover the whole range of output values, but this leads to higher average inaccuracies in the output. Concerning the determination coefficient R^2, SVR does not always provide the highest values despite providing the highest estimation accuracy in terms of mean errors. This behaviour was also observed in the case of the *Scatterometer* data set, characterized by a higher ambiguity. Thus different quality metrics may sometimes provide competing indications about the quality of the estimates of a regression technique. This aspect may become critical when dealing with the model-selection issue in real operational conditions.

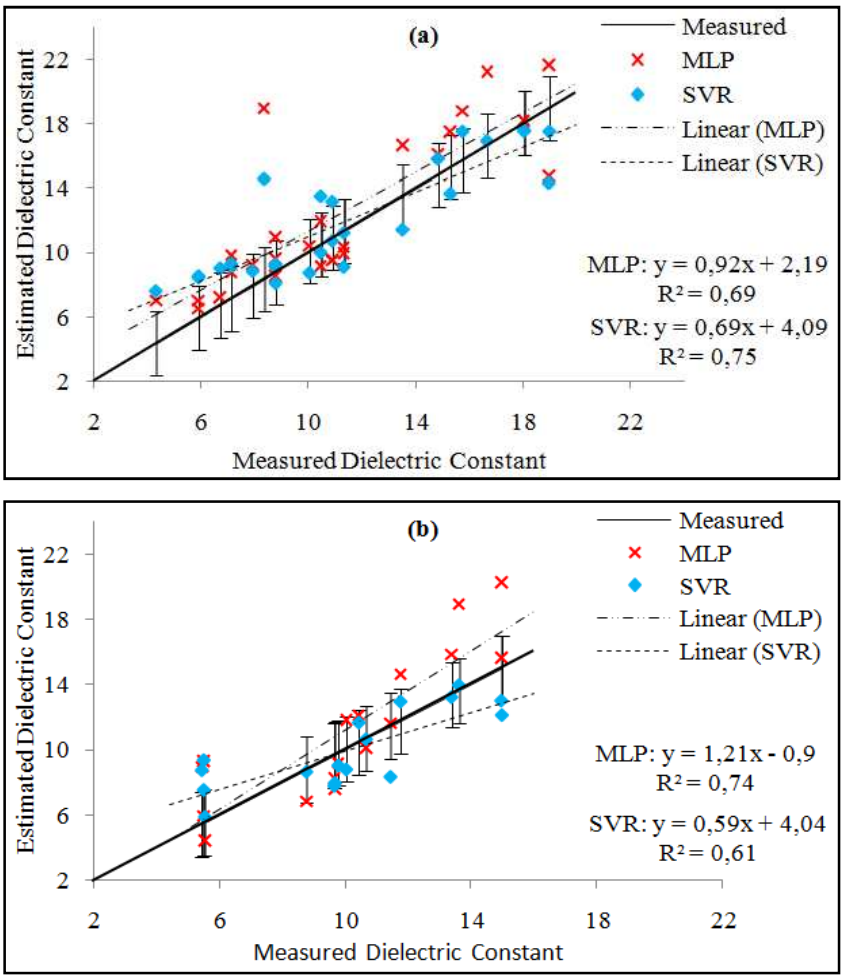


Figure 2.2. Scatter-plots of estimated versus measured dielectric constant values for Scenario 1 experiments: (a) ActPass HHVV4.6-HV4.6 configuration; and (b) Scatterometer HHVV 23°/40° configuration. Error bars refer to the standard deviation of the measured values.

Table 2.3. Estimation accuracy of the MLP NN and SVR methods on ActPass and Scatterometer datasets for Scenario 1 experiments.

	***ActPass* Dataset**						***Scatterometer* Dataset**			
	HH4.6-H4.6		***HH4.6-H2.5***		***HHVV4.6-HV4.6***		***HHVV 23°***		***HHVV 23/40°***	
Full diel Range										
Method	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***
MLP NN	8.92	18.55%	9.28	17.06%	8.28	15.64%	11.79	24.44%	6.79	18.8%
SVR	6.21	16.53%	5.67	15.61%	5.04	15.8%	4.98	15.16%	3.74	16.55%
diel < 10										
Method	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***
MLP NN	17.46	31.07%	16.95	26.5%	12.77	20.06%	16.02	32.09%	4.84	23.87%
SVR	9.85	24.5%	8.27	21.41%	7.24	22.5%	6.67	19.36%	4.63	20.92%
diel > 10										
Method	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***
MLP NN	3.05	9.93%	4.48	11.16%	5.19	12.6%	8.03	17.62%	8.52	14.3%
SVR	3.71	11.5%	3.88	11.62%	3.53	11.2%	3.48	11.42%	2.94	12.66%

2.4.2 Results: Scenario 2

Numerical performance achieved in the second scenario is reported in Table 2.4. SVR technique achieves better estimation accuracies with respect to MLP with both *ActPass* and *Scatterometer* data sets according to all the considered figures of merit. For some configurations, such as the *ActPass HHVV4.6-HV4.6* and the *Scatterometer HHVV 23°*, the difference between the two techniques is rather high. A more detailed analysis of the different figures of merit suggests that SVR is more robust with respect to MLP. The second one shows an unexpected degrade of the estimation accuracy for one configuration (i.e., the *ActPass HHVV4.6-HV4.6* configuration), while SVR is able to achieve good accuracies also in this case. It is worth noting that in this case one out-of-range estimate has been excluded from the computation of the figures of merit associated with the MLP regressor, thus indicating the difficulty of this technique to proper control the range of variability of the output estimates. This behaviour can be ascribed to the limited availability of training samples that characterize this scenario. Indeed, in this operative condition neural networks may be more affected by the problem of under-fitting/over-fitting. Moreover, the limited number of samples available for the tuning of the network connection weights makes this process instable and affected by the specific choice of the initial configuration (typically random).

A comparison of the results reported in Table 2.4 with those of the previous scenario (Table 2.3) indicates that for the *ActPass* data set the accuracies are worse to some extent, while for the *Scatterometer* data sets the accuracies are in general higher. In the first case, the trend is expected, due to the smaller number of training samples that characterize this operational scenario. In the second case the observed trend can be due to the high variability and high range of the roughness condition, close to the limit of validity of the IEM model.

The quantitative results in terms of figures of merit are confirmed by the scatter plots of estimated versus measured dielectric constant values (Figure 2.3 reports two examples). For the *ActPass*, both SVR and MLP follow the same trend, slightly underestimating low dielectric constant values and overestimating high dielectric constant values (with the exception of the *ActPass HHVV4.6-HV4.6* configuration). According to R^2 values, SVR estimates present a smaller variability around the linear trend line than the MLP ones. However, in general the performances are worse with respect to the previous scenario (Figure 2.2). For the *Scatterometer* data set we can observe a behavior similar to that analyzed in the previous subsection. However, here the highest R^2 values are reached by the SVR technique. Concerning the behaviour varying the input feature configurations, it is not possible to recognize a clear and common trend. SVR presents similar performances for the different configurations considered, with limited improvements corresponding to the cases where additional features are considered. The MLP NN technique shows stronger improvements in the estimation accuracy considering two frequencies and two acquisition angles (however, the accuracies are still lower than those of SVR), while a strong degrade of the estimation performance can be observed when two polarizations are used. This points out that the use of

diversified or additional input features can improve the estimation process, but to a less extent with respect to the first operational scenario. This effect can be due to: i) the noise associated to the additional features considered, that may introduce further ambiguity in the small reference set, and ii) the increased complexity of the estimation problem when the number of input features increases. SVR showed to be robust and effective also in this complex operational scenario.

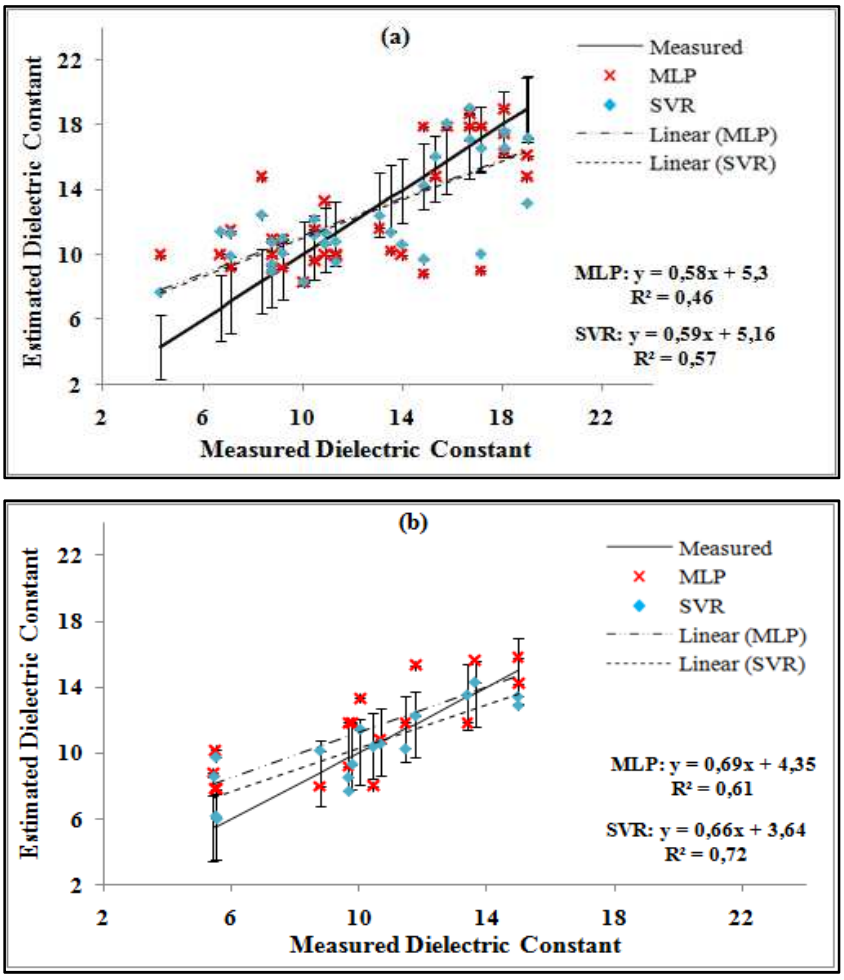


Figure 2.3. Scatter-plots of estimated versus measured dielectric constant values for Scenario 2 experiments: (a) ActPass HH4.6-H2.5 configuration; and (b) Scatterometer HHVV 23°/40° configuration. Error bars refer to the standard deviation of the measured values.

Table 2.4. Estimation accuracy of the MLP NN and SVR methods on ActPass and Scatterometer datasets for Scenario 2 experiments.

	***ActPass* Dataset**						***Scatterometer* Dataset**			
	HH4.6-H4.6		***HH4.6-H2.5***		***HHVV4.6-HV4.6***		***HHVV 23°***		***HHVV 23/40°***	
Full diel Range										
Method	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***
MLP NN	13.68	23.44%	9.51	21.25%	24.53	48.06%	9.23	20.03%	5.35	18.45%
SVR	7.25	17.99%	7.22	18.05%	7.39	17.86%	2.86	13.06%	2.76	12.93%
diel < 10										
Method	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***
MLP NN	15.45	25.31%	11.91	24.78%	37.75	52.34%	7.19	20.72%	6.71	24.21%
SVR	7.31	22.01%	8.43	23.68%	7.87	21.49%	3.99	18.37%	4.5	19.57%
diel >10										
Method	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***	***MSE***	***MRE***
MLP NN	12.91	22.63%	8.47	19.71%	18.45	46.11%	11.04	19.42%	4.15	13.33%
SVR	7.22	16.25%	6.71	15.61%	7.19	16.28%	1.51	8.34%	1.22	7.02%

2.5 Discussion and Conclusion

This chapter introduced the use of the ε-insensitive Support Vector Regression technique to address the challenging problem of soil moisture estimation at field/basin scale on bare soils from microwave remotely sensed data. More in detail, two distinct operational scenarios were investigated: i) the inversion of a theoretical based electromagnetic model; and ii) the inference of an empirical based mapping from few field measurements. The analysis of the results obtained confirmed the good theoretical features of the SVR method and its effectiveness for the considered estimation problem. In this framework, the SVR is a valid alternative to the more traditional MLP NN regression method since: i) it provided in general greater (or comparable) estimation accuracies; ii) it was more stable and robust to the outliers and noise present in the data; and iii) despite having similar computational complexity in the prediction phase, it was easier to handle (no architecture has to be defined) and faster to train. In particular, from the computational viewpoint SVR showed a significant reduction (on average, equal to about three orders of magnitude with respect to the MLP method) in the computation time required for the training phase. The difference can be explained by the fact that MLP requires the definition of the network architecture (typically addressed with a trial-and-error strategy) and has a learning strategy that should take into account the presence of many local minima in the cost function. In contrast, SVR takes advantage of the convex form of the cost function and simply requires the tuning of relatively few free parameters. All of these characteristics are of great importance in real operational scenarios and in the evaluation of effective soil moisture content processors for near real-time or large scale estimations.

Some difficulties were observed in reproducing the entire dynamic range of the output values, especially in the presence of high ambiguity in the input data. This aspect, which might be properly controlled by taking into account multiple criteria during the model-selection phase, requires further investigations. Other developments of this work will consider: 1) the study of the effectiveness of the SVR technique in estimating soil moisture when applied to distributed data (e.g., SAR images); and 2) the definition of a strategy for combining in an effective way the usually very few available ground collected reference samples with the electromagnetic model simulations for training the SVR.

Chapter 3

A Novel Multi-Objective Strategy for Tuning the Free Parameters of Non-Linear Regression Methods

This chapter deals with the tuning of the free parameters of non-linear regression methods for the retrieval of geo-/bio-physical variables from remote sensing data. We propose to address this task in the framework of the multi-objective optimization, in which a set of two or more metrics are jointly optimized according to the concept of Pareto optimality. The formulation of the proposed strategy is presented with the help of a synthetic data set. Then its effectiveness for tuning the free parameters of the SVR technique is assessed in the domain of soil moisture retrieval from microwave signals. The experimental results achieved on data sets associated with two different operational conditions are discussed and compared with those yield by the SVR tuned with a traditional strategy based on a single metric. This allows drawing indications about the usefulness and applicability of the proposed strategy in defining estimation systems for real application domains.

3.1 Introduction and Motivation

In Chapter 2, the effectiveness and usefulness of non-linear machine learning methods for the retrieval of geo-/bio-physical variables has been discussed and assessed. Generally, machine learning methods depend on a set of free parameters. With particular regard to the SVR formalism (since the promising theoretical properties and the good experimental results presented previously) they are the regularization parameter C, the width of the insensitive tube ε and the kernel function parameters. These parameters control the learning of the retrieval method and have a direct impact on its accuracy and generalization ability. The procedure for tuning the free parameters is often referred as *model-selection* and plays a crucial role in the definition of a retrieval algorithm. In real estimation problems this task is not straightforward. A large (hypothetically infinite) number of possible free parameters combinations exist for a given problem. At the same time, very few general guidelines are available. Several semi-heuristic rules have been proposed to relate the free parameters of

SVR with noise and statistics of input data (*Smola et al., 1998; Mattera and Haykin, 1999; Cherkassky and Yunquian, 2004; Chu et al., 2004*). For instance, Smola et al. (*1998*) suggested to use asymptotically optimal values of ε proportional to the noise variance in the training data. Mattera & Haykin (*1999*) proposed to select the *C* parameter equal to the range of target values. However, no agreement on the optimal setting of the SVR parameters can be found in the machine learning community. In addition, the available rules are often valid only under restrictive hypothesis (e.g., a certain loss function as for the approach proposed by Chu et al. (*2004*)) so that they require a specific expertise in handling machine learning methods and cannot be easily applicable in real geo/bio-physical variable estimation problems. This calls for the availability of robust, effective and practical strategies for addressing the model-selection issue.

In the remote sensing literature, the tuning of SVR parameters is usually addressed in terms of a numerical optimization problem, i.e., the parameters are tuned according to the minimization (or maximization) of a criterion function, evaluated on the available reference samples, by means of a numerical search algorithm. This requires the definition of: i) the criterion function and ii) the search algorithm.

Concerning the criterion function, some authors proposed the use of a theoretical bound. Chang and Lin (*2005*) generalized to regression problems the radius-margin and span bounds, which are well-known in the context of classification with Support Vector Machines (SVM). A similar approach was used by Moser and Serpico (*2009*) for tuning the parameters of a SVR and retrieving land and sea surface temperature from satellite data. Another used theoretical bound is the Vapnik measure, which represents an approximation of the upper bound of the prediction risk provided by the statistical learning theory (*Burges, 1998; Vapnik, 1998*). Since theoretical bounds are computed directly from the available reference samples without involving the training of the regression algorithm, their main advantage is the limited computational complexity. However, they represent just an approximation of a performance bound. Moreover, their physical meaning is not user-friendly and the definition typically involves complex analytical formulations difficult to be interpreted for non-expert in the SVR method.

Alternatively to theoretical bounds, the criterion function can be defined as an empirical metric, as it is usually done in geo/bio-physical variable estimation problems. The rationale is to train the regression technique with a given free parameter configuration on a subset of reference samples and then to assess its performance on another subset of independent samples (often referred as validation samples) according to the selected empirical metric. When few reference samples are available, a cross-validation scheme may be introduced to increase the representativeness of training and validation sets (*Cherkassky and Mulier, 1998*). Commonly adopted empirical metrics are the mean squared error *MSE* (which quantifies the accuracy of the estimates) and the determination coefficient R^2 (which expresses the amount of variability within the target values accounted for by the regression technique) (*Notarnicola*

et al., 2008; Pasolli et al., 2010). Empirical metrics are independent from the theoretical formulation of the regression technique adopted and thus can be exploited while tuning any regression method. Several studies indicate their effectiveness in comparison with theoretical bounds. For example, Smeets et al. (*2007*) compared the *MSE* empirical metric with several theoretical bounds showing good performance. The main drawback of empirical metrics is the intensive computational load (their use requires at each iteration the training and the application of the regression algorithm). However, this is typically compensated by their simplicity and ease of use.

Concerning the search algorithm, a typically used strategy is the grid search (*Scholkopf and Smola, 2001*). It consists in sampling the parameter search space (i.e., the SVR free parameters space) according to predefined steps, assessing the estimation performance in each node according to the selected criterion function and finally selecting, among the parameter configuration investigated, the one showing the best criterion function value. To reach good results, a dense sampling of the parameter space should be carried out. This results in a high computational load. Nevertheless, this strategy is usually adopted when dealing with geo/bio-physical variables retrieval and especially in combination with empirical criterion functions (*Zhan et al., 2003; Pasolli et al., 2010*). To reduce the computational burden and efficiently explore the parameter space, evolutionary optimization methods have been proposed. These methods start from a pool of initial random solutions that evolve iteratively till the convergence to the (near-) optimal parameter configuration. In this way a reduced number of trials are performed with respect to a grid-based approach, while the effects of local minima of the cost function within the search space are mitigated thanks to the random component that characterizes the search process. Examples are the particle swarm optimization (PSO) and the genetic algorithm (GA), which were successfully applied for addressing the parameter tuning of SVR with improvements in terms of both computational time and accuracy of the solution (*Zong et al., 2006; Lessmann et al., 2006*). As drawback, these algorithms typically require the user to set additional parameters, such as the number of iterations and the size of the searching population.

In this chapter, a novel method for tuning the free parameters of the SVR technique is presented. While model-selection strategies traditionally developed and used in the field of geo-/bio-physical variable retrieval exploit scalar criterion functions, i.e., a single empirical metric or a single theoretical bound, our method models the free parameters tuning process as a multi-objective optimization problem, in which the multi-objective function is made up of a set of *two* (or more) metrics. The rationale behind this idea is that different criteria (e.g., *MSE* and R^2) evaluate the goodness of a parameters configuration from different perspectives, and thus differently contribute to the task considered. When the available reference samples are affected by noise and ambiguity (it will be shown in the next section), relying on a single metric may affect the robustness and the stability of the model-selection process and different considered criteria may become slightly correlated one to each other. This implies that the

configuration of free parameters optimal according to one metric does not necessarily optimizes other metrics, thus limiting the possibility to effectively control the performance of the regression system and meet the application requirements. This issue is common in the retrieval of geo-/bio-physical variables from remotely sensed data, due to the influence of multiple target properties on remotely sensed signals, the effect of the atmosphere and other disturbing factors (*Luckman, 1998; Ulaby et al. 1979; Stamm et al., 1970*).

In the proposed method the multiple metrics considered for model-selection are *jointly* optimized according to the *Pareto optimality* (*Feldman, 1980*). This is a well assessed concept in economics and has been successfully exploited also in remote sensing for addressing various problems, such as the feature selection in hyper-spectral imagery (*Bruzzone and Persello, 2009*). In the Pareto optimality framework, the physical meaning of each metric is preserved and multiple solutions are obtained. Each identified solution represents an optimal trade-off between the metrics. The user has thus the possibility to easily and effectively select the solution (i.e., configuration of the free parameters) that best meets the specific requirements in terms of quality and accuracy of the estimates for the addressed retrieval problem. Moreover, the joint use of the metrics has the advantage of conveying and exploiting more and better the information contained in the available reference samples for driving the model-selection process. In our experiments the proposed method is applied to the challenging application domain of soil moisture retrieval from microwave remotely sensed data by using a SVR technique. However, it is general and can be used either for the estimation of other geo/bio-physical variables or with other regression methods. For comparison purposes, the results obtained by the proposed multi-objective model-selection strategy have been compared with those of a traditional model-selection based on a single criterion.

The rest of the chapter is organized into three sections. After a brief introduction of the notation and of the addressed problem with the help of a toy example, Section 3.2 describes the proposed novel model-selection approach. The experimental analysis carried out for assessing the effectiveness of the proposed methodology is illustrated in Section 3.3. Section 3.4 presents the discussion of the obtained results and draws the conclusion of the chapter.

3.2 Proposed Multi-Objective Model-Selection Strategy

3.2.1 Problem Formulation

The model-selection process for a regression technique (e.g., SVR) implies the concept of goodness of a parameter configuration $\boldsymbol{\omega} = [\omega_1, \omega_2, \dots, \omega_q] \in \Omega$, which may involve the sparseness or the robustness to the noise of a solution or the accuracy of the estimates provided by the regression algorithm. The relevance of each of these (or other) components depends on the specific application domain considered. Criterion functions used in standard

model-selection (e.g., theoretical bounds or empirical metrics) allow capturing and quantifying only one component of the goodness of a parameter configuration, ignoring the others. This represents an issue when noise and ambiguity affect the geo/bio-physical variable estimation problem.

In order to better clarify this concept, let us consider a simple 1-dimensional (i.e., with just one input feature) estimation problem in which the following mapping function has to be approximated:

$$y = f(x) = x^3 + 2 \tag{3.1}$$

where $x, y \in \Re$ are the input feature and the corresponding output target variable, respectively. Two sets $\boldsymbol{R}_{LOW}$ and $\boldsymbol{R}_{HIGH}$ of 100 reference samples (x_i, y_i) each are defined. The input feature values x_i are generated by randomly sampling the input feature space $[-1; 1]$, while the corresponding target values y_i are computed according to the mapping function in (3.1) and by further adding a Gaussian noise term $e \sim N(0, std_e)$. std_e is the noise standard deviation, which has been set to a small (0.05) or large (0.2) value (compared to the range of variability of the target values) in the case of the $\boldsymbol{R}_{LOW}$ and $\boldsymbol{R}_{HIGH}$ reference sets, respectively. The noise term models in a simple way all the disturbances and ambiguities that may affect a real geo-/bio-physical variable estimation problem, such as the influence of the atmosphere or the sensitivity of remotely sensed data to multiple target properties. A representation of the generated reference samples as well as of the mapping function in (3.1) is shown in Figure 3.1.

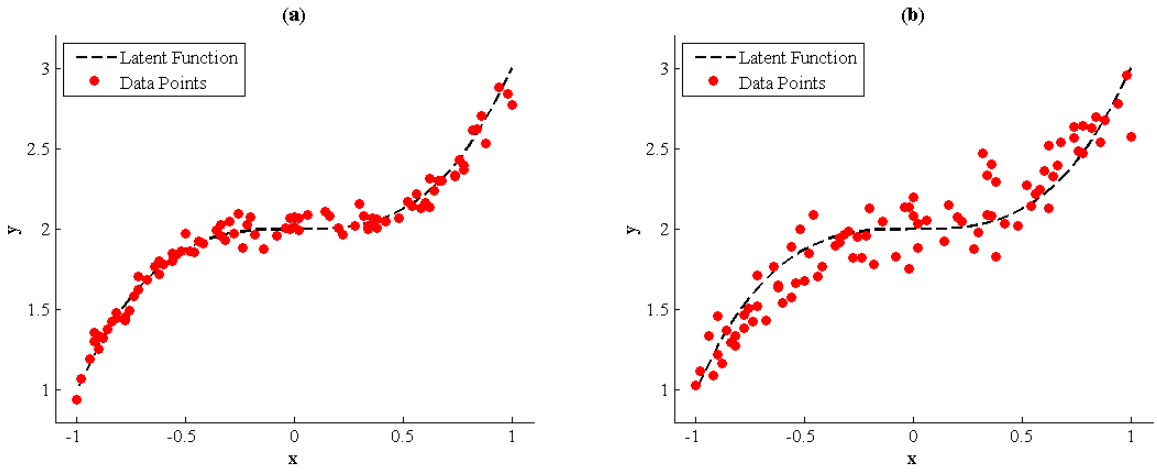


Figure 3.1. Reference samples generated for the toy example under the hypothesis of (a) low and (b) high noise affecting the data ($\boldsymbol{R}_{LOW}$ and $\boldsymbol{R}_{HIGH}$ data sets, respectively).

In this example, the mapping *f(.)* is approximated by means of the SVR technique. The tuning of the free parameters $\boldsymbol{\omega}$ is carried out according to a traditional model-selection strategy based on a single criterion function. Due to the limited number of samples available, a *t*–fold cross-validation procedure is implemented. The latter consists in randomly dividing the available reference samples into *t* subsets. Iteratively, *(t-1)* subsets are used as training while the remaining subset is used as validation. In this way, all the samples are used for both training and validating the algorithm, thus increasing the representativeness of training and validation sets and reducing the likelihood of biased results of the model-selection process. In the experiments *t* is set equal to 3, while the ranges for the free parameters are: $\varepsilon \in [10^{-3}, 10]$, $C \in [10^{-3}, 10^{3}]$ and $\gamma \in [10^{-3}, 10^{3}]$, where γ is the RBF kernel width. Two experiments are run in parallel using two empirical metrics: the mean squared error (*MSE)* and the statistical determination coefficient R^2. The search within the parameter space is carried out with the grid search strategy considering the following quantization for the search space: $[10^{-3}, 10^{-2.5}, 10^{-2}, \ldots, 10^{3}]$ for *C* and γ, and $[10^{-3}, 10^{-2.5}, 10^{-2}, \ldots, 10]$ for ε. At the end of the model-selection procedure, the trained and tuned SVR algorithm is run on a new set of samples, the test set, in order to evaluate the estimation performance on independent samples. The entire procedure is repeated for both the low noise and high noise reference data.

Table 3.1 summarizes the achieved results. Under the assumption of a low noise level in the reference data ($\boldsymbol{R_{LOW}}$ data set**)**, both criterion functions considered lead to very similar solutions of the model-selection problem. The SVR estimation performance evaluated on the test samples are almost identical. In case of high noise affecting the reference data ($\boldsymbol{R_{HIGH}}$ data set) rather different solutions of the model-selection problem are obtained. When *MSE* is optimized, the best performance in terms of accuracy is achieved (*MSE* = 0.032), while 84% of the variability inside the target values is explained (R^2 = 0.84). By optimizing the R^2 metric, the regression algorithm explain 90% of the variability inside test samples, but with a reduced accuracy (*MSE* = 0.18) and an over- and under-estimation behavior for low and high target values, respectively.

This simple example shows that a standard model-selection strategy based on a single criterion function, which is effective in relatively simple operational conditions, decreases its reliability when the complexity of the problem increases, i.e., when the noise and ambiguity inside the reference samples increase. Since this condition often occurs in geo/bio-physical variable estimation problems, it is crucial to develop a methodology being able to deal with this challenging condition.

Table 3.1. Performance achieved by the SVR technique when run on test samples under the hypothesis of low (R_{LOW} data set) and high (R_{HIGH} data set) noise affecting the data. Each solution is associated with the SVR parameters configuration that leads to the optimal MSE or R^2 values according to a single-objective grid-search strategy.

Criterion Function	**SVR parameters**			**SVR Estimation Performance**			
	γ	**C**	**ε**	***MSE***	***R²***	***Slope***	***Intercept***
R_{LOW} Data Set							
MSE	0.152	284.8	0.017	0.0054	0.98	0.86	0.27
R^2	0.152	284.8	0.001	0.0056	0.98	0.85	0.27
R_{HIGH} Data Set							
MSE	0.001	81.11	0.059	0.032	0.84	0.96	0.14
R^2	23.1	0.001	0.774	0.18	0.90	0.001	1.88

3.2.2 Proposed Multi-Objective Strategy

For a general geo/bio-physical variable estimation problem, let us assume that a set $\boldsymbol{Q} = \{Q_1(\boldsymbol{\omega}), Q_2(\boldsymbol{\omega}), \dots, Q_u(\boldsymbol{\omega})\}$ of u (where $u \geq 2$) criterion functions has to be optimized. Each criterion quantifies the goodness of a given configuration of free parameters $\boldsymbol{\omega}$ from a different (and sometimes competing) perspective. As an example, the set $\boldsymbol{Q}$ may consist of the *MSE*, the R^2, the *slope* of the linear trend line and other criteria. In order to face the model-selection issue based on all the criteria in $\boldsymbol{Q}$, a simple strategy could be the definition of a new scalar metric $E(\boldsymbol{\omega})$ combining the desired set $\boldsymbol{Q}$ of criterion functions through a weighted average:

$$E(\boldsymbol{\omega}) = \sum_{i=1}^{u} c_i Q_i(\boldsymbol{\omega}) \qquad (3.2)$$

where $c_1, c_2, \dots, c_u$ are the weights of the average function. The configuration $\boldsymbol{\omega}^*$ that optimizes the expression in (3.2) will represent the solution of the model-selection problem. Despite its simplicity, this formulation has an important drawback, i.e., the definition of the weights c_i of the average function, which should be done by the user. This task is very critical, since i) it should be carried out empirically; ii) it significantly affects the final results; and iii) the criteria involved may have intrinsic different scales. Another important limitation is the fact that the physical information conveyed by the resulting new scalar metric does not reflect the original metrics and becomes difficult to interpret.

To avoid these problems one can express the optimization problem according to a *u'*-dimensional multi-objective function $\boldsymbol{Q}'(\boldsymbol{\omega})$ which is made up of the $u' \leq u$ objectives $Q_1(\boldsymbol{\omega}), Q_2(\boldsymbol{\omega}), \dots, Q_{u'}(\boldsymbol{\omega})$ that represent the set of adopted criterion functions:

$$\boldsymbol{Q}'(\boldsymbol{\omega}) = [Q_1(\boldsymbol{\omega}), Q_2(\boldsymbol{\omega}), \dots, Q_{u'}(\boldsymbol{\omega})] \quad (3.3)$$

All the metrics of $\boldsymbol{Q}'(\boldsymbol{\omega})$ are jointly optimized and are considered equally important. Thus the multi-objective optimization problem can be formulated as follows:

$$\underset{\boldsymbol{\omega}}{argmin} \{\boldsymbol{Q}'(\boldsymbol{\omega})\} \text{ subject to } \boldsymbol{\omega} = [\omega_1, \omega_2, \dots, \omega_q] \in \Omega \quad (3.4)$$

This problem is characterized by a vector valued objective function; thus it cannot be solved deriving a single optimal solution as in scalar optimization problems. Instead, a set $\boldsymbol{P}$ of optimal solutions can be obtained following the concept of *Pareto dominance* (Feldman, 1980). A configuration $\boldsymbol{\omega}^*$ is said to be *Pareto optimal* if it is not dominated by any other configuration in the search space Ω, i.e., there is no other $\boldsymbol{\omega}$ such that:

$$Q_i(\boldsymbol{\omega}) \leq Q_i(\boldsymbol{\omega}^*) \; \forall i = 1,2, \dots, u \quad (3.5)$$

and

$$Q_i(\boldsymbol{\omega}) < Q_i(\boldsymbol{\omega}^*) \; for \; at \; least \; an \; i = 1,2, \dots, u \quad (3.6)$$

In other words, $\boldsymbol{\omega}^*$ is *Pareto* optimal if there exists no other configuration of parameters that would decrease a criterion function without increasing at least another one at the same time. The set $\boldsymbol{P}$ of all optimal solutions is called Pareto-optimal set, while the plot of the objective function of all the solutions in the Pareto set is often referred as Pareto front $\boldsymbol{PF} = \{\boldsymbol{Q}'(\boldsymbol{\omega}) | \boldsymbol{\omega} \in \boldsymbol{P}\}$.

Because of the dimensionality of the search space and the complexity of the multi-objective optimization problem, an exhaustive search of the set $\boldsymbol{P}$ of optimal solutions is typically unfeasible. Instead of identifying the true set of Pareto optimal solutions, one may estimate a set $\boldsymbol{P}^*$ of non-dominated solutions with values of the objective functions as close as possible to the Pareto front. This task is typically handled by using evolutionary algorithms, which allow an efficient and effective investigation of the parameter space also in the complex case of the multi-objective optimization. Various evolutionary approaches have been proposed for this task in the last few years in the literature (*Fonseca and Fleming, 1998; Deb, 2001*).

The main advantage of the joint *u'*-dimensional optimization approach consists in the fact that it avoids aggregating metrics thus capturing and conveying different and heterogeneous information (in our case concerning the goodness of a parameter configuration) into a single measure. Thanks to the multi-dimensional formulation of the optimization problem, it preserves the physical meaning of each metric and allows easily and effectively identifying various possible optimal trade-offs among different criterion functions. The final selection of the optimal solution to the model-selection problem is demanded to the user, who can identify

the best trade-off among the considered criteria on the basis of the specific requirements of the considered retrieval problem.

For the sake of clarity, let us consider again the toy example presented in Section 3.2.1. The experimental set-up remains the same, with the exception of the model-selection for the SVR technique, which is addressed by jointly optimizing the two considered empirical criterion functions (*MSE* and the R^2) according to the concept of Pareto optimality. For implementation requirements, the optimization was carried out on *MSE* and $(1-R^2)$ values, so that the ideal optimal value is 0 for both metrics. The subset of dominated (non-optimal) solutions investigated during the evolutionary process as well as the estimated Pareto front are shown in Figure 3.2. The shape of the front clearly shows the competing behavior between the two considered criteria: the minimization of the first index implies an increase of the second one and vice-versa. The extremes of the front (indicated as Solution 1 and Solution 2) are associated with the highest *MSE* and R^2 values, respectively. It is worth noting that these two optimal (in the Pareto sense) solutions are not necessarily found by traditional model-selection strategies based on a single metric. As shown in the lower-right part of the plot, multiple (sub-optima) solutions show nearly the same R^2 performance (around 0.9), but rather different *MSE* accuracy (ranging from 0.11 to 0.17). Looking at the R^2 performance only, these solutions appear equivalent. Thus, an automatic selection of the solution showing the highest value of R^2 while minimizing the *MSE* is possible only when a joint optimization of both the criteria is carried out. Depending on the application domain considered, Solutions 1 or 2 may not represent the best choice for meeting the user requirements. Our model-selection approach provides the user also with different solutions, the remaining set of non-dominated solutions lying on the Pareto front, any of them representing a possible optimal trade-off among the optimized criteria. For example, the application could impose an upper bound on the estimation accuracy (such as $MSE \leq 0.05$). Looking at the Pareto front, the user has the possibility to select the SVR parameter configuration that meets this requirement while maximizes at the same time the R^2 metric (Solution 3). This kind of solutions to the model-selection process can be systematically identified only if a multi-objective optimization that preserves the physical meaning of the criterion functions is carried out.

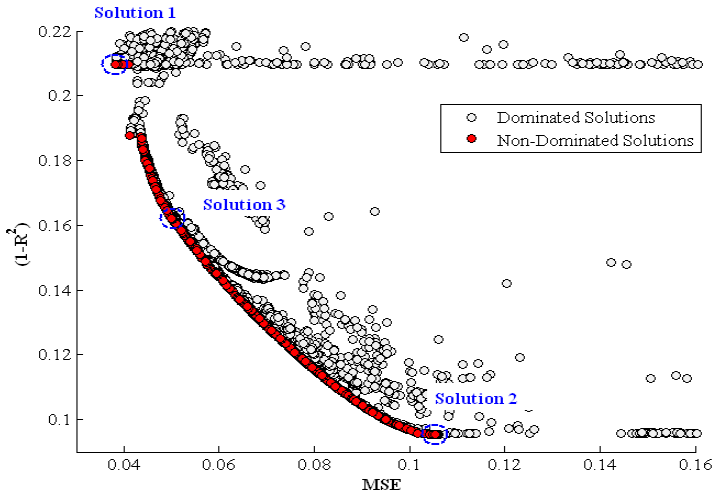


Figure 3.2. Estimated Pareto front using MSE and R^2 metrics under the hypothesis of high noise level affecting the toy data ($\boldsymbol{R_{HIGH}}$ data set).

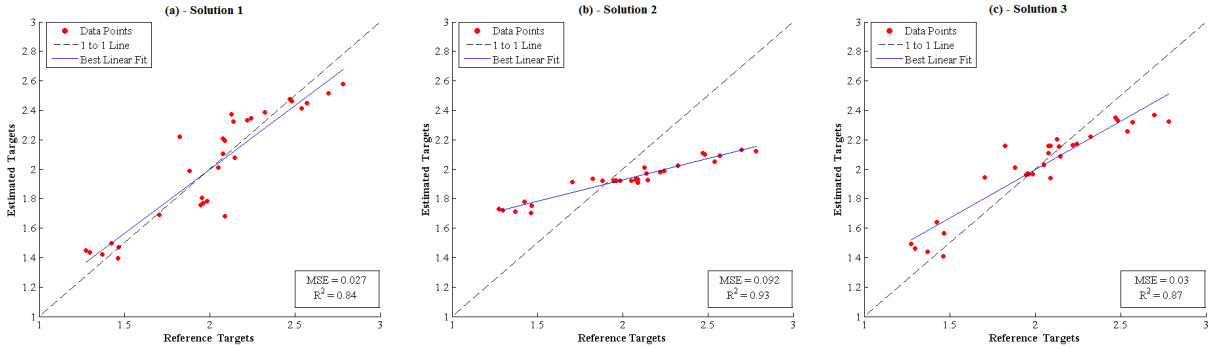


Figure 3.3. Scatter plots of estimated versus measured test target values associated with three possible Pareto-optimal solutions taken from the estimated Pareto Front of Figure 3.2: (a) Solution 1, showing the lowest MSE; (b) Solution 2, with the highest R^2; and (c) Solution 3, a possible trade-off among the two previous solutions.

3.3 Experimental Analysis

3.3.1 Data Set Description

For assessing the effectiveness of the presented model-selection approach, experiments on the estimation of soil moisture content from microwave remote sensing data with the SVR technique are carried out. As pointed out in the previous chapters, this application domain is particularly challenging due to the non-linearity of the relationship between the microwave signals and the target parameter as well as the sensitivity of microwave signals to different

target properties, such as surface roughness and vegetation. These features make this domain suitable to investigate the effectiveness of the proposed model-selection strategy.

Two data sets are considered in our analysis. They present quite heterogeneous characteristics (in terms of data set composition, signal properties, investigated area, etc.) and are associated to two different operative scenarios (Table 3.2). The first data set, referred as *Scatterometer* data set (the same considered for the analysis presented in Chapter 2), consists of 17 field soil moisture content measurements paired with the corresponding backscattering coefficients at HH polarizations and 23°/40° incidence angles acquired with a C-band field scatterometer. (*Mattia et al. 2003*). In addition to the field samples, a set of three hundred simulated samples were generated according to the Integral Equation Model IEM (*Fung et al., 1992*) run under the same conditions (in terms of signal properties and soil characteristics) of the field measurements. Please note that for this experiment the generation of the simulated samples do not include too wide ranges for the model input variables, as done for the experimental analysis in Chapter 2. Since simulated samples are used for training/tuning the regression system, this allowed reducing the computational burden of the experimental analysis, at the price of a reduced generality of the inferred relationship. For the aim of this analysis, this issue is not relevant. The assessment of the estimation performance is done on the experimental field samples.

The second data set (referred as *SAR* data set) is representative of a completely different operational scenario. It consists of dielectric constant measurements acquired in June and July 2010 in 75 heterogeneous vegetated sites (meadows and pastures) in a small Alpine valley located in Alto Adige/South Tyrol, Italy. The average moisture content of the area is around 15%-20%, but wet patterns with higher soil moisture (up to 50%) can be observed mainly due to irrigation in managed meadows and small superficial rivers on the valley side. For a more detailed description of the measurement procedure and the study area, we refer the reader to Chapter 5. Remotely sensed microwave data were acquired contemporary to the field campaigns and consisted of two C-band RADARSAT2 SAR images. After the pre-processing of the data, backscattering coefficients were extracted and associated to the ground measurements. The data set was completed by including features extracted from ancillary data (this issue will be described thoroughly in the next Chapters). No theoretical model simulations were considered in the case of the *SAR* data set, in order to assess the effectiveness of the proposed model-selection method in case of limited availability of reference samples. This realistic scenario is particularly challenging since the complexity of effectively training/tuning a regression method is further increased. Part of the available field samples (i.e., 57 samples) were assigned to the reference set used for training/tuning purposes. The remaining 18 samples were used to quantitatively assess the estimation performance. The sampling was carried out manually to ensure that both training/validation and test sets were well representative of the whole range of target variable values.

Table 3.2. Main characteristics of the two data sets considered in the experimental analysis.

Data Set	***Scatterometer***	***SAR***
Study area	Bare agricultural fields	Alpine meadows and pastures
Microwave Sensor	Field Scatterometer	Fully Polarimetric Satellite SAR
# Samples	17 field samples (test) + 300 simulated (IEM) samples (training/validation with 3-fold CV)	75 field samples (18 test + 57 training/validation with 3-fold CV)
Input Features	Backscattering coefficients HH polarization at 23°/40° incidence angles	Backscattering coefficients HH/HV/VV polarizations + ancillary information (NDVI, land-use class, local incidence angle, local height)
Target value range (dielectric constant)	$5 \div 15$	$3 \div 23$

3.3.2 Design of Experiments

For all the experiments, a Gaussian RBF kernel was adopted for the SVR method. The ranges of variability of the free parameters to tune were set to $[10^{-3}; 10^{3}]$ for the kernel width γ, $[10^{-3}; 10^{3}]$ for the regularization parameter *C* and $[10^{-3}; 10]$ for the width of the insensitive tube ε. For the implementation of the SVR algorithm we considered the LibSVM library, freely available online (*Chang et al., 2011*).

Concerning the multi-objective model-selection process, four empirical metrics were considered: the mean squared error (*MSE* ϵ [0,∞] ideally *MSE* = 0); the statistical determination coefficient (R^2 ϵ [0,1], ideally R^2 = 1); the *slope* (*slope* ϵ [-∞,∞], ideally *slope* = 1) and the *intercept* (*intercept* ϵ [-∞,∞], ideally *intercept* = 1) of the linear fit between estimated and measured target values (which provides useful information about under- or over-estimation trends). To solve the multi-objective optimization problem and find out Pareto-optimal solutions, a variation of the Non Dominated Sorting Genetic Algorithm II (NSGA-II) was implemented. This algorithm showed promising performance both in terms of accuracy in the approximation of the Pareto front and computational time in several optimization problems and in comparison with other search algorithms (*Deb et al. 2002*). For this reasons it has been selected for our experiments. We refer the reader to this paper for details on the algorithm. In order to take into account the wide range of variability of the free parameters, in our experiments the initialization of the chromosomes has been modified by means of an exponential random function. The population size was set equal to 1000 and the maximum number of iterations to 10. These settings have been selected after preliminary experiments aiming at identifying the most suitable trade-off among computational complexity and stability of the results. All the criterion functions were expressed in such a

way their ideal value was 0 (i.e., $R^2 \rightarrow (1 - R^2)$, $slope \rightarrow (1 - |slope|)$, $intercept \rightarrow |intercept|$, respectively).

In order to simulate and investigate different possible operative scenarios, the model-selection process has been addressed with several combinations of the adopted metrics as objectives of the proposed multi-objective strategy. For comparison purposes, a traditional mono-objective model-selection procedure has been also implemented. More in detail, the four empirical metrics considered for the proposed multi-objective model-selection were exploited independently to drive the model-selection in combination with the well assessed grid-search strategy. The quantization of the search space was defined according to a logarithmic strategy, which divided the search space into 4913 combinations. In each experiment, SVR was trained/tuned with the available reference samples in order to infer the mapping between input features and target variable values. For increasing the representativeness and stability of training and validation sets, a 3-fold cross-validation procedure has been adopted for both data sets. After the training/tuning phase, the SVR with the selected parameter configuration was run on the independent test samples, in order to quantitatively assess the estimation performance. For the sake of brevity, in this analysis we focus the attention on two sets of experiments: Experiment 1, in which all the four criteria are included in the optimization problem; and Experiment 2, where only two criteria are exploited for the multi-objective model-selection.

3.3.3 Experiment 1: Model-Selection Using All the Criterion Functions

This first set of experiments is aimed at investigating the effectiveness and usefulness of the proposed multi-objective model-selection strategy in the complex scenario of numerous (in this case four) criteria to optimize. Among all Pareto optimal solutions identified by the proposed algorithm, we selected (as explanatory example) those lying at the extremes of the 4-dimensional Pareto front. Each extreme is characterized by the best value found for one of the four metrics while minimizing the others. These solutions were compared with those achieved by the mono-objective model-selection. Tables 3.3 and 3.4 summarize the results of these experiments, showing the SVR parameter configurations found for each solution and the performance (in terms of the four empirical metrics, the computational time taken on a Intel® Core™ 2 duo 2.53GHz CPU and 3GB RAM personal computer and the number of SVR calls) achieved on test samples for both the *Scatterometer* and *SAR* data sets, respectively.

Concerning the *Scatterometer* data set, a first analysis of the results clearly shows a competing behavior among all the metrics considered in the analysis. Indeed, the optimization of each metric leads to the worsening of the others. This result stands for both the standard grid search and the proposed multi-objective approach. For example, the parameter configurations selected for minimizing the error (solutions 1a and 1e) are associated with both an increased spread (i.e., lower R^2) and a slightly flattening of the target value dynamic (as indicated by the *slope* and *intercept* values). Analogous observations can be drawn for the

other selected solutions. This behavior is motivated by the intrinsic ambiguity in the backscattering coefficients due to the quite heterogeneous conditions in terms of soil roughness that characterize the *Scatterometer* data set. Table 3.3 also shows that the proposed model-selection strategy outperforms the standard mono-objective approach, despite the slightly higher computational time required. Beside the slightly better performance for the metric optimized in each solution (e.g., the *MSE* in solutions 1a and 1e), which is an expected result due to the use of the more effective genetic algorithm for addressing the optimization process, our method shows better performance with respect to the mono-objective model-selection also looking at the other metrics. For example, very similar values for the *intercept* metric are achieved by both model-selection methods in solution 1d and solution 1h (-0.7 and -0.64 for the grid search and the proposed strategy, respectively). However, the solution found by the multi-objective approach shows much better performance in terms of all the other metrics with respect to the corresponding solution of the traditional mono-objective method (6.55 vs. 10.35 for *MSE*, 0.6 vs. 0.49 for R^2 and 0.93 vs. 0.79 for *slope*, respectively). This aspect points out the great advantage of the joint optimization of multiple criteria, which allows effectively and easily optimizing all the metrics involved in the model-selection process at one time.

With respect to the *SAR* data set (Table 3.4), a competing behavior between metrics measuring the accuracy and those indicative of the dynamic of the target values is evident for both model-selection strategies. This suggests that a rather high ambiguity characterizes the backscattering coefficients of the *SAR* data set. The ambiguity in this case is explained by the heterogeneous land-cover and topographic conditions typical of the Alpine environment, which persist despite the use of ancillary data as additional input features. A one-to-one comparison between the solutions of the standard mono-objective strategy and those selected for the proposed strategy points out again that our method leads to a slight improvement in terms of the target metric and to rather better performance in terms of the other (competing) metrics. This result is particularly evident for the solution 2g, which shows a slightly higher value of the *slope* metric (0.83) with respect to the corresponding solution of the standard approach (0.82) while exhibits a sharply lower mean error (*MSE* = 9.57 vs. *MSE* = 11.22, respectively). Such a behavior confirms again the effectiveness of the proposed multi-objective approach with respect to a traditional model-selection strategy that relies on a single metric.

Table 3.3. SVR parameters configurations and corresponding estimation performance evaluated on test samples (Experiments 1, Scatterometer data set): solutions 1(a) – (d) are associated with the optimization of each metric separately trough a grid-search algorithm; solutions 1(e)-(h) are identified by the proposed multi-objective approach jointly using all the metrics and choosing the extremes of the estimated Pareto front. The metrics showing the best values are highlighted.

Solution ID	**SVR parameters**			**SVR Estimation Performance**				**Computation Time**
	γ	**C**	**ε**	***MSE***	***R^2***	***Slope***	***Intercept***	
Mono-Objective Model-Selection								
1a	0.03	1.0	3.16	*3.23*	0.64	0.66	2.85	2232[sec] (4913 SVR calls)
1b	0.03	74.99	3.16	10.82	*0.69*	1.38	-5.42	
1c	0.18	31.6	0.32	7.97	0.56	*1.10*	-0.76	
1d	0.001	31.6	3.16	10.35	0.49	0.79	*-0.70*	
Proposed Multi-Objective Model-Selection								
1e	0.059	1.73	2.89	*2.93*	0.65	0.61	3.68	2425 [sec] (~4517 SVR calls)
1f	0.068	0.54	4.08	4.89	*0.7*	0.24	7.39	
1g	0.077	5.28	0.45	7.4	0.54	*1.03*	-0.65	
1h	0.038	0.78	1.37	6.55	0.6	0.93	*-0.64*	

Table 3.4. SVR parameters configurations and corresponding estimation performance evaluated on test samples (Experiments 1, SAR data set): solutions 2(a) – (d) are associated with the optimization of each metric separately trough a grid-search algorithm; solutions 2(e)-(h) are identified by the proposed multi-objective approach jointly using all the metrics and choosing the extremes of the estimated Pareto front. The metrics showing the best values are highlighted.

Solution ID	**SVR parameters**			**SVR Estimation Performance**				**Computation Time**
	γ	**C**	**ε**	***MSE***	***R^2***	***Slope***	***Intercept***	
Mono-Objective Model-Selection								
1a	0.18	31.6	0.01	*8.81*	*0.79*	0.73	1.85	1842[sec] (4913 SVR calls)
1b	0.18	31.6	0.01	*8.81*	*0.79*	0.73	1.85	
1c	0.18	1000	0.001	11.22	0.73	*0.82*	*1.1*	
1d	0.18	1000	0.001	11.22	0.73	*0.82*	*1.1*	
Proposed Multi-Objective Model-Selection								
1e	0.23	19.02	0.001	*8.57*	*0.8*	0.73	1.9	1867 [sec] (~4423 SVR calls)
1f	0.23	19.02	0.001	*8.57*	*0.8*	0.73	1.9	
1g	0.13	991.7	0.21	9.57	0.76	*0.83*	1.06	
1h	0.2	790.1	0.16	11.06	0.73	0.83	*1.01*	

3.3.4 Experiment 2: Model-Selection Using Two Criterion Function

In this second set of experiments we investigate the operative condition in which two competing metrics are considered in the model-selection problem, in order to study in greater detail the behavior of the proposed approach. This could be the case in which end-users give priority to some metrics with respect to others.

In the case of the *Scatterometer* data set, Experiments 1 showed that almost all the considered criteria present a competing behavior one to each other. For this reason, in this experiment we carried out several trials considering different possible pairs of objective functions to be optimized. For the sake of brevity, we report the example in which the mean squared error (*MSE*) and the determination coefficient (R^2) are used in the multi-objective model-selection process. The estimated Pareto front is shown in Figure 3.4. As expected, the graph clearly points out the competing behavior among the two considered metrics. A first subset of optimal solutions is associated with performance of the SVR algorithm quite similar in terms of *MSE*, yet relatively different in terms of determination coefficient. Similarly, a second subset of solutions determines similar performance of SVR in terms of R^2 values, yet significantly different behavior in terms of *MSE*. These two subsets meet in the knee of the estimated Pareto front. Thanks to this visual inspection of the front, the user can easily identify and select the optimal (in the Pareto sense) configuration of parameters that addresses the requirements of the considered application. For example, instead of selecting the solutions at the extremes of the Pareto front (Solutions 1 and 2), which lead to the best performance in terms of *MSE* and R^2, respectively, the user could choose a solution near the knee of the front (e.g., solution 3). A visual analysis of the scatter plots associated with the three identified configurations of parameters (Figure 3.5) confirms that solution 3 results in a good trade-off among the two considered metrics.

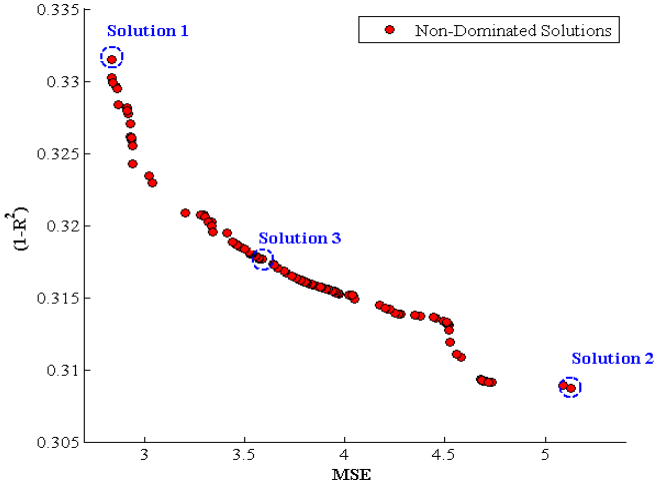


Figure 3.4. Estimated Pareto front using MSE and R^2 metrics (Experiments 2, Scatterometer data set).

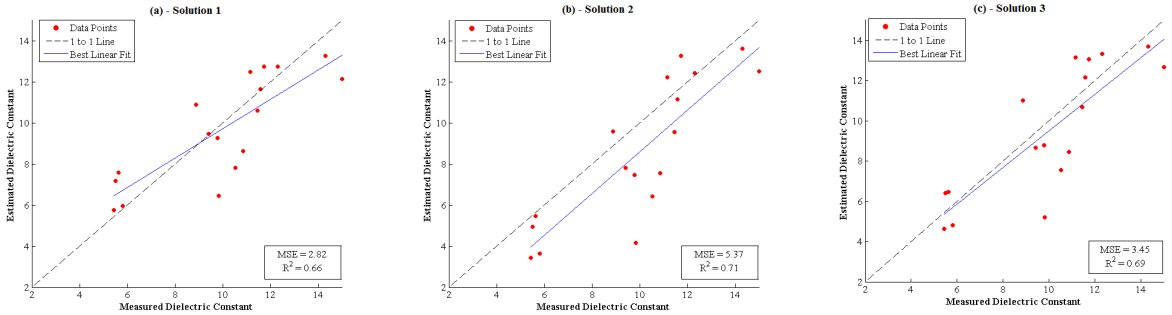


Figure 3.5. Scatter-plots of estimated versus measured test dielectric constant values associated with three possible optimal solutions taken from the estimated Pareto Front of Figure 3.4: (a) Solution 1, showing the lowest MSE; (b) Solution 2, with the highest R^2; and (c) Solution 3, a possible trade-off among the two previous solutions (Experiments 2, Scatterometer data set).

Concerning the *SAR* data set, the analysis in Experiments 1 has shown a clear competing behavior among the metrics quantifying the error and those expressing the under- and over-estimation tendency of the estimations. Here we discuss the example in which the *MSE* and the *slope* of the linear tendency line are involved in the optimization process. However, similar results are obtained with other combinations of competing metrics used for the optimization (e.g., *MSE* and *intercept*, etc.). The estimated Pareto front is shown in Figure3.6. Also in this case three different regions can be identified: i) the subset of solutions which result in similar accuracy but different capability to reproduce the dynamic of the target values; ii) the subset of solutions which result in similar capability to reproduce the dynamic of the target values but different accuracy; and iii) the subset of solutions lying near the knee of the front, which are associated with performance quite close to the optimal one according to both the metrics. For our illustrative example, we selected three possible SVR parameter configurations belonging to each subset of solutions (Figure 3.7). Again, the achieved results suggest that with the help of a visual inspection of the estimated front, the desired trade-off among the considered criterion functions can be easily and effectively identified. These results are in agreement with those obtained in other experiments (not reported here for the sake of brevity) carried out with other metric configurations.

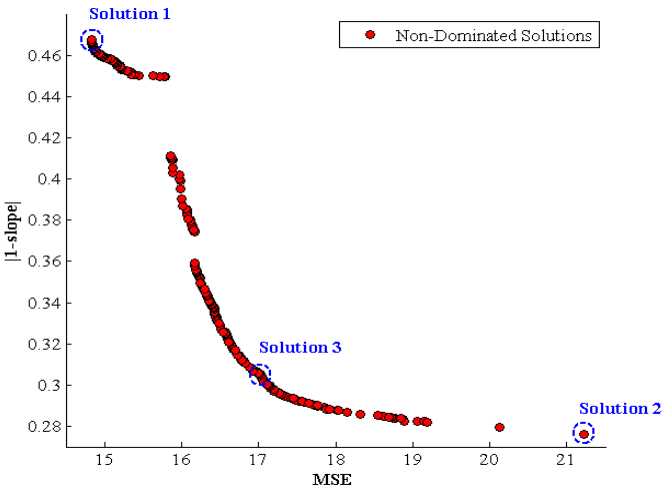


Figure 3.6. Estimated Pareto front using MSE and slope metrics (Experiments 2, SAR data set).

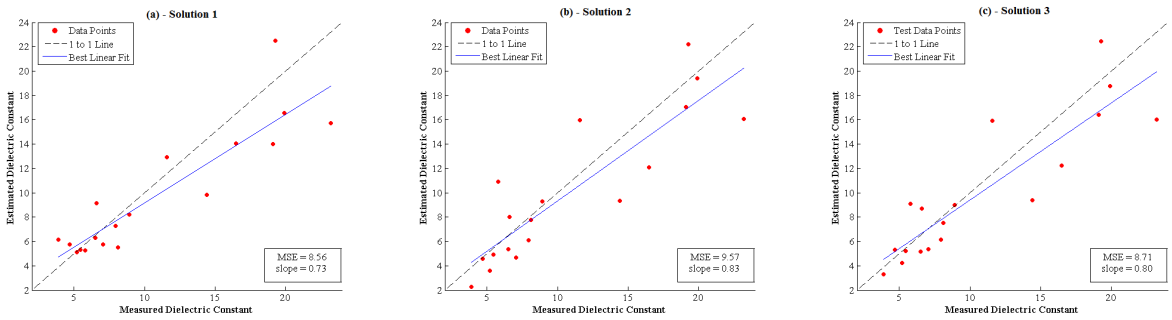


Figure 3.7. Scatter plots of estimated versus measured test dielectric constant values associated with three possible optimal solutions taken from the estimated Pareto Front of Figure 3.6: (a) Solution 1, showing the lowest MSE; (b) Solution 2, with the best slope; and (c) Solution 3, a possible trade-off among the two previous solutions (Experiments 2, SAR data set).

3.4 Discussion and Conclusion

In this chapter a novel approach has been presented which can tune the free parameters of the Support Vector Regression technique to be applied in the estimation of geo-/bio-physical variables from remotely sensed data. The model-selection is addressed as a multi-objective optimization problem, in which the multi-objective function is made up of a set of two (or more) metrics. The metrics are jointly optimized according to the concept of Pareto optimality. In this framework, the physical meaning of each metric is preserved and multiple solutions (i.e., configurations of SVR parameters) are obtained. Each solution leads to a different optimal trade-off between the metrics considered.

The proposed method has been assessed in the specific and challenging application domain of soil moisture retrieval from microwave remotely sensed data. Several experiments were carried out with two data sets representative for two rather different operative conditions, i.e., the retrieval of soil moisture in bare agricultural fields with the availability of numerous and well representative (simulated) reference samples and the retrieval of soil moisture in vegetated mountain pastures and meadows with limited availability of reference samples. Despite the different operative conditions associated with the two data sets considered, the proposed method achieved in both cases promising results and outperformed the traditional mono-objective model-selection. This suggests that the proposed approach is effective for facing the model-selection issue, as: 1) it is characterized by an intrinsic increased robustness with respect to traditional approaches, since each solution obtained is the result of the optimization of multiple metrics; 2) it allows one deriving and choosing effectively trade-off solutions that jointly optimize the metrics selected; and 3) it provides the user with a simple tool for the selection of the parameter configuration that best meets the requirements and the constraints of the specific retrieval problem considered. Concerning the last point, it is worth noting that the possibility to visually individuate the solution that best meets the user requirements by means of the inspection of the Pareto front is feasible only when two (as in our experiments) or three objectives are considered. When more objectives have to be optimized, the method has shown to work properly, despite more effort is required to the end-user to search among all the Pareto optimal solutions the one (or those ones) that satisfies the application constraints. An interesting development of this work could be the definition and assessment of (semi-) automatic strategies for the selection of the optimal trade-off given the requirements of the specific application domain considered, in order to reduce the effort required to the end-user in this phase when more than three metrics are considered.

From the computational viewpoint, the proposed multi-objective strategy only showed a slight increase in the time required for the processing with respect to the mono-objective approach run in the same experimental conditions. This is due to roughly the same number of calls to the SVR algorithm. Actually, the majority of the computational effort was required for running the SVR algorithm and computing the empirical metric values. Such a result points out that the advantages of the multi-objective model-selection discussed previously can be achieved without a dramatic increase of the computational overhead with respect to traditional mono-objective strategies. For the reduction of the computational time, which is usually an issue in real application problems, one may act on the population size and the number of iterations of the genetic search algorithm. Another interesting possibility could be the use of multi-core system architectures, in order to exploit the intrinsic parallel nature of the model-selection process.

It is worth noting that, even though the proposed model-selection strategy has been experimentally assessed in the application domain of soil moisture retrieval from microwave

remotely sensed data with the Support Vector Regression technique, it is general and can be integrated in the design of any system for geo-/bio-physical variable estimation based on a regression method that requires the tuning of free parameters. Moreover, since no constraints are given on the choice of the considered metrics, the user can define the model-selection process in the way that best fits the requirements of quality and accuracy of the addressed application problem.

A further developments of this work could be the inclusion in the model-selection process not only of continuous valued free parameters (as done in this work) but also of discrete valued parameters, such as the choice of the kernel function type in the case of kernel estimation methods. Moreover, the proposed multi-objective approach could be extended to address other crucial steps of the geo-/bio-physical variable estimation process, such as the choice of the features to give as input to the regression technique. All the variable terms of a regression system have a strong impact on the quality and accuracy of the final estimates. Thus their tuning may take advantage from the proposed multi-objective strategy to further improve the robustness and reliability of the estimation. Finally, another important development of the analysis carried out would be a further validation of the proposed method in different application domains (e.g., the estimation of other geo/bio-physical variables from remote sensing data), also by including different sets of empirical metrics and (more computationally efficient) theoretical bounds to optimize.

Chapter 4

A Novel Hybrid Approach to the Retrieval of Geo-/Bio-Physical Variables from Remote Sensing Data

This chapter presents a novel hybrid approach to the estimation of geo-/bio-physical variables from remotely sensed data. This approach integrates theoretical analytical models and empirical relationships based on field reference samples to increase the reliability and the accuracy of the estimation. The estimation process is modeled by two terms: the first one expresses the relationship between the input features and the target geo-/bio-physical variable according a theoretical model based on the physics of the considered problem. The second term corrects the deviation between theoretical model estimates and true target values according to an empirical data-driven model. The latter is derived by exploiting the available (typically few) field reference samples. In this way the robustness and generality of theoretical model based estimates, which stem from the rigorous theoretical foundation, is preserved, while the bias and imprecision (due to simplifications in the analytical formulations of the model with respect to the real estimation process) are reduced. The experimental analysis is carried out in the challengin field of soil moisture content retrieval from microwave remotely sensed data. Two different correction strategies are investigated. The results achieved with two data sets point out the effectiveness and potential of the proposed hybrid estimation approach.

4.1 Introduction and Motivation

As discussed and investigated in the previous chapters of this thesis, the retrieval of geo-/bio-physical variables from remote sensing data can be addressed following two main approaches: i) the derivation of empirical data-driven relationships; and ii) the inversion of physical based analytical models. Both approaches present strengths but also limitations.

The first approach relies on the availability of a set of reference samples, i.e., couples of *in-situ* measurements of the desired target variable associated with the corresponding measurements of the remote sensor. These samples are exploited for directly deriving an empirical mapping between remotely sensed data and target geo-/bio-physical variable. Often, statistical regression techniques in combination with parametric (linear, logarithmic or polynomial) functions are exploited. Then the relationship identified is extended to the whole satellite image (*Colombo et al. 2003; Heiskanen, 2006; Teodoro et al., 2007*). When the complexity of the retrieval problem increases, analytically more sophisticated parametric relationships (*O'Relley et al.; 1998*) or even advanced non-parametric regression methods (*Bruzzone and Melgani, 2005; Moser and Serpico, 2009*) are used. Empirical relationships are appealing since they are typically fast to derive and quite accurate. Moreover, they abstract complex physical phenomena to a higher level, which can be easily addressed by non-expert without a specific background in the field. The main drawback is the need of a sufficient number of possibly well representative reference samples. This condition is not verified in many practical geo-/bio-physical variable estimation problems, such as near real time monitoring. Indeed, the collection of ground measurements requires the human intervention and is usually a time-consuming and expensive task. Moreover, errors may occur for various reasons during the measurement process, thus negatively affecting the quality and quantity of reference samples available. Another important issue is the fact that empirical relationships are typically site and sensor dependent, since derived from samples collected under specific operational conditions. This limits the possibility to extend their use to different areas and different remote sensing systems, since they remain valid only under the conditions in which reference samples have been collected (*Colombo et al., 2003; Meroni et al., 2004*).

The second approach demands the definition of the desired mapping function to an analytical electromagnetic model. Such models are based on a solid physical based description of the mechanisms involving the interaction of the electromagnetic radiation and the target object of interest. In the direct operational way, they simulate the response of a target object as function of: 1) the target characteristics (i.e., structural, chemical and biophysical variables); and ii) the signal characteristics (i.e., wavelength, incidence/reflection angle, etc.). Thus in the inverse operational way they can be used to represent the mapping between the measurements at the remote sensor and the variable of interest. A wide variety of analytical electromagnetic models has been proposed in the literature, with different levels of complexity and generality (*Fung et al., 1992; Jacquemond and Baret, 1990;Verhoef, 1984*) Thanks to the solid physical foundation and the wide range of applicability (in terms of both target properties and system characteristics), electromagnetic models can operate in more general scenarios that are difficult to represent through the collection of *in-situ* measurements. For this reason, they are particularly appealing to address the estimation of geo-/bio-physical variables from remote sensing data. A major concern is related to the fact that they rely on hypothesis and assumptions that simplify their representation of real phenomena. This issue is

intrinsic in the modelization process and can be reduced (but not completely elimintated) by increasing the complexity of the model, at the price of a reduced generalization ability. Another drawback of electromagnetic models is the fact that they do not take into account issues of the specific estimation problem considered, such as sensor noise or calibration errors.. A random contribution (with predefined distribution derived from the prior knowledge on the estimation problem) could be applied to the theoretical model in order to handle the variability and non-linearity of the inaccuracies in different regions of the feature space (*Notarnicola et al., 2008*). However, often the results are not satisfactory.

In order to convey the strengths of both approaches, namely emprical relationships inference and analytical models inversion, a novel hybrid approach to the estimation of geo-/bio-physical variables is formulated. The main rationale behind this approach is the integration and exploitation of both a set of (few) field reference samples and a theoretical analytical model. In greater details, the theoretical model is used for derive the mapping between input features and the target geo-/bio-physical variable, while the field reference samples are exploited for modeling in the feature space domain the deviation between the theoretical model predictions and the true target values. The aim is to overcome or at least reduce the limitations (such as the analytical simplifications) of theoretical models while keeping at the same time their characteristics in terms of generalization ability and robustness coming from the rigorous theoretical foundation. Concerning the characterization of the deviation between theoretical model predictions and true target values, two possible strategies are proposed: i) the *global deviation bias*, which assumes a constant behavior of the deviation in the whole feature space for modeling biases between theoretical models and real world phenomena espeicially in case of very limited availability of field reference samples; and ii) the *local deviation bias*, which takes into consideration the more realistic condition in which the deviation is variable within the input space under the assumption of the availability of slightly more reference samples. In our experiments the proposed approach is applied to the challenging application domain of soil moisture retrieval from microwave remotely sensed data. However, it is general and can be used in other application domains where the limited availability of field reference samples and the spatial and temporal variability of the target geo-/bio-physical variable represent a strong limitation for the inference and applicability of empirical models.

The rest oft he chapter is organizeed into three more sections. Section 4.2 presents the genearal formulation of the proposed hybrid retrieval approach and explains the details about the two proposed strategies for modeling the deviation between model predictions and target values in the feature space. The experimental analysis carried out for assessing the effectiveness of the proposed methodology is illustrated in Section 4.3. Finally, Section 4.4 draws the conclusions of this work.

4.2 Proposed Hybrid Retreival Approach

Let us recall the general formulation oft he retrieval problem discussed in Chapter 1. Given $\boldsymbol{x} = [x^1, x^2, \dots, x^m]$ the vector of input features extracted from the remotely sensed data (e.g., optical or SAR images) and y the desired continuous target variable (the geo-/bio-physical variable of interest), the retrieval problem can be expressed as:

$$y = f(\boldsymbol{x}) \tag{4.1}$$

where f represents the desired mapping between the input space and the target parameter. T is a theoretical forward model, which will simulate the behavior of the remotely sensed signals as a function of the target variable (under certain assumptions and simplifications intrinsic in its analytical formulation). To address the estimation problem, one may resort on the functiona $g(.)$ which represents the inverse mapping between $\boldsymbol{x}$ and y according to T . In this case, the estimation will be affected by a deviation $\delta(.)$ which stems from the simplifications and approximations intrinsic within the theoretical model formulation and depends, in general terms, from the input features vector $\boldsymbol{x}$, i.e.:

$$y = g(\boldsymbol{x}) + \delta(\boldsymbol{x}) \tag{4.2}$$

Let us assume now to consider a set $\boldsymbol{R}$ of reliable field reference samples, i.e., couples of measurements of the target geo-/bio-physical variable and the corresponding features extracted from the remotely sensed signals. The idea of the proposed approach is to exploit the information conveyed by the field reference samples to characterize the deviation $\delta(\boldsymbol{x})$. Accordingly, the estimation associated with a generic unknown sample $\boldsymbol{x}_*$ is the sum of two contributions: the first one derived from the inverse theoretical model functional $g(\boldsymbol{x})$ and the second one provided by the estimated deviation $\hat{\delta}(\boldsymbol{x}, R)$, i.e.:

$$\hat{y}_* = g(\boldsymbol{x}_*) + \hat{\delta}(\boldsymbol{x}_*, R) \tag{4.3}$$

In this formulation the term $\hat{\delta}(\boldsymbol{x}, R)$ represents a correction to the systematic error due to the model analytical formulation and the intrinsic ambiguities of the data. In this way it is possible to improve the accuracy of the estimation process.

From the operational viewpoint, the proposed technique can be split into two different phases: i) the training phase, in which the available field reference samples and the corresponding estimates of the theoretical model are provided to the system for the definition of the deviation functional $\delta(\boldsymbol{x})$ (see Figure 4.1); and ii) the operational phase, where the technique estimates $\delta(\boldsymbol{x})$ for unknown samples and corrects the target value estimate according to (4.3). For the characterization of $\delta(\boldsymbol{x})$ during the training phase, one may resort on different strategies. In this work we present two possible strategies: i) the *global deviation bias*; and ii) the *local deviation bias*.

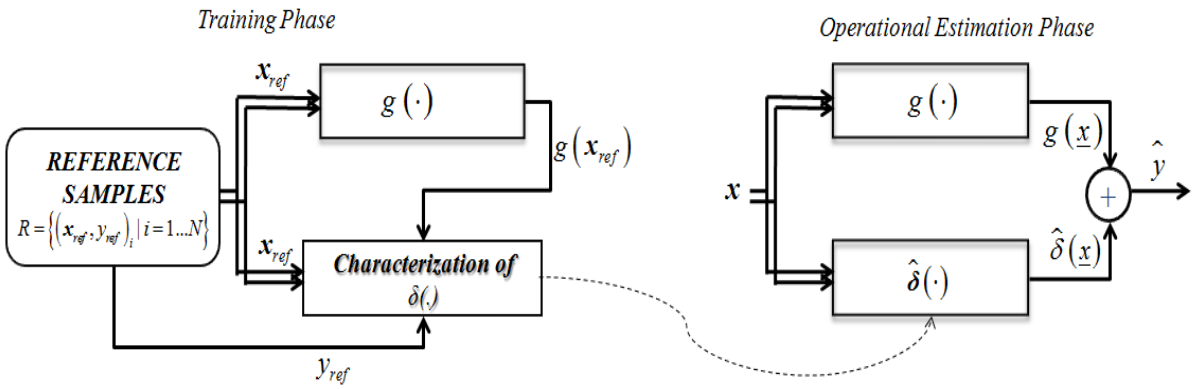


Figure 4.1. Block diagram of the proposed hybrid approach.

4.2.1 Global Deviation Bias (GDB) Strategy

The easiest way for the characterization of the deviation functional $\delta(\boldsymbol{x})$ is to assume a constant behavior in different regions of the input feature space, i.e., $\delta(\boldsymbol{x}) = Const$. A simple and reliable estimation of the constant $Const$ can be obtained by computing the average deviation between the estimation of the inverse theoretical model functional $g(.)$ and the target measurements of the set R of available field reference samples according to:

$$\hat{\delta}(\boldsymbol{x}) = Const = \sum \frac{\delta(\boldsymbol{x}_i)}{|\boldsymbol{R}|} \quad s.t.\ \boldsymbol{x}_i \in R \tag{4.4}$$

where $\delta(\boldsymbol{x}_i)$ is the deviation associated to the i-th field reference sample and $|\boldsymbol{R}|$ is the cardinality of the reference set $\boldsymbol{R}$. This strategy relies of the inverse theoretical model for handling the complexity and non-linearity of the retrieval problem, while it simply corrects with a global bias term the estimates of the theoretical model. Thanks to its simplicity, this strategy is particularly suitable for the operational conditions in which very few field reference samples are available.

4.2.2 Local Deviation Bias (LDB) Strategy

The simple strategy previously described can be improved to take into consideration the more general and realistic case in which $\delta(\boldsymbol{x})$ is variable within the input space. Indeed, it is expected a different accuracy of the theoretical model in different portions of the input feature space. The idea in this case is to locally approximate the deviation between the estimates of the theoretical model and the true target sample values according to:

$$\hat{\delta}(\boldsymbol{x}) = Const_{\boldsymbol{x}} \tag{4.5}$$

where the value of $Const_{\boldsymbol{x}}$ depends on the position within the input space. In greater details, $Const_{\boldsymbol{x}}$ can be calculated as the average deviation associated to the subset $\boldsymbol{R}_{neighborhood}(\boldsymbol{x})$ of field reference samples in the neighborhood of the generic sample $\boldsymbol{x}$:

$$C_{\boldsymbol{x}} = \sum \frac{\delta(\boldsymbol{x_j})}{|\boldsymbol{R}_{neighborhood}(\boldsymbol{x})|} \quad s.t.\ \boldsymbol{x_j} \in \boldsymbol{R}_{neighborhood}(\boldsymbol{x}) \subseteq \boldsymbol{R} \tag{4.6}$$

This corresponds to a local analysis of the input feature space based only on the $N_{neighborhood} = |\boldsymbol{R}_{neighborhood}(\boldsymbol{x})|$ field reference samples in the neighborhood of $\boldsymbol{x}$. The local neighborhood can be defined according to two procedures: a) the *fixed local neighborhood*; or b) the *adaptive local neighborhood*.

The *fixed local neighborhood* procedure applies a deterministic partition of the input space and selects the $\boldsymbol{S}$ referece samples falling within the *m*-dimensional quantization cell of the generic sample $\boldsymbol{x}$. This strategy can be seen as a simplified application of the Parzen estimation approach, in which the neighborhood used for the estimation of a distribution is fixed and has the same quantization size for all cells. It is worth noting that in some cases (too fine quantization or very few field reference samples) no neighborhood samples might be available for some quantization cell. In this situation, one may resort on a very rough quantization or on the global deviation bias described in the previous paragraph.

The *adaptive local neighborhood* strategy defines the neighborhood in an adaptive way by a K-nearest-neighbor (K-NN) strategy. The latter selects the $K = N_{\text{neighborhood}}$ samples closer to the generic target $\boldsymbol{x}$ according to a distance function (e.g., the Euclidean distance). In this way it is possible to consider a fixed number of reference samples for the determination of $Const_{\boldsymbol{x}}$. This results in an adaptation of the shape and the size of the quantization cell to the local distribution of the samples.

From a theoretical viewpoint, there is no limit to the definition of strategies for the characterization of $\delta(\boldsymbol{x})$ in the proposed approach. For example, one may consider advanced non-linear and non-parametric estimation techniques (such as the Support Vector Regression) for a global estimation of the deviation in the whole feature space. However, it is worth noting that there is an important trade-off between the complexity of the empirical estimation strategy and the number of field reference samples available. Increasing the complexity (e.g., the non-linearity) of the functional type associated with $\delta(\boldsymbol{x})$ typically increases the number of required field reference samples to ensure generality and avoid over-fitting phenomena.

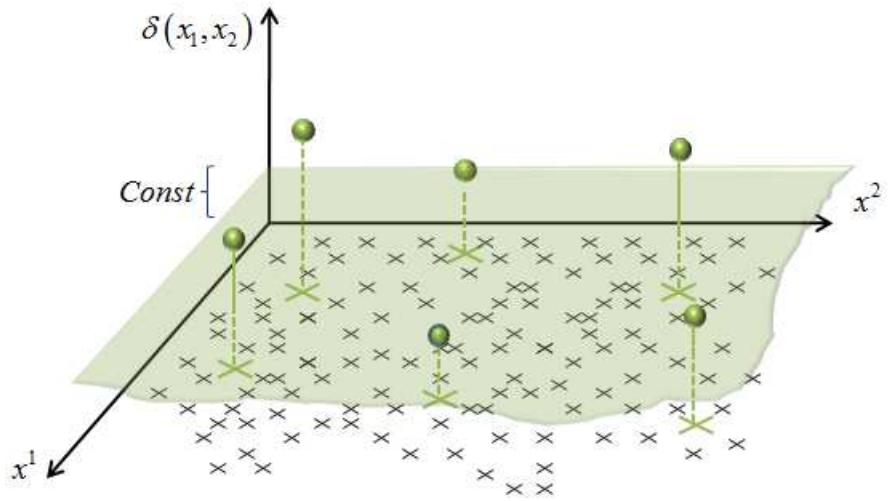


Figure 4.2. Graphical representation of the global deviation bias strategy in the case of two input features. Black crosses represent general samples in the 2-D input feature space. Green crosses represent the field reference samples, each one associated with the corresponding deviation between theoretical model prediction and target value (gree dots along the vertical axis). The gray plane represents the constant deviation bias in the whole feature space computed as the average of the deviations associated with the field reference samples.

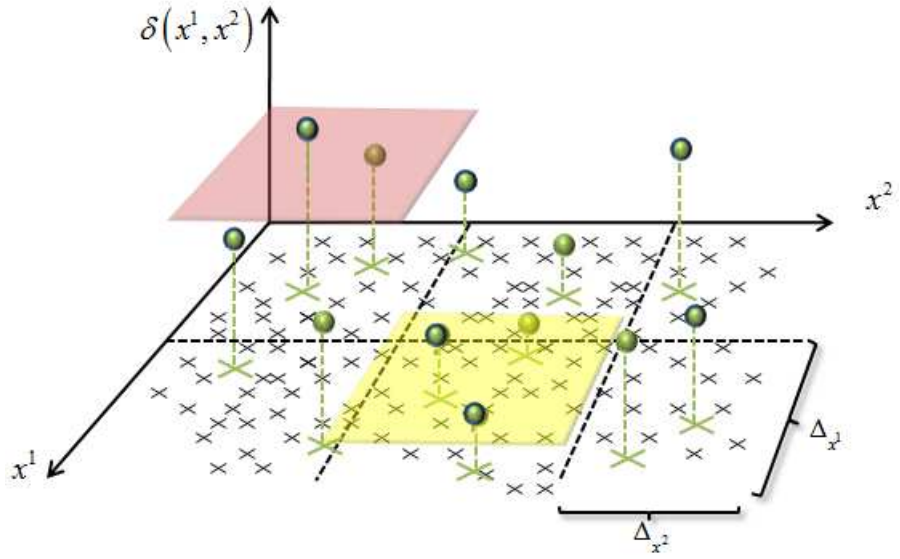


Figure 4.3. Graphical representation of the local deviation bias strategy with a fixed local neighborhood in the case of two input features. The input feature domain is quantized according to Δ_{x1} and Δ_{x2}. Then for each cell identified a constant bias is computed on the basis of the deviations associated with the reference samples falling within the cell boundaries. Different cells may show different biases (red and yellow squares).

4.3 Experimental Analysis

4.3.1 Data Set Description

In order to evaluate the effectiveness of the proposed hybrid approach in real retrieval problems, it was applied to the estimation of soil moisture content from microwave remotely sensed data. This application domain is particularly suitable for assessing the the proposed hybrid approach, due to the limited availability (usually) of field reference samples and the complexity and non-linearity of the retrieval problem, as discusded in the previous chapters. However, it is worth stressing that the proposed methodology is general and can be applied in any application domain when there is the need to integrate theoretical model based estimations with field measurements.

Two data sets have been considered for the experimental analysis. They present quite heterogeneous characteristics in terms of signal propertries and investigated area (Table 4.1). The first data set is the *Scatterometer* data set already presented in Chapters 2 and 3. It consists of 17 field soil moisture content measurements paired with the corresponding backscattering coefficients at HH polarizations and 23°/40° incidence angles acquired with a C-band field scatterometer on two bare agricultural fields near Matera, southern Italy (*Mattia et al., 2003*).

The second data set (referred as *SMEX* data set) consists of a subset of data acquired during the Soil Moisture Experiment 2002 (SMEX'02), which took place in Iowa, USA, from June 24th to July 12th, 2002 (*WWW4*). During the experiment, several soil parameters, among the others soil moisture content and surface roughness, together with optical and SAR remote sensing data were collected over crop fields. The aim was to validate the soil moisture algorithm of the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) and studing the effect of disturbing target properties on the soil moisture accuracy, with particular regard to the effect of vegetation. In grater detail, in this study we consider 35 field measurements collected over four soybean sites (which showed reduced biomass values) and coupled with remote sensing backscattering measurements acquired by the AirSAR sensor (Table 4.1). The SAR sensor operated at L- and C-bands in polarimetric mode and imaged the study area with a nominal incidence angle of 40°, in order to be comparable with other orbiting satellite systems. However, only L-band co-polarized HH and VV backscattering coefficients are used in this study. SAR images were processed and calibrated by the AirSAR operational processor and post-processed with a multi-look procedure in order to reduce speckle effects in the data. For more details on the SMEX'02 experiment and on the soil data set, we refer the reader to (*WWW4; WWW5; Notarnicola et al., 2006*).

Table 4.1. Main characteristics of the two data sets considered in the experimental analysis.

Data Set	***Scatterometer***	***SMEX***
Study area	Bare agricultural fields	Vegetated soybean fields
Microwave Sensor	Field Scatterometer	Airborne SAR
# Samples	17 field samples	35 field samples
Input Features	Backscattering coefficients C-Band, HH polarization, 23°/40° incidence angles	Backscattering coefficients L-Band, HH/VV polarizations, 40° incidence angle
Dielectric Constant Range	$5 \leq \text{diel} \leq 15$	$5 \leq \text{diel} \leq 13$
Surface Roughness Range (*s* and *l*)	$1.3\text{cm} \leq s \leq 2.1\text{cm}$ $5.0\text{cm} \leq l \leq 13.0\text{cm}$	$1.3\text{cm} \leq s \leq 2.5\text{cm}$ $2.0\text{cm} \leq l \leq 10.0\text{cm}$

4.3.2 Design of Eperiments

The retrieval of the target variable, i.e., the soil dielectric constant, was first addressed according tot he inversion of the well-known and validated Integral Equation Model (IEM) (*Fung et al., 1992*). This model is mainly used to simulate the microwave response of bare soil surfaces. However, it has been exploited also for the experimental analysis with the *SMEX* data set. On the one hand, the contribution of vegetation on the SAR backscattering was rather limited at the wavelength considered (L-band) especially for the soybean fields due to the low biomass values. On the other hand, this experimental scenario allowed us to investigate the effectiveness of the proposed hybrid approach also in this challenging and rather common operational condition. Due to the high complexity and analytical non-tractability of the IEM model inversion process, the inference of the inverse functional $g(.)$ has been carried out with the Support Vector Regression (SVR) technique according to the strategy discussed in Chapter 2. More in detail, two sets of 300 and 250 simulated reference samples were generated using the IEM model in the forwar operational way and under the same conditions (in terms of signal frequency, polarization, acquisition angle and soil characteristics) of the *Scatterometer* and *SMEX* field measurements, respectively. Simulated backscattering values were then corrupted by a Gaussian additive noisy contribution with mean and variance equal to 0 and 0.1, respectively. This was done in order to increase the variability within the simulated data. Please note that this noisy contribution did not take into account the characteristics in terms of noise of the field data as done in the previous chapters. Then SVR has been trained on the simulated reference samples. A standard grid search procedure aimed at minimizing the meand squared error (MSE) on validation samples was used for tuning the SVR free parameters (*Scholkopf and Smola, 2001)*. A Gaussian RBF

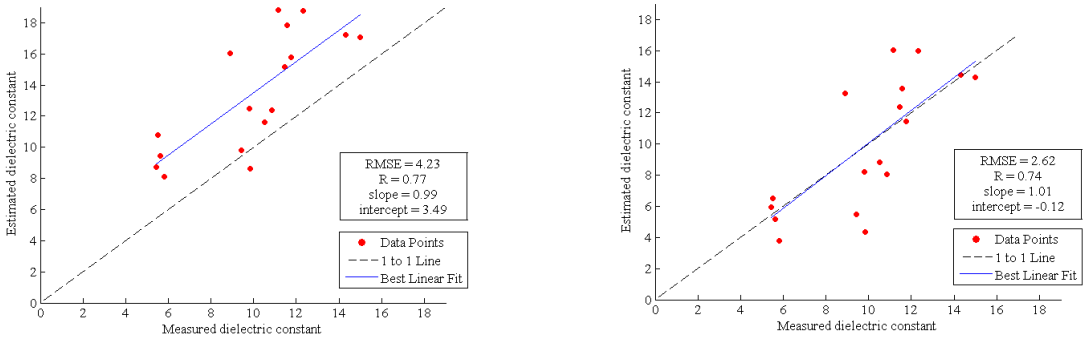
kernel and the following ranges were considered in this phase: $[10^{-3}; 10^{3}]$ for γ (the kernel width) and *C*, respectively; $[10^{-4}; 10]$ for ε.

After the training/tuning oft he SVR regressor on simulated reference samples, the deviation between theoretical model estimates and field reference measurements was investigated and characterized according to two strategies: i) the *global deviation bias* (GDB); and ii) the *local deviation bias* (LDB), considering the *local fixed neighborhood* solution described previously. Finally, the estimated target values of unseen samples obtained with the inverse theoretical functional $g(.)$ were corrected according to the deviation contribution $\hat{\delta}(\boldsymbol{x}, R)$ estimated previously (see euqtion 4.3). The resulting dielectric constant values were compared with the measured reference values and evaluated in terms of root mean squared error (*RMSE*), statistical correlation coefficient *R*, *slope* and *intercept* of the linear trend line between estimated and reference target values. Due to the limited number of field reference samples available for this experimental analysis, at this stage a *t*-fold cross-validation procedure has been adopted (*Cherkassky and Mulier, 1998*). Field reference samples were divided into *t* subsets. Iteratively, *(t-1)* subsets were used for the characterization of the deviation functional $\delta(.)$, while the remaining subset was exploited as independent test set for assessing the performance of the proposed approach. The procedure was repeated until all field samples were used as test. Then the numerical performance were evaluated. As benchmark, the performance achieved with the inverse theoretical functional without applying the correction procedure was considered.

4.3.3 Experiment 1: Correction with the Global Deviation Bias (GDB) Strategy

In this first experiment the available field reference samples were exploited for the estimation (and thus the correction) of the global deviation bias between theoretical model based estimates and true target samples. This kind of misalignment could be due to biases in the theoretical model or calibration errors associated to the remote sensor.

Let us first analyze the results obtained with the *Scatterometer* data set. For this experiment, we considered as input features the backscattering coefficient with HH polarization acquired at 23° incidence angle and two folds for the cross-validation procedure (i.e., half samples were used for characterizing the deviation $\delta(.)$ and half for assessing the estimation performance). The comparison between the target value estimates obtained without and with the integration of the field reference samples in the retrieval process according to the proposed hybrid approach is shown in Figure 4.4. From the analysis of the two plots it clearly emerges a noteworthy improvement of the estimation performance after the application of the simple GDB correction strategy, mainly concerning the *RMSE* and *intercept* metrics. This suggests a biased behavior of the theoretical model, which is effectively corrected by the proposed methodology through the use of the field reference samples.



(a) *(b)*

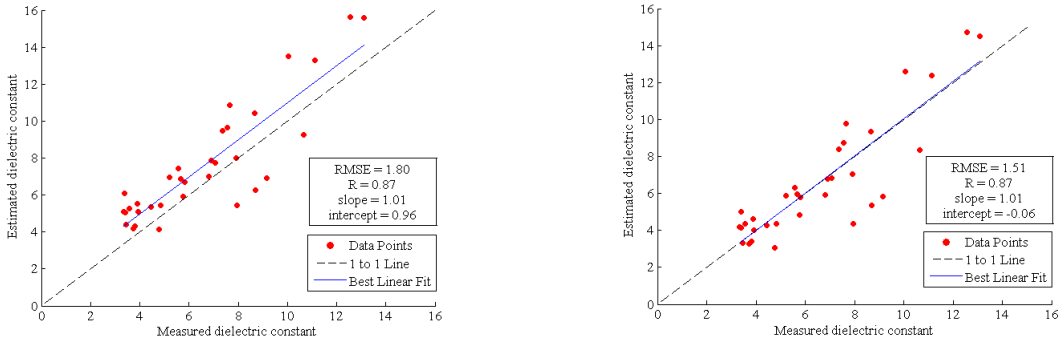
Figure 4.4. Scatter plots of estimated versus measured dielectric constant values (a) before and (b) after the GDB correction according to the proposed hybrid approach. Experiment 1, Scatterometer data set.

Table 4.2. Estimation performance obtained before and after the proposed global deviation bias (GDB) correction strategy. Different proportions of reference and testing samples are considered. Experiment 1, Scatterometer data set.

Method	**RMSE**	**R**	**Slope**	**Intercept**
Theoretical Model	4.23	0.77	0.99	3.49
GDB Corr. (2 folds CV)	2.62	0.74	1.013	-0.12
GDB Corr. (5 folds CV)	2.54	0.75	0.99	0.036
GDB Corr. (LOO CV)	2.53	0.75	0.99	0.0008

In order to assess the sensitivity to the number of fiels reference samples available for the integration, different trials were performed with the same experimental condition and simply changing the number of folds of the cross-validaiton procedure. The results are reported in Table 4.2. The estimation performance show a quite stable behavior while increasing the number of reference samples used for the characterization of $\delta(.)$. This is an expected behavior, since the GDB strategy is particularly simpe and just one parameter has to be estimated, i.e., the global bias.

Concerning the *SMEX* data set, the achieved results are reported in Figure 4.5. Again the proposed hybrid approach outperforms the standard theoretical model inversion in terms of both *RMSE* and *intercept* metrics. This indicates a biased behavior of the theoretical model that can be ascribed to the effect of vegetation on the SAR backscattering coefficients. A slight overestimation of high dielectric constant values can be observed in the final output estimates also after the bias correction. This points out a limit (expected) of the the simple GDB strategy, i.e., it does not take into consideration the possibility of a deviation variable within the input feature domain.



(a) *(b)*

Figure 4.5. Scatter plots of estimated versus measured dielectric constant values (a) before and (b) after the GDB correction according to the proposed hybrid approach. Experiment 1, SMEX data set.

Table 4.3. Estimation performance obtained before and after the proposed global deviation bias (GDB) correction strategy. Different proportions of reference and testing samples are considered. Experiment 1, SMEX data set.

Method	RMSE	R	Slope	Intercept
Theoretical Model	1.80	0.87	1.01	0.96
GDB Corr. (2 folds CV)	1.51	0.87	1.01	-0.06
GDB Corr. (5 folds CV)	1.50	0.87	0.99	-0.01
GDB Corr. (LOO CV)	1.50	0.87	0.99	0.0003

With regard to the sensitivity to the number of available reference samples (see Table 4.3), again the results achieved are almost stable, confirming the effectiveness of this strategy also when very few field reference samples are available.

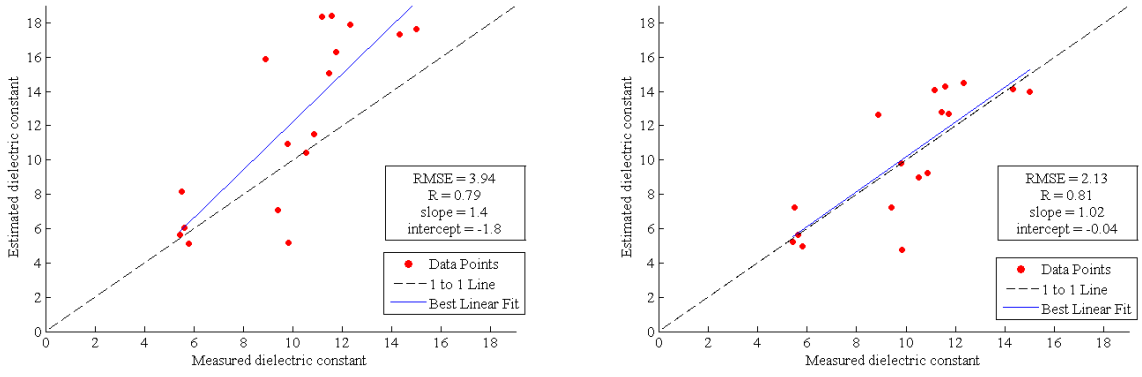
4.3.4. Experiment 2: Correction with the Local Deviation Bias (LDB) Strategy

The second experiment presented aimed at evaluating the capability of the proposed hybrid approach to handle the (more general and realistic) condition in which the deviation between theoretical model estimates and true target is modeled with a different behavior in the input feature space. To this purpose, the deviation $\delta(\boldsymbol{x})$ has been approximated with the local deviation bias estimated by considering the *fixed local neighborhood* strategy (in the input feature space) described previously. Then the correction has been applied to the unknown sample estimates provided by the inverse theoretical model. Different sizes (i.e., 1x1, 3x3 and 5x5) of the neighborhood window were tested. Moreover, a 5-fold cross-validation procedure

has been adopted to avoid issues with regard to the choice of the reference and testing samples.

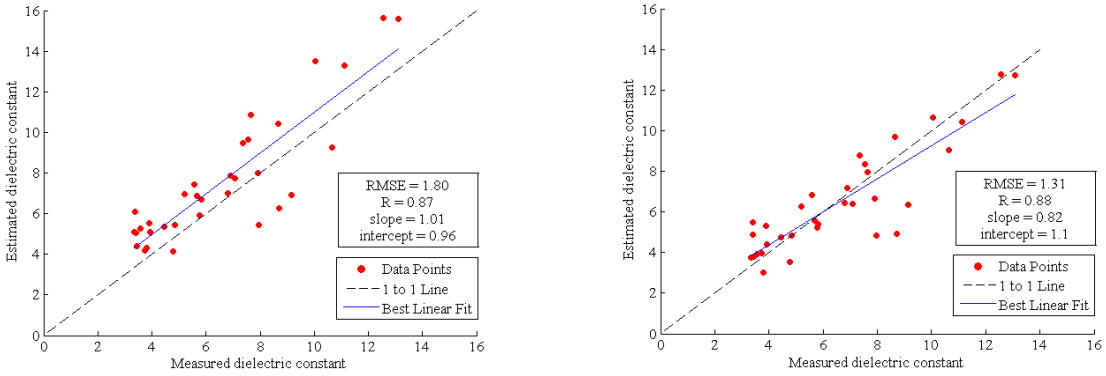
Concerning the Scatterometer data set, we considered for this experiment the case of 2 input features, i.e., the backscattering coefficients with HH polarization and 23°/40° incidence angles. The results reported in Figure 4.6 indicate a general significant improvement in terms of accuracy (i.e., *RMSE*) with respect to the case where no correction is applied. This is mainly due to the reduction of the overestimation trend for high dielectric constant values shown by the estimates obtained with the model inversion without the exploitation oft he field reference samples. Indeed, after the correction process the *slope* and *intercept* values become closer to the ideal ones (1 and 0, corresponding to the one-to-one line, respectively). This results suggest the effectiveness of the proposed approach to properly handle the variability of the functional within the input feature space.

Similar results were achieved for the *SMEX* data set. As shown in Figure 4.7, the overestimation trend for high dielectric constant values pointed out in the previous subsection was reduced by exploiting the capability of the LDB strategy to locally model the deviation functional $\delta(\boldsymbol{x})$ within the input feature space. Consequently the *RMSE* value is further reduced. Despite this improvement, the *slope* and *intercept* metrics show slightly worse performance, which can be probably ascribed to the presence of some outlier samples (the samples with measured dielectric constant values between 8 and 9 and slightly under-estimated by the retrieval system) which affect these metrics.



(a) *(b)*

Figure 4.6. Scatter plots of estimated versus measured dielectric constant values (a) before and (b) after the LDB with fixed local neighborhood (3x3 window) correction according to the proposed hybrid approach.
Experiment 2, Scatterometer data set.



(a) *(b)*

Figure 4.7. Scatter plots of estimated versus measured dielectric constant values (a) before and (b) after the LDB with fixed local neighborhood (3x3 window) correction according to the proposed hybrid approach.
Experiment 2, Scatterometer data set.

With regard to the size of the quantization window, the achieved results on both the considered data sets indicate a strong influence of this parameter on the estimation performance of the LDB strategy (Table 4.4 and 4.5). For both data sets, the best performances are achieved by considering the 3x3 window, which leads to a strong improvement of the accuracy and a significant reduction of the overestimation trend for high dielectric constant values. The small window (1x1) preserves more the locality of the correction applied. However, typically no or very few reference samples fall within the neighborhood region, thus affecting the correction applied. On the contrary, the 5x5 quantization window includes a high number of reference samples (far one to each other) in the local correction process, at the price of a reduced capability to follow the variations of the deviation functional $\delta(\boldsymbol{x})$. Thus it emerges that the choice of the best window size strictly depends on the characteristics of the estimation process considered (in particular the characteristics of the deviation $\delta(\boldsymbol{x})$) and the number of available field reference samples. An automatic tuning of this parameter is not straightforward and further research is required in order to define effective procedures for the selection of this parameters.

Table 4.4. Estimation performance obtained before and after the proposed local deviation bias (LDB) with local fixed neighborhood correction strategy. Different sizes of the quantization window are considered.
Experiment 2, Scatterometer data set.

Method	RMSE	R	Slope	Intercept
Theoretical Model	3.94	0.79	1.4	-1.8
LDB Corr. (1x1)	3.64	0.8	1.54	-5.25
LDB Corr. (3x3)	2.13	0.81	1.02	-0.04
LDB Corr. (5x5)	2.77	0.79	1.25	-2.41

Table 4.5. Estimation performance obtained before and after the proposed local deviation bias (LDB) with local fixed neighborhood correction strategy. Different sizes of the quantization window are considered.
Experiment 2, SMEX data set.

Method	RMSE	R	Slope	Intercept
Theoretical Model	1.80	0.87	1.01	0.96
LDB Corr. (1x1)	1.47	0.85	0.78	1.32
LDB Corr. (3x3)	1.31	0.88	0.82	1.1
LDB Corr. (5x5)	1.53	0.87	1.005	-0.04

4.4 Conclusion

In this chapter, a novel hybrid approach to the estimation of biophysical variables from remotely sensed data based on the integration of field reference samples and theoretical analytical models has been presented. The core idea bethind the proposed method is the exploitation of the available (typically few) field reference samples to model (and thus correct) the deviation between the theoretical model based estimates (affected by biases and approximations intrinsic within the analytical formulation of theoretical models) and the measured target values. Two different strategies for the estimation and correction of such a deviation have been presented and analitically formulated. Then their effectiveness has been evaluated in the specific application domain of soil moisture estimation from microwave remotely sensed data. The achieved results are promising and indicate the proposed integration approach as effective for the estimation of biophysical parameters from remotely sensed data since: i) it allows increasing the accuracy of the estimates by overcoming biases and simplifications intrinsic within the analytical formulation of theoretical models; ii) it is

capable to handle the variability of the deviation functional $\delta(\boldsymbol{x})$ within the input space domain; and iii) it is simple, easy to implement and fast during the processing.

Despite the experimental analysis has been carreid out in one specific application domain, it is important to stress the fact that the proposed hybrid approach is general and can be applied also in other operative conditions, i.e., with other correction strategies, in combination with other model inversion techniques (e.g., iterative methods, look-up tables) and in different application fields.

As future work, a first improvement to the proposed hyrbid approach will regard the investigation and implementation of novel strategies for the estimation and correction of the deviation functional $\delta(\boldsymbol{x})$. As pointed out in the experimental analysis presented in this chapter, the effectiveness of the proposed hybrid retrieval method stritctly depends on the capability to properly exploit the available field reference samples for this task. Another issue which deserves further investigations is the development of automatic and robust strategies for the adaptive tuning of the correction strategy parameters. Finally, despite the promisin results achieved in this experimental analysis, further investigations are required for assessing the effectiveness ot the proposed methodology in other challenging real estimation problems for which the limited availability of field reference samples represents an issue, such as the retrieval of snow pack geo-/bio-physical variables in mountain areas.

Chapter 5

Retrieval of Soil Moisture Content from New Generation RADARSAT2 SAR Imagery in an Alpine Catchment

In this chapter, the effectiveness of RADARSAT2 SAR imagery for the estimation of soil moisture content in an Alpine catchment is investigated. In greater detail, a sensitivity analysis of SAR backscattering coefficient to the moisture content of soil together with other target properties (e.g., local topography and the presence of vegetation) is first carried out. Then a system for the retrieval of soil moisture is proposed. The main features of this algorithm are: i) the exploitation of advanced state-of-the-art methods, namely the support vector regression and the multi-objective parameter optimization introduced in the previous chapters; and ii) the integration of ancillary data in the retrieval process. Discussion on accuracy and capability to reproduce spatial patterns of soil moisture provide indications on the effectiveness of new generation SAR imagery together with the proposed retrieval system to deal with the soil moisture retrieval problem in the challenging Alpine environment.

5.1 Introduction and Motivation

As introduced in Chapter 2, soil moisture content is related with many hydrological and meteorological processes. The accurate mapping of this variable is thus crucial for both hydrological applications and earth sciences. This is particularly true for the mountain environment. Here, the scale of spatial and temporal variability of geo-/bio-physical variables reduces, due to the heterogeneity and the variability of the environment (Rodriguez-*Iturbe et al., 1995; Gebremichael et al., 2009*). This aspect makes the knowledge of accurate and reliable information on soil moisture status more complex (*Grayson et al., 1997*). At the same time, information on soil moisture status is becoming crucial. Historical climate observations have proved that the climate in the Alps has changed significantly. For the future strongest climatic change in the Alps can be expected for the summer months with much drier and warmer conditions in all regions, particularly in the southern part (*Brunetti et al. 2001*). In

addition, climate models agree on a higher inter-annual variability (*Graham et al. 2007*). This means on the one hand increasing drought periods (summer), while on the other hand higher probability of heavy rain (winter). These variations may have a strong impact on the water availability (*Barnett et al. 2005*) for agricultural and human purposes and may be strongly related to natural hazards such as floods and landslides (*Horton et al. 2006*).

In the last few years, the increasing number of space-borne sensors, with complete and frequent coverage of the Earth's surface, has determined an increasing interest for the estimation of geo-/bio-physical surface variables from remotely sensed data. In this field, one of the most challenging problems is related to the estimation of soil moisture content from microwave sensors, in particular Synthetic Aperture Radars (SARs). The sensitivity of SAR backscattering coefficients to soil moisture at low microwave frequencies (P- to L-band) has been investigated and demonstrated by many scientists (*Paloscia, 2002; Notarnicola et al., 2006*). Nowadays, however, most of the orbiting sensors (ERS-2, RADARSAT 2, ENVISAT ASAR) operate at higher frequencies. C-band backscattering is still sensitive to soil moisture (*Macelloni et al., 1999*) but is also affected by other surface characteristics, such as the roughness of the soil surface (*Mattia et al., 1997*), the land-cover heterogeneity (Lakhankar *et al., 2009*) and the presence of vegetation (*Ulaby et al., 1979*). All these factors may strongly affect the sensitivity of SAR backscattering coefficient to the desired target parameter.

Topography is another important aspect to be taken into consideration when dealing with the estimation of soil parameters. Satellite systems, in particular SAR systems, are strongly affected by the topography of the area. Distortion effects (i.e., foreshortening, layover and shadowing) may occur due to the side-looking acquisition geometry (specific of the SAR sensor) and the local topography on the ground (slopes, valleys and hills). Even if these extreme distortion effects do not occur, the SAR signal is affected by the local incidence angle and the distance between the target area and the sensor antenna. These topographic effects are usually taken into consideration during the calibration of the data. However, when dealing with mountain areas, such as the Alps, it is fair to expect to have a non-negligible residual contribution within the signal due to the extreme topographic conditions (*Luckman, 1998*). Also this contribution may significantly influence the sensitivity of the microwave signal acquired by the satellite sensor to the moisture content of the soil and consequently could further increase the complexity of the estimation problem. However, limited effort has been devoted to this challenging aspect in the assessment of soil moisture in Alpine areas. For example, Paloscia et al. (*2010*) investigate the effectiveness of ASAR remotely sensed data in combination with optical images for the estimation of soil moisture in the Cordevole area (Veneto region, Italy). The analysis points out the significant influence of the vegetation coverage on the backscattering signal. However, the area of interest does not present significant variability in terms of topography, thus limiting the applicability of the presented analysis on other mountain areas with different topographic conditions.

All these aspects make the problem of the characterization of soil moisture in mountain areas from remotely sensed data extremely complex and challenging. With the prospective of the integration of soil moisture estimates in real applicative scenarios, like those cited in the previous chapters, it is important to have a clear comprehension of the possibilities and the limitations of the new generation satellite SAR sensors in combination with advanced state-of-the-art methodologies for the retrieval of soil parameters in the Alpine environment. Despite some works in this direction have started, further analysis is still required.

This chapter tries to overcome the limitations of the current state-of-the-art by presenting an analysis on the sensitivity of new generation RADARSAT2 satellite SAR imagery to the soil moisture content in the challenging mountain environment. The analysis takes advantage from the availability of ancillary data for better understanding the influence of vegetation, land-cover heterogeneity and topography on the SAR signal. Then an advanced retrieval system for the retrieval of soil moisture in this operational condition is proposed. The system exploits both the methodological developments presented in the previous chapters and the information derived from the ancillary data available to address the complex and highly non-linear retrieval problem. The effectiveness of the proposed solution is evaluated and discussed in terms of both numerical accuracy on field measurements and capability to reproduce expected spatial and temporal patterns of the investigated variable.

The rest of the chapter is organized as follows. Section 5.2 introduces the study area on which our analysis is focussed and describes the data set adopted. The analysis of the sensitivity of the RADARSAT2 data to the soil moisture content is presented in Section 5.3, while Section 5.4 is devoted to the description of the proposed estimation algorithm and the experimental setup for its validation. Section 5.5 shows the experimental results achieved. Finally, Section 5.6 draws the conclusion of the work.

5.2 Study Area and Data Set Description

5.2.1 Study Area

The study area chosen for the analysis presented hereafter is the Mazia valley (Figure 5.1). It is a small side valley in the north-western part of South Tyrol region (Northern Italy). It covers an area of around 100 km^2 with altitudes ranging from 920 meters a.s.l. (Sluderno) to 3738 meters a.s.l. (Palla Bianca). Despite the relatively small dimension, the Mazia valley is well representative in terms of geomorphology and topography for the whole South Tyrol. The area is constantly monitored by 16 meteorological stations distributed along the valley in locations representative of different elevation, slope, aspect, soil type and land-cover conditions. The stations measure the soil status (moisture content) and meteorological variables (air temperature and humidity, precipitation, wind speed and direction, solar radiation). In addition, several research project are currently ongoing in the area and

numerous and accurate ancillary information regarding soil composition, land-use distribution, topography and vegetation are available (*Bertoldi et al., 2010*).

Excluding forests, the most diffuse land-cover/use types are meadow and pasture. They present quite heterogeneous characteristics in terms of vegetation, spatial distribution and human usage. In more detail, meadows are located in the valley bottom and on the first slopes up to 1700 m a.s.l.. They are intensively used in terms of cutting (twice a year), manuring (twice a year) and irrigation (several times during summer drought). The soil surface is typically homogeneous and flat in terms of roughness while the vegetation type is determined as Trisetum flavescentis with a high percentage of grasses. Cut events during the summer period determine variations in the biomass of the vegetation coverage. Pastures have completely different characteristics. First of all, they are located on the slopes of the valley at medium/high altitudes (from 1700 to 2400 m a.s.l.) where the terrain becomes very steep. Due to cattle and sheep grazing the soil surface is heterogeneous, with the presence of bare soil, stones and in some cases of large rocks in areas on higher altitudes. The vegetation is characterized by hemicryptophytes and determined as Sclerantho-Sempervivetum arachnoidei in the lower parts and Sieversio-Nardetum strictae in the upper parts.

From the hydrological viewpoint, the Mazia valley is relatively dry with respect to other parts of South Tyrol. The mean annual precipitation of 525 mm (Mazia, 1580 meters a.s.l.) determines a significant difference in terms of soil water availability between meadows and pasture. Meadows are typically wet (40/45% of soil moisture content) due to the irrigation practice especially in dry summer periods. Pastures on the contrary are typically dry, with the exception of stripy wet buffers along small rivers going down from small water sources at medium altitudes (around 2000 m a.s.l.).

Regarding the soil composition that can influence the behaviour of water in soils, organic content, grain size distribution and bulk density are highly variable even within areas of the same land-cover/use type. On meadows and pastures the dominant soil type is brown soil while above the tree line combinations of brown soils and ranker appear. In streamlet areas, also gley is detected. Regarding the soil texture of fine earth, the fraction of sand is dominant (45-75%), the fraction of silt is quite variable (10-40%) and the fraction of clay is low (5-15%).

5.2.2 Satellite Images

During the summer 2010 (3rd June and 21st July), two SAR images were acquired by the RADARSAT2 satellite sensor over the Mazia valley. The acquisition mode was Standard Quad Polarization, right looking, with a nominal incidence angle of 45° and an ascending orbit. This acquisition geometry has been carefully selected by performing several simulations with the help of a SAR image simulator. The aim was to minimize the geometrical distortions (shadowing and layover/foreshortening) that occur in presence of mountain reliefs due to the side looking view of SAR systems. Geometrical distortions prevent the possibility of any geo-

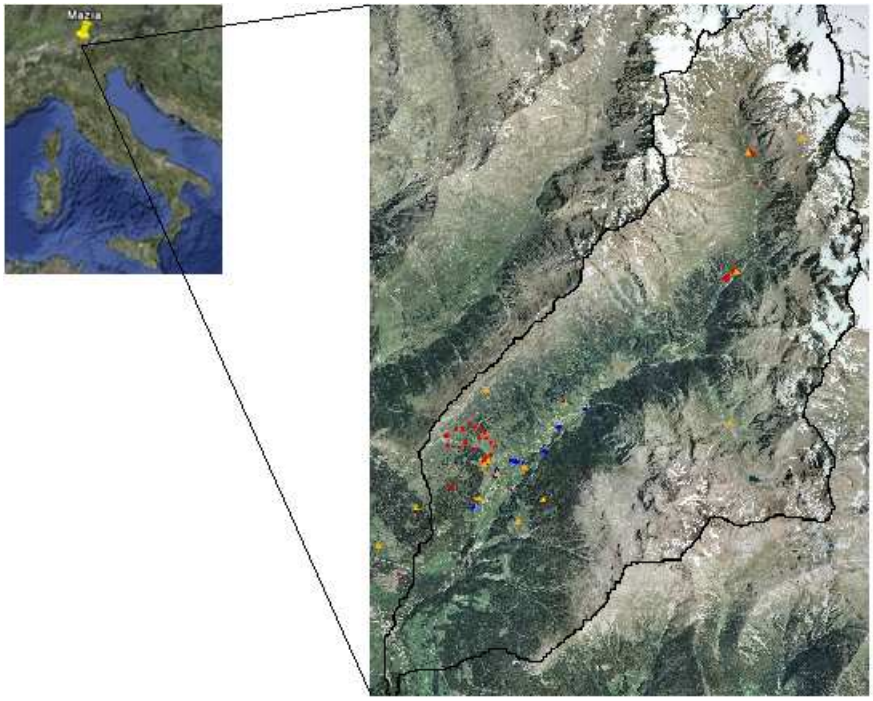


Figure 5.1. The Mazia Valley, South Tyrol, Italy. Orange triangles indicate the position of the fixed meteorological stations while circles represent the locations of the field measurements in both meadow (blue) and pasture (red) areas.

/bio-physical variable retrieval since the backscattering coefficient is mainly dominated by these effects. In our specific case, the selected acquisition geometry allowed an optimal imaging of the west side of the valley, where a larger extension of pasture/meadow areas and higher number of meteorological stations are present with respect to the east side.

Original images were provided in single look complex (SLC) format with pixel size of 4.93 m and 17.48 m in azimuth and ground range directions, respectively. The data have been multi-looked, de-speckled with a Frost filter (5x5 window size), calibrated and geocoded (UTM WGS-84) with the help of a high spatial resolution (2.5x2.5 m^2) digital elevation model. The final resolution of the processed images is 20x20 m^2. All pre-processing steps have been carried out with the SARscape® software (www.sarmap.ch). Figure 5.2 shows the results of the pre-processing for the 21st July image. Polarimetric configurations (HH, HV and VV) have been combined in this RGB image in order to enhance the different information content of each channel. Vegetated areas are shown in green, with different intensities depending on the density of the vegetation. Violet areas are associated to rocks at the highest altitudes. On the east side of the valley the effects of geometric distortions (in particular layover/foreshortening, in white) are particularly evident. These effects are minimized on the west side, thanks to the careful selection of the SAR acquisition geometry.

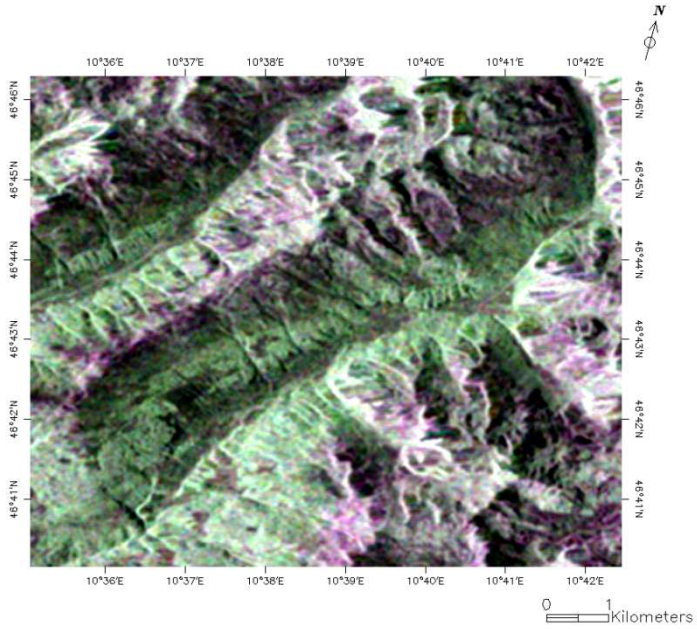


Figure 5.2. RADARSAT2image acquired on July 21st, false color RGB composition (R=HH, G=HV, B=VV).

5.2.3 Field Measurements

Contemporary to the two satellite acquisitions, field measurement campaigns have been carried out in the study area. The aim was to acquire information on the moisture status of soils in meadows and pastures. Non-destructive measurements were acquired using the mobile Delta T WET 2 TDR sensor (www.delta-t.co.uk). The TDR instrument measures only the real part of the top five centimetres soil dielectric constant, but has the advantage to be easy to carry in the field and fast in measuring. The sampling strategy was as follows. Moving along two predefined transects (the vertical transect along the west slope and the horizontal transect parallel to the valley floor), sampling areas of approximately 30x30 m^2 were identified and selected for their homogeneity and representativeness in terms of local topography, vegetation status and land-cover/use type. For each sampling area the actual position was registered with a GPS device. Then repeated measurements (three to five) of soil dielectric constant distant few meters one to each other were collected and averaged. Soil measurements in more than 100 sampling areas were collected. Table 5.1 reports the main characteristics of the soil measurements on meadows and pastures during the two field campaigns. As can be observed, meadows present higher target variable values with respect to pastures, which are in general dryer. Moreover, both pastures and meadows are dryer in July with respect to June. Measurements of July over meadows present a high standard deviation, due to the heterogeneity in the scheduling of silage and irrigation.

Along with TDR measurements, destructive measurements were collected in few representative sampling areas. They were performed by physically taking a soil sample with a

5 centimetres height metallic cylinder. The soil sample was weighted, sealed and carried in laboratory in order to be dried according to standard soil measurements protocols (*Cihlar et al., 1987*). Then soil moisture, soil composition and density analysis have been started. This kind of sampling was necessary to have accurate measurements of soil gravimetric moisture and bulk density and to derive a calibration curve for the non-destructive TDR measurements. The latter task was accomplished by comparing TDR measurements and soil moisture values derived from gravimetric samples. This resulted in a linear relationship as shown in Figure 5.3. According to this relationship, the dielectric constant values measured with the WET sensor have been converted into soil moisture content values.

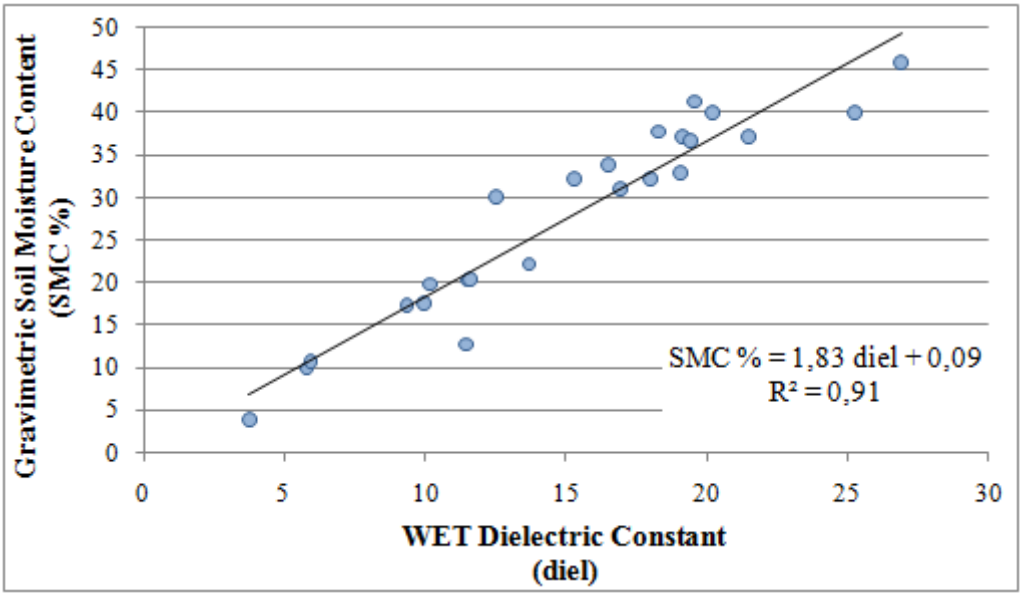


Figure 5.3. Empirical relationship between gravimetric soil moisture content and WET dielectric constant measurements (linear fit).

Table 5.1. Main characteristics of the measures acquired with the Delta T WET 2 TDR sensor during the field campaigns. Values are expressed in SMC %.

	June 2010		**July 2010**	
	Pasture	**Meadow**	**Pasture**	**Meadow**
Min SMC %	12.36	7.05	11.91	5.95
Max SMC %	42.59	49.61	32.61	16.71
Average SMC %	*30.72*	*28.44*	*21.75*	*10.83*
Std SMC %	*21.75*	*7.05*	*5.46*	*2.79*

5.2.4 Ancillary Data

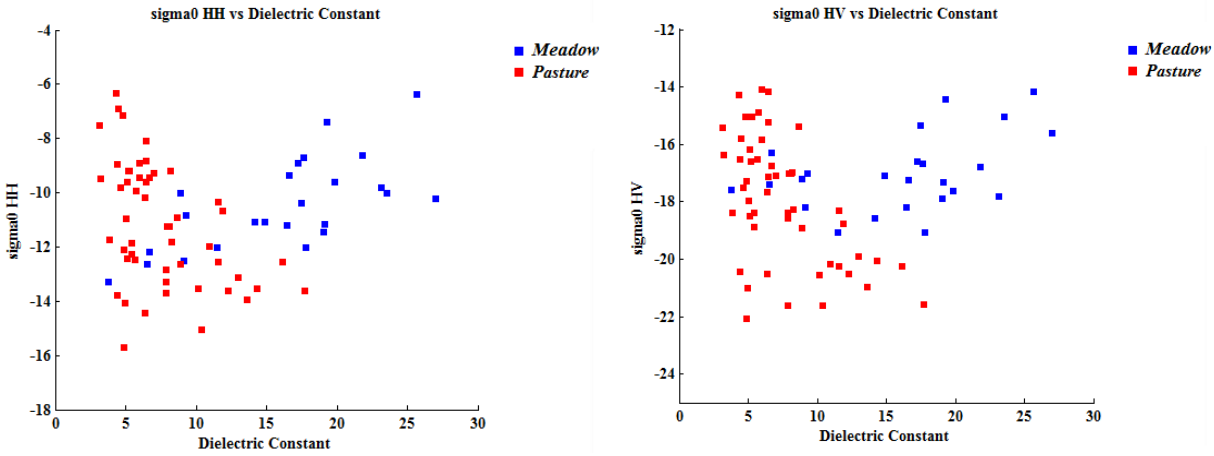
To carry out the analysis presented in this work, ancillary data have been exploited. In greater detail:

- A *digital elevation model* (DEM) with high spatial resolution (2.5x2.5 m^2) obtained from the processing of airborne LIDAR acquisitions over the whole South Tyrol area during a measurement campaign in 2008. The DEM can provide information on the local topography of the area, in particular the local altitude and the local incidence angle of the SAR signal.
- Two *normalized difference vegetation index (NDVI) maps* extracted from two cloud free images acquired by the NASA MODIS sensor onboard the Terra satellite. MODIS images have been selected as close as possible to the RADARSAT2 satellite overpasses (i.e., within ±2 days from the RADARSAT2 acquisition) depending on the cloud coverage over the Mazia valley. This was possible thanks to the high revisit frequency of the MODIS sensor. The spatial resolution in the red and near-infrared bands (the portions of the spectrum necessary for the computation of NDVI) is 250x250 m^2. NDVI is a well-known vegetation index sensitive to the vegetation phenology and growth status. Even though the nominal resolution is coarse, MODIS NDVI provides an average indication on the status of the vegetation coverage in meadows and pastures.
- A high resolution (25x25 m^2) land-cover map of the Mazia valley derived from ortophotos, ground surveys and visual interpretation. This information allows distinguishing between pasture and meadow areas, which differ not only for the vegetation type but also for the soil surface characteristics, especially in terms of roughness, due to the different human management.

Ancillary data have been geocoded using the same projection of the SAR imagery (UTM WGS-84) and re-sampled with a bilinear convolution method in order to be completely overlapped with the RADARSAT2 images.

5.3 Sensitivity Analysis

To carry out the analysis presented hereafter, a reference data set was first build. It consisted of couples of soil dielectric constant measurements and the corresponding features extracted from the SAR images. The latter task was addressed by defining a 3x3 pixels region of interest of the area corresponding to each field measurement. Pixel values of each region were averaged and then associated with the soil measurement acquired on the ground. Samples associated with foreshortening and layover were discarded from the analysis. Considering both acquisition dates and both pasture and meadow land-cover types, a data set of 75 samples was finally obtained.



(a) (b)

Figure 5.4. Scatter-plots of backscattering coefficients extracted from RADARSAT2 imagery versus dielectric constant measurements in case of (a) HH polarization configuration and (b) HV polarization configuration.

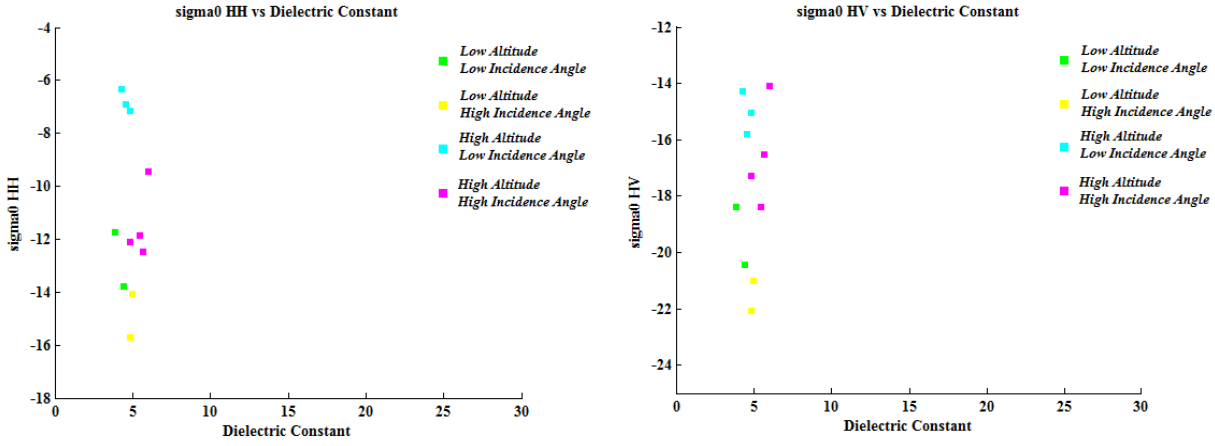
The data set was exploited for investigating the sensitivity of RADARSAT2 backscattering coefficients to different levels of soil moisture content with the help of scatter-plots. Figure 5.4 shows the plots in case of HH and HV backscattering coefficients (on the vertical axis) as function of dielectric constant values measured on the ground (on the horizontal axis). Analogous results have been achieved for the VV and VH configurations. The points are differentiated according to the land-use class derived from the ancillary data.

From a first analysis it is possible to observe that the points associated to meadows present an expected increasing trend versus dielectric constant values (more evident in the case of the HH with respect to the HV polarization). On the contrary, no clear trend can be recognized in the samples associated to pastures. In greater detail, these samples show a high level of ambiguity (i.e., samples with similar dielectric constant values present significant differences in terms of backscattering coefficients) especially for low dielectric constant values. As explained previously, different target properties and external factors may affect the microwave signal acquired by the satellite sensor. Taking into account the environmental conditions observed during the field measurement campaigns, two factors can be considered as mainly responsible for the variability and ambiguity observed in the pasture samples: i) topography and ii) heterogeneity of vegetation/land-cover. In the following, these two aspects are better investigated with the help of ancillary data, in order to understand whether and to what extent they affect the RADARSAT2 measurements.

5.3.1 Effect of Topography

As explained previously, topography significantly affects the signal acquired by a satellite SAR system. In our case, despite the calibration of the signal was carried out with the help of an accurate digital elevation model, residual topographic effects are expected to introduce significant ambiguity in the backscattering coefficients. This is expected especially for pastures, since they extend over large portions of the valley sides, with altitudes ranging from 1600 to 2400 meters. On the contrary, meadows are mainly located in the valley floor, thus they present similar topographic conditions.

In order to investigate the effect of topography on the backscattering signal, the digital elevation model was exploited for the extraction of two features related with the local topography of the investigated scene: i) the local incidence angle of the SAR signal (i.e., the angle between the line of sight of the SAR sensor and the direction normal to the surface within the resolution cell); and ii) the local altitude. Samples corresponding to pasture areas (which demonstrated the highest ambiguity in the SAR signal, as shown in Figure 5.4) were divided into various dielectric constant classes (e.g., below 4.5, between 4.5 and 5.5, between 5.5 and 6.5, and so on) in order to keep constant this variable in the analysis. Then, samples of each class were grouped into four clusters: 1) low altitude/high incidence angle; 2) low altitude/low incidence angle; 3) high altitude/high incidence angle; 4) high altitude/low incidence angle. Intermediate conditions were excluded from the analysis. Figure 5.5 shows the resulting scatter plot for values of dielectric constant between 4.5 and 5.5 (which demonstrated the highest variability in the backscattering coefficients) and both HH and HV polarization configurations. In the plots it is possible to observe that samples with similar characteristics in terms of altitude and local incidence angle are quite close one to each other and located in specific portions of the feature space. Samples acquired in areas with low altitude and high local incidence angles of the SAR signal present the lowest values of backscattering coefficient. On the contrary, samples characterized by high altitude and low local incidence angle show the highest backscattering coefficients. The difference between these two extreme topographic conditions is particularly enhanced and can be quantified in 8-9 dB for both HH and HV polarization configurations. Analogous results were obtained for the other dielectric constant ranges. The samples with intermediate topographic characteristics, i.e., low altitude and low incidence angle and high altitude and high incidence angle, are located between these two extremes. It emerges that both the local incidence angle of the SAR signal and the local altitude of the investigated area affect the backscattering coefficient, introducing attenuation or increase of its value. However, a certain level of variability still remains in the data, as can be observed for example in the cluster of samples associated to high altitude and high local incidence angle. This suggests that topography is not the only factor that affects the SAR signal in these environmental conditions.



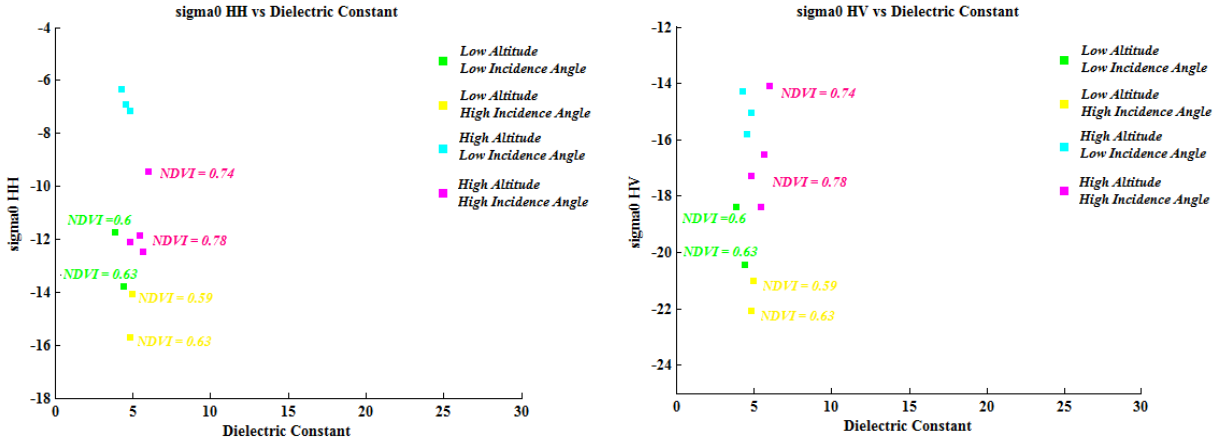
(a) (b)

Figure 5.5. Scatter-plots of backscattering coefficients versus dielectric constant measurements over pasture areas and dielectric constant values between 4.5 and 5.5: (a) HH polarization configuration; and (b) HV polarization configuration. Samples are grouped into 4 clusters according to local altitude and local incidence angle features derived from the DEM.

5.3.2 Effect of Vegetation/Land-cover Heterogeneity

As it was observed in the Mazia valley during field campaigns, the Alpine landscape is characterized by a high variability and heterogeneity in terms of vegetation/land-cover. Meadows, located in the valley floor, are intensively farmed and irrigated. The soil is typically homogeneous, flat in terms of roughness and the grass is typically thick. Cut events during the summer period determine variations in the biomass of the vegetation coverage. Pastures have completely different characteristics. First of all, they are located on the sides of the valley where the terrain becomes steep and the altitude increases. The soil is heterogeneous, with the presence of stones and in some cases of large rock's areas when the altitude becomes higher. Also the vegetation coverage is irregular, presenting areas with a significant presence of grass and others less vegetated or quite bare.

Vegetation influences the microwave signal by introducing an attenuation effect with respect to bare soils, as indicated in several studies (*Lakhankar et al. 2009*). On the contrary, the presence of stones and rocks as well as the irregularity of the surface may increase the backscattering coefficient values, due to both multiple reflections and the high irregularity of the surface. Thus, these two factors may explain the residual ambiguity and variability observed in the SAR signal after taking into account the topographic effects. In order to verify this hypothesis, we exploited the normalized different vegetation index (NDVI) extracted from two MODIS Terra satellite images acquired as close as possible to the RADARSAT2 overpasses. This index is sensible to variations in the green leaf vegetation and thus in biomass. For the purposes of our analysis, it can be exploited as proxy to quantify the land-



(a) (b)

Figure 5.6. Scatter plots of backscattering coefficients versus dielectric constant measurements over pasture areas and with dielectric constant values between 4.5 and 5.5: (a) HH polarization configuration and (b) HV polarization configuration. NDVI values are shown for the samples which show strong residual variability in the backscattering coefficient value.

cover heterogeneity and the presence of vegetation typical of the Alpine area. In particular, this index will have the highest values in presence of meadows with dense and tall vegetation, while the value will progressively decrease moving to cut meadows or pastures with lower vegetation coverage and an increasing presence of rocks. NDVI values were associated to the samples presenting similar characteristics in terms of dielectric constant value, topography and land-use class (meadow or pasture) but showing a residual variability in the backscattering values. For the sake of brevity, in this chapter we will present the analysis just for the samples of Figure 5.5, but good agreement was found also for the other cases.

Plots shown in Figure 5.6 suggest that the NDVI can explain the residual variability within the samples of each topographic cluster. In particular, for each class of topographic conditions (e.g., high altitude/high incidence angle), it is possible to observe that lower NDVI values are associated to higher backscattering values and vice-versa. This confirms the hypothesis that also the vegetation/land-cover heterogeneity affects the SAR signal in the investigated area. It is worth noting that the NDVI map considered for the analysis presented above is characterized by a quite coarse spatial resolution (250 meters) with respect to both the SAR images and the heterogeneity of the landscape. However, it provided useful indications (at least qualitative) for explaining the variability inside the SAR signal.

5.4 Proposed Retrieval System

The sensitivity analysis presented in the previous section suggests that the backscattering coefficient measured by the RADARSAT2 SAR sensor is sensitive to variations in the dielectric constant of soils, thus to variations in the moisture content. However, the microwave signal is also strongly affected by the topography of the area (also after standard topographic correction) and the heterogeneity of vegetation coverage and land-cover. These factors increase the complexity of the retrieval of soil moisture content in mountain areas. Estimation approaches based on the inversion of theoretical based models may be not effective in this operational condition. Despite advanced electromagnetic models have been proposed in the literature for handling complex targets (e.g., vegetated fields, *Guoquing and Rason, 1995*), they are not able to model accurately all the non-linearity and complexity typical of the Alpine environment. Moreover, the high number of input parameters typically required by these advanced models further increases the ill-posed nature of the inversion process.

In this chapter we propose address the retrieval problem by inferring an empirical mapping between SAR backscattering coefficient and target parameter from the set of available reference samples collected during the field campaigns. To deal with the ambiguity that characterizes the SAR signal, features extracted from ancillary data are provided as additional input to the retrieval technique. In particular, MODIS NDVI, land-cover class, local altitude and incidence angle are exploited, as they well explained the ambiguity in the SAR backscattering coefficient shown during the sensitivity analysis. The combined use of SAR backscattering coefficient and ancillary features will allow the retrieval algorithm to better disentangle the contribution of soil moisture to that of other soil properties on the SAR signal. With regard to the retrieval algorithm, the ε-insensitive Support Vector Regression technique is used. SVR has shown good accuracy, stability and generalization ability also while dealing with complex non-linear problems and in presence of a limited number of reference samples (see Chapter 2). Moreover, it easily handles high dimensional input spaces, also with features extracted from various sources and with different distributions. These features make it a good candidate to deal with the soil moisture retrieval in the challenging Alpine environment

Going more into detail, the retrieval process is divided into two phases: i) the training phase; and ii) the estimation phase (Figure 5.7). During the training, a subset (i.e., 60) of the available samples (i.e., the measurements acquired during the field campaign associated to the corresponding RADARSAT2 backscattering coefficient values and the ancillary features) are exploited as reference set and provided to the SVR regressor in order to learn the mapping between input features and target variable. Reference samples were selected in order to be well representative and equally distributed in terms of spatial position within the valley and variability of the target parameter values. Note that in this phase the target variable is the soil

dielectric constant, since it represents the physical measurement acquired on the ground during the field campaigns. Concerning the SAR backscattering features, only the HH polarization signal is considered at this stage. A thorough discussion about feature extraction and selection will be provided in the next chapter. The tuning of the free SVR parameters is addressed with the multi-objective model-selection strategy presented in Chapter 3 together with a 5-fold cross-validation procedure. This ensures the robustness of the model-selection process also with ambiguous data and limited availability of reference samples. *MSE* and *slope* are the metrics considered in the optimization process, while the ranges for the model parameters are set as follows: $[10^{-3}; 10^{3}]$ for the Gaussian RBF kernel width γ, $[10^{-3}; 10^{3}]$ for the regularization parameter C and $[10^{-3}; 10]$ for the width of the insensitive tube ε. The selection of the final parameter configuration is done on the basis of a visual inspection of the Pareto front.

After the training phase, the regressor can be used for the estimation of target variable values from both unseen input samples and remotely sensed images (the estimation phase). The output of the regressor is then converted to soil moisture content values according to the relationship shown in Figure 5.3. It is important to notice that in this phase the same input feature configuration of the training phase is needed.

Regarding the implementation of the proposed retrieval system, the LibSVM library (Chang and Lin, 2011) was used for the SVR technique, while an ad hoc Matlab code was developed for the multi-objective model-selection strategy.

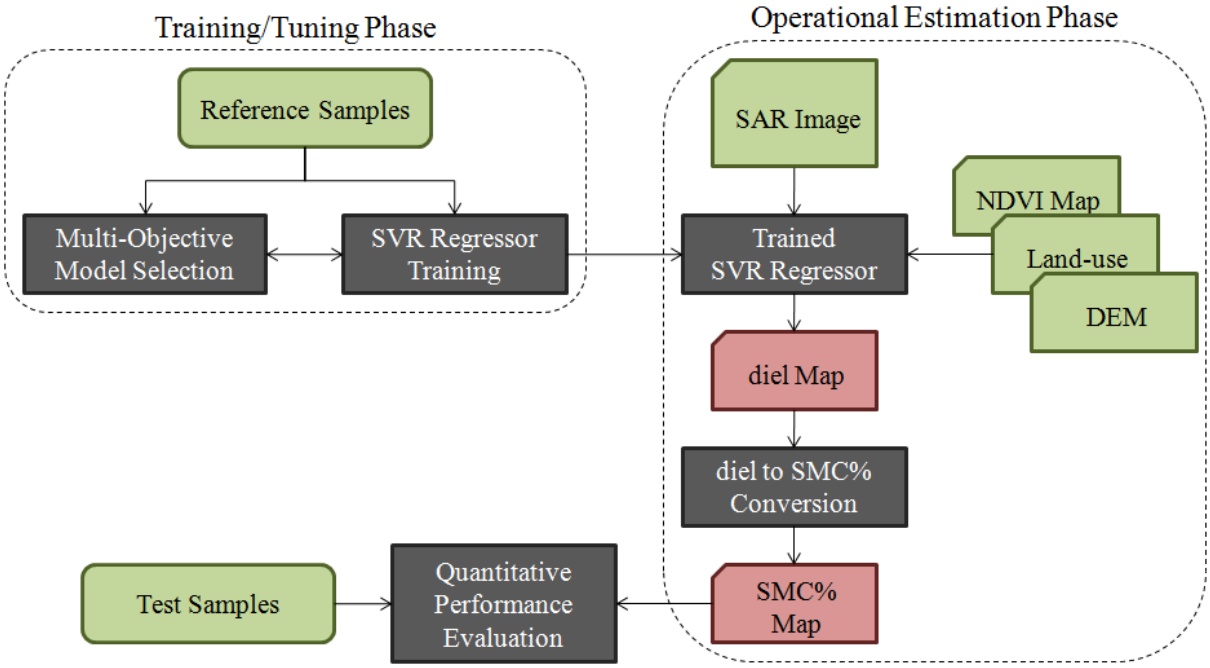


Figure 5.7. Block scheme of the proposed retrieval method.

5.5 Retrieval accuracy Assessment

5.5.1 Quantitative Assessment with Point Measurements

To assess the effectiveness of the proposed system for the retrieval of soil moisture content in mountain areas, we first analyzed its performance over point measurements. To this aim, 15 reference samples (independent from those used in the training phase) were provided to the system in the estimation operational phase. Estimated values were then compared with reference measurements through four quality metrics: Root Mean Squared Error (*RMSE*), determination coefficient (R^2), *slope* and *intercept* of the linear fit between estimated and reference target variable values. Table 5.2 reports the values achieved by each metric, while Figure 5.8 shows the scatter-plot of estimated versus measured soil moisture content values.

Table 5.2. Estimation performance achieved by the proposed retrieval system on independent (test) reference samples.

RMSE	5.38
R^2	0.79
Slope	1.20
Intercept	-2.60

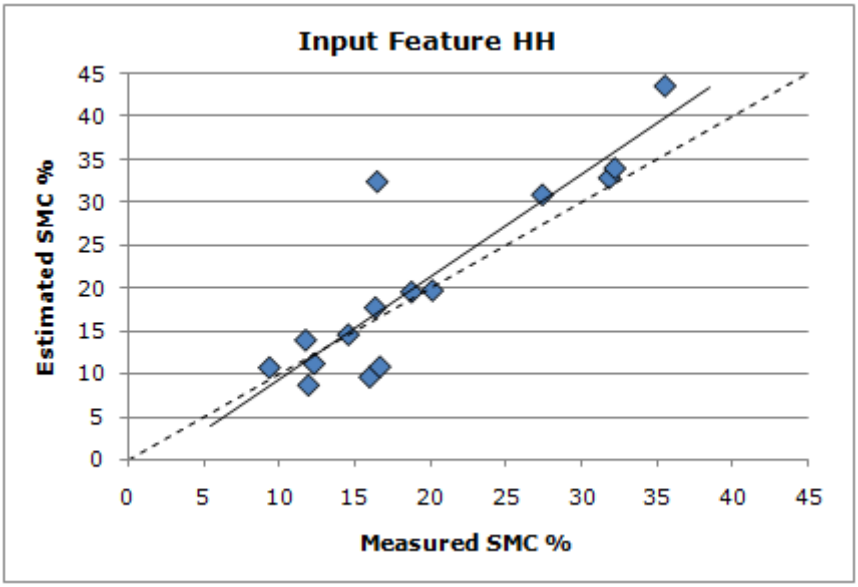


Figure 5.8. Scatter-plot of estimated versus measured soil moisture content (SMC %) associated with independent (test) reference samples. Linear fit is represented by the black line, while the dashed line shows the 1-to-1 relationship.

Globally, the achieved accuracy is very promising, with a RMSE of 5.38. Operational soil moisture retrieval algorithms consider an RMSE of 5.0 as target accuracy. Considering the complexity of the retrieval process in the mountain area, this result suggest the effectiveness of the proposed method. Also the linear fit parameters indicate the good performance of the method, with a determination coefficient close to 0.8. This trend is confirmed by the analysis of the associated scatter plot, which shows a limited spread of the estimated samples. A slight overestimation can be observed for high soil moisture content values, which are associated with irrigated meadow areas. This effect is probably due to: i) the range of variability of the target variable, which is much larger in the case of meadows with respect to pastures; and) the number of reference samples, which is lower in the case of meadows with respect to pastures (see Table 5.1). Both these factors may increase the complexity of the retrieval problem in case of meadows.

5.5.2 Qualitative Assessment with Distributed Maps

Together with point estimates, the proposed retrieval system was used for the generation of distributed maps of soil moisture content over the area of interest. The aim is to assess the effectiveness to reproduce expected spatial and temporal patterns of soil moisture content. Estimated maps associated with the two RADARSAT2 images acquired in June and July 2010 are shown in Figure 5.9. Black areas (no value) correspond to masked regions, mainly forest, water bodies, rocks and urban areas, according with the land-use mask.

From a qualitative viewpoint, the maps reproduce well the expected trend of soil moisture content, presenting high values near to the valley floor (where the irrigated meadows are located) and progressively decreasing values moving to the pastures at higher altitudes. At the same time, the humidity patterns are well recognized, as for example in the case of the small rivers going down to the valley floor along the side shown in the details of the maps (Figure 5.8 (a), Zoom 2 and Figure 5.8 (b), Zoom 2).

A comparison between the map of June and that of July indicates that the soil in the second date presents a drier behavior, especially in the lower part of the valley side, as can be observed in the details shown in Figure 4.8. This trend is confirmed by the field measurements carried out in the areas during the two campaigns, as indicated in Section 5.2.3. High variability of the moisture patterns can be observed especially in the July map, with extreme wet and dry conditions close to each other. This behavior can be ascribed to the irrigation practice, which may determine strong variations in the moisture content of soil between irrigated and non-irrigated fields. However, it is not excluded that residual ambiguity within the SAR signal still affects the retrieval process, determining too strong variations between dry and wet regions. This issue will be further investigated in the next chapter.

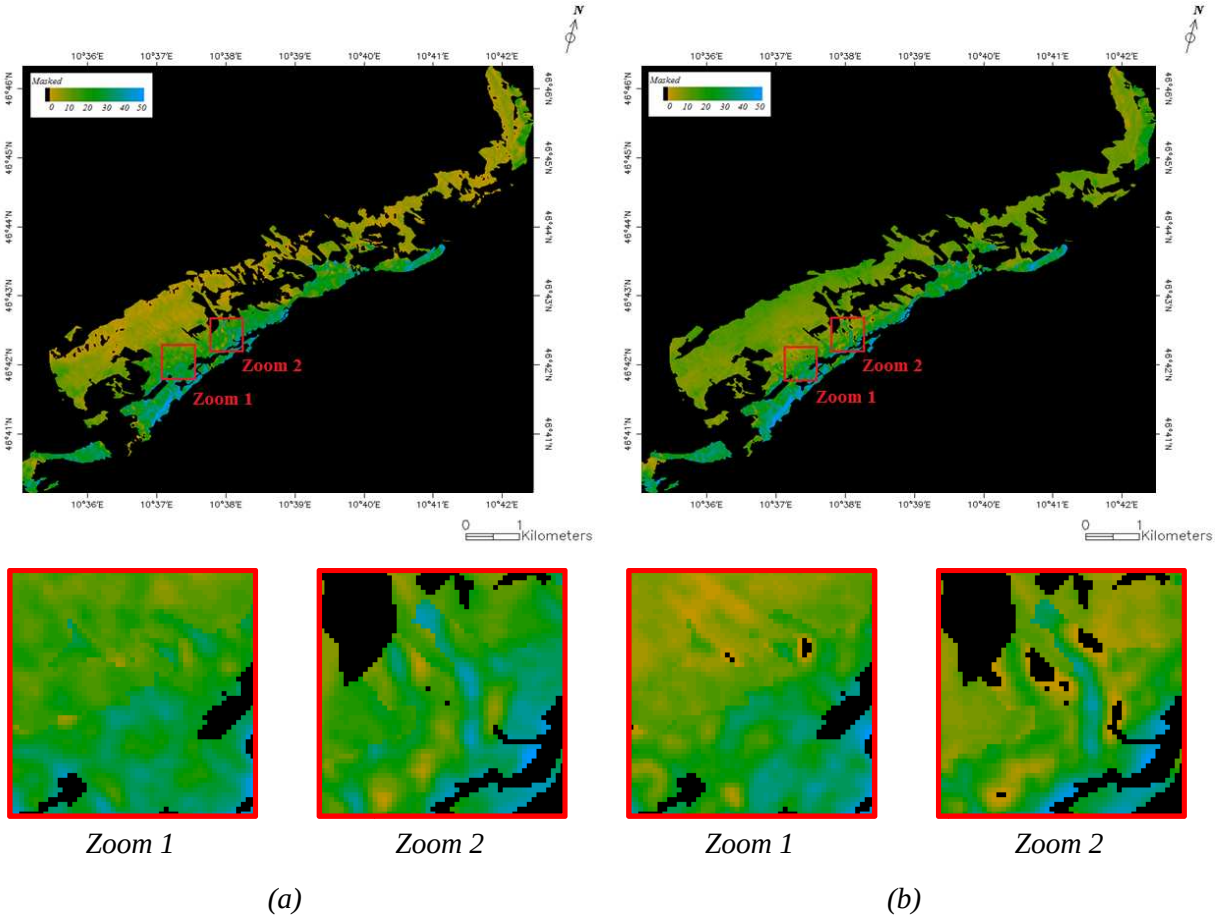


Figure 5.9. Maps of soil moisture content (SMC %) derived from satellite RADARSAT2 SAR imagery acquired over the Mazia valley with the proposed retrieval system: (a) 3rd June 2010 and (b) 21st July 2010. Red boxes depict details of each map.

5.6 Conclusion

In this chapter, new generation RADARSAT2 SAR images are exploited for the estimation of soil moisture content in an Alpine catchment. First a sensitivity analysis with the exploitation of field measurements of the target parameter and ancillary data is carried out. This analysis pointed out that both topography and vegetation/land-cover heterogeneity strongly affects the backscattering signal acquired over mountain areas, introducing a significant ambiguity in the data. The altitude, the local incidence angle and the NDVI revealed to be useful features to explain the high level of variability intrinsic in the SAR data.

The following step was the development of a retrieval system for the estimation of soil moisture content from the RADARSAT2 images. To this aim, the ε-insensitive Support

Vector Regression technique was used. Thanks to its formulation, this method is able to handle complex non-linear estimation problems with good generalization ability also when a limited number of reference samples is available. Moreover, it handles easily high dimensional input spaces, also containing heterogeneous features. These characteristics were important for exploiting the available reference samples for inferring an empirical relationship between SAR backscattering signal and soil moisture content values and integrating in the retrieval process the information extracted from ancillary data. The results achieved indicate that the proposed technique is promising in terms of: 1) capability to exploit the information provided by the ancillary data to reduce the ambiguity intrinsic into the SAR signal and address the complex estimation problem in mountain areas; 2) estimation accuracy over point reference measurements; and 3) capability to reproduce the soil humidity patterns when applied to distributed data.

Future development of this work regards first of all a better characterization of the effect of vegetation/land-cover heterogeneity on the SAR signal. This could be carried out with the help of high geometrical resolution data. In particular, the effect of rocks and stones on the microwave signal in relationship to the retrieval of soil parameters will be analyzed. A second interesting development is the exploitation of the polarimetric capability of the RADARSAT2 sensor by means of polarimetric decompositions of the signal, in order to improve the feature extraction/selection process and thus the retrieval of soil parameters. Moreover, an extended validation of the algorithm, by exploiting the measurements provided by the field stations in the Mazia valley and further RADARSAT2 SAR acquisitions over the whole Alto Adige area will be considered. Finally, the possibility to assimilate satellite soil moisture content maps into hydrological models for water resources and drought monitoring will be investigated.

Chapter 6

Effectiveness of Fully-Polarimetric SAR Data for Soil Moisture Retrieval in Mountain Areas

In this chapter, the effectiveness of fully-polarimetric SAR data for the retrieval of soil moisture in mountain areas is investigated. To this aim, three feature extraction strategies, namely the standard intensity & phase SAR processing, an advanced polarimetric decomposition method and a statistical data transformation technique, are implemented and applied to new generation RADARSAT2 satellite SAR images. The features extracted according to these strategies as well as their combination are assessed together with an advanced retrieval system for their capability to improve the retrieval of soil moisture in mountain areas. Both quantitative accuracy on a set of test reference samples and capability to reproduce the expected soil moisture content spatial patterns in a well-known test area are evaluated and discussed.

6.1 Introduction and Motivation

The sensitivity of SAR backscattering coefficients to soil moisture has been thoroughly presented and discussed in the previous chapters. A major problem when dealing with soil moisture retrieval from SAR data is the ambiguity in the signal introduced by other target properties such as soil roughness and vegetation. To deal with this issue, advanced retrieval system taking advantage from both robust regression techniques and the information derived from ancillary data can be used. Another possible solution consists in exploiting the features extracted from multi-modal acquisitions of the scene of interest.

In the literature, several studied have been proposed on the use of multi-angle and multi-frequency SAR signals for the retrieval of soil moisture (*Srivastava et al., 2009; Pierdicca et al., 2008; Notarnicola and Posa, 2007; Bindlish and Barros, 2000*). The rationale is that the SAR signal is sensitive to the target properties in a different way and to a different extent depending on the system frequency and incidence angle. To combine the information acquired with different system configurations leads to a better characterization of the factors

affecting the SAR signal and thus to a more accurate estimation of the variable of interest. More in detail, the combined use of C and L band microwave signals has shown to be particularly suitable in order to disentangle the vegetation contribution from that of the soil (*Notarnicola and Posa, 2007*), while the exploitation of both low and high incidence angle SAR images revealed useful to incorporate in the retrieval process the effects of soil roughness and texture (*Srivastava et al., 2009*). Despite the promising results achieved, multi-frequency and multi-angle scenes are very difficult to acquire from satellite platforms on a regular basis. This significantly limits the possibility to apply multi-frequency and multi-angle approaches over large areas with a frequent and constant spatial/temporal coverage.

More efficient is the acquisition of SAR imagery using various and multiple polarization configurations. Indeed, new generation satellite SAR systems offer the possibility to choose among different configurations of single-, dual- and (in some cases) quad-polarization acquisition modes. Polarimetry has shown to be useful for increasing the robustness and accuracy of soil moisture retrieval from SAR imagery. Paloscia et al. (*2008*) investigated the use of dual-polarization ASAR SAR images for the retrieval of soil moisture in the Po valley. The study showed that the use of both co- and cross- polarized backscattering coefficients could be effective for the reduction of the ambiguity within the signal due to soil roughness. Polarization coefficients can also be combined by means of simple ratios to obtain indices useful for handling the ambiguity due to vegetation in the signal (*Notarnicola et al., 2006*). When fully-polarimetric data are at disposal, one may rely on more sophisticated processing methods. For instance polarimetric decomposition techniques can be adopted. Such methods aim at separating the contribution of different scattering mechanisms within the scene of interest using both the intensity and phase information in the four polarimetric channels (*Claude and Pottier, 1996*). Among the others, the alpha/Anisotropy/Entropy (α/A/H) decomposition proposed by Cloude and Pottier (*1997*) has been widely investigated for land-use classification and geo-/bio-physical variables retrieval. A study carried out using airborne very high resolution polarimetric SAR data over a flat agricultural area pointed out the capability of the method to separate the contribution related to the humidity of the soil from that of the roughness (*Hajnsek et al., 2003*), suggesting its potential in soil parameter estimation applications.

Despite the rising interest in the use of SAR polarimetry for the retrieval of soil geo-/bio-physical variables and the promising results achieved, some issues remains still unexplored and further research is required. This is true especially in case of specific and challenging environmental conditions, such as the Alps. Here, the SAR signal is affected by the high heterogeneity of the land-coverage and the presence of topography (*Luckman, 1998*). Up to now, no investigations on the use of SAR polarimetry were carried out in this operational condition. Further research effort is thus required in the direction of a better characterization and exploitation of polarimetric SAR imagery for the retrieval of geo-/bio-physical variables in the challenging Alpine environment.

The work presented in this chapter inserts in this context. It extends the analysis presented in Chapter 5 on the retrieval of soil moisture from satellite SAR imagery in mountain areas by focussing on the fully polarimetric acquisition capability of the new generation RADARSAT2 system. The aim is to thoroughly investigate the effectiveness of polarimetry for the retrieval of the soil moisture content in mountain areas. For doing this, different feature extraction strategies are implemented, namely the standard Intensity&Phase SAR processing, an advanced polarimetric decomposition and a statistical feature transformation technique. The features extracted from the polarimetric data are then provided as input to the retrieval system already developed together with the ancillary data. The effectiveness of different feature combinations is assessed by means of quantitative accuracy metrics computed on a set of test field measurements. A visual inspection of the final soil moisture maps is also considered in order to verify the capability to reproduce soil moisture spatial patterns as observed during field surveys.

The rest of the chapter is organized as follows. Section 6.2 briefly recalls the main characteristics of the study area and the data set considered in the analysis. For more details we refer the reader to the previous chapter. An overview of the feature extraction strategies investigated in this work and the experimental setup is presented in Section 6.3. Section 6.4 discusses the experimental results obtained and Section 6.5 draws the conclusions of the analysis.

6.2 Study Area and Data Set

The study area for the analysis presented hereafter is the Mazia valley, located in the north-western part of South Tyrol region (Northern Italy). It covers an area of around 100 km^2 with altitudes ranging from 920 meters a.s.l. (Sluderno) to 3738 meters a.s.l. (Palla Bianca). Despite the relatively small dimension, the Mazia valley is well representative in terms of geomorphology and topography for the whole South Tyrol. Excluding forests, the most diffuse land-cover/use types are meadow and pasture. They present quite heterogeneous characteristics in terms of vegetation, spatial distribution and human usage.

During the summer 2010 (on the 3rd of June and the 21st of July), two SAR images were acquired by the RADARSAT 2 satellite sensor over the Mazia valley. The acquisition mode was Standard Quad Polarization, right looking, with a nominal incidence angle of 45° and an ascending orbit. This acquisition geometry has been carefully defined in order to minimize the geometrical distortions (shadowing and layover/foreshortening) that occur in presence of mountain reliefs due to the side looking view of SAR systems. The data have been multi-looked, de-speckled with a Frost filter (5x5 window size), calibrated and geocoded (UTM WGS-84) with the help of a high spatial resolution (2.5x2.5 m^2) digital elevation model. The final resolution of the processed images is 20x20 m^2. All pre-processing steps have been carried out with the SARscape® software (www.sarmap.ch).

Contemporary to the two satellite acquisitions, field measurement campaigns have been carried out in the study area. Non-destructive measurements were acquired using the mobile Delta T WET 2 TDR sensor (www.delta-t.co.uk), which measures the top five centimetres soil dielectric constant. The sampling was performed selecting homogeneous areas of approximately 30x30 m^2 representative in terms of local topography, vegetation status and land-cover/use type. Repeated measurements (three to five) of soil dielectric constant distant few meters one to each other were collected and averaged. Soil dielectric constant measurements in more than 100 sampling areas were collected. Along with TDR measurements, destructive measurements were collected in few representative sampling areas. They were performed by physically taking a soil sample with a 5 cm height metallic cylinder. The soil sample was weighted, sealed and carried in laboratory in order to be dried according to standard soil measurements protocols (*Cihlar et al., 1987*). These measurements were used to derive a calibration curve for the TDR sensor (we refer to Chapter 5 for more details on field measurements).

Ancillary data are also considered. As shown in the previous chapter, they help the retrieval system disentangling the contributions of topography and vegetation/land-cover heterogeneity to that of soil moisture in the SAR signal. More in detail, a digital elevation model (DEM), two NDVI maps derived from MODIS satellite imagery and a high resolution land-cover map are used.

6.3 Design of the Experiments

6.3.1 Feature Extraction Strategies

Feature extraction is a core step in the definition of an estimation system. It aims at extracting from the input data the most suitable variables to provide as input to the retrieval algorithm for the estimation of the desired target variable. Unsuited or not sufficient variables as input to the retrieval algorithm might lead to poor estimation performance. To fully take advantage of the available fully-polarimetric SAR data, three different feature extraction strategies have been implemented in this work:

- *Standard Intensity&Phase SAR Processing*
  SAR images have been processed according to a standard Intensity and Phase processing chain in order to extract standard polarimetric features from the scene of interest. In a greater detail, the following features have been considered: 1) the backscattering coefficients at four different polarimetric channels (HH, HV, VH and VV); 2) the ratios among co- and cross-polarized backscattering coefficients (the polarization ratio HH/VV and the linear depolarization ratio HV/VV); and 3) the polarimetric phase difference (PPD) among HH and VV channels.

- *Polarimetric Decompositions*
  The main idea behind theoretical decomposition theorems is to express the scattering of the investigated scene as a sum of independent elements and to associate a physical mechanism with each component. Among different decomposition strategies proposed in the literature, the alpha/Anisotropy/Entropy technique proposed by Cloude and Pottier (*1997*) has become popular for the analysis of polarimetric SAR data and has been implemented as feature extraction strategy in this analysis. It is based on the eigen-decomposition of the polarimetric coherency matrix and allows one to extract three parameters: the alpha angle α, the scattering entropy H and the scattering anisotropy A. These parameters express the characteristics of the scattering mechanisms and the degree of randomness of the scattering process. For the estimation of the α/A/H parameters from the polarimetric RADARSAT2 single look complex (SLC) images, the SARscape® software has been used, setting a 5x5 sliding window.

- *General Purpose Feature Extraction techniques*
  When dealing with feature extraction problems with multi-dimensional data (such as fully-polarimetric data) affected by multiple information sources (e.g., soil moisture, surface roughness, vegetation, etc.) one may resort on blind approaches. Such approaches are based on a statistical analysis of the signals. An example is the Independent Component Analysis (ICA), which is a powerful method for blind signal separation (*Comon, 1994*). This method has various practical applications, among the others change detection analysis (*Marchesi and Bruzzone, 2009*). The main rationale behind this method is the maximization of the statistical independence of the estimated components (i.e., the estimated information sources mixed in the input signal). This is achieved by eigenvector decomposition and dimension reduction of the data, while the statistical independence of the components is typically measured according to the central limit theorem. Various implementations of the ICA approach have been proposed in the literature. In this work we considered the FastICA algorithm (*Van Hateren and Van Der Schaar, 2000*), which has demonstrated high computational efficiency.

6.3.2 Experimental setup

SAR images were processed for deriving polarimetric features according to the three strategies described above. Standard Intensity&Phase processing and α/A/H polarimetric decomposition were performed with the SARscape® software, while the ICA feature extraction technique was implemented using Matlab and the FastICA algorithm. In each output map, the areas corresponding to the field measurements were identified and associated with a 3x3 pixels region of interest. The pixel values of each region were then averaged and associated with the dielectric constant values measured on the ground. The same was done for the features extracted from ancillary data (i.e., local altitude and incidence angle, NDVI and

land-cover class). Regions affected by residual geometric distortions (in particular layover) were excluded. At the end a data set of 75 samples (i.e., couples of feature values and corresponding soil dielectric constant measurements) was obtained.

The available data set has been divided into two subsets: 1) a reference set with 60 samples and 2) a test set with 15 samples. Reference and test samples were selected in order to be well representative and equally distributed in terms of spatial position within the valley and variability of the target parameter values.

The reference set was exploited for the training/tuning of the retrieval system proposed in the previous chapter (see Section 5.4 for more details). For consistency with the previous analysis, the same settings regarding the SVR technique and the tuning of the free parameters were maintained. After the training/tuning phase, the algorithm was run on the independent test samples. Then output target values were converted in soil moisture content. This allowed us to quantitatively assess the estimation performance by means of four different metrics: the mean absolute error *MAE*, the root mean squared error *RMSE*, the determination coefficient R^2, the *slope* and *intercept* of the linear trend between estimated and measured target values. Finally, the algorithm was applied to the satellite imagery for obtaining the target variable map.

The analysis carried out was organized in three sets of experiments, one for each feature extraction strategy investigated. For each set various combinations of features extracted from the polarimetric SAR imagery have been defined and used (together with the features derived from the ancillary data) as input to the retrieval algorithm for its training and testing as described above. The retrieval performance with the different input feature configurations are then analyzed and compared.

6.4 Experimental Results

The results achieved terms of estimation accuracy (*MAE* and *RMSE*) and goodness of fit (R^2, *slope* and *intercept* of the linear trend line) as well as the corresponding SVR parameter configurations (kernel width γ, regularization parameter C and insensitive tube width ε) for the most significant input feature combinations investigated are reported in Table 6.1. The metrics reported are evaluated on the independent test samples. For each feature extraction strategy, the best feature configuration investigated was selected and the corresponding scatter plot of estimated versus measured test target values is shown in Figure 6.1.

For comparison purposes, the case of a single polarimetric channel (the HH polarization, discussed in the previous chapter) was considered and shown as a benchmark. It is well established that HH polarization, among single polarization channels, is the most sensitive and suitable for soil moisture content retrieval (*Macelloni et al., 1999; Ulaby et al., 1979; Paloscia et al., 2008*). In the following, the results associated with each investigated feature extraction strategy are compared and discussed.

Table 6.1. *Performance achieved on test reference samples with various input features configurations extracted according to the three investigated extraction strategies.*

Input Feature Configuration	**SVR Parameters**			**Estimation Performance**				
	γ	C	ε	MAE	RMSE	R^2	Slope	Intercept
Reference								
HH	*0.11*	*5.07*	*0.012*	*3.56*	*5.38*	*0.79*	*1.20*	*-2.60*
Intensity & Phase Features								
HH + HV	0.055	9.98	0.078	3.70	5.27	0.78	1.17	-2.04
HH + HV/VV	*0.046*	*10.14*	*0.17*	*3.51*	*4.85*	*0.81*	*1.15*	*-1.66*
HH + PPD	0.21	50.89	0.12	4.37	6.67	0.59	0.93	3.55
HH + HV + VH + VV	0.024	23.1	0.069	3.72	5.42	0.79	1.22	-3.31
Polarimetric Decomposition Features								
H	0.002	91.12	0.18	4.44	5.67	0.6	0.84	3.89
α + H	0.001	328.07	0.17	4.41	5.64	0.6	0.84	3.62
A + H	*0.002*	*35.92*	*0.23*	*3.87*	*5.06*	*0.63*	*0.75*	*5.3*
α + A + H	0.002	31.35	0.26	4.07	5.16	0.63	0.77	5.28
Independent Component Analysis Features								
ICA1	0.29	2.25	0.004	4.95	7.11	0.79	1.4	-5.16
ICA3	0.002	22.73	0.129	4.09	5.09	0.69	0.85	1.02
ICA1 + ICA3 + ICA4	0.269	0.338	0.002	4.09	5.03	0.69	0.91	0.69
ICA1 + ICA2 + ICA3 + ICA4	*0.088*	*1.71*	*0.13*	*3.72*	*5.06*	*0.71*	*0.98*	*1.26*

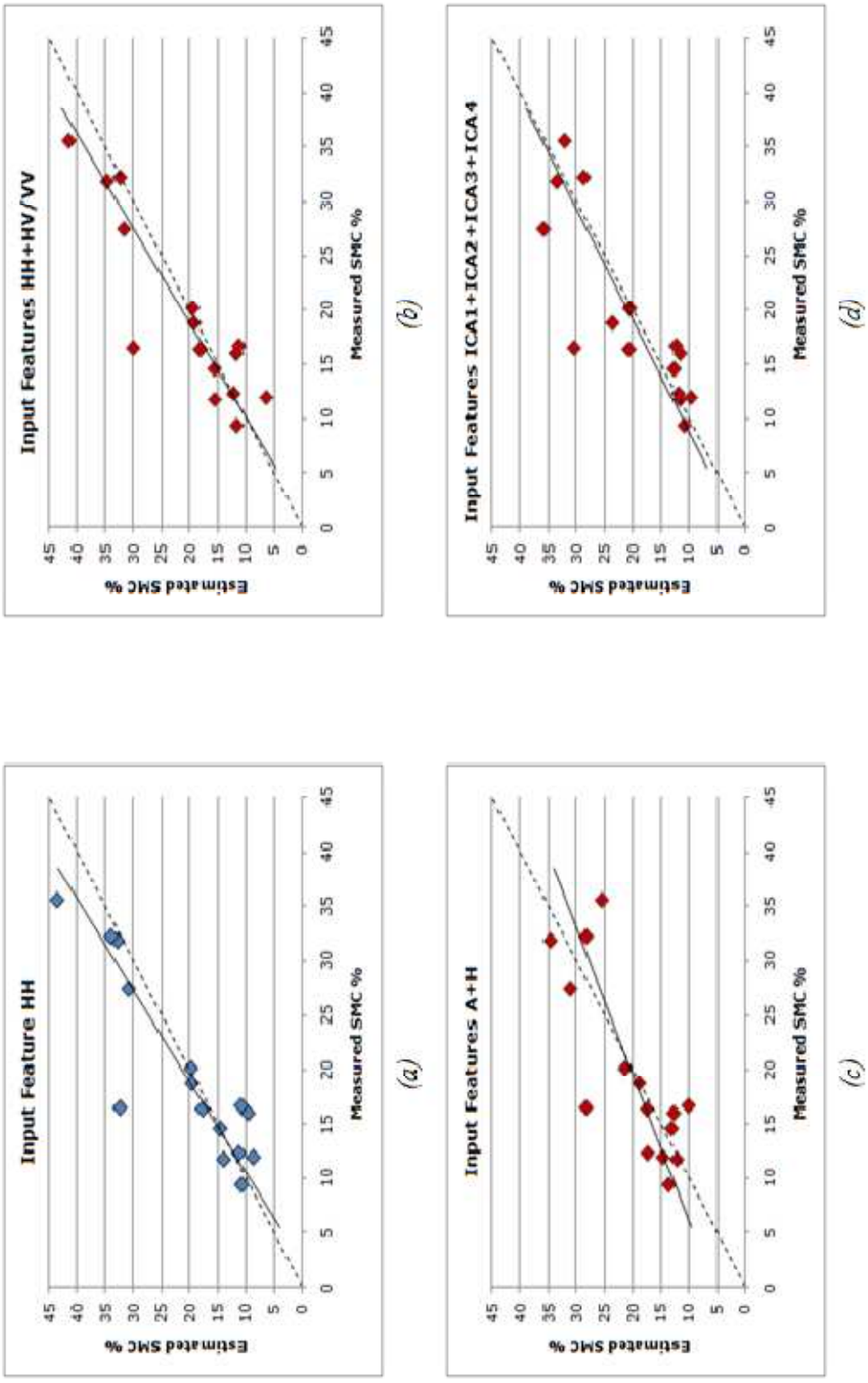


Figure 6.1. *Comparison of estimated versus measured soil moisture content by using: (a) HH feature, (b) HH+HV/VV features, (c) A+H features, and(d) ICA1+ICA2+ICA3+ICA4 features.*

6.4.1 Experiments 1: Intensity & Phase Processing

In case of standard Intensity & Phase polarimetric features, the joint use of multiple polarimetric features determines in general a slight improvement in the estimation performance with respect to the case in which only the HH polarization is available. The exception is represented by the use of the polarimetric phase difference (PPD), which allows reducing the overestimation trend shown for high soil moisture content values in the benchmark configuration (Figure 6.1(a)) but determines a degrade in terms of accuracy (both *MAE* and *MSE*) and capability to explain the variability of the measured target values (R^2). The best performance is achieved when the HH+HV/VV configuration is provided as input to the retrieval algorithm. This behaviour suggests and confirms the effectiveness of the linear depolarization ratio for the retrieval process in these specific environmental conditions, i.e., Alpine meadows and pastures. This feature helps the retrieval algorithm explaining the ambiguity introduced by the local scale vegetation and roughness conditions, while it minimizes the effects of the heterogeneity thanks to the ratio. Less effective seems to be the use of all the polarization channels (HH+HV+VH+VV) as input to the retrieval algorithm. This can be explained by considering that each channel introduces in the estimation process a certain level of noise and ambiguity without providing at the same time significant information on the target variable, i.e., the soil moisture. It is well recognized that HH is the most sensitive channel to soil moisture content. HV and VH channels are strongly correlated one to each other, thus they care the same information to the retrieval algorithm, but also a noisy contribution that could be not the same. The VV channel instead has shown to be problematic for the retrieval of soil moisture in vegetated areas due to the strong absorption of the signal in this channel determined by the vegetation layer (*Paloscia et al., 2010*). As a consequence, to use the simple polarization channels might introduce more ambiguity than information to the retrieval problem. This result suggests the importance not only of a proper feature extraction, but also of an effective feature selection for addressing a retrieval problem.

6.4.2 Experiments 2: α/A/H Polarimetric Decomposition

As can be observed in Table 6.1, the joint use of multiple features extracted with the Cloude-Pottier decomposition allows sometimes increasing the estimation accuracy of the retrieval algorithm, especially when anisotropy (A) together with other features is considered. However, what clearly emerges from the results is the reduced capability of the retrieval system to explain the variability and reproduce the dynamic of target values when α/A/H polarimetric features are considered with respect to the benchmark case. This is confirmed by the lower R^2 values and the under- and over-estimation trend for high and low target values, respectively, shown in Figure 6.1(c). For explaining this results one has to consider the fact that the α/A/H are computed by means of a 5x5 pixels sliding window, thus averaging the contributions coming from a relatively wide area on the ground. The outcome is a set of coefficients that are well representative of the mean scattering behaviour of the ground, but

less sensitive to small scale variability and patterns. This could affect the retrieval performance in the highly heterogeneous Alpine environment. Another possible factor affecting the computation of the α/A/H polarimetric features is the topography. Despite the alpha and entropy quantities are roll-invariant, the anisotrophy has shown a variable behaviour as a function of the local incidence angle (*Park and Moon, 2002*). This is the result of the different scattering mechanisms that are observed by the SAR system by changing the local incidence angle (*Touzi, 2007*). This effect could determine another source of ambiguity for the retrieval algorithm. The retrieval is in average quite accurate, but the deviation of the estimated target values increases moving to the lowest and highest values.

6.4.3 Experiments 3: Independent Component Analysis

Features extracted according to the ICA technique do not improve significantly the retrieval of the soil target variable. The third ICA component seems to be the most sensitive to the soil moisture content and leads to a slight improvement in terms of accuracy with respect to the benchmark configuration. When additional ICA component are considered, the accuracy does not vary significantly. The slope of the linear trend line improves becoming close to 1 for the ICA1+ICA2+ICA3+ICA4 configuration. This indicates a good capability to reproduce the dynamic range and the trend of the target variable without under- and over-estimation. As a blind signal separation method, ICA performs a statistical analysis of the signal and extracts those components that maximize the statistical independence. While doing this, the noisy components of the signal could be discarded. From one hand, this leads to a set of features able to well describe the average trend of the target variable. From the other hand, it is possible that a higher order variability within the signal due to soil moisture small differences is not recognized as signal and thus assumed as noise. This is realistic, considering the numerous factors affecting the SAR signal in an Alpine environment (e.g., topography and vegetation) and the limited number of channels provided as input to the ICA algorithm. This could explain the quite low determination coefficient values obtained with the ICA1+ICA2+ICA3+ICA4 feature configuration.

6.4.4 Qualitative Assessment with Soil Moisture Content Maps

The feature configuration showing the best quantitative performance on test samples have been selected for the comparison of the associated target variable maps. Again, the map obtained considering only the HH polarization channel has been considered as a benchmark. The maps corresponding to the RADARSAT2 acquisition on July, 21st, are shown in Figure 6.2. Black values represent areas that were excluded from the processing since out the study area or associated with forest land-cover.

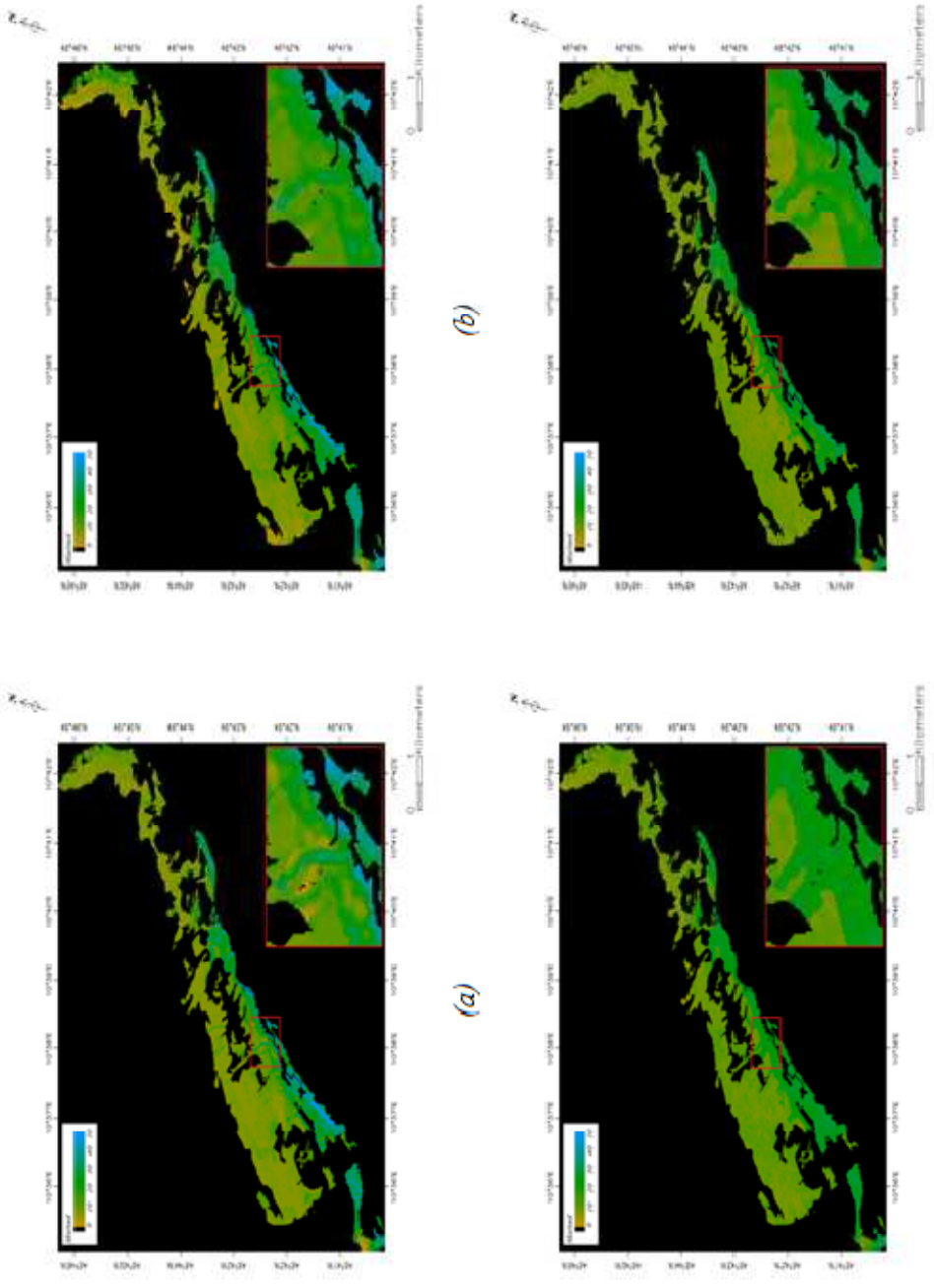


Figure 6.2. Soil moisture content maps obtained with: (a) HH input feature; (b) HH+HV/VV input features; (c) A+H input features; and (d) ICA1+ICA2+ICA3+ICA4 input features.

From a general viewpoint, the four maps show similar macro-patterns, with lower soil moisture content in pastures and at higher altitudes, while higher soil moisture content close to the valley floor in areas of highly managed meadows. This suggests a capability to individuate the average trends for all the feature extraction strategies. However, the map associated with the *A+H* features (Figure 6.2(c)) clearly shows a smoother behaviour, confirming the trend highlighted by the analysis of the test samples estimates.

Let us now move the attention on the local soil moisture patterns (the red box). This particular is associated to an area corresponding to a wet vertical buffer in a pasture area along a small river and some irrigated meadows near the floor of the valley. In the case of the polarimetric decomposition features *A+H*, only the average soil moisture values are represented. Small variations in the soil moisture content are completely flattened. Different is the case of the ICA features (Figure 6.2(d)). As can be observed, the patterns are well recognized and the small scale variability is quite well represented. However, the very wet patterns that are expected in correspondence of the small river are only slightly reproduced. The map associated with the Intensity & Phase features (Figure 6.2(b)) further improves the representation of the wet buffer along the river, showing at the same time a more realistic representation of the dry patterns in the surrounding pasture area with respect to the benchmark case (Figure 6.2(a)). Despite dramatic changes of soil moisture content can be expected in the pasture area, it is not feasible such a variation in few meters. Probably this artefact is imputable to the different vegetation cover in the area that attenuates or increases the SAR signal. This effect is mitigated when the polarization ratio HV/VV feature is included in the retrieval process.

6.5 Conclusion

In this chapter we presented an analysis of different feature extraction strategies for the exploitation of fully polarimetric RADARSAT2 SAR data in the retrieval of soil moisture content in the Alpine area. In greater detail, simple Intensity&Phase polarimetric SAR processing, α/A/H polarimetric decomposition and the Independent Component Analysis technique have been investigated. For each feature extraction strategy, different combinations of features have been defined and provided as input to an advanced retrieval algorithm together with a set of features extracted from ancillary data. The effectiveness of each feature configuration has been assessed quantitatively by means of performance metrics evaluated on a set of test samples and qualitatively by comparing the associated soil target variable maps and investigating the capability to reproduce expected soil moisture content patterns.

Both the quantitative and the qualitative assessment suggest that the use joint use of multiple polarimetric features leads to an improvement in the retrieval of soil moisture content in the Alpine area. In particular, standard Intensity&Phase polarimetric processing allows extracting the most suitable features for the estimation of the desired soil moisture

variable. The use of the linear depolarization ratio HV/VV has shown to be important for the performance improvement, thanks to its capability to provide information on the local vegetation and roughness status of the target area. The other feature extraction strategies investigated imply a more time demanding processing of the SAR imagery and do not lead to improvements in the accuracy or in the capability to reproduce the local patterns of soil moisture content. This can be probably ascribed to the high ambiguity and complexity that characterize the retrieval problem in the Alpine environment. In these operative conditions, not too sophisticated methods are preferable.

As future activity, we intend extending the analysis presented with additional RADARSAT2 SAR acquisitions. In particular, the use of very high spatial resolution polarimetric imagery will be investigated. The aim is to understand if the increased spatial resolution can improve the effectiveness of polarimetry in the Alpine environment. Further validation of the experimental results will be carried out with the data provided by the measurement stations located in the valley and new field measurement campaigns.

Chapter 7

An Improved Algorithm for the Retrieval of Green Area Index from MODIS Satellite Imagery in Mountain Grasslands in the Alps

In this chapter the focus is moved to the retrieval of another important geo-/bio-physical variable, namely the Green Area Index (GAI) of mountain grasslands. An improved algorithm for the estimation of this variable from moderate resolution satellite MODIS imagery is presented. The main features of the proposed algorithm, which is based on the inversion of a radiative transfer model, are: i) the higher spatial resolution (250 meters) with respect to currently available operational products; ii) the tuning on the spectral characteristics of mountain grasslands, for which standard products present limited accuracy and reliability; and iii) the accounting for local topography. The detailed description of the algorithm is followed by a thorough discussion of the experimental analysis carried out on a series of satellite MODIS imagery acquired over the central Alps in the period 2005-2007. GAI estimates are validated for both temporal consistency and accuracy with the use of time series of ground measurements collected at three different study sites in the investigated area. For comparison purposes, the standard MODIS product is also considered. The results obtained demonstrate the capability of the proposed algorithm to follow the expected temporal and range dynamics of the target variable in the challenging Alpine environment. At the same time spatial patterns are better reproduced. The outcome of this research opens a promising avenue for the exploitation of moderate resolution satellite data for novel and more accurate monitoring studies at regional scale in mountain environments.

7.1 Introduction and Motivation

The green Leaf Area Index (LAI), defined as half the surface area of green leaves per unit of ground horizontal surface area (*Curan, 1980*), is a key structural variable of vegetation canopies. It describes the potential surface area available for leaf gas exchange between

terrestrial biosphere and atmosphere and is related to many biological processes of plants, such as the interception of light and water, photosynthesis, evapotranspiration and carbon and nutrients cycles (*Sellers et al., 1997*). When totally green vegetation canopies are considered, such as grasslands or crops, this variable is more properly referred as Green Area Index (GAI), since not only leaves, but also stems and other above-ground plant organs are green and photosynthetically active and thus contribute to the exchange of CO_2, H_2O and other trace gases (*Wohlfahrt et al., 2001*).

The availability of spatially and temporally distributed information about this parameter on a regular basis is of crucial importance for climate change studies (*CGOS, 2006; Sellers et al. 1996*), as well as for ecosystem models that quantify carbon, water and energy fluxes (*Turner et al., 2004; Running et al., 1988*), which can take advantage of the spatial and temporal trends of LAI/GAI to force or adjust model predictions by means of assimilation techniques (*Quaife et al., 2008; Dorigo et al., 2007; Moulin et al., 1998*). Although several direct and indirect methods do exist for ground based measurements of canopy structural properties (*Weiss et al., 1994*), these are typically time consuming and unfeasible for large scale monitoring over time. A promising support to ground measurements, which is exciting a growing interest in the scientific community since many years, is the exploitation of satellite Earth Observation (EO) imagery.

The retrieval of vegetation biophysical variables from remote sensing data takes advantage of the causal relationship that exists between canopy characteristics and the electromagnetic radiation it reflects toward space. However, it is necessary to properly characterize and invert such a relationship. To this end several algorithms have been proposed in the literature, which fall into the two broad chategories defined in Chapter 1:

- *Definition of empirical relationships, by means of multiple regression and/or spectral vegetation indices*. This approach has been widely used in the remote sensing community due to its simplicity and relatively good performance. However, such relationships are site- and biome-specific and their application is limited by the representativeness of the set of reference samples used for calibration. Thus they cannot be extended over large areas and different time periods without the calibration on each EO image (*Colombo et al., 2003; Meroni et al., 2004*) due to changes that may occur in canopy structure, atmospheric conditions, viewing and sun geometry.
- *Inversion of canopy radiative transfer models (RTM) by means of minimization search algorithms, look-up tables or more sophisticated machine learning methods*. RTMs are based on a rigorous physical description of the interactions between electromagnetic radiation, the canopy elements and the underlying soil surface. RTMs can simulate a great variety of conditions in terms of vegetation type and characteristics as well as sensor acquisition geometry, thus ensuring typically a higher generalization and portability with respect to empirical models. However, the inversion process is the most

of the time computationally demanding and ill-posed, due to the high number of model input parameters. Moreover, uncertainties in remote sensing measurements and approximations of the model may result in variability of the results (*Combal et al., 2003*). To cope with these issues, prior information on the area investigated (e.g., regarding the biome type or the model input parameters) can be exploited to better constrain the inversion process (*Combal et al., 2002; Lavergne et al., 2007*).

While empirical relationships are mainly used in combination with high/very high spatial resolution EO data for retrieving LAI/GAI at local and sub-regional scale, radiative transfer model inversion is often exploited with medium resolution (MR) satellite imagery. Despite the quite coarse spatial resolution (ranging from 250 m to 1 km), MR data have the great advantage of being acquired globally, on a regular basis and very frequently (typically once a day) since many years. Thus they offer a valuable database for modeling and monitoring purposes. The Moderate Resolution Imaging Spectroradiometer (MODIS) sensor on board of the AQUA and TERRA satellites is probably the most used and well-known MR system within the remote sensing community. Since the launch of the first satellite, TERRA, in 2000 there has been an unprecedented effort in using satellite remote sensing data for characterizing many of the most important Earth processes (*Justice et al., 2002*). The system acquires images of the Earth surface in 36 spectral bands in the visible and infrared portions of the electromagnetic spectrum with a spatial resolution ranging from 250 m (red and near infrared bands) to 1 km (mainly in the thermal bands) at least twice a day. A suite of high-level standard products, including global LAI/GAI maps (the MODIS 15A2 and 15A3 products), has been developed by the MODIS Land Discipline Groups and is freely available to the user community.

MODIS 15A2 and 15A3 products are based on the inversion of biome specific radiative transfer models by means of look up tables. The algorithm compares atmospherically and bidirectional reflectance distribution function (BRDF) corrected 1-7 channels acquired over a certain area (1 by 1 km^2) and within a time window (4 to 8 days) with the values modeled for a broad range of canopy structure and soil patterns (*Myneni et al., 2000*). The mean value of all the acceptable solutions (i.e. the solutions for which the uncertainty between modeled and measured reflectance is less than the uncertainty of the measured reflectance (*Privette et al., 2002*) is retained as final output value, while a backup algorithm based on an NDVI biome specific relationship is used in case of the main algorithm failure. Since its first release, the MODIS product has been widely used and has received considerable validation from the scientific community, which allowed the development of significant improvements of the original algorithm. The latest version, referred as Collection 5, has been released in 2007. Major changes with respect to the previous version regarded: i) the re-calibration of the LUTs, by means of a new stochastic radiative transfer model able to better depict three dimensional effects (*Shabanow et al., 2005*); and ii) the definition and the number of biome types (8 instead of 6) with the introduction of evergreen and deciduous subclasses for

broadleaf and needleleaf forests (*Yang et al., 2006*). Several authors documented the improved accuracy of the Collection 5 product, especially in forested areas, as well as a much more realistic temporal dynamic and spatial consistency of the LAI/GAI estimates (*De Kauwe et al., 2011; Fang et al., 2012*). However, a bias towards low values was observed in the upper range of the target variable (*Jensen et al., 2011*). Accuracy issues together with rather general assumptions and simplifications about the retrieval algorithm (which are nevertheless necessary when working on a global scale) call for the need of LAI/GAI products customized for specific regions of the Earth and peculiar operational conditions (*Le Maire et al., 2011; Duveiller et al., 2011; Rochdi and Fernandes, 2010*). This is the case when dealing with the challenging Alpine environment.

As already discussed in the previous chapters, mountain regions, such as the Alps, present an extreme topographic variability, often being characterized by steep slopes and altitude variations of thousands of meters. The complexity of the landscape is also reflected in patchy land-cover and vertically structured ecosystems (*Becker et al., 2007*). These conditions may affect the retrieval of surface geo-/bio-physical variables from medium resolution remotely sensed data (*Garrigues et al., 2006; Ryu et al., 2008*). Apart from forest, the Alpine landscape is dominated by grasslands, which represent one of the most diverse man-made landscape formations and cover up to 40% of the agriculturally used area (*Maurer et al., 2006; Tasser et al., 2009*). While grasslands at higher altitudes are mainly used as summer pastures, those close to the valley floors and easily accessible are often managed (fertilized and irrigated) for hay or silage. Meadow canopies show typically very dense structure (with GAI values that can reach 10 [m^2/m^2]) and because of the vegetation cuttings (up to four in intensively managed meadows) followed by rapid plant re-growth, undergo multiple growing cycles within a single vegetation period (*Wohlfahrt and Cernusca, 2002*). Thanks to these characteristics, Alpine grasslands are ideal for studying the effects of climate change on carbon sequestration and energy partitioning, since a wide range of environmental conditions is reproduced at relatively small spatial and temporal scales (*Pauli et al., 2003*). Such analyses are of crucial importance, since the European Alps are assumed to be particularly sensitive to changes in the climate system (*Beniston, 2005; Rammig et al., 2010*). The monitoring of vegetation parameters of Alpine grasslands, and in particular the GAI, is also important for assessing the effects of human activities (such as fertilization and irrigation) on the Alpine ecosystem for resource management, agro-environmental and political issues, as suggested in the European Agriculture Policy Reg. CE 1782/03. Despite several attempts were made for retrieving structural and biochemical parameters of Alpine grasslands from remote sensing data (*Vohland and Jarmer, 2008; Vescovo and Gianelle, 2008; Darvishzadeh et al., 2008; Vescovo et al., 2011*), most studies concentrated on the use of field spectroradiometers or high resolution imagery (e.g., SPOT 5 and IRS). These data cannot accomplish the goal of a temporal frequent monitoring for the whole Alpine region. On the other hand, no existing LAI/GAI global product is designed to cope with the peculiarities and

challenging aspects of the Alpine environment, thus suggesting the need of further research in deriving robust and accurate estimation systems for these specific operational conditions.

In this chapter we present an improved algorithm for the retrieval of GAI from satellite MODIS imagery specifically calibrated for mountain grasslands in the Alps. To this aim, the well-known PROSAIL radiative transfer model has been used and inverted by means of a LUT procedure. The main advantages of the proposed algorithm with respect to the standard MODIS produc are: i) the improved spatial resolution (from 1km to 250 meters), which is necessary to handle the heterogeneity of Alpine areas; ii) the tuning of the radiative transfer model on the spectral characteristics of mountain grasslands; and iii) the exploitation of the characteristics of local topography in the retrieval process. The algorithm was applied to a temporal series of satellite MODIS images acquired over the central Alps during the period 2005-2007 and GAI estimates were validated for both temporal consistency and accuracy with the use of time series of ground measurements collected at three different study sites in the investigated area. For comparison purposes, the Collection 5 MODIS product was also considered.

7.2 Study Site and Data

The study area is the Euroregion Tyrol-South Tyrol-Trentino, formed by the Austrian State of Tyrol and the provinces of South Tyrol/Südtirol/Alto Adige and Trentino in northern Italy. The area is entirely located in the Eastern Alps, between 45°40'- 47°44' N and 10°5'-12°57'E, covering roughly 26,250 km^2 with altitudes that vary from 200m to 3,800 m a.s.l. (Figure 7.1). Flat areas account for only a small proportion and are mainly located in the valley floors. Topography is highly variable and the area is almost evenly distributed among the four main aspect directions. The climate is variable as well, from alpine to subcontinental, with cold and quite snowy winters. Spatial variation in precipitation is high, which is mostly due to the screening effects of topography. While forests represent the most widespread land-cover type, ranging from deciduous broadleaf to evergreen needleleaf forests with the increase of altitude, grasslands are highly present in the area. Grasslands show very heterogeneous characteristics in terms of human management practice, ranging from lightly used pastures at the highest altitudes to moderately/intensively managed meadows with up to 4 cut events per year close to the valley floors. All these characteristics together make the study of this area from satellite extremely challenging, while representative of the possible different environmental/management conditions that can be encountered in the Alpine region.

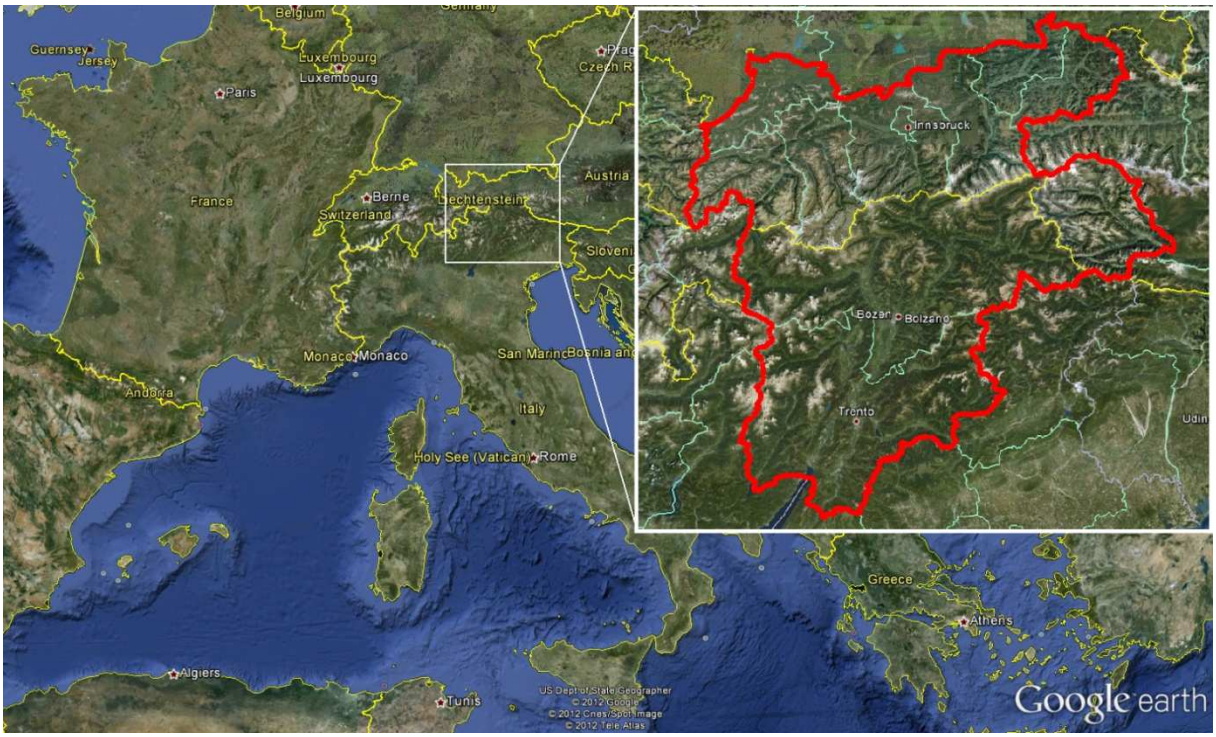


Figure 7.1. The Euroregion Tyrol-South Tyrol-Trentino study area.Credit: Google Earth.

7.2.1 Satellite Data

MODIS imagery acquired by AQUA and TERRA satellites is routinely processed at the Land Processes Distributed Active Archive Center (LP DAAC) within the NASA Earth Observing System Data and Information System (EOSDIS). Among the data products available, the MODIS Surface Reflectance was used in this work. It provides an estimate of the surface spectral reflectance corrected for atmospheric gases and aerosols. The Version-5 surface reflectance product is validated Stage 2, meaning that accuracy has been assessed over a widely distributed set of locations and time periods via several ground truth and validation efforts. More specifically, the data used in this work is the MOD/MYD 09GQ-LG2daily product (for TERRA and AQUA satellites, respectively), which provides bands 1 (red) and 2 (near infrared) together with auxiliary quality information in a 250-meter resolution grid in the sinusoidal projection. This product is meant to be used in conjunction with the MOD/MYD 09GA-LG2 daily product, which includes 500-meters reflectance values in the spectral bands of green (band 4), blue (band 3) and short wave infrared (band 5 to 7). Together with spectral information, 1-kilometer observation and geolocation statistics (such as solar and sensor angles and spectral acquisition quality flags) are provided. For comparison purposes, the Collection 5 MODIS MCD 15A3-L4, has been also considered. It is a 4-day temporal composite LAI/GAI product and is provided at 1 kilometer resolution together with

a standard deviation value and quality flags indicating, among other information, the algorithm (main or back-up) used for the calculation of the corresponding output value.

Both the surface reflectance and MCD 15A3 products have been downloaded from the LP DAAC data pool (https://lpdaac.usgs.gov/get_data/data_pool) to build a satellite data set for the period 2005-2007 over the study area. One advantage of MODIS level 2 (and derived) data products is the fact that they are stored in the same grid (the so called L2G grid, *Wolfe et al., 1998*) and distributed by tiles of ~1100x1100 km^2. This ensures good correspondence between pixels of different products without co-registration effort. In our case the tile h18v04, covering the full study area, was selected. Images were automatically pre-processed to extract useful layers and scale pixel values using the MODIS conversion toolkit, (*White, 2008*). Data were maintained in the native sinusoidal projection, in order to avoid misalignments due to re-projection and interpolation approximations, while the spatial coverage was limited to the Tyrol-South Tyrol-Trentino Euroregion.

7.2.2 Ground Measurements

Ground measurements were conducted at three study sites in the Northern part of the study domain: Neustift (970 m a.s.l., 47° 07’ 00’’N, 11° 19’ 07’’E), Längenfeld (1180 m a.s.l., 47° 03’ 50’’N, 10° 57’ 52’’E) and Leutasch (1110 m a.s.l., 47° 22’ 44’’N, 11° 09’ 56’’E). All three sites are managed grasslands characterized by two to three cutting events per year. For a more detailed description of the study sites we refer to Vescovo et al. (*2011*). The assessment of the GAI in the field measurement sites was performed by means of two different strategies: i) in a destructive fashion, by clipping of square plots of 30 x 30 cm^2 (usually 3 to 5 replicates) and subsequent plant area determination using a planimeter (Li-3100, Li-Cor, Lincoln, NE, USA); ii) in an indirect way, by relating GAI to maximum canopy height or phytomass according to predefined empirical relationships (*Hammerle et al., 2008; Wohlfahrt et al., 2008*). In order to derive continuous time series of GAI, direct and indirect measurements were fitted separately for each growing phase by using a generalized logistic model for each growing phase and a second-order polynomial model after the last cutting event, respectively. An overview of the interpolated temporal profiles for the three selected field measurement sites is provided in Figure 7.2.

7.2.3 Ancillary Data

To support the retrieval of GAI from MODIS imagery, a digital elevation model (DEM) and a land-cover/use map were exploited. The DEM consisted of the version 2.1 Shuttle Radar Topography Mission (SRTM) product, derived from the data of the STS-99 mission in February, 2000, and edited by the National Geospatial-Intelligence Agency (NGA) (*Farr et al., 2007*). The data, freely available online, are worldwide available with a spatial resolution of 90 x 90 m^2, except from the United States territory where the resolution is improved to 30 x 30m^2. As land-cover/use map, the GlobCorine 2009 was considered (*Bontemps et al., 2010*).

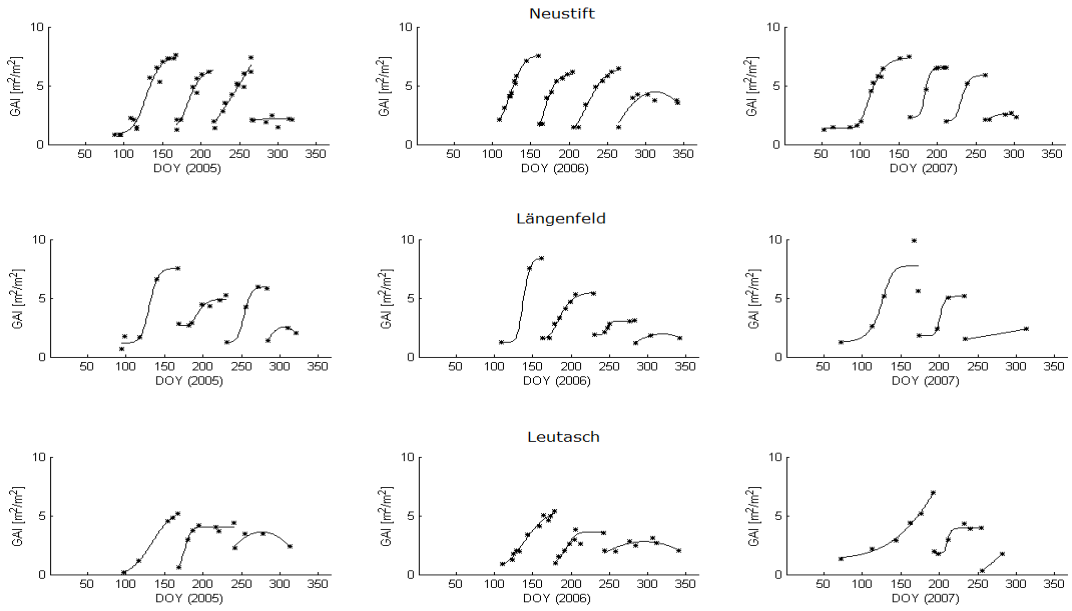


Figure 7.2. Measured (symbols) and interpolated (solid line) GAI values for the three field measurement sites and the three years considered.

This map has been generated in the framework of the ESA's GlobCorine project, covering a pan-European area corresponding to the 27 European Union countries extended to the Mediterranean basin and Western Russia. The map was derived from an automatic and regionally-tuned classification of a time series of Medium Resolution Imaging Spectroradiometer (MERIS) seasonal and annual mosaics over the period between January and December 2009. The final resolution of the classification map is 300 x 300 m^2. In order to fit with the MODIS satellite data, both the DEM product and the land-cover map were converted from the native projection into the sinusoidal MODIS projection and cut over the study area.

7.3 Methods

The main rationale behind the proposed GAI retrieval algorithm is the inversion of a canopy radiative transfer model, adapted to the characteristics of the Alpine environment, by means of a look-up table (LUT) procedure. A LUT is a pre-computed spectral library which contains different simulated spectra corresponding to different combinations of biophysical and geometrical parameter. During the operational estimation phase, a search procedure based on the match between simulated and remote sensing spectra is performed and the parameter configuration that yields the best performance is retained as output estimate. LUT inversion is particularly suitable for the processing of large amounts of data, such as time series of remote

sensing imagery, since it only involves searching procedures while the time consuming simulation of spectra is performed offline. In addition to this, LUT inversion shows interesting features with respect to alternative machine learning inversion strategies, such as neural networks, recently proposed for the retrieval of LAI/GAI from remote sensing data (*Duveiller et al., 2011; Baret et al., 2007*). Indeed: i) it permits a global search (thus avoiding local minima), while showing less unexpected behavior when the spectra characteristics of the targets are not well represented by the modeled spectra (*Schlerf and Atzberger, 2006*); ii) it does not require a training phase, which is necessary with machine learning methods and in most of the cases is complex and time consuming; iii) it easily allows constraining the inversion process by means of prior knowledge of the target variable; and iv) it is flexible and easy to adapt and extend to new operational conditions (e.g., the use of data with different spectral bands or different acquisition geometries) without additional training/tuning. To achieve good estimation accuracy, the dimension of the LUT must be sufficiently large (*Weiss et al., 2000*), which may slow down the estimation process. To deal with this problem, one may resort to efficient methods for the generation of the LUT (*Hedley et al., 2009*) or to a proper parameterization of the radiative transfer model, which can take into account prior information on the application domain considered. The second strategy was implemented in this work.

A general overview of the algorithm flowchart is shown in Figure 7.3, while a detailed description of its main components is given in the following.

7.3.1 Reference Spectral Library Generation

The generation of the reference spectral library is a crucial step for the effective retrieval of GAI from MODIS imagery. This task has been addressed in this work using the PROSAIL radiative transfer model. PROSAIL, which consists of the PROSPECT leaf optical model (*Jacquemond and Baret, 1990*) coupled to the SAILH canopy reflectance model (*Verhoef, 1984; Verhoef, 1985; Kuusk, 1991*), simulates the top-of-the-canopy bi-directional reflectance from the visible (400nm) to shortwave infrared (2500nm) wavelengths as a function of a series of input variables describing the leaf optical properties, the canopy structure, the background soil reflectance and the sun/observer system geometry. This results in the possibility to simulate various canopy and experimental scenarios. More in detail, leaf optical properties are quantified in terms of mesophyll structural parameter (S_M) and leaf chlorophyll a+b (C_{ab}), dry matter (C_m), water (C_w) and brown pigment (C_{bp}) content. Canopy structure is expressed in terms of canopy structural parameter (S_C) related to mean leaf angle inclination, hot spot parameter (*Hotspot*) and obviously the *LAI* (in this study *GAI* since grassland canopies are investigated). The contribution of the background soil is accounted for by the input soil spectral signature, while the sun/observer geometry is described by the solar/observer zenith and relative azimuth angles and the fraction of diffuse incoming solar radiation (*skyl*).

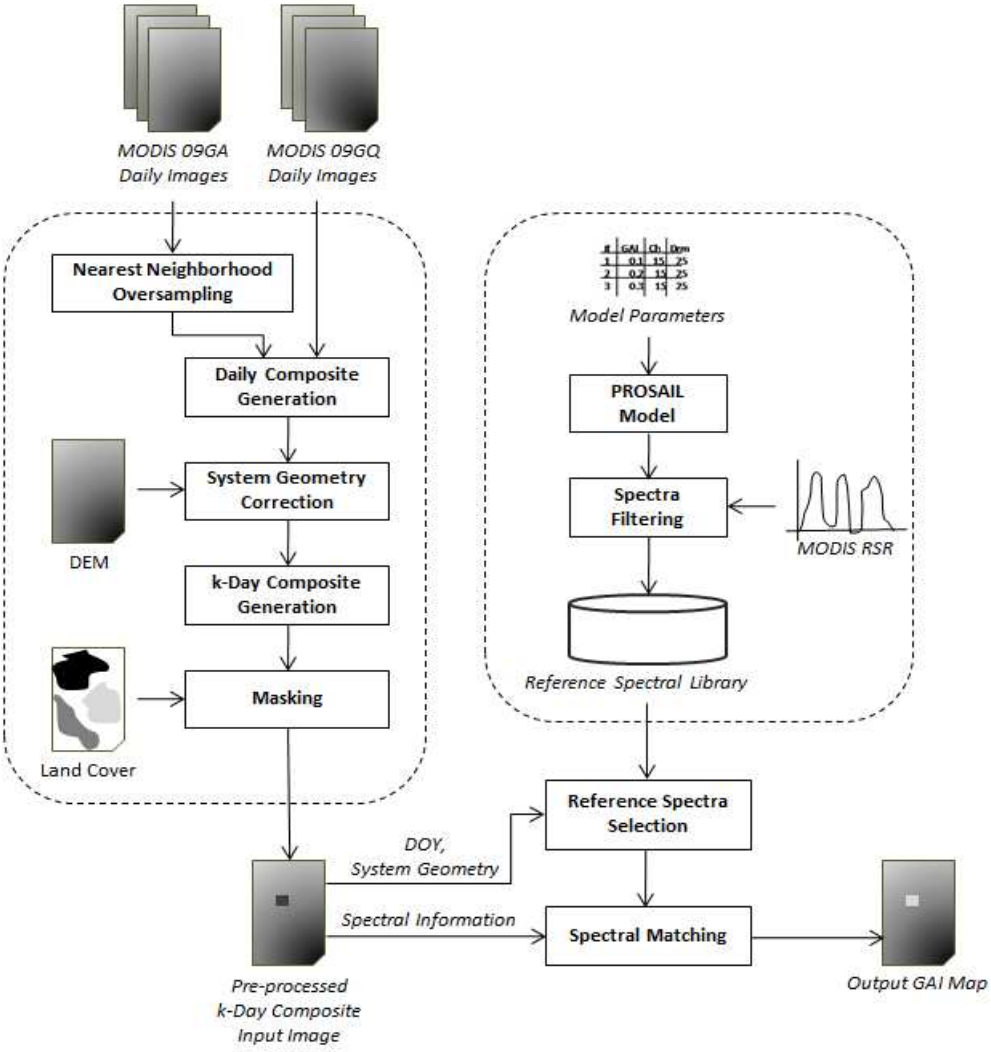


Figure 7.2. Flowchart of the proposed GAI retrieval algorithm.

PROSAIL has been widely used and validated by the scientific community for both studying plant canopy spectral reflectance and retrieving vegetation biophysical variables (*Jaquemoud et al., 2009; Meroni et al., 2004; Vohland and Jarmer, 2008; Duveiller et al., 2011; Darvishzadeh et al., 2008; Jaquemoud et al., 2000*). Although the basic physical assumptions (e.g., semi-finite horizontally homogenous plant canopies) represent a limitation for its application in some specific domains with high heterogeneity at plot scale (*Darvishzadeh et al., 2008*), the model demonstrated to be reliable, accurate and applicable to the study of both agricultural and natural vegetation canopies with medium and high resolution satellite imagery. Moreover, PROSAIL is relatively simple, easy to handle (thanks to the rather limited number of parameters to tune) and characterized by a reasonable computational time.

The spectral library (SL) consists of an ensemble of canopy reflectance simulations $\boldsymbol{CR^{sim}} = \left[CR^{sim}_{ch_1}, CR^{sim}_{ch_2}, \dots, CR^{sim}_{ch_n}\right]$ (where $\mathrm{ch}_1, \mathrm{ch}_2, \dots, \mathrm{ch}_\mathrm{n}$ represent different spectral channels) each associated with the corresponding model parameter configuration $\nu = [\nu_1, \nu_2, \dots, \nu_\mathrm{m}]$. Not only the target parameter (in this case the GAI), but also the other model parameters should be varied in this phase, in order to take into account the different possible conditions that can be encountered on the ground in different areas and different temporal periods. This will easily lead to a large amount of spectra in the SL. In order to constrain the behavior of the model to the vegetation canopy of interest in this study (i.e. mountain grasslands) while reducing parameter equifinality (the phenomenon whereby similar spectral signatures can be associated with different biophysical parameter configurations rendering the inversion problem ill-pose, *Beven & Freer, 2001; Beven 2006*) a specific parameterization for the PROSAILH inputs was adopted. More in detail, sensor/sun geometry angle configurations were derived from the pre-processed MODIS imagery (after a quantization process with a 5 degrees step), while a constant value (0.1) for the whole spectrum was considered for the *skyl* parameter. Indeed, this parameter showed limited influence on the simulated reflectance in several earlier analyses (*Schlerf and Atzberger, 2006; Clevers and Verhoef, 1991*). The soil spectral response was modeled averaging various signatures from the ASTER spectral library (*Baldrige et al., 2009*) typical for grasslands soils in the Alps. Concerning the canopy and leaf biophysical parameters, the ranges of variability for these parameters (except for the canopy structure S_C considered constant to 2) were first determined according to the available field measurements and the values typically used in the literature in similar operational conditions (*Darvishzadeh et al., 2008; Vohland and Jarmer, 2008; Vescovo and Gianelle, 2008*). Then a sensitivity analysis of the MODIS spectral channels (especially bands 1 and 2) to variations in leaf and canopy properties of Alpine grasslands was performed. More in detail, a set of simulated spectral responses was generated by iteratively varying one parameter within its range, while retaining the others fixed to the mean value. PROSAIL simulated spectra were then converted into MODIS spectral channels reflectance values by means of the sensor relative spectral response (RSR) and correlated with the corresponding biophysical parameter. Figure 7.4 depicts the different analysis steps for the GAI parameter. It is worth noting that a linear dependence is tested for each parameter, despite in some cases a non-linear behavior of the simulated spectra to variations in the input parameter emerged (e.g., MODIS band 1 vs. GAI). This approximation was considered acceptable at this stage. The resulting statistical correlation coefficient determined the parameterization for the model inputs: a uniform distribution within the corresponding range was used for parameters showing significant correlation values ($p<0.05$), while a fixed value (i.e., the mean) was retained for the other parameters. As reported in Table 7.1, except from the GAI parameter, only the chlorophyll a+b (C_{ab}) and dry matter (C_m) content showed significant correlation with MODIS bands 1 or 2 during the sensitivity analysis.

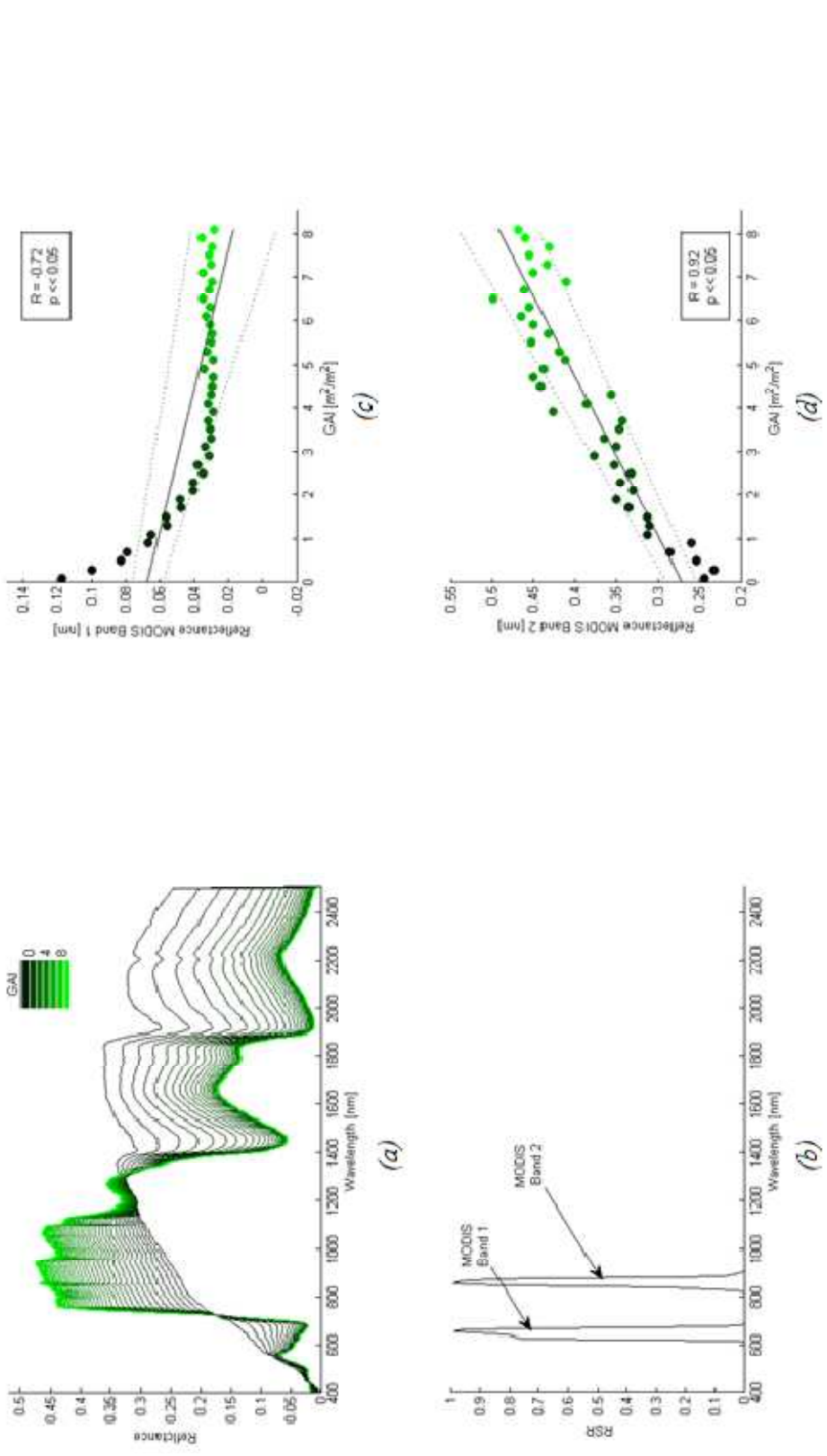


Figure 7.4. Sensitivity analysis to variations of Alpine grassland GAI of the MODIS spectral channels with the PROSAIL model: (a) simulated spectral signatures for different GAI values; (B) MODIS relative spectral Response (RSR) for bands 1 and 2; scatter plots of (c) MODIS band 1 and (d) MODIS band 2 reflectance values vs. GAI. Linear trend (solid line) and 95% confidence interval (dotted line) are also shown.

Table 7.1. Summary of the sensitivity analysis of MODIS spectral bands 1 and 2 to biophysical characteristics typical of Alpine grasslands. On the right, the final parameterization used for the generation of the reference spectral library with the PROSAIL model is reported.

Biophysical Parameter	Original Range	Correlation Band 1 MODIS	Correlation Band 2 MODIS	Final Model Parameterization
Mesophyll Structural Parameter (N)	1.5-1.9 [-]	r = 0.29 p = 0.06	r = 0.17 p = 0.29	1.75 [-]
Chlorophyll a+b (C_{ab})	15-55 [µg/cm²]	*r = -0.93* *p << 0.05*	r = -0.08 p = 0.62	15-55 [µg/cm²]
Dry Matter (C_m)	0.005-0.01 [g/cm²]	r = -0.12 p = 0.42	*r = -0.65* *p <<0.05*	0.005-0.01[g/cm²]
Water (C_w)	0.01-0.02 [g/cm²]	r = 0.18 p = 0.21	r = 0.15 p = 0.30	0.015 [g/cm²]
Brown Pigment (C_{bp})	0-0.1[g/cm²]	r = 0.042 p = 0.77	r = -0.03 p = 0.82	0 [g/cm²]
HotSpot	0.05-0.25 [m/m]	r = 0.35 p = 0.08	r = -0.1 p = 0.62	0.1 [m/m]
Green Area Index (GAI)	0.1-8 [m²/m²]	*r = 0.72* *p << 0.05*	*r = 0.92* *p << 0.05*	0.1-8 [m²/m²]

7.3.2 Inversion Algorithm

The core of the inversion algorithm is represented by the matching procedure. Given a pixel in the input image, the associated spectral signature $\mathrm{CR}^* = \left[\mathrm{CR}^*_{\mathrm{ch}_1}, \mathrm{CR}^*_{\mathrm{ch}_2}, \ldots, \mathrm{CR}^*_{\mathrm{ch}_{\mathrm{N_{ch}}}}\right]$ is compared with those belonging to the reference spectral library according to a metric M_D, which quantifies the matching degree between two spectral signatures. In this study the root mean squared error metric was used:

$$M_D = \sqrt{\sum_{i=1}^{N_{ch}} \frac{\left(CR^*_{b_i} - CR^{sim}_{b_i}\right)^2}{n}} \qquad (7.1)$$

where N_{ch} is the number of spectral bands considered. In order to preserve the spatial information of the input MODIS image, only the spectral information associated with bands 1 and 2 was considered. These bands are anyway the most suitable for GAI retrieval (*Curran and Williamson, 1987*). Together with the spectral values, spectral vegetation indices based on these two bands (i.e., the simple ratio SR (*Jordan, 1969*), the normalized difference vegetation index NDVI (*Rouse et al., 1974*) and the enhanced vegetation index EVI (*Huete et al., 1996*) were also considered. Their role was to increase the robustness of the matching procedure and enhance the sensitivity to the target biophysical parameter. Since spectral bands and spectral indices derived from these had different ranges of variability, their values were rescaled to the range [0,1] before computing the matching score.

Traditional LUT inversion procedures retain the value associated with the reference spectrum showing the highest degree of matching with the remotely sensed spectral signature as output target parameter. This procedure could suffer of instability and limited robustness when noise is affecting remote sensing data. In order to limit this issue, the first *z* best matching spectra were considered (*Darvishzadeh et al., 2008*) and the weighted mean of the corresponding *z* target parameter values was provided as output estimate:

$$\widehat{GAI} = \frac{\sum_{j=1}^{z} \eta_j * GAI_j}{\sum_{j=1}^{z} \eta_j} \qquad (7.2)$$

where

$$\eta_j = {}^{1}/_{M_{D_j}} \qquad (7.3)$$

and

$$S_{M_1} \leq S_{M_2} \leq \cdots \leq S_{M_t} \qquad (7.4)$$

Together with the mean value, the weighted standard deviation was computed as measure of the reliability and stability of the retrieved target parameter value:

$$Std_{\widehat{GAI}} = \sqrt{\frac{\sum_{j=1}^{t} w_j * \left(GAI_j - \widehat{GAI}\right)^2}{\sum_{j=1}^{t} w_j}} \qquad (7.5)$$

After some trials, the number of best matching spectra *z* was set equal to 10.

The exploitation of prior information on the addressed target variable was found to significantly improve the inversion process (*Combal et al. 2002; Meroni et al., 2004*). Instead of increasing the complexity of the matching score, as proposed by several authors, a prior selection and reduction of the reference spectra used in the spectral matching procedure for each processed pixel was implemented in the proposed algorithm. This strategy allowed easily constraining the retrieval process given the prior information while speeding up the inversion procedure, since a reduced number of evaluations of the matching score were performed. In the analysis, two kinds of prior information were considered: 1) the acquisition geometry of the system (i.e. the sun and sensor zenith and azimuth angles), which was assumed known from the MODIS image pixel by pixel; 2) a constraint on the range of variability for the GAI values as a function of the day of the year. This information was retrieved combining the envelope of the interpolated GAI measurement time series in different places within the study area and in different years. While the first information allowed significantly reducing the ambiguity introduced by the high variability in the acquisition geometry, the second was useful for constraining the GAI retrieval process especially during the winter season, when GAI values are typically very low (due to vegetation senescence and snow cover) while the occurrence of noisy pixels due to cloud cover or snow contamination on the ground is quite high.

7.3.3 Data Processing and GAI Map Generation

Before applying the retrieval algorithm described above, MODIS daily surface reflectance data required some processing. First of all, daily MOD/MYD 09GA images were oversampled to the 250 meters spatial resolution by means of a nearest neighborhood strategy and composed with the corresponding 09GQ images.

In the next step, the acquisition and illumination geometry information associated with the MODIS surface reflectance (i.e. the sensor and solar zenith and azimuth angles) were corrected in order to take into account the topography of the investigated area. Geometry information provided together with MODIS data refer to the ideal case of a locally flat horizontal surface. However, in case of complex topographic conditions, such as in mountain areas, this hypothesis is not fulfilled. Steep slopes and rapid changes of aspect direction typical of Alpine valleys cause significant variation in the acquisition geometry local reference system even from one pixel to the other. This strongly affects the sun and sensor geometry angles (Figure 7.5). To account for this effect, the local surface characteristics (local slope and local aspect angles) were extracted from the digital elevation model over the entire

study area. Then for each MODIS pixel, the sun illumination and sensor acquisition zenith angles were corrected to the local conditions of the area according to this equation:

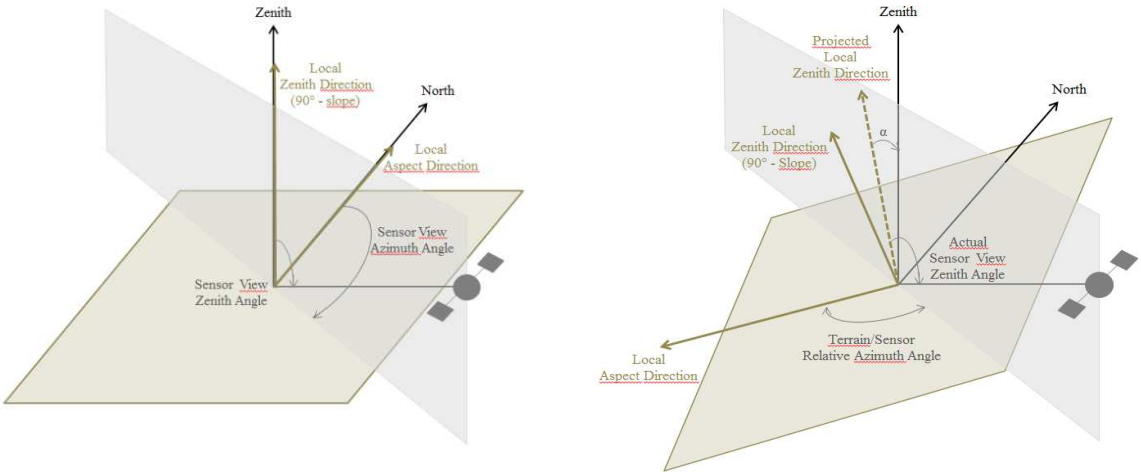
$$Zenith^{*} = Zenith + \chi \qquad (7.6)$$

where $Zenith^{*}$ is the actual zenith angle, *Zenith* represents the sun/sensor zenith angle before correction and χ is the local slope angle projected on the sun/sensor view plane and determined according to:

$$\chi = \cos^{-1}\left(\cos\left(\pi/2 - slope\right).\cos(\Delta Azimuth)\right) - \pi/2 \qquad (7.7)$$

where *slope* is the local slope of the area and *ΔAzimuth* is the relative angle between the local aspect direction and the sun/sensor azimuth direction (see Figure 7.5).

The third pre-processing step consisted in aggregating daily surface reflectance products in a composite product. This procedure aimed at reducing the amount of data to be processed and mitigating the occurrence of cloudy, aerosol contaminated or badly acquired (from the acquisition geometry viewpoint) pixels, a frequent problem in mountain areas. The aggregation strategy implemented was similar to that adopted by the MODIS 8-day composite surface reflectance product. Given a temporal buffer of *I* subsequent MODIS daily images, a score was computed pixel by pixel and for each image by accounting for the presence of cloud, cloud shadow and/or atmospheric aerosols (according to the daily product state flags) and the sun and sensor zenith incidence angles. Then for each position in the image, the pixel showing the highest score (ideally, without cloud coverage and aerosols contamination and showing the lowest sun and sensor incidence angles) was retained for the composite product.



(a) (b)

Figure 7.5. Representation of the sensor acquisition geometry in case of (a) flat horizontal reference cell and (b) inclined reference cell due to topography. Note that the same conditions hold for the sun illumination geometry.

According to this strategy a 4-day composite combining both AQUA and TERRA daily acquisitions was generated. This choice allowed a high temporal frequency of the output images maintaining at the same time a high number (8) of daily input observations among which to choose for each pixel.

Finally, a masking procedure was applied to the data. In this phase, both the Corine 2009 land-cover map and the composite product pixel quality were exploited in order to exclude pixels not belonging to the target biome or still showing poor quality (e.g., cloud or cloud shadow coverage, aerosols contamination) from the subsequent processing.

After the preliminary processing, data were ready and the inversion algorithm was applied pixel by pixel for generating the estimated output GAI maps. Together with the estimated GAI values each map contained additional layers with the standard deviation value, the day of the year (DOY) and the quality flags associated with each GAI estimate.

7.3.4 Validation

The validation of the proposed GAI retrieval algorithm was carried out according to a threefold strategy:

1. *Visual analysis of estimated GAI maps*. The GAI maps derived with the proposed algorithm were visually inspected and compared with the corresponding MODIS MCD 15A3 product. In the absence of a distributed ground reference map, this analysis aimed at qualitatively assessing the capability of the algorithm to reproduce spatially consistent and expected patterns of the target variable.

2. *Analysis of estimated GAI dynamics*. By exploiting the available temporal data set, GAI temporal signatures were extracted from both the proposed algorithm and the MCD 15A3 product maps over five test sites representative of different conditions on the ground: three corresponding to the ground measurement stations, one corresponding to a mid-altitude pasture (1850 meters a.s.l., 46° 42’ 14’’ N, 10° 37’ 15’’ E), and one to an high-altitude pasture (2300 meters a.s.l., 46° 41’ 5’’, 10° 38’ 57’’) both in Matscher Valley, South Tyrol. The capability to reproduce the GAI temporal and range dynamics typical of Alpine areas was then analyzed in terms of seasonal (number of growing cycles, start and end of the season) and statistical (median, quartiles and minimum/maximum) indicators. Where available, the GAI signatures were compared with the available reference measurements.

3. *Assessment of estimated GAI accuracy*. The estimated and the corresponding interpolated ground GAI values associated with the three field measurement stations were analytically compared in terms of the following statistical indicators:
    - Root Mean Squared Error (*RMSE*)

$$RMSE = \sqrt{\frac{1}{q}\sum_{i=1}^{q} e_i^2} \quad (7.8)$$

- Bias (B)

$$B = \frac{1}{q}\sum_{i=1}^{q} e_i \quad (7.9)$$

- Standard Deviation (SD)

$$SD = \sqrt{\frac{1}{q}\sum_{i=1}^{q} (e_i - B)^2} \quad (7.10)$$

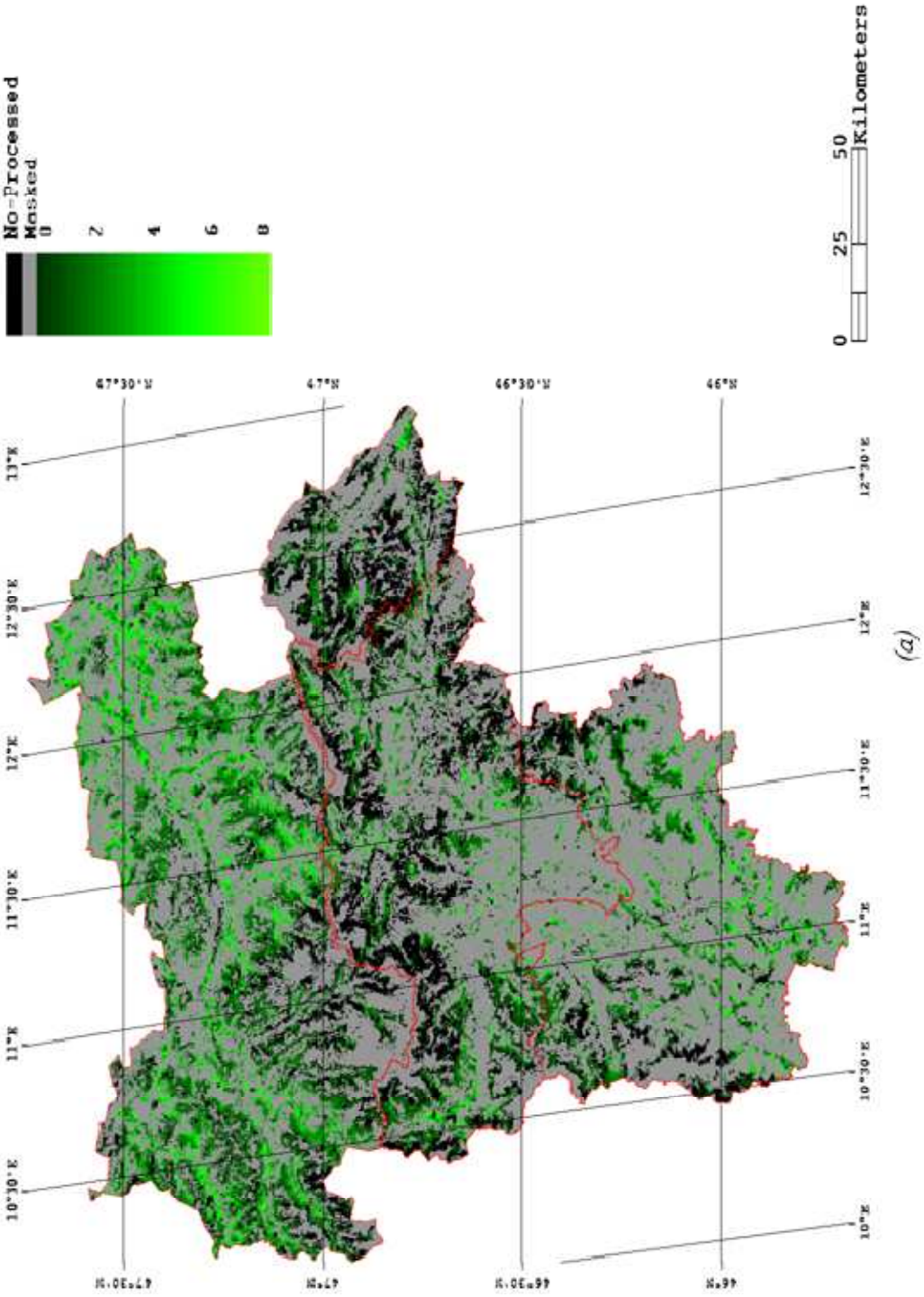
where e indicates the difference between estimated and reference GAI and q is the number of samples considered. In addition, the statistical correlation coefficient (r), the *slope* and *intercept* of the linear model between estimated and reference GAI values were calculated. The indicators were computed for both the proposed algorithm and the MODIS LAI standard product.

A major issue regarding points 2 and 3 of the validation process was the selection of the pixel in the map for the extraction of the estimated GAI value as resampling or geolocation errors may occur (*Fernandes et al., 2004; Weiss et al., 2007*). In order to minimize such effects, satellite estimates can be aggregated over a 3x3 (or more) pixels region before their comparison with ground reference measurements (*Morisette et al., 2006*). However, this approach was not applicable in the investigated study area due to the large heterogeneity of the landscape, in terms of topography and land-use. Moreover, the pixel aggregation strategy would prevent the assessment of the improved spatial resolution of the proposed algorithm, especially when evaluating GAI temporal trajectories. For these reasons, GAI values were extracted from a single pixel for each selected site according to a vicinity (in terms of geographic coordinates) and pixel homogeneity (in terms of topography and land-cover) rule. Since the same approach was applied to both the proposed algorithm and the MCD 15A3 product, this will not affect the validation results.

7.4. Results

7.4.1 Visual Analysis of Estimated GAI Maps

An example of GAI map obtained with the proposed algorithm relative to the last 4-day composite of June 2005 is shown in Figure 7.6 contrasted by the corresponding MCD 15A3 product.


No-Processed
Masked
0
2
4
6
8
0
25
50
Kilometers
47°30'N
47°N
46°30'N
46°N
13°E
12°30'E
12°E
11°30'E
11°E
10°30'E
10°E
(a)

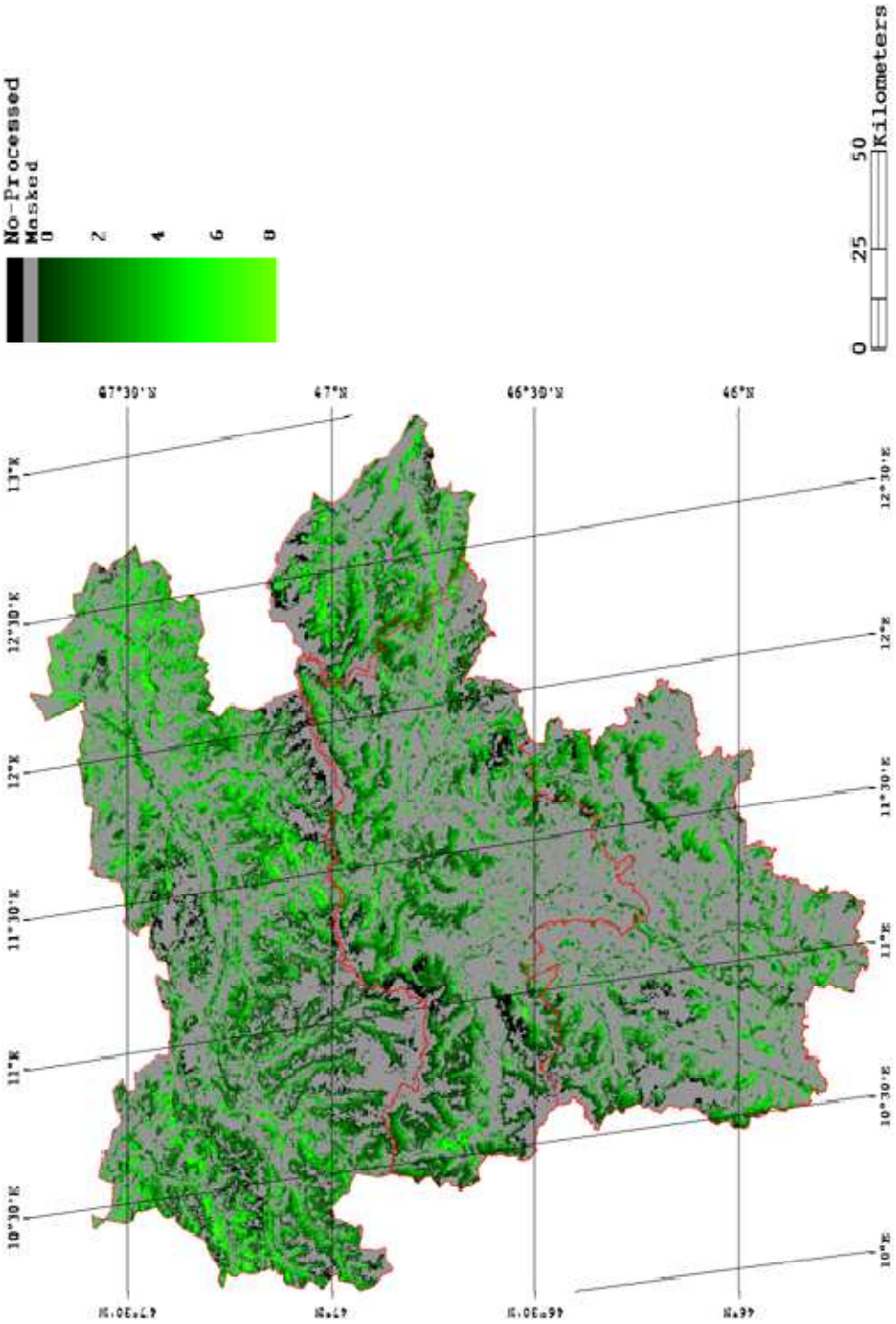
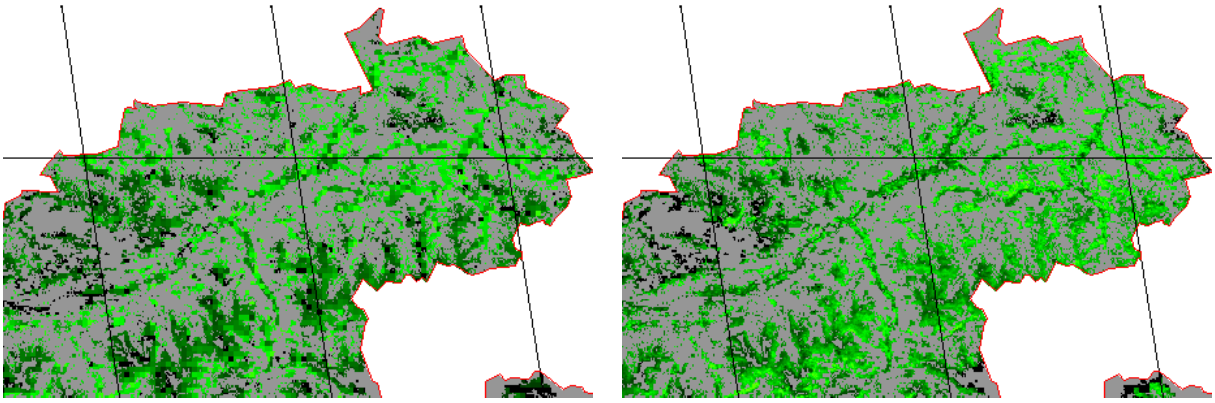


Figure 7.6. GAI maps over the Euroregion Tyrol-South Tyrol-Trentino study area for the end of June, 2005: (a) the MCD 15A3 product and (b) the proposed algorithm map.



(a) *(b)*

Figure 7.7. Detail of the GAI map over the Inn Valley, eastern North Tyrol: (a) the MCD 15A3 product and (b) the proposed algorithm map.

A first inspection of the map indicates the occurrence of high GAI values mainly in the valley floors, where intensively managed meadows are located, while GAI becomes progressively lower with the increase of the altitude or on the sides of the Alpine valleys, where pastures and less intensively managed meadows are typically present. These patterns are evident in both the maps. However, going more into spatial details, it is possible to observe the improved mapping of GAI patterns in the proposed algorithm map. This is mainly attributable to the increased spatial resolution with respect to the MCD 15A3 product as pointed out, for example, from the fact that the shape of Alpine valleys can be more easily identified. Moreover, small-scale spatial patterns can be differentiated within each valley, such as the distinction between highly and extensively managed meadows. Within the 1x1 km^2 MCD15A3 pixel their contributions are typically mixed, as shown for example in the top right part of the map, corresponding to the eastern part of the Inn valley in North Tyrol (Figure 7.7). A dark elongated spatial pattern in the middle of the valley from the center to the top right corner of the image is visible in the GAI map generated with the proposed algorithm. It corresponds to the Inn river, which crosses the valley and is surrounded by trees/shrubs and meadows. The same pattern is not visible in the standard MODIS product map, where the river and the surrounding meadows are mixed. Several other examples (not shown here for space constraints) can be found in other portions of the maps and are consistent with those observed in other periods of the year.

7.4.2 Analysis of Estimated GAI Dynamics

7.4.2.1 Temporal Dynamics

The temporal evolution of estimated GAI for the five selected reference sites is shown in Figure 7.8. From a first analysis the MCD 15A3 product and the proposed algorithm agree well with regard to the annual phenological cycles. The start and length of the green seasons, which are shorter and delayed in case of mid/high altitude pastures, longer and earlier in case of managed meadows near valley floors, are recognized properly. Differences start emerging when examining at the inter-annual dynamics of managed meadows. GAI estimates obtained with the proposed algorithm show multiple (up to three) growing phases during the vegetation period, in line with the temporal dynamics in the ground measurements (Figure 7.2). Each growing phase is followed by a sharp decrease of the estimated values. The only exception is noticeable for the Neustift site, where the descending phases of GAI values after the peaks are not fully reproduced for the year 2007. This is probably due to contamination by clouds and aerosols, which often occur in Alpine areas. On the contrary, the MCD 15A3 product mainly follows the annual envelope of GAI with some sporadic drops related to the backup algorithm (black points). A more detailed comparison with the ground measurement trajectory suggests that the proposed algorithm is also able to properly capture the temporal dynamic of GAI in terms of start and length of the growing phase. A shift in the start of the growth can be identified for the Neustift and Längenfeld sites after the first and especially the second cuts. This phenomenon was to some extent expected. Plot scale measurements used in this work, despite providing crucial information about temporal dynamics of GAI in Alpine meadows, cannot be considered representative of the spatial and temporal variability of the whole area covered by a MODIS 250x250 m^2 pixel. Cut events in the surrounding plots may differ by several days with respect to the considered site and this phenomenon affects the spectral signature of the MODIS pixel determining a mixture of high and low GAI contributions. This is confirmed by the fact that peaks show a certain variability in GAI values and grass cuts are sometimes associated with a gradual decrease lasting few days (e.g. the Längenfeld temporal trajectory). The pasture sites do not exhibit multiple growth periods, for both the standard and the proposed algorithms, in line with the generally low-intensity grazing at these sites.

7.4.2.2 Range Dynamics

Together with complex temporal trajectories, Alpine grasslands are characterized by typical range dynamics, strongly affected by human management practices. In order to assess the capability to capture these range dynamics, statistical indicators associated with each temporal profile derived from the MCD 15A3 product, the proposed algorithm estimates and (where available) the interpolated ground measurements were compared with the help of boxplots in Figure 7.9.

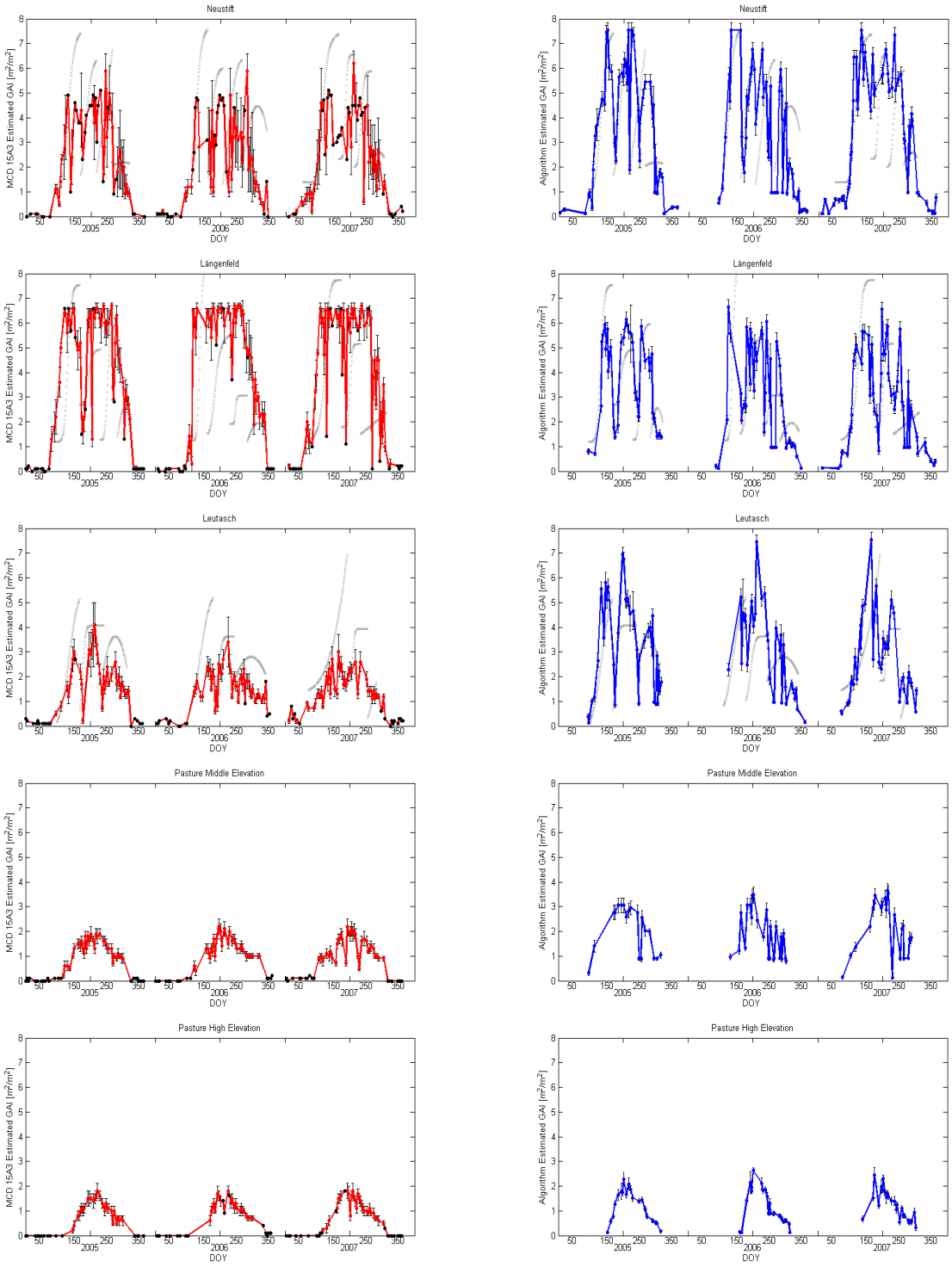


Figure 7.8. Temporal GAI signatures associated with the five selected sites on the ground for the period 2005-2007: (left column) MCD 15A3 estimates and (right column) proposed algorithm estimates.

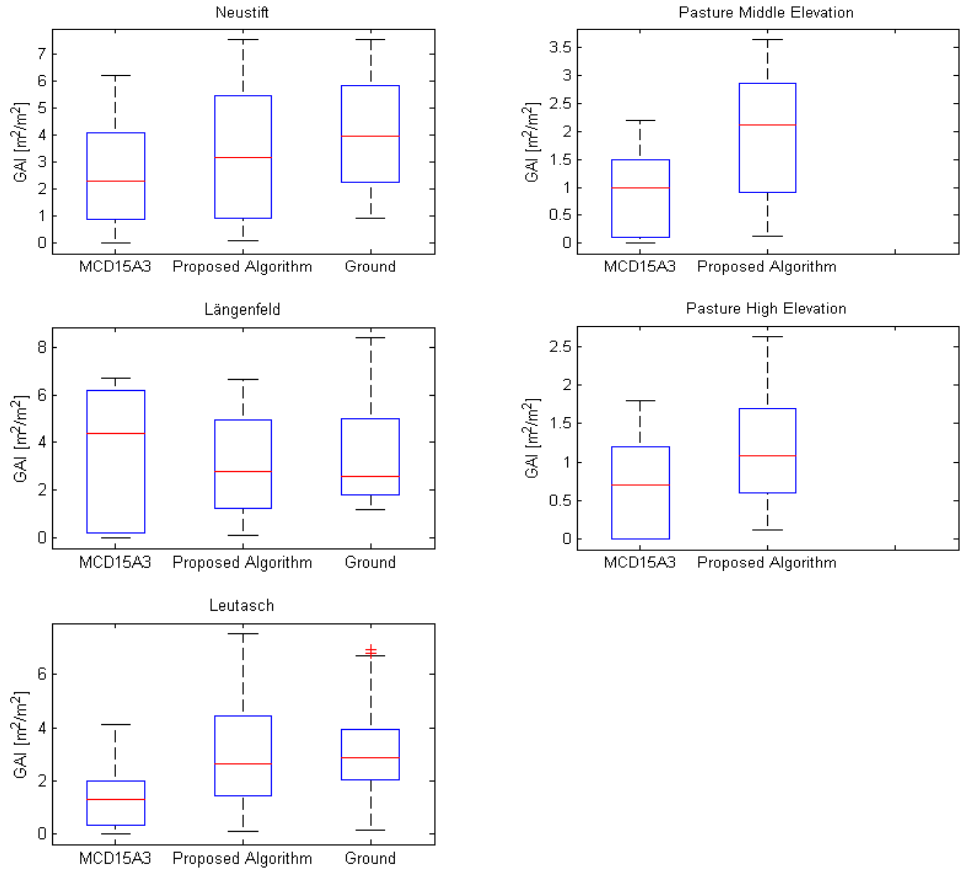


Figure 7.9. Boxplots representing median (red lines), quartiles (blue boxes) and min/max values (black lines) of GAI samples associated with the five selected test sites.

Managed meadows (Neustift, Längenfeld and Leutasch) present a wide range of variability, with maximum values that can easily exceed 7 m^2/m^2. Statistics associated with the estimates provided by the proposed algorithm, especially the median values, are well in agreement with the field measurements for all three sites. Slight over-estimation and under-estimation of the highest values are noticeable for the Leutasch and Längenfeld sites, respectively, which are probably attributable to remaining aerosols contamination or mixture effects within the MODIS 250x250 m^2 pixel. However, the results confirm the capability of the algorithm to properly capture the variability of the GAI variable. MODIS standard algorithm estimates show a different behavior. GAI values are underestimated at Neustift and especially Leutasch, as suggested by the median indicator. The latter site shows also a dynamic compressed in a very narrow range, unusual for managed meadows. On the contrary, for the Längenfeld site retrieved GAI values present a wide dynamic, with a median value above the expectation (mainly due to the absence of multiple growing phases during the year). Difficulties in properly capturing the GAI dynamics can be ascribed to the coarser spatial

resolution of the standard product with respect to the proposed algorithm. Mixture effects due to the variability typical of Alpine areas in land-use close to the selected site can often occur within the 1x1 km^2 resolution cell of the MODIS standard algorithm. The presence of needleleaf forests, as near to the Leutasch site, or meadows with different management practice or timing, as in the surrounding of the Längenfeld site, strongly affects the retrieval process and explains the observed GAI dynamics.

With regard to pastures, GAI values show a reduced dynamic range, with median values between 1 and 2 m^2/m^2 and peaks up to 3.5 m^2/m^2. Despite the lack of ground measurements at these sites, estimated values are in line with the expected dynamics for this grassland type. The proposed algorithm provides slightly higher estimates with respect to the MCD 15A3 product. Moreover, a wider GAI dynamic range can be observed for the mid-altitude pasture with respect to the high-altitude pasture, in line with the expected altitudinal gradient effect. Also the MODIS standard algorithm shows sensitivity to variations in altitude of the investigated site, but to a more limited extent.

7.4.3 GAI Accuracy Assessment

The summary of accuracy statistical indicators for the standard MODIS product and the proposed algorithm is reported in Table 7.2. and 7.3, respectively. *RMSE* values indicate a better agreement between the proposed algorithm estimates and the interpolated ground measurements as compared to the MCD15A3 product. In general the *RMSE* indicator is lower than 2 m^2/m^2 for the proposed algorithm, with the exception of the Neustift site for the year 2006, ranging from 1.35 to 1.85 [m^2/m^2]. Considering all the years and the sites together, the metric equals 1.68 [m^2/m^2], pointing out a 21% improvement compared to the standard MODIS product (2.13 [m^2/m^2]). Concerning the Bias component, the proposed algorithm shows generally good performance, with values very close to zero for both the global performance over years and the global performance for each site. This shows that the proposed algorithm is able to capture the mean target value with very low systematic errors. Consistent biases are noticeable instead in the MCD15A3 GAI values. They range from -1.41 to 2.32 [m^2/m^2] and vary in sign and magnitude from one site to the other, suggesting spatial variability of the systematic error.

Interesting indications about performance and accuracy of estimated GAI values are provided also by the scatter-plots shown in Figure 7.10, where all samples of each field measurement site have been grouped together. MCD15A3 estimates show the best performance at the Neustift site. However, the quite low slope value indicates a reduced dynamic range of the target values and the underestimation tendency for high target values discussed above. The estimates, when retrieved with the main algorithm, are characterized by a high standard deviation, indicating a limited confidence associated to these values from the retrieval algorithm. Better performance is achieved by the proposed algorithm. The linear tendency line is closer to the ideal 1-to-1 line and also the correlation coefficient *r* reaches a

higher value. The graph shows two subsets of samples, lying above and below the 95% confidence interval, respectively, for which the estimation error is rather high. These samples are mainly associated with the second and third growth of the grassland and are related to the non-perfect alignment between the measured and retrieved temporal profiles (which has been discussed in the previous paragraph). Apart from these samples, the agreement between estimated and interpolated ground GAI values is accurate along the whole range of GAI values.

The proposed algorithm exhibits similar performance for the Längenfeld site, despite a slight underestimation of high target values is noticeable. Again it is possible to identify a subset of samples with high estimated GAI values corresponding to low ground interpolated values, which are attributable to the misalignment between temporal profiles. The corresponding graph for the MCD 15A3 product shows clearly the overestimation trend for low and medium GAI values, resulting in a low slope (0.37) and high intercept (3.68) values. This trend is related to the seasonal growing phases, which are not captured by the standard algorithm.

In the Leutasch site, the proposed algorithm achieves the best performance, as shown by the slope, intercept and correlation coefficient values. The majority of the samples lie within the 95% confidence interval of the linear regression indicating that the temporal variability during the year is correctly predicted without large errors related to temporal misalignments as shown in the previous two plots. On the contrary, the dynamic shown by the MCD15A3 estimates is almost flat, as indicated by the low slope value, and points out a strong underestimation trend for high GAI values. The trend shown in this graph is in accordance with the previous considerations about numerical performance indicators and dynamic range analysis.

7.5 Discussion

Results presented in the previous section demonstrate that the proposed retrieval algorithm is effective for the retrieval of GAI of mountain grasslands in the Alps. In particular, estimated GAI values followed the expected spatial patterns and agreed reasonably well with the temporal and dynamic trends measured on the ground, being able in most of the cases to capture the multiple growing phases typical of managed meadows. Global RMSE accuracy reached 1.68 [m^2/m^2], with a lower limit of 1.40 [m^2/m^2] for the Leutasch site, which is a good achievement considering the range of variability of the target variable and the complex nature of the retrieval problem in the Alpine environment with moderate resolution remote sensing data. As confirmed by the comparative analysis, the proposed method represents a step forward with respect to the currently available MODIS LAI products and opens the path for new, specific and more accurate monitoring activities in mountain areas at regional scale.

Table 7.2. Summary of the accuracy performance achieved by the MCD 15A3 product when compared with interpolated GAI measurements in the three field measurements sites over the period 2005-2007.

	Neustift	Lägenfeld	Leutasch	Global
2005	RMSE = 1.73 Bias = -0.7 SD = 1.58	RMSE = 2.28 Bias = 0.97 SD = 2.07	RMSE = 1.42 Bias = -1.15 SD = 0.83	RMSE = 1.88 Bias = -0.22 SD = 1.87
2006	RMSE = 2.25 Bias = -1.41 SD = 1.75	RMSE = 2.93 Bias = 2.32 SD = 1.8	RMSE = 1.45 Bias = -1.16 SD = 0.87	RMSE = 2.30 Bias = -0.08 SD = 2.29
2007	RMSE = 1.98 Bias = -0.92 SD = 1.76	RMSE = 2.54 Bias = 1.27 SD = 2.21	RMSE = 1.83 Bias = -1.14 SD = 1.44	RMSE = 2.17 Bias = -0.19 SD = 2.16
Global	RMSE = 2.00 Bias = -1.01 SD = 1.73	RMSE = 2.60 Bias = 1.51 SD = 2.12	RMSE = 1.57 Bias = -1.15 SD = 1.07	RMSE = 2.13 Bias = -0.16 SD = 2.12

Table 7.3. Summary of the accuracy performance achieved by the MCD 15A3 product when compared with interpolated GAI measurements in the three field measurements sites over the period 2005-2007.

	Neustift	Lägenfeld	Leutasch	Global
2005	RMSE = 1.62 Bias = 0.31 SD = 1.59	RMSE = 1.41 Bias = -0.33 SD = 1.37	RMSE = 1.35 Bias = 0.37 SD = 1.30	RMSE = 1.47 Bias = 0.11 SD = 1.47
2006	RMSE = 2.22 Bias = -0.80 SD = 2.07	RMSE = 1.79 Bias = 0.50 SD = 1.72	RMSE = 1.46 Bias = 0.09 SD = 1.45	RMSE = 1.87 Bias = -0.08 SD = 1.87
2007	RMSE = 1.67 Bias = 0.51 SD = 1.60	RMSE = 1.85 Bias = -0.08 SD = 1.85	RMSE = 1.38 Bias = 0.11 SD = 1.37	RMSE = 1.66 Bias = 0.18 SD = 1.65
Global	RMSE = 1.86 Bias = 0.0004 SD = 1.86	RMSE = 1.71 Bias = 0.04 SD = 1.71	RMSE = 1.40 Bias = 0.19 SD = 1.38	RMSE = 1.68 Bias = 0.07 SD = 1.68

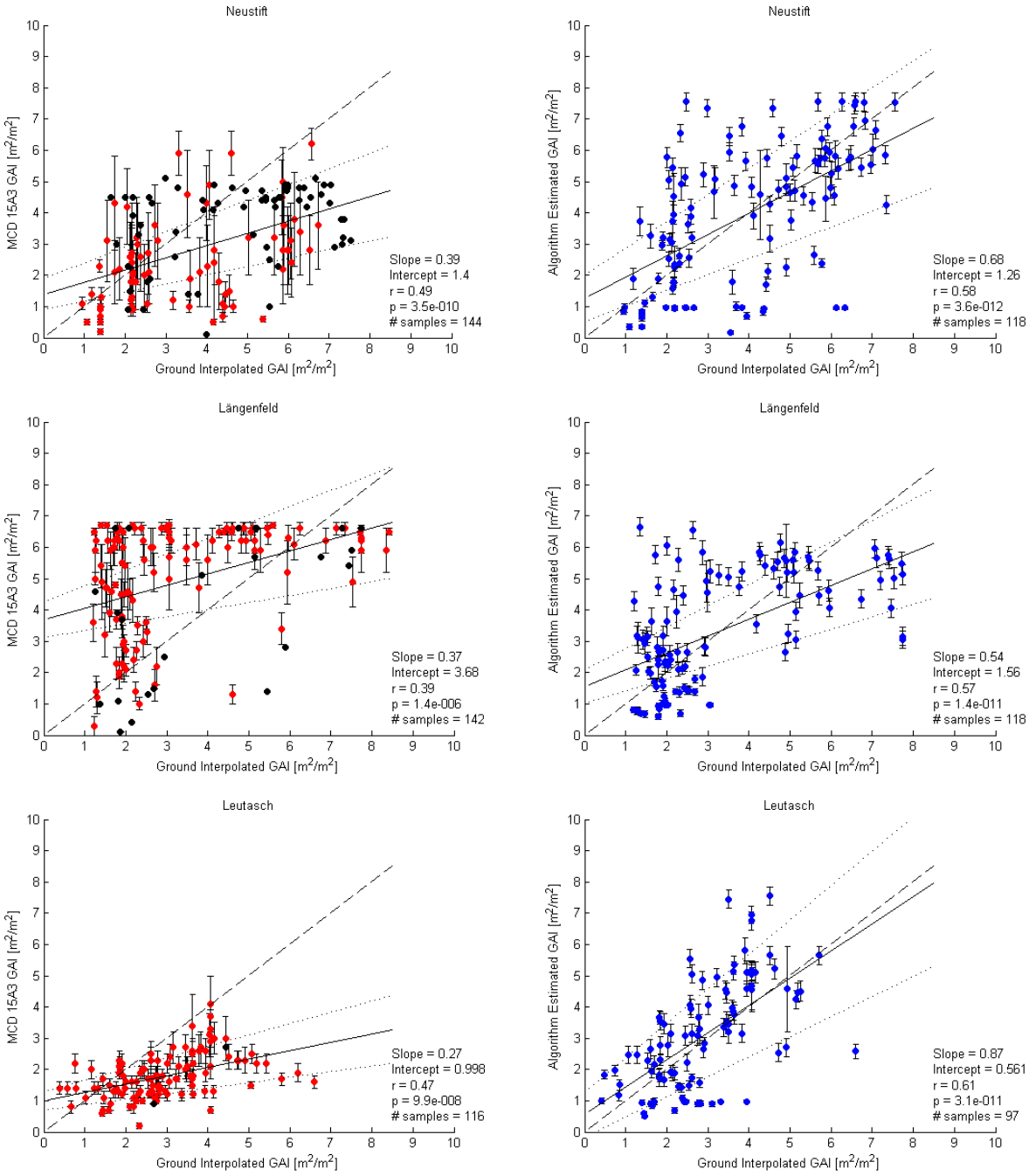


Figure 7.10. Scatter plots of estimated versus ground interpolated GAI values over the 2005-2007 period for the Neustift, Leutasch and Längenfeld sites: (left column) MCD 15A3 product and (right column) proposed algorithm. Linear trend line (solid line) and 95% intervals (dotted lines) are reported to be compared with the 1-to-1 line (dashed line). Black circles are associated with backup-algorithm estimates (only for the MCD 15A3 product).

This could be the case of the characterization of human management in mountain pastures and meadows (e.g. timing/frequency of harvesting events or timing/duration of grazing periods) or a more accurate mapping of carbon sequestration of grasslands (taking into account also inter-annual dynamics) (*Gianelle et al., 2009*). A limitation regarding the validation analysis presented is related to the considered reference measurements. Such data represent a valuable database on temporal and range dynamics of GAI in mountain grasslands. However, as they are acquired at plot scale, they cannot be considered representative for the whole footprint of the MODIS pixel. This issue is reflected in some misalignment between estimated and measured temporal trajectories causing sometimes consistent deviations in the scatter-plots for some samples. However, it is worth pointing out that the main objective of the validation analysis was to demonstrate the improved capability of the proposed algorithm to capture the dynamics (both in time and range) typical of mountain grasslands rather than an absolute validation of the estimated GAI values. The accuracy metric proposed aimed at a quantitative comparison between the proposed algorithm and the MCD 15A3 product in the same challenging operative condition for highlighting the different performance. Further validation activities of the proposed algorithm, with a more appropriate and spatially representative field sampling, are planned as a future development of this work.

The proposed algorithm is based on a relatively simple one-dimensional canopy radiative transfer model. This model is subject to approximations that have been partially overcome by more sophisticated 3-D approaches (*De Visser et al., 2002; Evers et al., 2007*). However, these approaches present the drawback of an increased complexity and a more complicated parameterization of the model. There is no clear proof that the increased accuracy of 3-D models exceeds the ambiguity introduced by additional model parameters, especially when the retrieval deals with medium resolution satellite remote sensing data. Given a limited number of spectral bands (as in this work) an easy parameterization of the model becomes a constraint for an effective inversion of the model. For this reason this strategy has been followed by many authors in similar operational conditions (*Duveiller et al., 2011; Jaquemoud et al., 2009*), suggesting that the 1-D approach is adequate also in this work.

The main features that allow the improved performance of the proposed algorithm with respect to the MCD 15A3 product are the calibration of the radiative transfer model to the specific characteristics of mountain grasslands and the increased spatial resolution. Model calibration is crucial in order to adapt the model spectral response to that of the target biome, thus allowing the algorithm to better follow the range of target values and to limit the ill-posed nature of the inversion process when only a few spectral bands (as in this case) are available as input. The higher spatial resolution helps reducing the ambiguity and spectral mixture effects related to the land-cover and topography heterogeneity. However, some of these effects still remain, as indicated by unexpected fluctuations of GAI values during the temporal trajectories in some plots especially in correspondence to bad acquisition geometries (high zenith angles) which determine an enlargement of the actual footprint of the sensor. The

interpolation of the temporal trajectory or a proper filtering based on the actual size and shape of the sensor footprint could represent valid strategies in order to mitigate the contamination of heterogeneous pixels and further improve the characterization of temporal dynamics of GAI.

The need of ancillary data (a digital elevation model and a land-cover map) as input does not represent a limitation to the portability of the proposed algorithm across different mountain landscapes. In this work we considered ancillary data easily available to the scientific community, thus demonstrating that very high resolution or customized products are not required. An issue could be represented by the availability of prior information on vegetation and soil characteristics needed for a proper parameterization of the radiative transfer model. While in this work values well representative for a wide range of field conditions in mountain environments have been used, the extension to other mountain landscapes across the world may require re-tuning the model input parameter ranges.

As final remark, it is worth pointing out that the algorithm proposed does not define specific constraints on the data to be used and can be extended to other medium resolution sensors. The use of MODIS was motivated by the fact that it is probably the best known product within Earth observation community. In addition, MODIS images are easily and freely available and show the highest spatial resolution (in the red and near infrared bands) currently available for daily mapping of the Earth surface from space. Despite MODIS is approaching the end of its expected lifetime, MODIS data represent a valuable long term record for retrospective analysis and its operational successor, VIIRS onboard of the NPP satellite, will have similar spectral characteristics, thus ensuring the applicability of the proposed algorithm without significant changes also in the future.

7.6 Conclusions

This study presented and discussed an improved algorithm for the retrieval of green area index (GAI) from moderate resolution satellite MODIS imagery in Alpine pastures and meadows. The proposed method, which is based on the inversion of a canopy radiative transfer model by means of a look-up table procedure, presents two main advantages with respect to currently available global LAI/GAI products, such as the MODIS Collection 5 product: i) the tuning of the radiative model on the spectral characteristics specific of Alpine grasslands; and ii) the improved spatial resolution. Moreover, local topography of the investigated area is taken into consideration to correct solar and sensor view geometries. Due to these features, the algorithm demonstrated to capture reasonably well the range and temporal dynamics of the target variable typical of intensively managed as well as natural grasslands while showing an increased accuracy when compared to ground measurements with respect to the standard MODIS product. These results open the path to the exploitation of

moderate resolution data for novel and improved monitoring activities in the challenging Alpine area at regional scale.

Due to its formulation, the algorithm does not require field reference measurements and is computationally fast, thus being suitable for the processing of large amounts of data and long temporal databases. Moreover, its adaptation to the use with sensors different from MODIS or on other study sites is straightforward and without time consuming training/tuning tasks. The only user demanding activity is the definition of the model parameterization on the basis of prior knowledge on vegetation and soil characteristics of the area.

An important future development of this work will regard the validation activity. For this task, the acquisition of a set of field measurements well distributed within the study area and representative for the spatial variability of the footprint of medium resolution systems is going to be planned. This will give also the opportunity for a finer analysis and comparison with existing medium resolution satellite products, such as the MODIS Collection 5, the SPOT Vegetation and the MERIS products. Another interesting development is represented by the integration within the retrieval process or as post-processing step of methods for smoothing and filling the temporal series of estimated GAI values. This might reduce the variability within estimated temporal trajectories and consequently improve the ingestion/assimilation of estimated GAI data into productivity and carbon sequestration models, which is another important topic that will be investigated in the future.

Chapter 8

Conclusions

This chapter concludes the thesis by providing a summary of the main novel contributions and experimental analysis presented in this document. Possible future developments are also discussed.

8.1 Summary and Discussion

In this thesis we investigated and developed advanced methods and systems for the retrieval of geo-/bio-physical variables from satellite remote sensing imagery. In particular, several issues related to different steps of the retrieval process as well as to its application to challenging real operational scenarios were addressed. For each considered topic, an analysis of the state-of-the-art was conducted. Starting from this analysis, novel solutions were proposed, implemented ad applied to remote sensing data to assess their effectiveness. The achieved results pointed out that the proposed solutions represent a valuable contribution for the exploitation of the remote sensing technology for mapping and monitoring natural resources and physical processes on the Earth surface, with particular regard to the mountain environment. This represents a hot topic in the scientific community especially in the last years, thanks to the potential offered by new generation and upcoming satellite remote sensing systems and the growing interest in the accurate and up-to-date mapping and monitoring of the Earth surface.

A summary of the main novel contributions and conclusions for each addressed topic is reported in the following.

Machine learning methods are an effective solution to address the geo-/bio-physical variable retrieval problem. In Chapter 2 we investigated this topic by introducing the use of an advanced state-of-the-art method, the ε-insensitive Support Vector Regression technique, for the retrieval of soil moisture content from microwave remote sensing data. At the time of this analysis no studies existed in the literature that investigated the effectiveness of this technique in this specific context. The effectiveness of the SVR method was assessed in different

operational scenarios: i) the inversion of a physical based electromagnetic model; and ii) the inference of an empirical based model from a limited number of reference samples measured in the field. Attractive properties of the SVR method were pointed out in this analysis, namely: i) the good estimation accuracy in both operational scenarios; ii) the robustness to outliers and noise present in the data; iii) the effectiveness in presence of a limited number of reference samples; and iv) the ease of use. Beside the methodological comparison *per se*, this analysis provided valuable indications about the use of advanced state-of-the-art machine learning methods for the retrieval of soil moisture in real operational scenarios. These considerations are of crucial importance when defining accurate and robust retrieval systems or building soil moisture estimation processors for upcoming satellite missions or near real time applications.

The investigation on machine learning regression methods was continued in Chapter 3, where the issue of tuning the free parameters that control the learning process was addressed. This task is crucial for achieving good accuracy and robustness from the considered method. The model-selection problem was formulated in the framework of the multi-objective optimization, where two or more quality metrics can be jointly exploited. Solutions of the optimization problem were derived according to the concept of Pareto optimality, which allows deriving multiple optimal trade-off solutions among which selecting the one that best meets the application constraints and requirements. The proposed strategy was assessed in the challenging application domain of soil moisture retrieval from microwave remotely sensed data with the SVR technique, showing superior performance with respect to the mono-objective model-selection. The main features of the multi-objective strategy that emerged are: i) the intrinsic increased robustness with respect to traditional approaches, since each solution obtained is the result of the optimization of multiple metrics; ii) the effectiveness in deriving trade-off solutions that jointly optimize the metrics selected; and iii) the possibility to select the parameter configuration that best meets the requirements and the constraints of the specific retrieval problem considered. This task was easily handled with a graphical representation of the estimated Pareto front in the case of two or three metrics to optimize, which is the condition that most often characterize practical estimation problems. Nonetheless, further developments are required for defining automatic strategies for the selection of the optimal trade-off solution in the case of four or more objectives to optimize. Finally, from the computational viewpoint, the method showed only a slight increase in the time required for solving the optimization problem with respect to standard mono-objective strategies. This is another interesting feature of the proposed method.

Chapter 4 addressed the problem of overcoming the dichotomy between physical based and empirical based approaches to the retrieval of geo-/bio-physical variables. To this aim a novel hybrid approach based on the integrated use of a set of (few) field reference samples and a physical based theoretical model was proposed. According to the hybrid approach, the estimation process is modeled by two terms: the former expresses the relationship between

the input features and the target variable according the model based on the physics of the considered problem; the latter corrects the deviation between theoretical model estimates (affected by biases and approximations intrinsic in the theoretical model formulation) and true target values according to an empirical data-driven relationship inferred from the (few) field reference samples. Two different strategies for the estimation and correction of such a deviation were presented: i) the global deviation bias strategy; and ii) the local deviation bias strategy. The effectiveness of the proposed hybrid approach was investigated for the domain of soil moisture retrieval from microwave signals. The achieved results were in general promising and indicated the proposed integration approach as effective for the estimation of biophysical parameters from remotely sensed data, since: i) it allows increasing the accuracy of the estimates by overcoming biases and simplifications intrinsic within the analytical formulation of theoretical models; 2) it was capable to handle the variability of the deviation between model estimates and field reference samples within the input space domain; and 3) it was simple, easy to implement and fast during the processing. It is worth noting that, despite the experimental analysis was carried out on one specific application domain, the proposed hybrid formulation is general and can be extended to other experimental scenarios where the limited availability of field reference samples and the spatial and temporal variability of the target geo-/bio-physical variable represent a limitation for the development and applicability of robust and general empirical based models.

The remaining chapters moved the attention to operational estimation scenario. In particular, Chapter 5 and 6 addressed the retrieval soil moisture content from new generation RADARSAT2 satellite SAR imagery in the complex mountain environment. Chapter 5 presented a sensitivity analysis of the SAR backscattering coefficient to soil moisture content in a small Alpine catchment. This analysis pointed out that both topography and vegetation/land-cover heterogeneity strongly affect the signal acquired over mountain areas, introducing a rather high ambiguity in the sensitivity of backscattering to soil moisture content. Altitude, local incidence angle and NDVI derived from ancillary data revealed to be able to explain to a large extent the high ambiguity intrinsic in the SAR signal. On the basis of these considerations, an advanced retrieval system was introduced taking advantage from the methodological developments discussed previously (namely the use of the SVR technique and the tuning of its free parameters according to the proposed multi-objective strategy) and exploiting the available ancillary data as additional inputs to the retrieval process. The experimental analysis conducted on two RADARSAT2 SAR images indicated the effectiveness of the proposed retrieval system in terms of: i) capability to both exploit the information provided by the ancillary data to reduce the ambiguity intrinsic into the SAR signal and address the complex estimation problem in mountain areas; ii) estimation accuracy over *in-situ* point reference measurements; and iii) capability to reproduce the soil humidity patterns when applied to distributed data.

Chapter 6 went a step further in the analysis and exploitation of RADARSAT2 SAR imagery by investigating the fully polarimetric capability of the acquisition system considered. The aim was to understand the potentialities and the limitations, of this advanced acquisition modality of new generation SAR systems with regard to the analysis and mapping of soil variables in the complex mountain environment. Three different feature extraction strategies were exploited in this analysis: i) the standard intensity and interferometric SAR processing; ii) an advanced polarimetric decomposition technique, namely the alpha-Anisotropy-Entrodpy decomposition; and iii) a general purpose feature transformation technique, i.e., the Independent Component Analysis. Different combinations of the extracted features were investigated and evaluated in terms of both accuracy on *in-situ* point measurements and capability to reproduce spatial patterns of the target variable. The achieved results pointed out that the joint use of multiple polarimetric features is important for improving the accuracy and reliability of soil moisture content estimates in the Alpine area. In particular the use of the linear depolarization ratio HV/VV together with the standard HH polarization determined the achievement of the best performance, thanks to its capability to provide information on the local vegetation and roughness status of the target area. The other feature extraction strategies investigated implied a more time demanding processing of the SAR imagery and did not lead to improvements in the accuracy or in the capability to reproduce the local patterns of soil moisture content. This points out that in the complex mountain environment, not too sophisticated feature extraction methods are preferable.

In Chapter 7 we focused the attention on the mapping and monitoring of another crucial geo-/bio-physical environmental variable, i.e., the green area index (GAI) of Alpine grasslands. In this context an improved retrieval algorithm was presented, based on the use of moderate resolution satellite MODIS imagery and specifically customized for Alpine meadows and pastures. The proposed algorithm, which is based on the inversion of a canopy radiative transfer model by means of a look-up table procedure, is characterized by two main advantages with respect to currently available global LAI producs, such as the MODIS collection 5 product: i) the tuning of the radiative model on the spectral characteristics specific of Alpine grasslands; and ii) the improved spatial resolution. Moreover, it takes into consideration the local topography of the investigated area to correct solar and sensor view geometries. The algorithm was applied to a temporal series of satellite MODIS images acquired over the central Alps during the period 2005-2007 and GAI estimates were validated for both temporal consistency and accuracy with the use of time series of ground measurements collected in three different study sites in the area. For comparison purposes, the collection 5 MODIS LAI product was also considered. The algorithm demonstrated to capture reasonably well the spatial, temporal and range dynamics of the target variable of intensively manages as well as natural grasslands. At the same time, it showed an increased accuracy when compared with ground measurements with respect to the standard MODIS product. It is worth noting that no investigations were carried out in Alpine areas at regional scale before

this study. The achieved results open thus the path to the exploitation of moderate resolution data for new and improved monitoring activities in this challenging environment at regional scale.

8.2 Final Remarks and Future Developments

In this thesis we investigated and developed advanced methods and systems which can significantly improve the retrieval of geo-/bio-physical variables from satellite remote sensing imagery. Several issues related with the retrieval problem were discussed and addressed. Some of them were specific of the application domain investigated, with particular attention to the mountain environment. Some others were more general. In all the cases the proposed solutions demonstrated to be successful and contributed to a more effective use of the remote sensing technology for an accurate mapping and monitoring of natural resources and physical processes on the Earth surface. This is a valuable outcome especially in the optic of the exploitation of the new and upcoming satellite remote sensing missions.

Following the direction toward the exploitation of new and upcoming satellite remote sensing data for monitoring activities, several issues remain still open and need further investigations. Issues specific of the topic addressed were briefly discussed at the end of each chapter of this thesis. Here we identify (among others) three general topics that deserve future studies:

- Development of retrieval methodologies that can fully exploit the high temporal frequency of new generation and upcoming satellite remote sensing systems to improve the temporal consistency and accuracy of the estimation process
- Study of automatic methods for the adaptation of the retrieval system to different domains (e.g., several study areas with slightly different topographic and phenological conditions)
- Generalization of the proposed methods and systems to the retrieval of other geo-/bio-physical variables (e.g., snow pack characteristics) from new generation satellite remote sensing imagery.

List of Publications

Peer Reviewed Papers in International Journals

Pasolli, L., Notarnicola, C., Bruzzone, L. (2011). Estimating soil moisture with the support vector regression technique. *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 6, pp. 1080-1084.

Pasolli, L., Notarnicola, C., Bruzzone, L., Bertoldi, G., Della Chiesa, S., Hell, W., Niedrist, G., Tappeiner, U., Zebisch, M., Del Frate, F., and Vaglio Laurin, G. (2011). Estimation of soil moisture in an alpine catchment with RADARSAT2 images. *Applied and Environmental Soil Science*, vol. 2011, Article ID 175473, 12 pages.

Pasolli, L., Notarnicola, C., Bruzzone, L., Bertoldi, G., Della Chiesa, S., Niedrist, G., Tappeiner, U., Zebisch, M. (2012). Polarimetric RADARSAT 2 imagery for soil moisture retrieval in Alpine areas. *Canadian Journal of Remote Sensing*, in press.

Pasolli, L., Notarnicola, C., Bruzzone, L. (2012). Multi-objective Parameter Optimization in Support Vector Regression: General Formulation and Application to the Retrieval of Soil Moisture from Remote Sensing Data. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, accepted for publication.

Pasolli, L., Notarnicola, C., Bruzzone, L., Zebisch, M., Wohlfahrt, G. Retrieval of green area index in mountain grasslands in the alpes from modis satellite imagery. *Remote Sensing of Environment*, submitted for publication in March 2012.

Pasolli, L., Bruzzone, L., Notarnicola, C. A novel hybrid approach to the retrieval of geo-/bio-physical variables from remotely sensed data. *IEEE Transaction on Geoscience and Remote Sensing*, in preparation.

Papers in International Conferences and Workshops

Pasolli, L., Notarnicola, C., Bruzzone, L. (2009). Soil Moisture Estimation from Microwave Remote Sensing Data with Non-Linear Machine Learning Techniques. In *Image and Signal Processing for Remote Sensing XV*, edited by Lorenzo Bruzzone, Claudia Notarnicola, Francesco Posa, Proceedings of SPIE Vol. 7477 (SPIE, Bellingham, WA, 2009) 7477 1C.

Pasolli, L., Notarnicola, C., and Bruzzone, L. (2010). Soil parameters estimation from multi-frequency polarimetric SAR images with e-Insensitive Support Vector Regression. *Proc. of the IEEE Gold conference*, Livorno, Italy, April 2010.

Pasolli, L., Notarnicola, C., Bruzzone, L., Bertoldi, G., Niedrist, G., Tappeiner, U., Zebisch, M., Del Frate, M., and Vaglio Laurin, G. (2010). Exploiting fully polarimetric RADARSAT2 images for retrieval of soil moisture in alpine grassland and pasture. *Proc. of the URSI Microwave Signature Conference*, Firenze, Italy, April 2010.

Notarnicola, C., Bertoldi, G., Del Frate, F., Della Chiesa, S., Hell, V., Pasolli, L., Santi, E., Shellenberger, T., Tappainer, U., Vaglio Laurin, G., Ventura, G., and Zebisch, M. (2010). SOFIA, SOil and Forest Information Retrieval by using RADARSAT2 images: first results of the project. *ESA workshop LIVING PLANET*, Bergen, Norway, June 2010.

Notarnicola, C., Ventura, B., Pasolli, L., Di Giuseppe, F., and Zebisch, M. (2010). Towards an operational daily soil moisture index derived from combination of MODIS, ASAR and AMSR-E data. *Proc. of the IEEE-International Geoscience and Remote Sensing Symposium IGARSS 2010*, Honolulu, Hawaii, July 2010.

Pasolli, L.,Notarnicola, C., Bruzzone, L. (2010). Multiobjective model selection for non-linear regression techniques. *Proc. of the IEEE-International Geoscience and Remote Sensing Symposium IGARSS 2010*, Honolulu, Hawaii, July 2010.

Notarnicola, C., Di Giuseppe, F., Ventura, B., Pasolli, L., Petitta, M., Bonafè, G., Caporaso, L., and Bitelli, M. (2010). Exploitation of C and X band SAR images for soil moisture change detection estimation in agricultural areas (Po valley-ITALY). *Proc. of the SPIE Remote Sensing Conference*, Toulouse, France, September 2010.

Pasolli, L., Notarnicola, C., Bruzzone, L., Bertoldi, G., Niedrist, G., Tappeiner, U., Zebisch, M., Del Frate, F., and Vaglio Laurin. G. (2010). Analysis of polarimetric RADARSAT2 images for soil moisture retrieval in an alpine catchment. *Proc. of the SPIE Remote Sensing Conference*, Toulouse, France, September 2010.

Pasolli, L., Notarnicola, C., Bruzzone, L., Bertoldi, G., Della Chiesa, S., Tappeiner, U., Zebisch, M., Del Frate, F., Vaglio Laurin. G. (2011). Soil Moisture Retrieval in Alpine Areas by using Support Vector Regression Techniques and polarimetric RADARSAT2 Images. *Proc. of the PolInSAR 2011 Workshop*, ESA ESRIN, Frascati, Roma, Italy, January 2011.

Pasolli, L., Notarnicola, C., Bruzzone, L., Zebisch, M. (2011). Spatial and temporal mapping of leaf are index in alpine pasture and meadows with satellite MODIS imagery., *Proc. of the IEEE Multi-temp workshop 2011*, Trento, Italy, July 2011.

Pasolli, L., Bruzzone, L., Notarnicola, C. (2011). A novel hybrid approach to the estimation of biophysical parameters from remotely sensed data. *Proc. of the IEEE-International Geoscience and Remote Sensing Symposium IGARSS 2011*, Vancouver, Canada, July 2011.

Pasolli, L., Notarnicola, C., Bruzzone, L., Bertoldi, G., Niedrist, G., Tappeiner, U., Zebisch, M., Del Frate, F., Vaglio Laurin, G. (2011). Spatial and temporal mapping of soil moisture content with polarimetric RADARSAT 2 SAR imagery in the Alpine area. *Proc. of the IEEE-International Geoscience and Remote Sensing Symposium IGARSS 2011*, Vancouver, Canada, July 2011.

Notarnicola, C., Di Giuseppe, F., Pasolli, L., Temimi, M., Ventura, B., Zebisch, M. (2011). A synergetic use of observations from MODIS, SEVIRI MSG, ASAR and AMSR-E to infer daily soil moisture index. *Proc. of the IEEE-International Geoscience and Remote Sensing Symposium IGARSS 2011*, Vancouver, Canada, July 2011.

Pasolli, L., Notarnicola, C., Bruzzone, L. (2011). Comparison of L and C band Polarimetric SAR Data for the Retrieval of Soil Moisture in Alpine Areas. *Proc. of the SPIE Remote Sensing Conference*, Prague, Czech Republic, 19-22 September 2011.

Bibliography

Ahmad, S., Kalra, A., Stephen, H. (2010). Estimating soil moisture using remote sensing data: a machine learning approach. *Advances in Water Resources*, vol. 33, pp. 69-80.

Baghdadi, N., Gauthier, Y., and Bernier, M. (1997). Capability of multitemporal ERS-1 SAR data for wet-snow mapping. *Remote Sensing of Environment*, vol. 60, no. 2, pp. 174-186.

Baldridge, A. M., S.J. Hook, C.I. Grove, and G. Rivera, 2009.. The ASTER Spectral Library Version 2.0. Remote Sensing of *Environment*, vol. 113, pp. 711-715.

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berhelot, B., Nino, F., Weiss, M., Samain, O., Roujean, J.L., & Leroy, M. (2007). LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: Part 1: Principles of the algorithm. *Remote sensing of Environment*, vol. 110, pp. 275-286.

Barnett, T.P., Adam, J.C., and Lettenmeier, D.P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, vol. 438, pp. 303–309.

Bastiaanssen, W.G.M., and Bos, M.G. (1999). Irrigation performance indicators based on remotely sensed data: a review of literature. *Irrigation and Drainage Systems*, vol. 13, pp. 291-311.

Bastiaanssen, W.G.M., E.J.M. Noordman, H. Pelgrum, G. Davids, B.P. Thoreson, R.G. Allen, (2005). SEBAL model with remotely sensed data to improve water-resource management under actual field conditions. *ASCE J. Irrig. Drain. E.*, vol. 131, pp. 85-93.

Becker, A., Korner, C., Brun, J.J., Guisan, A., Tappeiner, U. 2007. Ecological and land-use studies along elevational gradients. *Mountain Research and Development*, vol. 27(1), pp. 58–65.

Beniston, M. (2005). Mountain climates and climatic change: An overview of processes focusing on the European Alps. *Pure and Applied Geophysics*, vol. 162(8-9), pp. 1587-1606.

Bertoldi, G., Della Chiesa, S., Niedrist, G., Rist, A., Tasser, E., and Tappeiner, U. (2010). Space-time evolution of soil moisture, evapotranspiration and snow cover patterns in a dry alpine catchment: an interdisciplinary numerical and experimental approach. *Geophysical Research Abstracts*, EGU General Assembly 2010, Vienna, Austria, vol. 12 EGU2010-12109.

Beven, K. & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, vol. 249, pp. 11-29.

Beven, K. (2006). Manifesto of the equifinality thesis. *Journal of Hydrology*, vol. 320, pp.18-36.

Bindlish, R., Barros, A.P. (2000). Multifrequency soil moisture inversion from SAR measurements with the use of IEM. *Remote Sensing of Environment*, vol. 71, no. 1, pp. 67-88.

Bolten, J.D., Crow, W.T., Zhan, X., Jackson, T.J., Reynolds, C.A. (2010). Evaluating the utility of remotely sensed soil moisture retrievals for operational agricultural drought monitoring. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 3, no. 1.

Bontemps, S., Defourny, F., Van Bogaert, E., Weber, J.L., Arino, O. (2010). GlobCorine – A joint EEA-ESA project for operational land dynamics monitoring at pan-European scale. In *Proceedings of Esa Living Planet Symposium*, Bergen, Norway.

Borengasser, M., Hungate, W.S., and Watkins, R. (2008). *Hyperspectral Remote Sensing Principles and Applications*. C. Press, Ed.

Brunetti, M., Maugeri, M., and Nanni. T. (2001). Changes in total precipitation, rainy days and extreme events in northeastern Italy. *Int. J. of Climatology*, vol. 21, pp. 861–871.

Bruzzone, L., and Melgani, F. (2005). Robust multiple estimator system for the analysis of biophysical parameters from remotely sensed data. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 1, pp. 159-174.

Bruzzone, L., and Persello, C. (2009). A novel approach to the selection of spatially invariant features for the classification of hyperspectral images with improved generalization ability. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 47, no. 9, pp. 3180-3191.

Burges, C. (1998). *A tutorial on support vector machine for pattern recognition*, Boston, MA: Kluver.

Camp-Valls, G., Gomez-Chova, L., Munoz-Marì, J., Vila-Frances, J., Amoros-Lopez, J., Calpe-Maravilla, J. (2006). Retrieval of oceanic chlorophyll concentration with relevance vector machines. *Remote Sensing of Environment*, vol. 105, no. 1, pp. 23-33.

Chang, M.W., and Lin, C.J. (2005). Leave-one-out bounds for support vector regression model selection. *Neural Comput.*, vol. 17, no. 5, pp. 1188-1222.

Cherkassky, V., and Mulier, F. (1998). *Learning from data: Concepts, theory and methods*. New York: Wiley.

Cherkassky V., and Yunquian, M. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, vol. 17, pp. 113-126.

Chang, C.C., and Lin, C.J. (2011). LIBSVM: A library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 27, pp. 1-27.* Software available at: httwp://www.csie.ntu.edu.tw/~cjlin/libsvm.

Chen, H.S. (1997). *Remote Sensing Calibration Systems: An Introduction*. Deepak, A Publishing.

Chu, W., Keerkhi, S., Ong, C.J. (2004). Bayesian support vector regression using a unified loss function. *IEEE Transaction on Neural Networks*, vol. 15, no. 1, pp. 29-44.

Cihlar, J., Dobson, M.C., Schmugge, T., Hoogeboom, P., Janse, A.R.P., Baret, F., Guyot, G., Le Toan, T., and Pampaloni, P. (1987). Procedures for the description of agricultural crops and soils in optical and microwave remote sensing studies. *International Journal of Remote Sensing*, vol. 8, no. 3, pp. 427-439.

Clevers, J.G.P.W., & Verhoef, W. (1991). Modeling and synergic use of optical and microwave remote sensing . Report 2: LAI estimation from canopy reflectance and WDVI: A sensitivity analysis with the SAIL model. *BCRS Report*, vol. 90-39.

Colombo, R., Bellingeri, D., Fasolini, D., & Marino, C.M. (2003). Retrieval of leaf area index in different vegetation types using high resolution satellite data. *Remote Sensing of Environment*, vol. 86(1), pp. 120-131.

Combal, B., Baret, F., & Weiss, M. (2002). Improving canopy variables estimation from remote sensing data by exploiting ancillary information. Case study on sugar beet canopies. *Agronomie*, vol. 86(1), pp. 120-131.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Mace, D., Pragnere, A., et al. (2003). Retrieval of canopy biophysical variables from bidirectional reflectance: using prior information to solve the ill-posed inverse problem. *Remote Sensing of Environment*, vol. 84(1), pp. 1-15.

Comon, P. (1994). Independent component analysis: a new concept?. *Signal Processing*, vol. 34, pp. 287-314.

Cloude, S.R., and Pottier, E. (1996). A review of target decomposition theorems in radar polarimetry. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 34, pp. 498-518.

Cloude, S.R., and Pottier, E. (1997). An entropy based classification scheme for land applications of polarimetric SAR. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 35, pp. 68-78.

Curran, P. J. (1980). Multispectral remote sensing of vegetation amount. *Prog. Phys. Geogr.*,vol. 4, pp. 315–341.

Curran P. J., and Williamson, H.D. (1987). Estimating the green leaf area index of grassland with airborne multispectral scanner data. *Oikos, vol. 49, no. 2, pp. 141-148.*

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C. (2008). Inversion of a radiative trasfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sensing of Environment*, vol. 112, pp. 2592-2604.

Deb, K. (2001). *Multi-Objective optimization using evolutionary algorithms*. Chichester, U.K.: Wiley.

Deb, K., Pratab, A., Agarwal, S., and Mejarivan, T. (2002). A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Transaction on Evolutionary Computation*, vol. 6, no. 2, pp. 182-197.

De Kauwe, M.G., Disney, M.I., Quaife, T., Lewis, P., Williams, M. (2011). An assessment of the MODIS collection 5 leaf area index product for a region of mixed coniferous forest, *Remote Sensing on Environment*, vol. 115, pp. 767-780.

Del Frate, F., Ferrazzoli, P., and Schiavon, G. (2003). Retrieving soil moisture and agricultural variables by microwave radiometry using neural networks. *Remote Sensing of Environment*, vol. 84, no. 2, pp. 174-183.

Derksen, C., Wulder, M., Ledrew, E., and Goodison, B. (1998). Associations between spatially autocorrelated patterns of SSM/I-derived praire snow cover and atmospheric circulation. *Hidrol. Process.*, vol. 2, no. 15, pp. 2307-2316.

De Visser, P., Marcelis, L., van der Heijden, G., Vos, J., Stuick, P., & Evers, J. (2002). 3D modeling of plants: a review. Tech. Rep. Wageningen: *Plant Research International B.V.*

Dobson, M.C., and Ulaby, F.T. (1981). Microwave backscatter dependence on surface roughness, soil moisture and soil texture: Part-III: Soil tension. *IEEE Trans. Geosci. Electron.*, vol. GRS-19, no. 1, pp. 51-61.

Dorigo, W., Zurita-Milla, R., de Wit, A., Brazile, J., Singh, R., & Schaepman, M. (2007). A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation*, vol. 9, pp. 165-193.

Dubois, P.C., van Zyl, J., and Engman, T. (1995). Measuring soil moisture with imaging radars. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 33, no. 4, pp. 915-926.

Durba, A.A., King, R.L., Younan, N.H. (2007). Support vector regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sensing of Environment*, vol. 107, pp. 348-361.

Duvellier, G., Weiss, M., Varet, F., & Defourny, P. (2011). Retrieving wheat green area index during the growing season from optical time series measurements based on neural network radiative transfer inversion. *Remote Sensing of Environment*, vol. 115(3), pp. 887-896.

Engman E T (1991). Application of microwave remote sensing of soil moisture for water resources and agriculture. *Remote Sensing of Environment*, vol. 35, pp. 213–226.

Engman, E.T., and Gurney, R.J. (1991). *Remote Sensing in Hyrdology*, Chapman and Hall, London.

Evers, J.B., Bos, J., Fournier, C., Andrieu, ., Chelle, M., & Strik, P.C. (2007). An architectural model of spring wheat: evaluation of the effects of population density and shading on model parametrization and performance. *Ecological Modeling*, vol. 200 (3-4), pp. 308-320.

Fang, H., Wei, S., Liang, S. (2012). Validation of MODIS and CYCLOPES LAI products using global field measurement data. *Remote Sensing of Environment*, vol. 119, pp. 43-54.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D. (2007). The Shuttle Radar Topography Mission. *Reviews of Geophysics*, vol. 45, pp. 1-33.

Feldman, A.M. (1980). *Welfare Economics and Social Choice Theory*. Kluwer, Boston.

Fernandes, R.A., Miller, J.R., Chen, J.M., & Rubinstein, I.G. (2004). Evaluating image-based estimates of leaf area index in boreal conifer stands over a range of scales using high resolution CASI imagery. *Remote Sensing of Environment*, vol. 89, pp. 200-216.

Fonseca, C.M., and Fleming, P.J. (1998). Multiobjective optimization and multiple constraints handling with evolutionary algorithms – Part I: a unified formulation. *IEEE Transaction Syst., Man, Cybern. A., Syst., Humans*, vol. 28, no. 1, pp. 26-37.

Fung, A.K., Li, Z., and Chen, K.S. (1992). Backscattering from a randomly rough dielectric surface. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 30, no. 2, pp. 356-369.

Fung, A.K. (1994). *Microwave scattering and emission models and their application*. Norwood, MA: Artech House.

Garrigues, S., Allard, D., Baret, F., Weiss, M. (2006). Influence of landscape spatial heterogeneity on the non-linear estimation of leaf area index from moderate spatial resolution remote sensing data. *Remote Sensing of Environment*, vol. 105, pp. 286-198.

GCOS (2006). Systematic observation requirements for satellite based products for climate. *Report of United Nations Environment Programme*. GCOS-107, WMO/TD No. 1338.

Gebremichael, M., Rigon, R., Bertoldi, G., and Over, T.M. (2009). On the scaling characteristics of observed and simulated spatial soil moisture fields. *Nonlinear Processes in Geophysics*, vol. 16(1), pp. 141–150.

Gianelle D., Vescovo L., Marcolla B., Manca G., Cescatti A., 2009. Ecosystem carbon fluxes and canopy spectral reflectance of a mountain meadow. *International Journal of Remote Sensing*, vol. 30(2), pp. 435-449.

Girard, C.M. and Girard, M.C. (2003) *Processing of Remote Sensing Data*. Taylor & Francis.

Goodison B.E. and Louie, P.Y.T. (1986). Canadian methods for precipitation measurements and correction. in *Proc. Workshop Correction Precipitation Measurements*, 1-3 April, 1985, Zurich, Switzerland, pp. 141-145.

Graham, L.P., Hagemann, S., and Beniston, M., (2007). On interpreting hydrological change from regional climate models. *Clim. Change*.

Grayson, R. B., A. W. Western, F. H. S. Chiew, and G. Bloschl (1997). Preferred states in spatial soil moisture patterns: Local and nonlocal controls. *Water Resour. Res.*, vol. 33, pp. 2897– 2908.

Guoquing, S., and Ranson, K.J. (1995). A three dimensional radar backscatter model of forest canopies. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 33, no. 2, pp. 372-382.

Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., Strachan, I.B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, vol. 90, pp. 337-352.

Hadjimitsis, D.G., Clayton, C.R.I., and Hope, V.S. (2004). An assessment of the effectiveness of atmospheric correction algorithms through the remote sensing of some reservoirs. *International Journal of Remote Sensing*, vol. 25, no. 18, pp. 3651-3673.

Hajnsek, I., Pottier, E., Cloude, S.R. (2003). Inversion of surface parameters from polarimetric SAR. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 41, no. 4, pp. 727-744.

Hall, D.K., and Martinec, J. (1985). *Remote Sensing of Ice and Snow*. Chapman & Hall, London – New York.

Hallikainen, M.T., Ulaby, F.T., Dobson, M.C., El-Rayes, M.A., and Wu, L. (1985). Microwave dielectric behavior of wet soil – Part I: Empirical models and experimental observations. *IEEE Transaction on Geoscience and Remote Sensing*, vol. GRS-23, no. 1, pp. 25-34.

Hammerle, A., Haslwanter, A., Tappeiner, U., Cernusca, A., and G. Wohlfahrt (2008). Leaf area controls on energy partitioning of a temperate mountain grassland. *Biogeoscience*, vol. 5, pp. 421-431.

Heathman, G.C., P.J. Starks, L.R. Ahuja, T.J. Jackson, (2003). Assimilation of surface soil moisture to estimate profile soil water content. *Journal of Hyrdology*, vol 279, pp. 1-17.

Hedley, J., Roelfsema, C., Phinn, S.R. (2009). Efficient radiative transfer model inversion for remote sensing application. *Remote Sensing of Environment*, vol. 113, pp. 2527-2532.

Heiskanen, J. (2006). Estimating aboveground tree biomass and leaf area index in a mountain birch forest using ASTER stellite data. *International Journal of Remote Sensing*, vol. 27, no. 6, pp. 1135-1158.

Hino, M., Odaka, Y., Nadaoka, K., Sato, A. (1988). Effect of initial soil moisture content on the vertical infiltration process – a guide to the problem of runoff-ration and loss. *Journal of Hydrology*, vol. 102, no. 1-4, pp. 267-284.

Horton, P. et al. (2006). Assessment of climate-change impacts on alpine discharge regimes with climate model uncertainty. *Hydrological Processes,* vol. 20, pp. 2091–2109.

Huete, A.R. Liu, H.Q., Batchily, K., & vanLeeuwen, W. (1997). A comparison of vegetation indices global set of TM images for EOS-MODIS. *Remote Sensing of Environment,* vol. 40, pp. 231-239.

Jackson T J, Schmugge T J (1991). Vegetation effects on the microwave emission of soils. *Remote Sensing of Environment*, vol. 36, pp. 203–212.

Jackson, T. (1997). Soil moisture estimation using special satellite microwave/imager satellite data over a grassland region. *Water Resour. Res.*, vol. 33, pp. 1475–1484.

Jacquemoud, S., and Baret, F. (1990). PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment*, vol. 34(2), 75-91.

Jaquemoud, S., Bacour, C., Poilvè, H., Frangi, J.P. (2000). Comparison of four radiative transfer models to simulate plant crop reflectances: direct and inverse mode. Remote *Sensing of Environment*, vol. 74, pp. 471-481.

Jaquemoud, S., Verhoef. W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., Francois, C., Ustin, S.L. (2009). PROSPECT + SAIL models: a review of use for vegetation characterization. *Remote Sensing of Environment*, vol. 113, pp. 56-66.

Jensen, J.L.R., Humes, K.S., Hudak, A.T., Vierling, L.A., Delmelle, E (2011). Evaluation of the MODIS LAI product using independent lidar-derived LAI: A case study in mixed conifer forest. *Remote Sensing of Environment*, vol. 115, pp. 3625-3639.

Jordan, C.F. (1969). Derivation of leaf area index from quality of light on forest floor. *Ecology*, vol. 50, pp. 663.

Justice, C.O., Townshend, J.R.G., Vermote, E.F., Masuika, E., Wolfe, R.E., Saleous, N., Roy, D.P., Morisette, J.T. (2002). An overview of MODIS land data processing and product status. *Remote Sensing of Environment*, vol. 83, no. 2, pp. 3-15.

Jia, X., and Richards, J.A. (2004). *Remote Sensing Digital Image Analysis*, 4th edition. Springer-Verlag, Ed. New York, USA.

Karam, M.A., Fung, A.K., Lang, R.H., and Chauhan, N.S. (1992). A microwave scattering model for layered vegetation. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 30, no. 4, pp. 767-784

Kuusk, A. (1991). The hot-spot effect in plant canopy reflectance. In R.B. Myneni, & J.Ross (Eds.), *Poton-vegetation interactions*, pp. 139-159. New York: Springer-Verlag.

Lakhankar, T., Ghedira, H., Temimi, M., Azar, A.E., Khanbilvardi, R. (2009). Effect of land cover heterogeneity on soil moisture retrieval using active microwave remote sensing data. *Remote Sensing*, vol. 1, pp. 80-91.

Lavergne, T., Kaminski, T., Pinty, B., Taberner, M., Gobro, N., Verstrate, M.M., Vossbeck, M., Widlowski, J.-L., & Giering, R. (2007). Application to MISR land products of an RPV model inversion package using adjoint and Hessian codes. *Remote Sensing of Environment*, vol. 107(1-2), pp. 362-375.

Le Maire, G., Marsden, C., Verhoef, W., Ponzoni, F.J., Lo Seen, D., Begue, A., Stape, J.-L., Nouvellon, Y. (2011). Leaf area index estimation with MODIS reflectance time series and model inversion during full rotations of Eucalyptus plantations. *Remote Sensing of Environment*, vol. 115, pp. 586-599.

Lessmann, S., Stahlbock, R., Crone, S.F. (2006). Genetic Algorithms for Support Vector Machine Model Selection. in *Proceedings of International Joint Conference on Neural Networks*, Vancouver, Canada, pp. 3063-3069.

Lin, D.S., Wood, E.F., Bevan, K., and Saatchi, S. (1994). Soil moisture estimation over grass-covered areas using AIRSAR. *International Journal of Remote Sensing*, vol. 15, no. 11, pp. 2323-2343.

Luckman, A.J., (1998). The effect of topography on mechanisms of radar backscatter from coniferous forest and upland pasture. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 36, no. 5.

Macelloni, G., Paloscia, S., Pampaloni, P., Sigismondi, S., de Matthæis, P., Ferrazzoli, P., Schiavon, G., Solimini, D. (1999). The SIR-C/X-SAR experiment on Montespertoli: sensitivity to hydrological parameters. *International Journal of Remote Sensing*, vol. 20, n. 13, pp. 2597-2612.

Marchesi, S., and Bruzzone, L. (2009). ICA and kernel ICA for change detection in multispectral remote sensing images. in *Proceedings of Geoscience and Remote Sensing Symposium IGARSS 2009*, 12-17 July, Cape Town, South Africa, pp. II-980 – II-983.

Mattera D., and Haykin, S. (1999). Support vector machines for dynamic reconstruction of a chaotic system. In *Advances in kernel methods: support vector machine,* Chambridge, MA: MIT Press.

Mattia, F., Le Toan, T., Souyris, J.C., De Carolis, G., Floury, N., Posa, F., Pasquariello, G. (1997). The effect of surface roughness on multifrequency polarimetric SAR data. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 954-966.

Mattia, F., and Le Toan, T. (1999). Backscattering properties of multi-scale rough surfaces. *J. Electromagn. Waves Appl.*, vol. 13, no.4, pp. 493-527.

Mattia, F., Le Toan, T., Picard, G., Posa, F., D'Alessio, A.C., Notarnicola, C., Gatti, A.M., Rinaldi, M., Satalino, G., and Pasquariello, G. (2003). Multitemporal C-band measurements on wheat field. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 41, no. 7, pp. 1551-1560.

Mattia, F., Satalino, G., Dente, L., and Pasquariello, G. (2006). Using a priori information to improve soil moisture retrieval from ENVISAT ASAR AP data in semiarid regions. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 44, no. 4, pp. 900-912.

Maurer, K., Weyand, A., Fischer, M., Stocklin, J. (2006). Old cultural traditions, in addition to land-use and topography, are shaping pland diversity of grasslands in the Alps. *Biol. Conserv.*, vol. 130, pp. 438-446.

Mätzler, C. (1987). Microwave sensors for measuring avalanche-critical snow parameters. in *Proceedings of the Davos Symposium*, September 1986, no. 162.

Meroni, M., Colombo, R., & Panigada, C. (2004). Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. *Remote Sensing of Environment*, vol. 92(2), pp. 195-206.

Morisette, J.T., Baret, F., Privette, J.L., Myneny, R.B., Nicheson, J., Garrigues, S., Shabanov, N.V., Weiss, M., Fernandes, R., Leblanc, S., Kalacska, M., Sanchez-Azofeifa, G.A., Chubey, M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A., Lewis, T., Iiames, J., Colombo, R., Meroni, M., Busetto, L., Cohen, W., Turner, D., Warner, E.D., Petersen, G.W., Seufert, G., & Cook, R. (2006). Validation of global moderate-resolution LAI products: a framework proposed within the CEOS land product validation subgroup. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 44, pp. 1804-1817.

Moser, G., Serpico, S.B. (2009). Automatic Parameter Optimization for Support Vector Regression for Land and Sea Surface Temperature Estimation from Remote Sensing Data. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 909-921.

Moulin, S., Bndeau, A., & Delecolle, R. (1998). Combining agricultural crop models and satellite observations: From field to regional scales. *International Journal of Remote Sensing*, vol. 19(6), pp. 1021-1036.

Myneni, R., Running, S.W., Glassy, J., & Votova, P. (2000). User's Guide: fPAR, LAI (ESDT: MOD15A2) 8-day composite. *NASA MODIS Land Algorithm*.

Notarnicola, C., and Posa, F., 2006. Combination of X, C and L band for retrieval of surface parameters. *in Proc. SPIE Europe Remote Sensing,* September 2007, Florence Italy.

Notarnicola, C., Angiulli, M., and Posa, F. (2006). Use of radar and optical remotely sensed data for soil moisture retreival over vegetated areas. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 44, no. 4.

Notarnicola, C., Angiulli, M., and Posa, F. (2008). Soil moisture retrieval from remotely sensed data: neural network approach versus Bayesian method. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 2, pp. 547-557.

Oh, Y. (2006). Robust inversion technique for retrieving soil moisture from multi-polarised backscatter of bare surface. *Electron. Letter*, vol. 42, no. 7, pp. 414-415.

Oliver, c., and Quegan, S. (2004). *Understanding Synthetic Aperture Radar images*. Raleigh, NC, USA: SciTech Publishing, Inc.

O'Reilly, J.E., Maritorena, S., Mitchell, B.G., Siegel, D.A., Carder, K.L., Garver, S.A., Kahru, M., McClain, C. (1998). Ocean Color Chlorophyll Algorithms for SEAWIFS. *Journal of Geophysical Research – Oceans*, vol. 103, pp. 24937–24953.

Paloscia, S., (2002). A summary of experimental results to assess the contribution of SAR for mapping vegetation biomass and soil moisture. *Canadian Journal of Remote Sensing*, vol. 28, n. 2, pp. 246-261.

Paloscia, S., Pampaloni, P., Pettinato, S., and Santi, E. (2008). A comparison of algorithms for retrieving soil moisture from ENVISAT/ASAR images. *IEEE Transaction on Geoscience and Remote* Sensing, vol. 46, no. 10, pp. 3274-3284.

Paloscia, S., Pampaloni, P., Pettinato, S., Santi, E., (2010). Generation of soil moisture maps from ENVISAT/ASAR images in mountainous areas: a case study. *International Journal of Remote Sensing*, vol. 31, no. 9, pp. 2265-2276.

Park, S.E., and Moon, W.M. (2002). Polarimetric target decomposition and physical interpretation of NASA (JPL) AIRSAR data in mountainous terrain. in *Proceedings of Geoscience and Remote Sensing Symposium IGARSS*, Toronto, Canada, vol. 5, pp. 2605-2607.

Pasolli, L., Melgani, F., and Blanzieri, E. (2010). Gaussian Process Regression for estimating chlorophyll concentration in subsurface waters from remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 3.

Pathe, C., Wagner, W., Sabel, D., Doubkova, M., and Basara, J.B. (2009). Using ENVISAT ASAR global mode data for surface soil moisture retrieval over Oklahoma, USA. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 47, no. 2, pp.468-480.

Pauli, H., Gottfried, M., and Grabherr, G., (2003). Effects of climate change on the alpine and nival vegetation of the Alps. *Journal of Mountain Echology*, vol. 7, pp. 9-12.

Penuelas, J., Gamon, J.A., Fredeen, A.L., Merino, J., & Field, C.B. (1994). Reflectance indices associated with physiological changes in nitrogen and water limited sunflower leaves. *Remote Sensing of Environment*, vol. 48, pp. 135-146.

Perry, C.A. (2000). Significant floods in the United States during the 20th century - USGS Measures a Century of Floods. *U.S. Geological Survey*.

Pierdicca, N., Castracane, P., Pulvirenti, L. (2008). Inversion of electromagnetic models for bare soil parameter estimation from multifrequency polarimetric SAR data. *SENSORS*, vol. 8 (12), pp. 8181-8200.

Privette, J.L., Myneni, R.B., Knyazikhin, Y., Mukelabai, M., Roberts, G., Tian Y., Wang, Y.,& Leblanc, S.G. (2002). Early spatial and temporal validation of MODIS LAI product in the Southern Africa Kalahary. *Remote Sensing of Environment*, vol. 83, pp. 232-243.

Quaife, T., Lewis, P., De Kauwe, M., Williams, M., Law, B.E., Disney, M. & Bowyer, P. (2008). Assimilating canopy reflectance data into an ecosystem model with an ensemble Kalman filter. *Remote Sensing of Environment*, vol. 112, pp. 1347-1364.

Rammig, A., Jonas, T., Zimmermann, N.E. and Rixen, C. (2010). Changes in alpine plant growth under future climate conditions. *Biogeosciences*, vol. 7(6), pp. 2013-2024.

Reale, B., and Jackson, T. (1991). *Neural Computing: an Introduction*. Bristol, U.K.: Adam Hilger.

Richards, J.A., and Jia, X. (2006). *Remote Sensing Digital Image Analysis: an Introduction*. Springer Verlag.

Ripley,B. (1996). *Pattern Recognition and Neural Networks.* Cambridge University Press, Cambridge.

Rochadi, N., Fernandes, R. (2010). Systematic mapping of leaf area index across Canada using 250-meter MODIS data. *Remote Sensing of Environment*, vol. 114, pp. 1130-1135.

Rodriguez-Iturbe, I., Vogel, G., Rigon, R., Entekhbi, D., and Castelli, F. (1995). On the spatial organization of soil moisture fields. *J. Geophys. Res.*, 22(20):2757–2760.

Rodriguez-Iturbe, I., D'Odorico, P., Porporato, A., et al. (1999). On the spatial and temporal links between vegetation, climate, and soil moisture. *Water Resources Research*, vol. 35-12, pp. 3709-3722.

Rouse, J.W., Haas, R.H., Schell, J.A., & Deering, D.W. (1974).Monitoring vegetation systems in the Great Plains with ERTS. *Third ERTS-1 Symposium*, Washington, DC: NASA, pp. 309-317.

Ruf, C.S., and Zhang, H. (2003). Performance evaluation of a single and multi-channel microwave radiometers for soil moisture retrieval. *Remote Sensing of Environment*, vol. 75, no. 7, pp. 1551-1560.

Running, S.W., & Choughlan, J.C. (1998). A general model of forest echosystems processes for regional applications I. Hydrologic balance, canopy gas exchange and primary production processes. *Ecological Modeling*, vol. 42, pp. 125-154.

Ryu, Y., Kang, S., Moon, S.K., Kim, J. (2008). Evaluation of land surface radiation balance derived from moderate resolution imaging spectroradiometer (MODIS) over complex terrain and heterogeneous landscape on clear sky days. *Agricultural and Forest Meteorology*, vol. 148, pp. 1538-1552.

Sandholt, I., K. Rasmussen, J. Andersen, (2002). A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sensing of Environment*, vol. 79, pp. 213-224.

Schanda, E. (1986). *Physical Foundaments of Remote Sensing*. Springer-Verlag: New York.

Schlerf, M., & Atzberger, C. (2006). Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. *Remote Sensing of Environment*, vol. 100, pp. 281-294.

Scholkopf, B., and Smola, A.J. (2001). *Learning with Kernels*, Cambridge: MIT Press.

Schulz, K., Seppelt, R., Zehe, E., Vogel, H.J., Attinger, S. (2006). Importance of spatial structure in advancing hydrological sciences. *Water Resources. Research*, vol. 42 (3).

Sellers, P.J., Los, S.O., Tucker, C.J., Justice, C.O., Dazlich, D.A., Collatz, G.J., & Randall, D.A. (1996). A revised land surface parameterization (SiB2) for atmospheric GCMs: Part II. The generation of global fields of terrestrial biophysical parameters from satellite data. *Journal of Climate*, vol. 9, pp. 706-737.

Sellers, P.J., Randall, D.A., Betts, A.K., Hall, F.G., Berry, J.A., Collatz, G.J., Denning, A.S., Mooney, H.A., Nobre, C.A., Sato, N., Field, C.B., & Henderson-sellers, A. (1997). Modeling the exchanges of energy , water and carbon between continents and the atmosphere. *Science*, vol. 275, pp. 502-509.

Smets, K., Verdonk, B., Jordaan, E.M. (2007). Evaluation of performance measures for SVR Hyperparameter selection. in *Proceedings of International Joint Conference on Neural Networks*, Orlando, Florida, USA, pp.637-642.

Shabanov N.V., Huang, D., Yang, W., Tan, B., Knyzikhin, Y., Myneni, R.B., Ahl, D.E., Gower, S.T., & Heute, A.R. (2005). Analysis and optimization of the MODIS leaf area index algorithm retrievals over broadleaf forests. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 43, pp. 1855-1865.

Showengerdt, R. (2007). *Remote Sensing, Models, and Methods for Image Processing*, 3rd Edition. Academic Press.

Slater, P.N. (1980). *Remote Sensing: Optics and Optical Systems*. Addison-Wesley Publishing.

Smola, A., Murata, N., Scholkipf, B., Muller, K. (1998). Asymptotically optimal choice of ε-loss for support vector machines. in *Proceedings of International Conference on Artificial Neural Networks*, Skövde, Sweden.

Song, K., Zhou, X., and Fan, Y. (2009). Empirically adopted IEM for retrieval of soil moisture from radar baskscattering coefficients. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 47, no. 6, pp. 1662-1672.

Srihari, S.N., Covindaraju (1993). *Pattern recognition.* Chapman &Hall, London.

Srivastava, H.S., Patel, P., Sharma, Y., and Navalgund, R.R. (2009). Large-area soil moisture estimation using multi-incidence-angle RADARSAT-1 SAR data. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 47, no. 8.

Stamm, A.J. and VonderHaar, T.H. (1970). *Atmospheric effects on remote sensing*. University of Wisconsin, Madison.

Stimson, H.C., Breshears, D.D., Ustin, S.L., & Kefauver, S.C. (2005). Spectral sensing of foliar water conditions in two co-occurring conifer species: Pinus edulis and Juniperus monosperma. *Remote Sensing of Environment*, vol. 96, pp. 108-118.

Tabatabaeenejad, A., and Moghaddam, M. (2009). Inversion of subsurface properties of layered dielectric structures with randomly slightly rough interfaces using the method of simulated annealing. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2035-2046.

Tasser, E., Ruffini, F.V., Tappeiner, U. (2009). An integrative approach for analyzing landscape dynamics in high diverse cultural and natural mountain areas. *Landscape Ecol.*, vol. 24, pp. 611-628.

Teodoro, A.C., Veloso-Gomes, F., Goncalves, H. (2007). Retrieving TSM concentration from multispectral satellite data by multiple regression and artificial neural networks. *IEEE Transcation on Geoscience and Remote Sensing*, vol. 45, no. 5, pp. 1342-1350.

Touzi, R. (2007). Target scattering decomposition in terms of roll-invariant target parameters. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 45, no. 1.

Tsang, L., Chen, Z., Oh, S., Marks II, R.J, Chang, A.T.C. (1992). Inversion of snow parameters from passive microwave remote sensing measurements by a neural network trained with a multiple scattering model. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 30, no. 5, pp. 1015-1024.

Turner,D.P., Ollinger, S.V., & Kimball, J.S. (2004). Integrating remote sensing and ecosystem processes models for landscape- to regional-scale analysis of the carbon cycle. *BioScience*, vol. 54, pp. 573-584.

Twomey, S. (1977). *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements.* Oxford, UK: Elsevier Scientific.

Ulaby, F.T. (1974). Radar measurements of soil moisture content. *IEEE Trans. Antennas Propag.*, vol. AP-22, no. 2, pp. 257-265.

Ulaby, F.T., Vatlivala, P.P., and Dobson, M.C. (1978). Microwave backscatter dependence on surface roughness, soil moisture and soil texture, Part-I: Bare soil. *IEEE Trans. Geosci. Electron.*, vol. GE-16, no. 4, pp. 286-295.

Ulaby, F.T., Bradley, G.A., and Dobson, M.C. (1979). Microwave backscatter dependence on surface roughness, soil moisture and soil texture, Part-II: Vegetation covered soil. *IEEE Trans. Geosci. Electron.*, vol. GE-17, no. 2, pp. 33-40.

Ulaby, F.T., Moore, R.K., Fung, A.K. (1986a). *Microwave Remote Sensing: Active and Passive. Volume-1: Microwave Remote Sensing Fundamentals and Radiometry*. Longman Higher Education.

Ulaby, F.T., Moore, R.K., Fung, A.K. (1986b). *Microwave Remote Sensing: Active and Passive. Volume-2: Radar Remote Sensing and Surface Scattering and Emission Theory*. Longman Higher Education.

Ulaby, F.T., Moore, R.K., Fung, A.K. (1986c). *Microwave Remote Sensing: Active and Passive. Volume-3: From Theory to Applications*. Longman Higher Education.

Van Hateren, J.H., and Van Der Schaar, A. (2000). Independent component analysis: Algorithms and applications. *Neural Netw. v*ol. 13, no. 4/5, pp. 411-430.

Vapnik, V. (1995). *The nature of statistical learning theory*. New York: Springer.

Vapnik, V.N. (1998). *Statistical Learning Theory*, ser. Adaptive and Learning Systems for Signal Processing, Communications and Control, S. Haykin, Ed. John Wiley & Sons, Inc.

Verhoef, W. (1984). Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sensing of Environment*, vol. 16(2), pp. 125-141.

Verhoef, W. (1985). Earth observation modeling based on layer scattering matrices. *Remote Sensing of Environment*, vol. 17(2), pp. 165-178.

Vescovo, L., and Gianelle, D. (2008). Using the MIR bands in vegetation indices for the estimation of grassland biophysical parameters from satellite remote sensing in the Alps region of Trentino (Italy). *Advances in Space Research*, vol. 41, pp. 1764-1772.

Vescovo L., Wohlfahrt G., Balzarolo M., Pilloni S., Rodeghiero M., Sottocornola M., Gianelle D. (2011) The use of new spectral vegetation indices based on the near infrared shoulder wavelengths for remote detection of grassland structural variables. *International Journal of Remote Sensing*, DOI:10.1080/01431161.2011.607195.

Vohland, M., Jarmer, T. (2008). Estimating structural and biochemical parameters from spectroradiometer data by radiative transfer modeling (PROSPECT + SAIL). *International Journal of Remote Sensing*, vol. 29 (1), pp. 191-209.

Walker, J.P., Willgoose, G.R., Kalma, J.D. (2004). In situ measurements of soil moisture: a comparison of techniques. *Journal of Hydrology*, vol. 293, no. 1-4, pp. 85-99.

Wang, J.R. (1980). The dielectric properties of soil-water mixtures at microwave frequencies. *Radio Sci*., vol. 15, no. 5, pp. 977-985.

Wang, J.R., and Choudhury, B.J. (1995). Passive microwave radiation form soil: Examples of emission model and observations. in *Passive Microwave Remote Sensing of Land-Atmosphere Interactions*. Zeist, The Netherlands: VSP.

Wegmueller, U., Maetzler, C., Hueppi, R., and Schanda, E. (1994). Active and passive microwave signature catalog on bare soil (2-12 GHz). *IEEE Transaction on Geoscience and Remote Sensing*, vol. 32, no. 3, pp. 698-702.

Weiss, M., Baret, F., Myneny, R.B., Pragnere, A., & Knyazikhin, Y. (2000). Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie*, vol. 20, pp. 3-22.

Weiss, M., Baret, F., Smith, G.J., Jonckheere, I., & Coppin, P. (2004). Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling. *Agricultural and Forest Meteorology*, vol. 121 (1-2), pp. 37-53.

Weiss, M., Baret, F., Garrigues, S., & Lacaze, R. (2007). LAI and faPAR CYCLOPES global products derived from VEGETATION. Part 2: validation and intercomparison with MODIS Collection 4 products. *Remote Sensing of Environment*, vol. 110(3), pp. 317-331.

White, D.A. (2008). *The MODIS conversion toolkit (MCTK) user's guide*. Online. Available: http://nsidc.org/data/modis/tools.html.

Wohlfahrt, G., Sapinsky, S., Tappeiner, U., & Cernusca, A. (2001). Estimation of plant area index of grassland from measurements of canopy radiation profiles. *Agr. Forest Meteorol.*, vol. 109, pp. 1-12.

Wohlfahrt, G., & Cernusca, A. (2002). Momentum transfer by a mountain meadow canopy: a simulation analysis based on Massman's (1997) model. *Bound-Lay. Meteorol.*, vol. 103, pp. 391-407.

Wohlfahrt, G., A. Hammerle, A. Haslwanter, M. Bahn, U. Tappeiner, and A. Cernusca (2008), Seasonal and inter-annual variability of the net ecosystem CO2 exchange of a temperate mountain grassland: Effects of weather and management, *J. Geophys. Res.*, vol. 113, D08110, doi:10.1029/2007JD009286.

Wolfe, R.E., Roy, D.P.,& Vermote, E. (1998). MODIS land data storage, gridding, and compositing metholodogy: level 2 grid. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 36, pp. 1324-1338.

WWW1. COSMO-SkyMed Website. [Online]. http://www.asi.it/en/activity/earth_observation/cosmoskymed

WWW2. Sentinel Website. [Online]. http://www.esa.int/esaLP/SEM097EH1TF_LPgmes_0.html

WWW3. Global Climate Observing System Website. [Online]. http://www.wmo.int/pages/gcos/index.php

WWW4. SMEX Experiment. [Online].
http://sidc.org/data/amsr_validation/soil_moisture/smex02/

WWW5. SMEX Experiment. [Online]
http://nsidc.org/data/amsr_validation/soil_moisture/smex02/

Yang, W., Huang, D., Tan, B., Stroeve, J.C., Shabanov, N.V., Knyzikhin, Y., Nemani, R.R., & Myneni, R.B. (2006). Analysis of leaf area index and fraction vegetation absorbed PAR products from the TERRA MODIS sensor: 2000-2005. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 44, pp. 1829-1842.

Zahn, H., Shi, P., Chen, C. (2003). Retrieval of oceanic chlorophyll concentration using support vector machines. *IEEE Transaction on Geoscience and Remote Sensing*, vol. 41, no. 12, pp. 2947-2951.

Zong, Q., Wenjing, L., and Dou, L. (2006). Parameters selection for SVR based on PSO. in *Proceedings of the 6th World Congress on Intelligent Control and Automation*, Delian, China, pp. 2811-2814.