Research Article

Estimation of Soil Moisture in an Alpine Catchment with RADARSAT2 Images

L. Pasolli,[1,2] C. Notarnicola,[2] L. Bruzzone,[1] G. Bertoldi,[3] S. Della Chiesa,[3,4] V. Hell,[3] G. Niedrist,[3,4] U. Tappeiner,[3,4] M. Zebisch,[2] F. Del Frate,[5] and G. Vaglio Laurin[5]

[1] Department of Information Engineering and Computer Science, University of Trento, Via Sommarive, 14, 38123 Trento, Italy
[2] EURAC-Institute for Applied Remote Sensing, Viale Druso, 1, 39100 Bolzano, Italy
[3] EURAC-Institute for Alpine Environment, Viale Druso, 1, 39100 Bolzano, Italy
[4] Institute of Ecology, University of Innsbruck, Sternwartestr. 15, 6020 Innsbruck, Austria
[5] Department of Computer Science, Systems and Production Engineering, Tor Vergata University, Via del Politecnico, 1, 00133 Rome, Italy

Correspondence should be addressed to L. Pasolli, luca.pasolli@eurac.edu

Soil moisture retrieval is one of the most challenging problems in the context of biophysical parameter estimation from remotely sensed data. Typically, microwave signals are used thanks to their sensitivity to variations in the water content of soil. However, especially in the Alps, the presence of vegetation and the heterogeneity of topography may significantly affect the microwave signal, thus increasing the complexity of the retrieval. In this paper, the effectiveness of RADARSAT2 SAR images for the estimation of soil moisture in an alpine catchment is investigated. We first carry out a sensitivity analysis of the SAR signal to the moisture content of soil and other target properties (e.g., topography and vegetation). Then we propose a technique for estimating soil moisture based on the Support Vector Regression algorithm and the integration of ancillary data. Preliminary results are discussed both in terms of accuracy over point measurements and effectiveness in handling spatially distributed data.

1. Introduction

Soil moisture content is a key parameter in many hydrological processes. It controls the infiltration rate during precipitation events, runoff production, and evapotranspiration [1]. 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. When we move the attention to the mountainous environment, such as the Alps, the scale of the spatial and temporal variability reduces, due to the heterogeneity and the variability of the environment [2, 3]. This aspect makes the knowledge of accurate and reliable information on soil moisture status much more complex and at the same time important and critical for all the applications cited above [4].

In the last few years, the increasing number of spaceborne sensors, with complete and frequent coverage of the Earth's surface, has determined an increasing interest for the estimation of bio-geophysical surface parameters 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 microwave signals to the soil moisture content depends on the influence of water on the dielectric constant and has been well established in several studies [5–7]. The challenge in the moisture content retrieval from microwave signals is represented by the complexity and nonlinearity of the estimation process. Moreover, several studies pointed out the sensitivity of the microwave signal to other target properties, such as the roughness of the soil and

the presence of vegetation, which introduce additional ambiguities and nonlinearity in the retrieval process [8, 9]. In order to reduce these effects, several studies have been carried out on the use of microwave data acquired with multiple incidence angles, frequencies, and polarization configurations. In particular, 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 [10]. However, most operative satellite systems (e.g., ERS-2, RADARSAT, and ENVISAT) have onboard a C-band SAR sensor only, thus limiting the possibility of applying multifrequency approaches in operative conditions. Another possible solution is the integration in the retrieval process of data acquired by optical sensors, which may provide useful information for reducing the ambiguity due to the presence of vegetation [11]. Concerning the polarization features, the use of both co- and cross-polarized backscattering coefficients has shown to be effective for the reduction of the ambiguity in the signal due to roughness [12]. However, even though the polarimetric approach has demonstrated to be very promising, it has not been fully exploited yet due to only recent availability of fully polarimetric satellite orbiting sensors. Some recent papers deal with the use of polarimetric RADARSAT2 images. Hendrickx et al., 2009, validate the RADARSAT2 retrieved soil moisture values against ground measurements and optical indices in semiarid areas providing promising results.

Topography is another important aspect (in addition to the effects of vegetation and surface roughness) 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 presence of topography on the ground. 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 nonnegligible residual contribution within the signal due to the extreme topographic conditions [13]. 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. Heitz et al., 2010, correlated RADARSAT2 backscattering coefficients to ground measurements indicating that retrieved soil moisture values are able to recognize the topographic soil wetness gradient.

From the methodological viewpoint, the retrieval of soil moisture content can be considered as a mapping problem from the space of the measured signal (i.e., the backscattering signal) to the space of the desired biophysical parameter (i.e., the soil moisture content). This task is commonly addressed by means of the inference of the desired mapping from theoretical forward models, such as the Integral Equation Model (IEM), with the use of iterative methods or nonlinear machine learning techniques [12, 14]. Theoretical models can describe a great variety of experimental conditions in terms of acquisition parameters and target properties. They ensure a high degree of generality to the estimation process and the possibility to handle operative conditions in which no (or very few) field ground truth is available. However, the formulation of theoretical models is typically extremely complex and involves a certain number of input parameters, thus making the inversion process nonlinear, analytically nontractable, and ill posed. Another critical point is the fact that theoretical models may relay on simplifications and approximations of the physical phenomena which may not be completely verified in the field especially in presence of complex environmental conditions [15]. This could be the case of the Alpine environment, due to the presence and heterogeneity of the vegetation coverage together with the effect of topography. These issues could significantly affect the accuracy and reliability of the estimation.

All these aspects make the problem of the characterization of soil moisture in alpine 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 above, it is important to have a clear comprehension of the possibilities, but also 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. Although some works in this direction have started, further analysis is required. The SOFIA project (SOil and Forest Information retrieval with RADARSAT2 images) inserts in this context and aims at investigating the capability of new generation polarimetric RADARSAT2 satellite SAR sensors in combination with advanced state-of-the-art methods for the estimation of soil and forest biophysical parameters in the Alpine environment. This paper introduces the rationale behind the experimental analysis carried out in the context of the SOFIA project for the specific topic of soil moisture estimation. The main objectives of the proposed work are

(i) to present the test area and the setup for the ground measurements,

(ii) to analyze the sensitivity of the RADARSAT2 polarimetric data on the soil moisture content in an Alpine catchment and the necessity to integrate SAR images with ancillary data,

(iii) to present the first results of soil moisture estimation derived from the inversion procedure based on the Support Vector Regression technique.

—— Watershed
⊙ Altitudinal transect stations
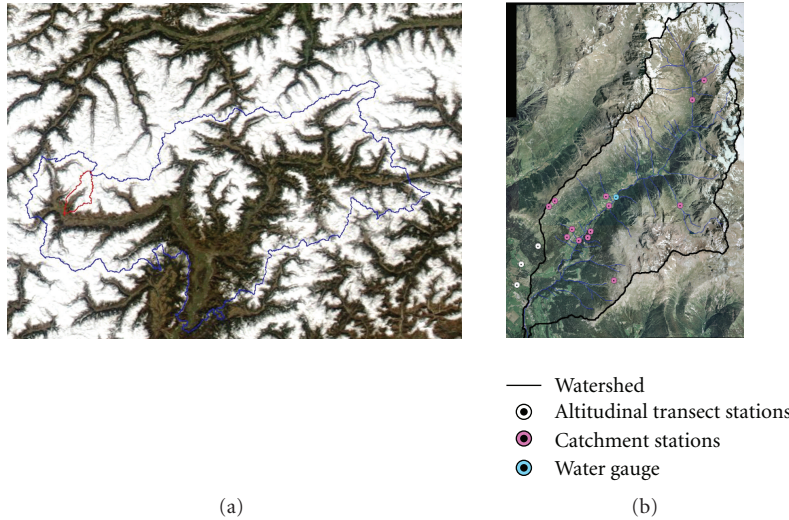● Catchment stations
⊙ Water gauge

(a)

(b)

FIGURE 1: Study area of the SOFIA project: (a) Alto Adige Province and (b) Mazia Valley, with the localization of the fixed measurement stations. The stations called "Transect" are the most complete ones, including 4 soil water content sensors in each station at two depths (5 and 20 cm). The stations called "Catchment" include one soil water content sensor at two depths (5 and 20 cm).

The rest of the paper is organized as follows. Section 2 introduces the study area on which our analysis is focused and describes the dataset adopted. The analysis of the sensitivity of the RADARSAT2 data to the soil moisture content is presented in Section 3, while Section 4 is devoted to the proposed estimation algorithm and to the experimental setup for its validation. Section 5 shows the first experimental results achieved. Finally, Section 6 draws the conclusion of the work.

2. Study Area and Dataset Description

2.1. Study Area. The study area of the SOFIA project is the Alto Adige Province, located in Northern Italy (see Figure 1(a)). Alto Adige covers an area of about 7400 km^2 with a lowest altitude of 220 m and a highest one of 3900 m. Historical climate observations have proved that the climate in the Alps has changed significantly. In the future, the 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 [16]. In addition, climate models agree on a higher interannual variability [17]. 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 [18] for agricultural and human purposes and may be strongly related to natural hazards such as floods and landslides [19].

Thus, Alto Adige represents an interesting test site for the following reasons:

  (i) high vulnerability to climate change in fields highly connected to the projects objectives (drought, lack of water, natural hazards, yield),

 (ii) representativeness at least for the central and southern Alps,

(iii) high diversity of land use with almost all types of land use of central European mountain areas,

(iv) good data supply, good contact to partners and access to the results of several scientific projects.

Within the Alto Adige area, the Mazia valley (Figure 1(b)), and represented by the red contour in Figure 1(a), a small side valley into the Venosta valley, has been chosen for the first investigations on soil moisture content estimation. Mazia valley covers an area of ca. 100 km^2 with altitudes that vary from 920 meters a.s.l. (Sluderno) to 3738 meters a.s.l. (Palla Bianca). The area is almost dry, with mean annual precipitation of 525 mm (Mazia, 1580 meters a.s.l.). However, wet patterns with higher soil moisture can be observed mainly due to irrigation practice in highly intensively managed meadows (in the valley floor) and the presence of wet buffers along small rivers going down from the top of the mountains. The land use types present in the area are well representatives for the whole South Tyrol, thanks also to the high variability in altitude. Meadows and pastures present heterogeneous characteristics in terms of vegetation species and human usage, becoming less intensively managed moving from the lower to the higher altitudes.

The valley is equipped with 16 fixed stations for the measurement and monitoring in time of soil parameters (moisture content at 5 and 20 cm depth) and meteorological data (air temperature and humidity, precipitation, wind speed and direction, solar radiation) [20]. The stations are distributed along the valley in locations representative of different elevation, slope, aspect, soil type, and land cover conditions (see Figure 1). Meadows and pastures are a significant presence in the valley. All these conditions make this area particularly suitable for sampling the high spatial variability typical of the mountain environment.

FIGURE 2: RADARSAT2 image acquired on July 21st, false color RGB composition (R = HH, G = HV, B = VV).

TABLE 1: Ranges of variability of the dielectric constant (real part) values measured during the field campaigns.

	Meadow		Pasture	
	June 2010	July 2010	June 2010	July 2010
Min dielectric constant value	6.7	3.8	6.4	3.2
Max dielectric constant value	23.2	27	17.7	8.7
Average dielectric constant value	16.7	15.4	11.6	5.7

2.2. Satellite Imagery. During the summer of 2010, two images were acquired by RADARSAT 2 over the Mazia valley on 3rd June and 21st July. The sensor acquisition mode was Standard Quad Polarization, with a mean incidence angle of 45° and an ascending orbit. The acquisition geometry has been selected such that the area of interest, characterized by a highly variable topography, was imaged minimizing the layover and shadowing effects on the east side of the valley, where a higher number of field measurement stations are present. 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. Thus the data have been multilooked, calibrated, and geocoded with the help of a high-geometrical resolution (2.5 meters) digital elevation model and filtered with a Frost filter (window size 5 × 5) in order to reduce the effect of speckle noise. The final resolution of the processed images is 20 m. All the preprocessing has been carried out with the SARscape software (http://www.sarmap.ch/). Figure 2 shows the results of the preprocessing in the case of the 21st July image. Polarimetric features have been composed in this RGB image in order to enhance the different information content of each channel. On the west side of the valley, the effects of geometric distortions (i.e., foreshortening, layover, and shadowing) are particularly evident. These effects are minimized in the east side, thanks to the specific acquisition geometry selected.

2.3. Field Measurement Campaign. Contemporary to the satellite acquisitions, two field measurement campaigns have been carried out in the Mazia valley. The aim was to acquire information on the soil parameters (moisture content and roughness) and on the vegetation status (biomass and vegetation water content) of meadow and pasture areas. These measurements have been exploited during the project for different purposes: (1) the calibration of the fixed measurement stations located in the valley, in order to have consistent information at these locations also in correspondence to future satellite overpasses and acquisitions, (2)

the analysis of the sensitivity of RADARSAT2 measurements to the properties of soils and vegetation in alpine areas, and (3) the development and validation of the algorithm for the estimation of the soil parameters from the satellite images.

Two different kinds of measurements have been performed: (1) destructive measurements of both vegetation and soil samples, by physically taking a sample of grass and soil. This kind of sampling was necessary to have accurate measurements of biomass, vegetation water content, soil gravimetric moisture, and bulk density. All the samples have been acquired, weighted, and then sealed in order to be dried in the laboratory according to standard measurement protocols [21]; (2) nondestructive measurements, which where possible thanks to the use of a mobile sensor (the Delta T WET 2 sensor, http://www.delta-t.co.uk/); these last measurements regarded only the soil dielectric constant, but had the advantage to be easier and faster with respect to the destructive measurements, so that it was possible to collect a higher number of samples. Sampling areas were selected in order to ensure a good representativeness in terms of local topographic and land use conditions. Moreover, repeated measurements (3 to 5) were collected in each sampling area and then averaged, in order to increase their spatial representativeness. More than 350 dielectric constant measurements were collected in more than 100 different sampling areas. Both destructive and nondestructive field measurements were concentrated on the west side of the valley, due to the better imaging properties of the selected acquisition geometry. Table 1 reports minimum, maximum, and average values of the dielectric constant measured on meadows and pastures during the two field campaigns. As can be observed, meadows present higher and much more variable dielectric constant values with respect to pastures, which are in general drier. This is probably due to the irrigation practice in some areas and to the differences in the soil type and vegetation coverage of meadows with respect to pastures. In fact, soil is quite heterogenous, ranging from Cambisols, Humic Leptosols, and Podsols to locally limited Planolsols and Histosols in hydromorphic areas. Also organic content, grain size distribution, and bulk density are highly variable even within areas of the same land cover type. On meadows and pastures, the dominant soil type is brown soil. Above the tree line, combinations of brown soils and ranker appear. In the forest in contrary also semipodzols are common, partly also the overlapping transition in
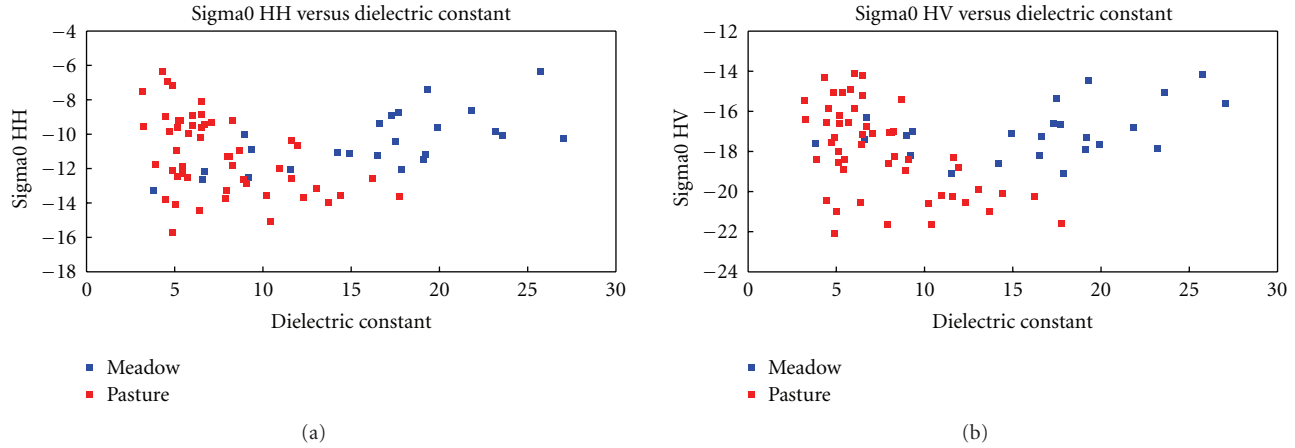
FIGURE 3: Scatter plots of backscattering coefficients extracted from the RADARSAT2 images versus dielectric constant measurements in the case of (a) HH polarization configuration and (b) HV polarization configuration.

semipodzolidation of brown soils. Podzols are predominant for coniferous forests. In the vicinity of streamlets also gley may appear. 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 mostly low (5–15%). Therefore, soil moisture measurements might be an additional information to validate soil maps as well as to understand the effect of soil texture and organic matter.

In this paper, we address the real part of dielectric constant because it represents the dielectric properties to which the SAR e.m. waves are particularly sensitive. The imaginary part of dielectric constant is in general very low and in most cases can be considered negligible [5].

2.4. Ancillary Data. To carry out the analysis presented in this work, ancillary data already available or extracted from satellite optical sensors have been considered. In greater detail,

(1) a digital elevation model (DEM) with high spatial resolution (2.5 m) obtained from the processing of airborne lidar acquisitions over the whole Alto Adige area during a measurement campaign in 2008,

(2) two normalized difference vegetation index (NDVI) maps extracted from two images acquired by the NASA MODIS sensor onboard the Terra satellite as close as possible to the RADARSAT2 satellite overpasses (i.e., within ±1 day from the RADARSAT2 acquisition). MODIS is a multispectral sensor with 36 spectral channels which acquires information in the visible and infrared portions of the spectrum with daily coverage of the whole Earth's surface. The high temporal resolution of this system allows extracting useful information of the area of interest maximizing the probability to have cloud-free acquisitions as close as possible to the date of interest. The spatial resolution of the sensor is 250 m in the red and near-infrared bands, the portions of the spectrum considered for the computation of the NDVI values,

(3) a high-resolution (25 m) land-cover map of the Mazia valley derived from ortho-photos, ground surveys, and visual interpretation.

Ancillary data have been geocoded and resampled (bilinear convolution) in order to be completely superimposed with the RADARSAT2 images.

3. Sensitivity Analysis

In order to understand the sensitivity of the RADARSAT2 signal to the moisture content of the investigated area, scatter plots of the backscattering coefficients at different polarization configurations versus the dielectric constant values were generated. To this purpose, in the two satellite images a small 3×3 pixels region was considered in correspondence of each field measurement point. Then the backscattering values were averaged and the resulting mean value was associated to the corresponding field measurement. Samples associated to foreshortening and layover areas were discarded from the analysis. Finally, considering both the acquisition dates and both meadow and pasture land cover types, 75 samples were used in the analysis. Figure 3 shows the plots in the case of HH and HV backscattering coefficients (analogous results have been achieved for the VV and VH configurations).

From a first analysis, it is possible to observe that the points associated to meadows present an expected increasing trend versus the 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 the 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
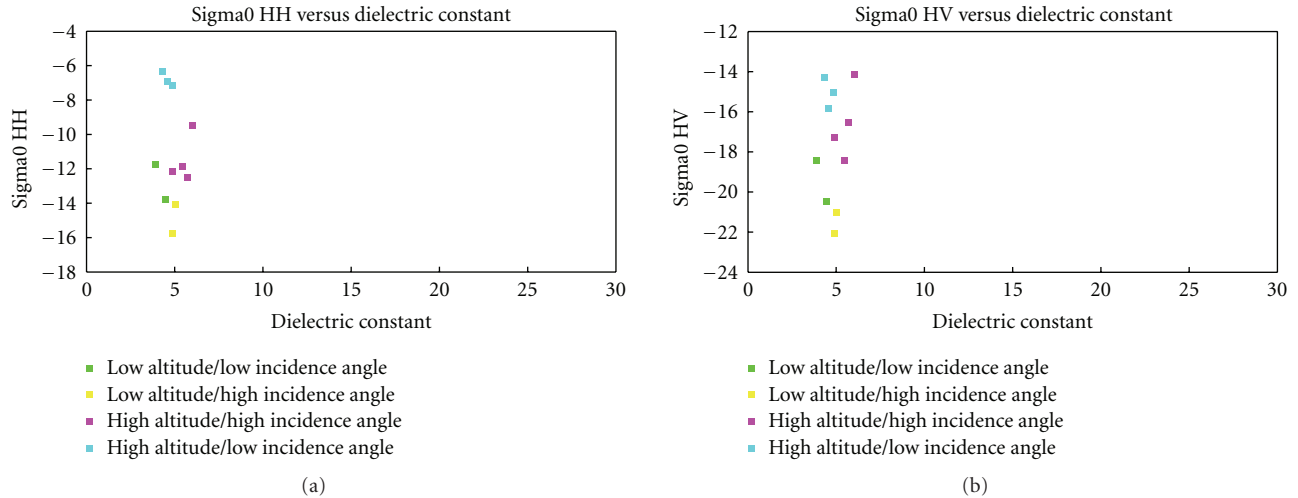
FIGURE 4: Scatter plots of backscattering coefficients extracted from the RADARSAT2 images versus dielectric constant measurements over pasture areas and with dielectric constant values between 4.5 and 5.5 in the case of (a) HH polarization configuration and (b) HV polarization configuration. The samples are grouped into 4 clusters according to the topographic features extracted from the DEM.

as mainly responsible for the variability and ambiguity observed in the pasture samples: (1) the topography and (2) the heterogeneity of the vegetation/land-cover. In the following, these two aspects are better investigated with the help of ancillary data, in order to understand if and to what extent they affect the RADARSAT2 measurements.

3.1. Effect of Topography. As explained previously, topography significantly affects the signal acquired by a satellite SAR system. In our case, although the calibration of the signal was carried out with the help of a detailed 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 1200 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 has been exploited for the extraction of two topographic features: 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, which takes into account the local topography of the area) and the local altitude. The samples associated to the pasture (which demonstrated the highest ambiguity in the SAR signal, as shown in Figure 3) were divided into different dielectric constant classes (e.g., below 4.5, between 4.5 and 5.5, between 5.5 and 6.5, and so on until 12.5; after this value the number of samples is reduced and the variability limited, as shown in Figure 3) in order to keep constant this variable in the analysis. Then, according to the topographic features, the 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, and

(4) high altitude/low incidence angle. Intermediate conditions were excluded from the analysis. Figure 4 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. Analogous results were obtained for the other dielectric constant ranges.

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. In greater detail, samples acquired in areas with low altitude and high local incidence angles of the SAR signal present the lowest values of the backscattering coefficient. On the contrary, samples associated to areas with high altitude and low local incidence angles are characterized by 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. The samples with intermediate topographic characteristics, that is, 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.

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,
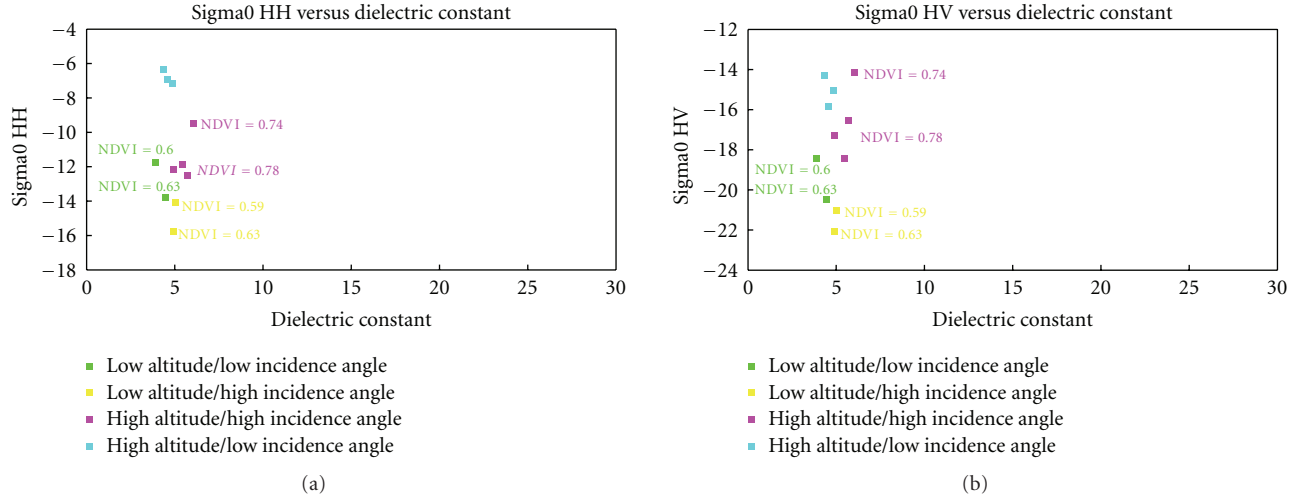
FIGURE 5: Scatter plots of backscattering coefficients extracted from the RADARSAT2 images versus dielectric constant measurements over pasture areas and with dielectric constant values between 4.5 and 5.5 in the case of (a) HH polarization configuration and (b) HV polarization configuration. NDVI values are shown for the samples which show strong residual variability in the backscattering coefficient value.

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 [22]. 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 vegetation/land-cover heterogeneity 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 paper, we will present the analysis just for the samples of Figure 4, but good agreement was found also for the other cases.

Plots shown in Figure 5 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. Further and more detailed analysis will be carried out on this point, with the help of higher geometrical resolution images.

The sensitivity analysis presented in this sections suggests that the backscattering coefficients measured by the RADARSAT2 SAR sensor are 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 the vegetation/land-cover. These factors should be properly taken into consideration for the retrieval of the moisture content of soils in presence of these challenging environmental conditions.

4. Soil Moisture Estimation Technique

Due to the effect of topography and vegetation/land-cover heterogeneity on the SAR signal, the retrieval of soil moisture content in alpine areas becomes particularly challenging and complex. Estimation approaches based on the inversion of

theoretical models may be not effective. Due to the high complexity and heterogeneity of the physical phenomena that affect the microwave signal, it is fair to expect that theoretical models (which introduce in their formulation several approximations and simplifications) will be not reliable and accurate in the estimation. In order to deal with this issue, a possible solution is the direct exploitation of the information contained in the data acquired during the field campaigns by means of nonlinear machine learning techniques. In particular, in this work we propose to address the estimation problem with the ε-insensitive Support Vector Regression [23], which presents properties suitable for the challenges and constraints of the estimation problem of interest.

Thanks to its formulation, SVR is able to handle complex nonlinear estimation problems with good intrinsic generalization capability also in presence of a limited number of training samples [24, 25]. Moreover, it easily handles high-dimensional input spaces, also with features extracted from different sources. These properties allow us to effectively exploit the samples collected during the field campaigns to infer the mapping between the SAR images and the target variable and at the same time to integrate in the retrieval process the information extracted from ancillary data. The latter is required to properly take into account the effects of topography and vegetation/land-cover heterogeneity on the input SAR data.

4.1. ε-Insensitive Support Vector Regression. Let us consider a generic estimation problem. We would like to retrieve a continuous variable y (e.g., the soil moisture content), given a set $\mathbf{x} = (x^1, x^2, \ldots, x^m)$ of m features extracted from the signals acquired using remote sensors. From an analytical viewpoint, the estimation problem can be expressed as

$$y = f(\mathbf{x}) + e, \tag{1}$$

where f denotes the desired and unknown input-output mapping and e is a Gaussian random variable with zero mean and unitary variance gathering all the noisy contributions affecting the considered estimation problem. The estimation of y corresponds to the problem of determining the function f' as close as possible to the true mapping f for the task considered.

Given a set of N reference samples $\{\mathbf{x}_i, y_i \mid i = 1, \ldots, N\}$, the goal of the ε-insensitive SVR technique is to find a smooth function f' that approximates f while keeping at most a deviation ε from the targets y_i [23]. To this purpose, the original m-dimensional input domain is mapped into a higher dimensionality feature space, where the function underlying the data is supposed to have an increased flatness. Thus it is approximated in a linear way:

$$f'(\mathbf{x}) = \mathbf{w} \cdot \Phi(\mathbf{x}) + b, \tag{2}$$

where $\mathbf{w}$ 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.
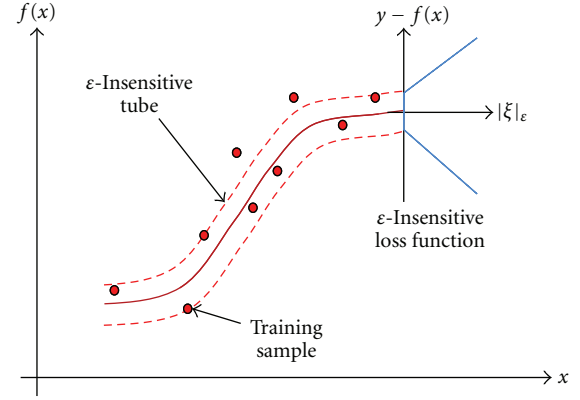

FIGURE 6: Example of a possible choice of the ε-insensitive loss function characterizing the SVR learning approach.

The optimal linear function in the transformed feature space is selected minimizing a cost function, which is the combination of the training error (empirical risk) and the model complexity (structural risk). The first term is calculated according to a ε-insensitive loss function, for example,

$$\lambda(\varepsilon) = \begin{cases} 0 \iff |y - f'(\mathbf{x})| \leq \varepsilon, \\ |y - f'(\mathbf{x}) - \varepsilon| \iff |y - f'(\mathbf{x})| > \varepsilon, \end{cases} \tag{3}$$

where ε is the tolerance to errors, that is, it allows one to define an insensitive tube surrounding the function f' (see Figure 6). Equation (3) means that losses smaller than this tolerance are neglected (thus increasing the robustness of the technique to the small errors and to the noise in the training set), whereas a penalty is assigned to estimates lying outside the tube. Equivalently, the penalty is expressed by means of nonnegative slack variables ξ, ξ^*, which measure the deviation of the training samples outside the ε-insensitive tube and are defined as follows:

$$\begin{aligned} \xi &= \lambda(\varepsilon) \iff \lambda(\varepsilon) > 0, \\ \xi^* &= \lambda(\varepsilon) \iff \lambda(\varepsilon) < 0. \end{aligned} \tag{4}$$

The second term is expressed through the Euclidean norm of the weight vector $\mathbf{w}$, 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. Thus, the cost function to minimize becomes

$$\Psi(\mathbf{w}, \xi) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*), \tag{5}$$

and it is subject to the following constraints:

$$\begin{aligned} y_i - [\mathbf{w} \cdot \Phi(\mathbf{x}_i) + b] &\leq \varepsilon + \xi_i, \\ [\mathbf{w} \cdot \Phi(\mathbf{x}_i) + b] - y_i &\leq \varepsilon + \xi_i^*, \quad i = 1, 2, \ldots, N, \\ \xi_i, \xi_i^* &\geq 0. \end{aligned} \tag{6}$$

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

The constrained optimization problem in (5) 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 (global minimum of the cost function). Leaving out mathematical details (for those we refer the reader to [23]), the final result of the estimation problem, in the original input domain, becomes

$$
\begin{aligned}
f'(\mathbf{x}) &= \sum_{i \in N} (\alpha_i - \alpha_i^*) \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}) + b \\
&= \sum_{i \in N} (\alpha_i - \alpha_i^*) k(\mathbf{x}_i, \mathbf{x}) + b,
\end{aligned}
\tag{7}
$$

where α_i and α_i^* represent the nonzero 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 to some type of inner product in the highly dimensional feature space (i.e., $k(\mathbf{x}_i, \mathbf{x}) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x})$). Thus, the kernel function allows one to evaluate the similarity between a couple of samples in the transformed feature space as a function of the samples in the input space, that is, without the explicit definition of the mapping function $\phi(\cdot)$. This strongly reduces the analytical complexity related to the latter issue. Commonly adopted kernels are polynomial functions and Gaussian radial basis functions [24]. Lagrange multipliers weight each training sample according to its importance in determining the solution function f'. Samples associated to 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 (according to the definition of the ε-insensitive loss function). 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.

4.2. Estimation Algorithm and Experimental Setup. The retrieval process is divided into two phases: (1) the training of the SVR algorithm and (2) the estimation phase.

During the training, the available training samples (i.e., the measurements acquired during the field campaign associated to the corresponding values of the microwave signal extracted from the RADARSAT2 images) are provided to the technique in order to learn the underlying relationship between the input features and the output target value. Typically, the samples are divided into two subsets: the first is used as training and the second is used as validation to assess the estimation performance of the technique (in terms of accuracy or other quality metrics) with different configurations of the free model parameters. In our analysis, in order to avoid problems related to the choice of the training and validation sets, we applied a k-fold cross validation procedure. Training samples are divided into k subsets. Iteratively, $k - 1$ subsets are used for the training of the regressor while the remaining subset is exploited for the validation.

At the end of the k iterations, the performance over the validation sets is averaged. In this way, all the samples are considered for both training and validation of the algorithm, thus ensuring a high robustness and good generalization of the training procedure. The selection of the best model among different possible configurations of the free model parameters (model selection issue) has been carried out by means of a multiobjective model selection strategy, which allows one to jointly optimize different and competing quality metrics. In this way the model selection process becomes more robust, since it relies on multiple criteria and not just one. Moreover, multiple optimal solutions are obtained according to the concept of Pareto optimality. Each one represents a different tradeoff among the considered quality metrics. The user has thus the possibility to choose the configurations which meets the requirement in terms of estimation quality related to the application considered. For further details we refer the Reader to [26].

After the regressor is trained, it is applied to the multi-dimensional image (which shall contain the same features considered during the training of the technique) in order to obtain the estimated moisture content map.

In our experiments, we considered a 5-fold for the cross validation procedure and the mean squared error (MSE) and the slope of the linear trend of estimated versus true target values as quality metrics to drive the multiobjective model selection. The optimal solution is selected on the basis of a visual inspection of the estimated Pareto front (i.e., the set of optimal solutions of the multiobjective model selection problem). Concerning the SVR technique, we selected an RBF Gaussian kernel and the following ranges for the model parameters: $[10^{-3}; 10^3]$ for γ, the kernel width, $[10^{-4}; 10^3]$ for C, and $[10^{-4}, 10]$ for ε.

As input features of the estimation system, we considered the four polarimetric configurations of the RADARSAT2 image: the altitude and the local incidence angle extracted from the DEM as topographic features and the NDVI and land-cover maps as features for the characterization of the vegetation/land-cover heterogeneity. Different experiments were carried out with different combinations of these features selected according to a sequential forward selection (SFS) strategy, in order to define the subset of them that provides the best results in terms of estimation accuracy.

From an operative viewpoint, for the implementation of the SVR algorithm, we considered the LibSVM software, freely available online [27]. The multiobjective model selection and the sequential forward feature selection strategies were implemented on our own using Matlab.

5. Experimental Results

5.1. Quantitative Assessment with Punctual Measurements. In order to evaluate the estimation performance of the SVR algorithm, different quality metrics were considered: the mean squared error (MSE) (or equivalently the Root MSE (RMSE)), which provides an information on the average error over the estimates; the slope and intercept of the linear regression line between estimated and true target values,

TABLE 2: Estimation accuracies achieved by the proposed algorithm with the best input feature configuration.

|	Global	Meadow	Pasture
RMSE	2.68	4.05	1.68
R^2	0.79	0.58	0.75
Slope	0.78	0.58	0.7
Intercept	2.26	7.15	2.3

FIGURE 7: Scatter plot of estimated versus measured dielectric constant values obtained with the proposed algorithm with the best input features configuration.

which indicate whether and to what extent the retrieval algorithm under- and overestimates the target variable with respect to the ideal case of a one-to-one line; the determination coefficient (R^2), which provides a measure about the spread of the estimates around the linear regression line (in the ideal case of a one-to-one line, this metric equals one). These metrics were evaluated over the available reference samples according to the 5-fold cross validation scheme described before. As previously explained, different input feature configurations were considered in the experiments according to the SFS strategy. Here, due to space constraints, we show and discuss the case with the input feature configuration that provided the best performances, that is, the configuration containing 2 polarimetric features (HH and HV), the 2 topographic features (Altitude and Local Incidence Angle), the NDVI, and the land-cover map. Table 2 presents the accuracies achieved by the proposed algorithm in this case, while Figure 7 shows the scatter plot of estimated versus measured dielectric constant values.

Globally, the achieved accuracies are promising, with an RMSE of 2.68 and a determination coefficients near to 0.8. Analyzing in more detail the results, it is possible to observe that the retrieval algorithm provides better performance over pastures with respect to meadows. In the latter case, the error is slightly higher and the algorithm tends to overestimate low values and underestimate high values of the dielectric constant. This effect is probably due to (1) the range of variability of the target variable, which is much larger in the case of meadows with respect to pastures and (2) the number of reference samples, which is lower in the case of meadows with respect to pastures (see Table 1). Both these factors may increase the complexity of the retrieval problem in the case of meadows. Further effort will be put on this issue, in order to better understand and, if possible, overcome the limitations of the estimation over meadows.

5.2. Soil Moisture Content Maps. After the training phase and the assessment over point measurements, the SVR algorithm was tested over the distributed dataset available, that is, the RADARSAT2 images acquired in June and July over the Mazia valley. The two images were provided in input to the trained SVR with in addition ancillary data according to the input features configuration considered for the training of the algorithm. The results of this processing step are two maps representing the estimated dielectric constant values over the area of interest and are shown in Figure 8. The masked values correspond mainly to forest, water bodies, rocks, and urban areas, according to the land use mask.
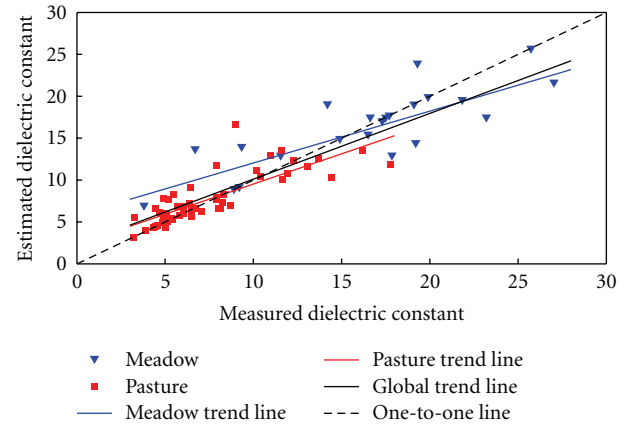
From a qualitative viewpoint, the maps reproduce well the expected trend of soil moisture content, presenting high values near to the valley floor (where the irrigated meadows are located) and progressively decreasing values moving to the pastures at higher altitudes. At the same time, the humidity patterns 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 (Figures 8(a) and 8(b)).

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 8. This trend is confirmed by the field measurements carried out in the areas during the two campaigns, as indicated in Section 2.3. In the upper part of the valley side, the maps indicate a slightly drier condition in the case of the June 2010 acquisition. This behavior will be better validated with the help of the soil and meteorological measurements provided by the stations located in the valley, as soon as the data will be available and properly calibrated.

6. Conclusion

In this paper, polarimetric RADARSAT2 SAR images are exploited for the estimation of soil moisture content in an alpine catchment. We first carried out a sensitivity analysis with the help of field measurements of the target parameter and ancillary data. This analysis pointed out that both topography and vegetation/land-cover heterogeneity strongly affect the backscattering signal acquired over alpine areas, introducing a significant variability and 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 technique for the estimation of soil moisture content from the RADARSAT2 images. We opted for an algorithm based on the ε-insensitive Support Vector Regression technique.
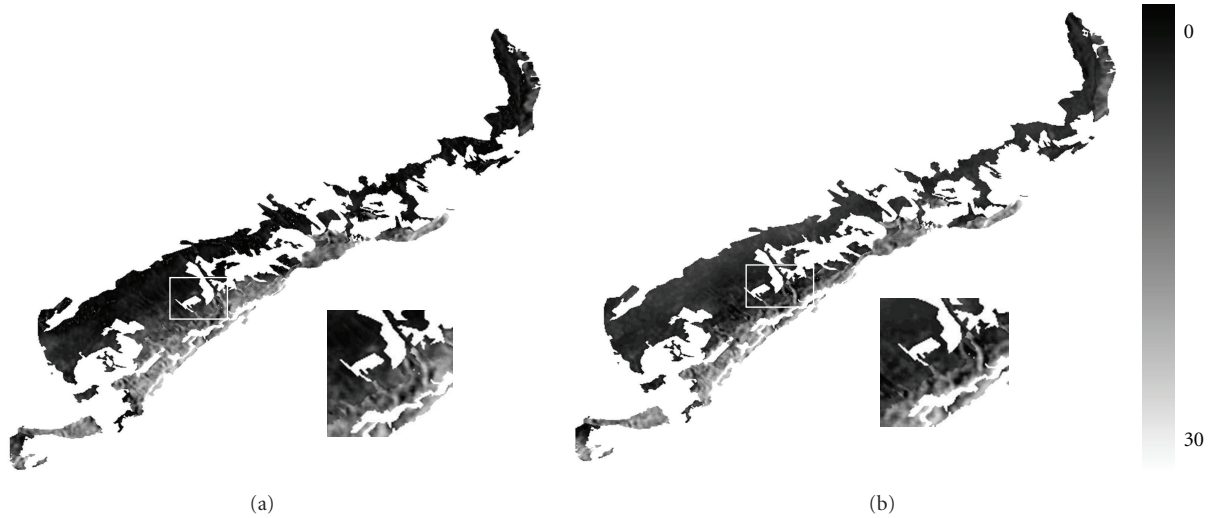
FIGURE 8: Maps of the dielectric constant of the east side of the Mazia valley: (a) 3rd June 2010 and (b) 21st July 2010. The small squares represent a zoom over particular areas extracted from the maps of June and July (indicated with the white square).

Thanks to its formulation, this method is able to handle complex nonlinear 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. The latter characteristic is important in order to integrate in the retrieval process the information extracted from ancillary data. Preliminary 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 alpine areas, (2) estimation accuracy over punctual measurements, and (3) capability to reproduce the soil humidity patterns when applied on 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 will 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 availability of high resolution spatially distributed surface soil moisture maps coming from the RADARSAT2 sensor can represent a major improvement for the validation of distributed hydrological models.

References

[1] I. Rodriguez-Iturbe, P. D'Odorico, A. Porporato, and L. Ridolfi, "On the spatial and temporal links between vegetation, climate, and soil moisture," *Water Resources Research*, vol. 35, no. 12, pp. 3709–3722, 1999.

[2] I. Rodriguez-Iturbe, G. K. Vogel, R. Rigon, D. Entekhabi, F. Castelli, and A. Rinaldo, "On the spatial organization of soil moisture fields," *Geophysical Research Letters*, vol. 22, no. 20, pp. 2757–2760, 1995.

[3] M. Gebremichael, R. Rigon, G. Bertoldi, and T. M. Over, "On the scaling characteristics of observed and simulated spatial soil moisture fields," *Nonlinear Processes in Geophysics*, vol. 16, no. 1, pp. 141–150, 2009.

[4] R. B. Grayson, A. W. Western, F. H. S. Chiew, and G. Blöschl, "Preferred states in spatial soil moisture patterns: local and nonlocal controls," *Water Resources Research*, vol. 33, no. 12, pp. 2897–2908, 1997.

[5] F.T. Ulaby, R. K. Moore, and A. K. Fung, *Microwave Remote Sensing: Active and Passive*, vol. 2, Artech House, Norwood, Mass, USA, 1986.

[6] J. R. Wang, "The dielectric properties of soil-water mixtures at microwave frequencies," *Radio Science*, vol. 15, no. 5, pp. 977–985, 1980.

[7] T. J. Jackson, "Soil moisture estimation using special satellite microwave/imager satellite data over a grassland region," *Water Resources Research*, vol. 33, no. 6, pp. 1475–1484, 1997.

[8] F. T. Ulaby, P. P. Vatlivala, and M. C. Dobson, "Microwave backscatter dependence on surface roughness, soil moisture and soil texture, Part-I: bare soil," *Search Results IEEE Transactions on Geoscience Electronics*, vol. 16, no. 4, pp. 286–295, 1978.

[9] F. T. Ulaby, G. A. Bradley, and M. C. Dobson, "Microwave backscatter dependence on surface roughness, soil moisture and soil texture, part-II: vegetation covered soil," *IEEE Transactions on Geoscience Electronics*, vol. 17, no. 2, pp. 33–40, 1979.

[10] C. Notarnicola and F. Posa, "Combination of X, C and L band SAR images for retrieval of surface parameters," in *SAR Image Analysis, Modeling, and Techniques IX*, vol. 6746 of *Proceedings of SPIE*, Florence, Italy, September 2007.

[11] C. Notarnicola, M. Angiulli, and F. Posa, "Use of radar and optical remotely sensed data for soil moisture retrieval over vegetated areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 4, pp. 925–934, 2006.

[12] S. Paloscia, P. Pampaloni, S. Pettinato, and E. Santi, "A comparison of algorithms for retrieving soil moisture from ENVIS AT/AS AR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 10, Article ID 4637966, pp. 3274–3284, 2008.

[13] A. J. Luckman, "The effects of topography on mechanisms of radar backscatter from coniferous forest and upland pasture," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 5, pp. 1830–1834, 1998.

[14] C. Notarnicola, M. Angiulli, and F. Posa, "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, 2008.

[15] F. Mattia and T. Le Toan, "Backscattering properties of multi-scale rough surfaces," *Journal of Electromagnetic Waves and Applications*, vol. 13, no. 4, pp. 493–527, 1999.

[16] M. Brunetti, M. Maugeri, and T. Nanni, "Changes in total precipitation, rainy days and extreme events in Northeastern Italy," *International Journal of Climatology*, vol. 21, no. 7, pp. 861–871, 2001.

[17] L. P. Graham, S. Hagemann, S. Jaun, and M. Beniston, "On interpreting hydrological change from regional climate models," *Climatic Change*, vol. 81, no. 1, pp. 97–122, 2007.

[18] T. P. Barnett, J. C. Adam, and D. P. Lettenmaier, "Potential impacts of a warming climate on water availability in snow-dominated regions," *Nature*, vol. 438, no. 7066, pp. 303–309, 2005.

[19] P. Horton, B. Schaefli, A. Mezghani, B. Hingray, and A. Musy, "Assessment of climate-change impacts on alpine discharge regimes with climate model uncertainty," *Hydrological Processes*, vol. 20, no. 10, pp. 2091–2109, 2006.

[20] G. Bertoldi, S. Della Chiesa, G. Niedrist, A. Rist, E. Tasser, and U. Tappeiner, "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*, vol. 12, 2010, EGU2010-12109, 2010, EGU General Assembly, Vienna, Austria.

[21] J. Cihlar, M. C. Dobson, T. Schmugge et al., "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, 1987.

[22] T. Lakhankar, H. Ghedira, M. Temimi, A. E. Azar, and R. Khanbilvardi, "Effect of land-cover heterogeneity on soil moisture retrival using active microwave remote sensing data," *Remote Sensing*, vol. 1, pp. 80–91, 2009.

[23] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.

[24] L. Bruzzone and F. Melgani, "Robust multiple estimator systems for the analysis of biophysical parameters from remotely sensed data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 1, pp. 159–174, 2005.

[25] L. Pasolli, C. Notarnicola, and L. Bruzzone, "Soil moisture estimation from microwave remote sensing data with non-linear machine learning techniques," in *Image and Signal Processing for Remote Sensing XV*, vol. 7477 of *Proceedings of SPIE*, Berlin, Germany, September 2009.

[26] L. Pasolli, C. Notarnicola, and L. Bruzzone, "Multiobjective model selection for non-linear regression techniques," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS '10)*, pp. 268–271, Honolulu, Hawaii, USA, July 2010.

[27] C.-C. Chang and C. J. Lin, "LIBSVM-A library for Support Vector Machines," http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

Journal of
Waste Management

Journal of
Ecosystems

International Journal of
Atmospheric Sciences

International Journal of
Biodiversity

Journal of
Earthquakes

The Scientific
World Journal

Journal of
Geological Research

Journal of
Computational
Environmental Sciences

Journal of
Environmental and
Public Health

Scientifica

International Journal of
Forestry Research

International Journal of
Geophysics

International Journal of
Ecology

International Journal of
Oceanography

Journal of
Marine Biology

Applied &
Environmental
Soil Science

Advances in
Meteorology

Journal of
Climatology

Advances in
Environmental
Chemistry

Hindawi

Submit your manuscripts at
http://www.hindawi.com