PhD Dissertation

International Doctorate School in Information and Communication Technologies

DISI - University of Trento

Development of new data fusion techniques for improving snow parameters estimation

Ludovica De Gregorio

Advisor:
Prof. Lorenzo Bruzzone
University of Trento

Co-advisor:
Dr. Claudia Notarnicola
Institute for Earth Observation, Eurac Research

2019

A Jacopo

Abstract

Water stored in snow is a critical contribution to the world's available freshwater supply and is fundamental to the sustenance of natural ecosystems, agriculture and human societies. The importance of snow for the natural environment and for many socio-economic sectors in several mid- to high-latitude mountain regions around the world, leads scientists to continuously develop new approaches to monitor and study snow and its properties.

The need to develop new monitoring methods arises from the limitations of in situ measurements, which are pointwise, only possible in accessible and safe locations and do not allow for a continuous monitoring of the evolution of the snowpack and its characteristics. These limitations have been overcome by the increasingly used methods of remote monitoring with space-borne sensors that allow monitoring the wide spatial and temporal variability of the snowpack. Snow models, based on modeling the physical processes that occur in the snowpack, are an alternative to remote sensing for studying snow characteristics.

However, from literature it is evident that both remote sensing and snow models suffer from limitations as well as have significant strengths that it would be worth jointly exploiting to achieve improved snow products. Accordingly, the main objective of this thesis is the development of novel methods for the estimation of snow parameters by exploiting the different properties of remote sensing and snow model data. In particular, the following specific novel contributions are presented in this thesis:

i. A novel data fusion technique for improving the snow cover mapping. The proposed method is based on the exploitation of the snow cover maps derived from the AMUNDSEN snow model and the MODIS product together with their quality layer in a decision level fusion approach by mean of a machine learning technique, namely the Support Vector Machine (SVM).
ii. A new approach has been developed for improving the snow water equivalent (SWE) product obtained from AMUNDSEN model simulations. The proposed method exploits some auxiliary information from optical remote sensing and from topographic characteristics of the study area in a new approach that differs from the classical data

assimilation approaches and is based on the estimation of AMUNDSEN error with respect to the ground data through a k-NN algorithm.
The new product has been validated with ground measurement data and by a comparison with MODIS snow cover maps. In a second step, the contribution of information derived from X-band SAR imagery acquired by COSMO-SkyMed constellation has been evaluated, by exploiting simulations from a theoretical model to enlarge the dataset.

Keywords*: snow cover area, snow water equivalent, optical remote sensing, active Syntetic Aperture Radar (SAR), snow model, machine learning, Support Vector Machine (SVM), k-Near Neighbor (k-NN) algorithm*.

Acknowledgments - Ringraziamenti

Se sono arrivata a questo punto, a scrivere i ringraziamenti per tutto questo, è sicuramente grazie a tante persone che fanno parte della mia vita e che mi hanno stimolato e aiutato in questo lungo (e anche faticoso) percorso.

La prima persona che voglio ringraziare è Claudia, colei che ha creduto in me prima e durante questo dottorato, che mi ha proposto di intraprendere questo percorso, che mi ha aiutato nei momenti più difficili, che con la sua disponibilità mi ha aiutato a crescere professionalmente, che ha fatto le nottate per leggere i miei articoli, che anche se dall'altra parte del mondo risponde in meno di 12 ore alle mails, che è rimasta fino a sera il venerdì della consegna della tesi, qualora "mi servisse qualcosa"…insomma, è grazie a te Claudia se sono riuscita ad arrivare qui e per questo ti sono infinitamente grata!

Vorrei poi ringraziare Lorenzo, che ha accettato di essere il mio tutor durante questo percorso, che anche se a distanza mi ha sempre fornito ottimi consigli sullo sviluppo di questo lavoro oltre alle utili (sebbene spesso di difficile interpretazione, viste che scritte a mano!) correzioni degli articoli. Grazie per avermi messo a disposizione la tua competenza e la tua esperienza di cui sicuramente farò tesoro.

Un'altra persona che mi ha supportato in questo lavoro è l'ormai caro amico Rudi, che oltre a darmi un grande supporto nell'utilizzo e l'interpretazione di tutti i dati provenienti dall'Ufficio Idrografico della Provincia che ho usato in questo lavoro, è riuscito anche a sopportarmi con le mie infinite ed estenuanti richieste di aiuto. Insomma, è grazie a te, Rudi, che gran parte di questo lavoro è stato validato! ☺ Voglio anche ringraziarti per le stimolanti e utili discussioni sulla neve che mi hanno chiarito dubbi e aiutato nell'interpretazione di alcuni risultati di questo lavoro.

Un grazie poi alle mie amiche, nonché colleghe, Daniela e Paola che sono riuscite a darmi la forza, farmi sorridere e ad alleggerire anche i momenti più pesanti e difficili. Grazie per la vostra sincera amicizia e anche per i grandi match a tennis, per gli allenamenti, per quell'ora settimanale di libertà e svago a cui difficilmente e raramente ho rinunciato durante tutto il periodo del dottorato.

Un grazie ai miei genitori e ai miei fratelli che mi hanno sempre appoggiato nelle mie scelte e che mi hanno fatto forza, anche se distanti, nei momenti di sconforto. Grazie mamma per essere riuscita

ad essere presente come nonna anche all'ultimo momento, prendendo il primo treno disponibile per venire a darmi una mano.
Un grazie speciale a Tom, che soprattutto nell'ultimo periodo è riuscito a sostenermi nel migliore dei modi. Grazie per le cenette che mi hai fatto trovare pronte, per le commissioni che hai sbrigato tu al mio posto per non farmi perdere tempo prezioso e infine grazie perché contribuisci alla felicità di nostro figlio Jacopo, che è rimasta l'ultima (volutamente) persona da ringraziare per questa esperienza.
Grazie a te caro Jacopo, che sei nato poco dopo aver iniziato questo percorso e che sei cresciuto con la mia tesi. È grazie a te se quando uscivo da lavoro, anche se stressata e stanca, riuscivo a distrarmi e a staccare del tutto la spina perché, volente o nolente, tu eri lì a richiedere attenzioni, che quando mi vedevi arrivare all'asilo mi saltavi al collo con una felicità contagiosa che mi permetteva di non pensare ad altro che a te e a quanto è bello che, nonostante la fatica, tu ci sia. Grazie.

Contents

List of Figures

List of Tables

Chapter 1

1. INTRODUCTION

1.1 Background

Snow cover is a critical geophysical parameter for Earth climate and hydrological systems [1]. Snow cover contributes to regulate the Earth surface temperature, and once it melts, the melting water helps fill rivers and reservoirs in many regions of the world.

In terms of spatial extent, seasonal snow cover is the largest single component of the cryosphere and has a mean winter maximum areal extent of 18.1 million square miles, about 98% of which is located in the Northern Hemisphere [2].While on large scale snow cover changes affect the energy exchange between Earth's surface and the atmosphere and are, thus, useful indicators of climatic variation, on a smaller scale, variations in snow cover can affect regional weather patterns. Therefore, snow cover is an important climate change variable at both large and small scales because of its influence on energy and moisture budgets. The high snow albedo (which measures how much sunlight is reflected back into the atmosphere) implies that a snow-covered surface may reflect up to 90% of incoming solar radiation, whereas vegetation and soil may reflect only 10-20% of sunlight. A decrease in snow cover results in a decrease of reflected energy and thus in an increase of solar radiation absorption, by adding heat to the system and self-powering the process. Surface temperature is highly dependent on the presence or absence of snow cover, and temperature trends have been linked to changes in snow cover [3] [4] [5]. Thus, since the beginning of the satellite era in the 1960s, the areal extent of snow cover has been a key satellite observation target for the purposes of daily weather forecasting and a better understanding of the Earth's climate system and hydrological cycle [6] [7] [8].

Figure 1.1 shows the average snow extent by month relative to the Norther Hemisphere. Data refer to the period between November 1966 and January 2017. August has the lowest value of average monthly snow cover extent with about 1.1 million square miles and with most of the snow cover observed at the high latitudes of the Arctic and in the highest elevations at lower latitudes. Snow

cover typically reaches its largest extent in mid-winter, with a January average snow cover extent of about 18.1 million square miles, covering 31% of Earth's land surface. The springtime snow melt is particularly important in terms of spring river discharge, permafrost thaw, and the length of the vegetation growing season.

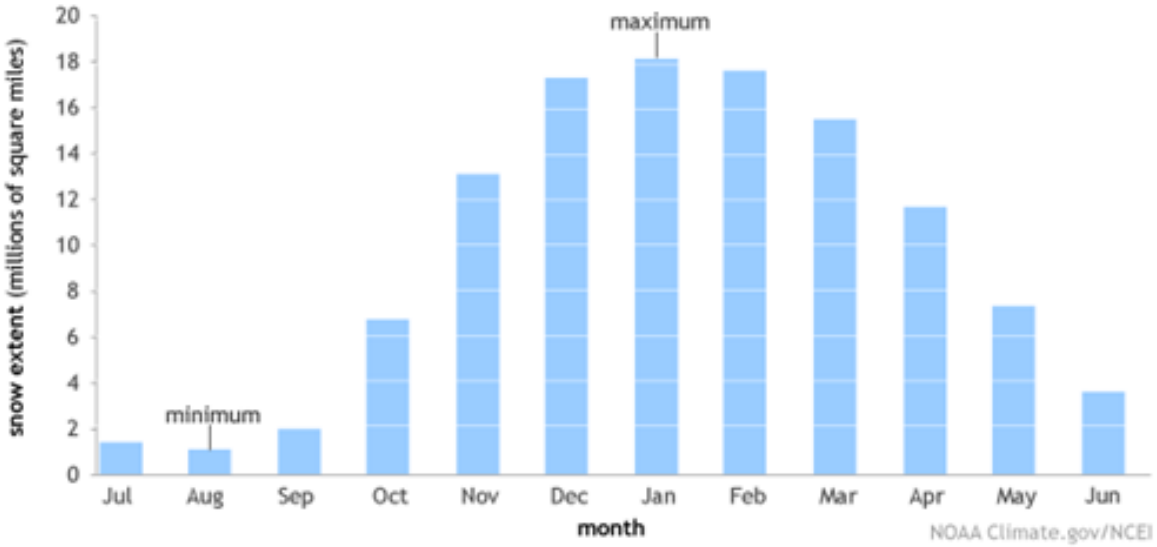

Figure 1.1 The annual snow cycle in the Northern Hemisphere, based on Rutgers Snow Lab data provided by Jake Crouch, NCEI. Graph by NOAA Climate.gov

Recent studies of snow climatology suggest an overall tendency toward decreases in several metrics of snow such as snow cover extent, snow water equivalent (SWE), and snow depth (SD) for the winter period from 1960/1961 to 2014/2015 [9]. In the study, the negative trend is especially strong in North America and apparently less evident in Europe, but this is mainly due to limited data coverage for Eurasia.

Because of the strong consequences of changes in snow amount on Earth's environment and population, scientists have developed ways to continuously measure and monitoring snow and its properties. While over the long-term changes in snow cover are useful as indicators of climatic variations, in the short term, information about snow amount can help for more operational purposes by providing indications about water availability, which are useful for water management or flood risk forecasting. The next sections show an overview on available methods for snow parameter retrieval.

1.2 Snow Observation

The traditional snow observations consist of in situ measurements during periodic field campaigns at fixed sites or through networks of automatic nivological stations recording snow parameters and often coupled with weather stations. The objective of field campaigns is to characterize the day from a nivometeorological point of view through the collection of a series of significant parameters extracted with simple and fast procedures. The measurements concern parameters such as: snowpack depth, fresh snow depth, air temperature, weather conditions, cloudiness, visibility and wind activity at high altitude. Nevertheless, since this thesis will focus on snow coverage and snow water equivalent (SWE), in the following only the description of the measurement techniques used for these two parameters will be reported.

The ideal site for measuring snow is a flat and wind sheltered area where the snow cover and the base surface are relatively homogeneous. Moreover, the measurement area should also represent the surrounding landscape as much as possible [10]. In alpine terrain, a flat area large enough to make a representative measurement without encountering edge effects should be chosen, avoiding basins, slopes and ridges.

1.2.1 Snow Depth

Snow depth is defined as the vertical distance from the snow surface to a stated reference level (the base surface, typically the ground) and is reported in full centimeters (cm). Unless the measurement area is very homogeneous, the automatic snow depth measurements at the point scale not necessarily are representative of the surrounding landscape. Vice versa, for manual snow depth measurements the observer has the possibility of averaging multiple point measurements and reporting the mean value. Manual snow depth measurements are generally performed every 24 hours between 06.00 and 08.00 UTC; depending on the site, fixed stakes or portable rulers are used by observers (Figure 1.2). The observers avoid creating disturbances in the measurement field around the stake. To prevent incorrect measurements, values are read as horizontally to the surface as possible. Measured values are reported in full centimeters.

Automatic snow depth measurements are performed by means of instruments employing either sonic or optical (laser) technology, depending on the measurement sites and on the country. Laser instruments have a higher degree of precision (approx. 0.1 cm) than sonic ones (approx. 2 cm), but they require more power for operation. Laser snow depth sensors emit a modulated beam of light

in the visible part of spectrum and determines the distance to a target by analyzing the phase information from the reflected beam; this kind of sensors are usually mounted 2 m above the ground [10]. Sonic snow depth sensors, instead, transmit an ultrasonic pulse towards the target and listen for a return signal reflected from that target. Sonic instruments usually measure the distance to the highest obstacle within the instrument response area, i.e. a conical footprint, the radius of which depends on the height of the instrument above the target; these sensors are usually located 4‑6 m above the ground [10], Figure 1.3.

Figure 1.2 (a) Manual measurement field, with a snow stake and a fresh snow board (Source: Meteomont Carabinieri, Italy). (b) Manual snow depth measurement with a graduated snow rod in a measurement (Source: Corpo Forestale Regionale Valanghe, SSCV, Friuli Venezia Giulia). (c) An extendible 1-cm-graduated snow rod (Source: Institute for Environment and Climate Change Canada, ECCC, Canada)

Figure 1.3 Automatic snow depth instruments: (a) sonic instrument with artificial turf below (Source: Finnish Meteorological Institute, FMI, Finland); (b) laser instrument with an artificial target below (Source: Deutscher Wetterdienst, DWD, Germany).

1.2.2 Snow Water Equivalent (SWE)

The SWE is defined as the vertical depth of water that would be obtained if the snow cover melted completely. SWE is the product of the snow depth in meters and the vertically integrated density of the snow in kilograms per cubic meter [11]. Generally, all manual measurement techniques of this parameter involve a snow sampler which collects a known (or calculable) volume of snow from which snow density is derived. During the field campaigns, a snow pit is manually dug down to the ground and used both for snowpack stratigraphy observations and for calculating the snow density. Field campaigns with snow pits are theoretically performed once per week, but in practice longer intervals are common. Using a graduated snow cylinder (in aluminum or steel) with a certain cross-sectional area (in m^2) and a certain length (in m), a snow sample is extracted vertically from the snowpack. The snow cylinder is then attached to a spring scale to measure the total weight of the snow (in kg). The corresponding SWE of the sample is calculated by dividing the measured weight by the cross-sectional area of the snow cylinder. The final SWE of the snowpack is calculated by summing the SWE values of all sampled layers. The snow tube measurement is based on the same concept of snow cylinder with the difference that in this case no snow pits are needed because the snow tube is inserted vertically into the snowpack until it reaches the base surface (using tube extensions where required) and a snow core is extracted. Figure 1.4 shows the two manual methods for estimating SWE.

The most common automatic measurement techniques for SWE retrieval are (1) weighing mechanisms (e.g. snow pillows or snow scales) and (2) passive gamma radiation instruments (Figure 1.5). The measurement principle of weighing mechanisms is similar to that of the manual instruments that measure the weight of the snowpack, by converting it in density and subsequently in SWE. Passive gamma radiation instruments apply the concept that snow attenuates natural gamma emissions from the soil, and the magnitude of attenuation is related to the mass of the water between the soil and the detector [12]. However, in this thesis no automatic measurements of SWE have been employed because the few available are not validated

Figure 1.4 (a) Snow cylinders, with volumes of 0.0001 and 0.0005 m3, and a spring scale (valid for a maximum weight of 0.5 kg) used to manually measure SWE in snow pits (Source: Provincia Autonoma di Trento, PAT, Italy). (b) A snow cylinder being weighed with a spring scale (Source: WSL Institute for Snow and Avalanche Research SLF, Switzerland). (c) Snow tube attached to a scale (Source: Agenzia Regionale per la Protezione dell'Ambiente Valle d'Aosta, ARPAVA, Italy).

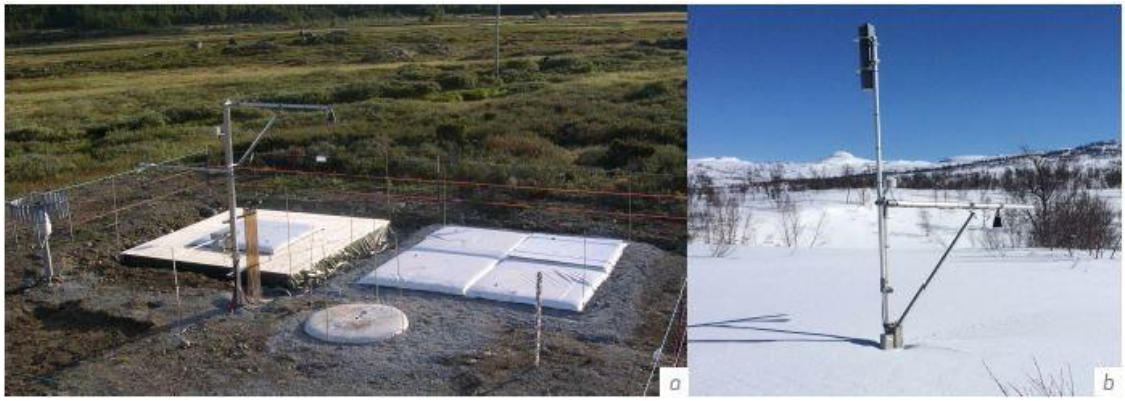

Figure 1.5 (a) Snow pillow and snow scale at the Filefjell station. (b) Gamma radiation sensor at the Breidvatn station (Source: the Norwegian Water Resources and Energy Directorate, NVE, Norway).

1.3 Satellite Remote Sensing of Snow

Satellite remote sensing represents an important tool for monitoring snow properties at large scale and in remote or inaccessible areas where in-situ measurements may be expensive and dangerous. The global coverage and regular repeatability of measurements provided by satellite remote sensing allow monitoring the wide spatial and temporal variability of the snowpack [7]. The interaction between snow cover and electromagnetic radiation at different frequencies makes snow distinguishable from other land covers by allowing its detection through the use of both active and passive remote sensing techniques. All these techniques present some limitations due to different factors, such as cloud presence, forest cover or complexity of mountainous terrain with its heterogeneity.

1.3.1 Visible and Near-Infrared Sensors

Due to the high snow reflectance in the visible part of the electromagnetic spectrum (0.4-0.8 μm), optical remote sensing is very suitable to detect the snow extent (i.e. the presence or absence of snow) regardless of snow amount. However, some limitations exist: firstly, darkness or low illumination conditions are problematic for optical sensors that can provide visible imagery only in that portion of the surface illuminated by sunlight. Secondly, the presence of clouds does not allow the optical signal to transit by reducing the possibility of acquisition to the only cloud-free conditions. Almost all clouds, indeed, reflect a significative part of visible radiation, preventing any visible radiative information about the surface from reaching the satellite [7]. Moreover, due to the similarity between albedo values of snow and some type of clouds, the discrimination between cloud and snow-covered surfaces may be difficult. For this reason, near-infrared bands

can be used to distinguish snow from most of clouds because at these wavelength, reflectance of most clouds is high while that of snow is low.

Finally, forest cover and in general vegetation obscure the underlying surface and lower the surface reflectance [13], by making difficult the snow detection below the trees. Figure 1.6 shows an example of optical image over the alpine area obtained from the Moderate-resolution Imaging Spectroradiometer (MODIS) aboard the NASA satellite Terra.

Figure 1.6 MODIS image over the alpine arc on 23 March 2019 (source: NASA https://worldview.earthdata.nasa.gov/).

1.3.2 Microwave sensors

When snow covers the ground, some of the microwave energy emitted by the underlying soil is scattered by the snow grains. Therefore, microwave emission from a snow-covered surface is diminished with respect to a snow-free surface, and the presence of snow can be identified [14], [15]. The signal attenuation depends on the microwave wavelength and the snowpack properties, such as the amount of snow, the grain size, the snow density, presence of ice lenses and the amount of liquid water [16]. Microwave data are mainly exploited for estimating volumetric snow parameters as SD and SWE [17], [18], [19]. Clifford [20] provides a review of global estimates of snow water equivalent from passive microwave. However, also passive microwave sensors suffer

of some limitations: one of the major limitations is the presence of liquid water in the snowpack that affects the snowpack dielectric constant, by increasing the absorption, masking the microwave emission signal from the snow and thus inhibiting the ability of microwave sensors to detect wet snow. In the last decades, scientists have also extensively investigated the potential of active Synthetic Aperture Radar (SAR) data for deriving SWE [21], [22], [23], [24].

Unlike visible and infrared sensors, microwave sensors do not depend on the presence of sunlight and thus are a valid alternative at high latitudes; moreover, microwave sensors are not affected by the presence of clouds, by offering the potential to estimate snow cover properties also in cloudy conditions. An example of microwave image is shown in Figure 1.7.

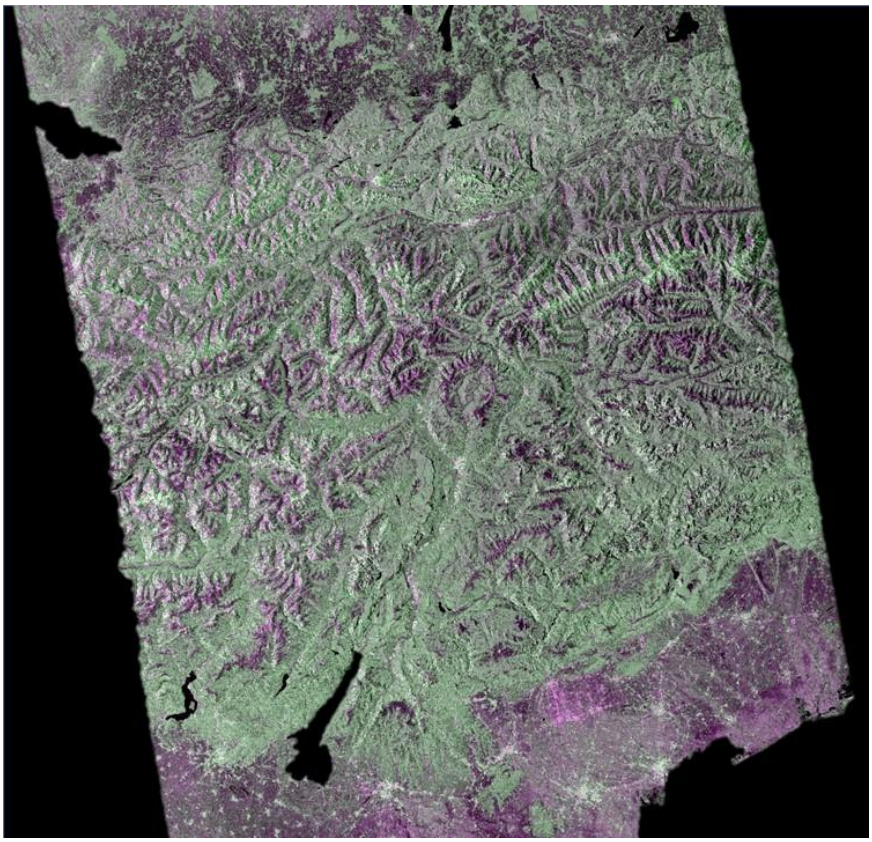

Figure 1.7 RGB composition of preprocessed Sentinel-1 backscatter data (Red: VV; Green: VH Blue: VV) from Track 117 on 24 January 2016. Source: Contains modified Copernicus Sentinel data [2016]/Eurac research

1.4 Snow Modelling

During the last decades many snow models have been developed. The modelling of physical processes that occur in the snowpack are used in hydrological forecasting, in numerical weather prediction and climate modelling.

Snow models calculate the energy and mass balances of snow on the ground by considering that snow, in some cases, may be partially obscured by tree canopy, which may itself hold intercepted snow. Moreover, it can occur that ground is partially snow-free. As a consequence, radiative fluxes beneath canopies are modified by interception of shortwave radiation and emission of longwave radiation by the tree canopy. Snow and rain can be partially intercepted by vegetation canopies and subsequently can be removed by evaporation or sublimation to the atmosphere or fall down to the underlying surface. Rain or meltwater at the snow surface can percolate into the snow, where a certain amount of water can be held in the liquid form or refreeze, releasing latent heat. Melt water reaching the base of the snow is partitioned into infiltration into the soil, runoff or basal ice formation. Snow models express the energy and mass balances through equations characterizing the temperature and water content of the canopy, snowpack and soil, coupled with terms that describe the evaporation, sublimation and melt processes [25]. Most simple models approximate the snowpack as a single layer or a combined snow and soil layer, but there has been increasing use of multi-layer snow models with 3 to 5 snow layers [26] [27]. In the single layer approach the surface temperature between the snowpack and the atmosphere above is modelled in a relatively straightforward way, by avoiding modelling the uncertainty of the processes within the snowpack. Indeed, one of the primary reasons of poor performance of single-layer models is the poor representation of internal snowpack heat transfer processes [28], [29]. The parametrization of surface properties and processes is fundamental in snow modelling: the energy exchange between snow and atmosphere is controlled by numerous factors and albedo plays a key role. Snow albedo is generally parametrized as a function of surface temperature and snow age. The effect of snow age on albedo differs at different wavelengths: in the near-infrared spectrum region, snow albedo decreases with age because snow grain size increases due to the metamorphism, whereas in the visible spectrum region snow albedo decreases with age due to the accumulation of impurities related to aerosols and dust deposition on snow surface [25]. For these reasons, some snow models consider albedo in two or more spectral bands. Another parameter that snow models consider is the surface roughness: indeed, snow reduces the surface roughness by covering vegetation and filling the topographic depressions. This can be represented in models by decreasing the surface roughness length as a function of snow depth, down to a minimum value for deep snowpack [25].

However, literature about snow models reveals that a small number of parameterizations are used in different combinations from different models, so the models are not all truly independent [30].

Intercomparison studies have shown that models differ greatly in their predictions of snow accumulation and ablation. Nevertheless, Essery et al. [30] showed that there is no "best" model, and increasing model complexity is no guarantee of improved model performance; well-established empirical parameterizations often give results that are as good as physically-based parameterizations. Although the physical processes within the snowpack are increasingly well parameterized, uncertainties still exist, by affecting the energy and mass balance simulation at the snowpack surface.

1.5 The study area

In this section a description of the investigation area is provided in order to characterize it from the meteorological and topographic point of view. The study area of this work is the European Region Tyrol – South Tyrol – Trentino, which consists of the Austrian federal state of Tyrol (12.648 km^2) and the Italian region of Trentino-Alto Adige (6.207 km^2 for Trentino and 7.398 km^2 for Alto Adige). The elevation range in the entire area is remarkable. Indeed, after having deeply shaped the landscape, the Sarca river flows into the waters of Lake Garda at an altitude of only 65 m a.s.l., while the Ortles peak reaches the 3.905 m altitude. Half of the study area surface is between 1.000 and 2.000 meters above sea level; only 20% is at lower altitudes and the remaining 30 % at higher altitudes. The average altitude is 1.620 m, while the mean slope is 23° and only 5 % of the area is considered flat (slope angle ≤3°).

Trentino includes the Garda Lake, the largest lake basin in Italy and the only large inland lake in the area considered in the study. Table 1. 1 shows the minimum and the maximum elevation and the percentage of area covered by the forest in each region. To have an idea of the dominating meteorological conditions for the snow accumulation, Figure 1.8 shows the mean annual air temperature (a) and precipitation (b) and the mean DJF air temperature and precipitation relative to the winter period of December-January-February (c and d).

*Table 1. 1 Topographic characteristics of the study area. *Source: http://www.europaregion.info/it/cifre-euregio.asp*

Parameter	Tyrol	South Tyrol	Trentino
Maximum elevation [m a.s.l.]	3.798	3.905	3.769
Minimum elevation [m a.s.l.]	465	207	65
Percentage of area covered by forest (%)*	36,9	39,5	63

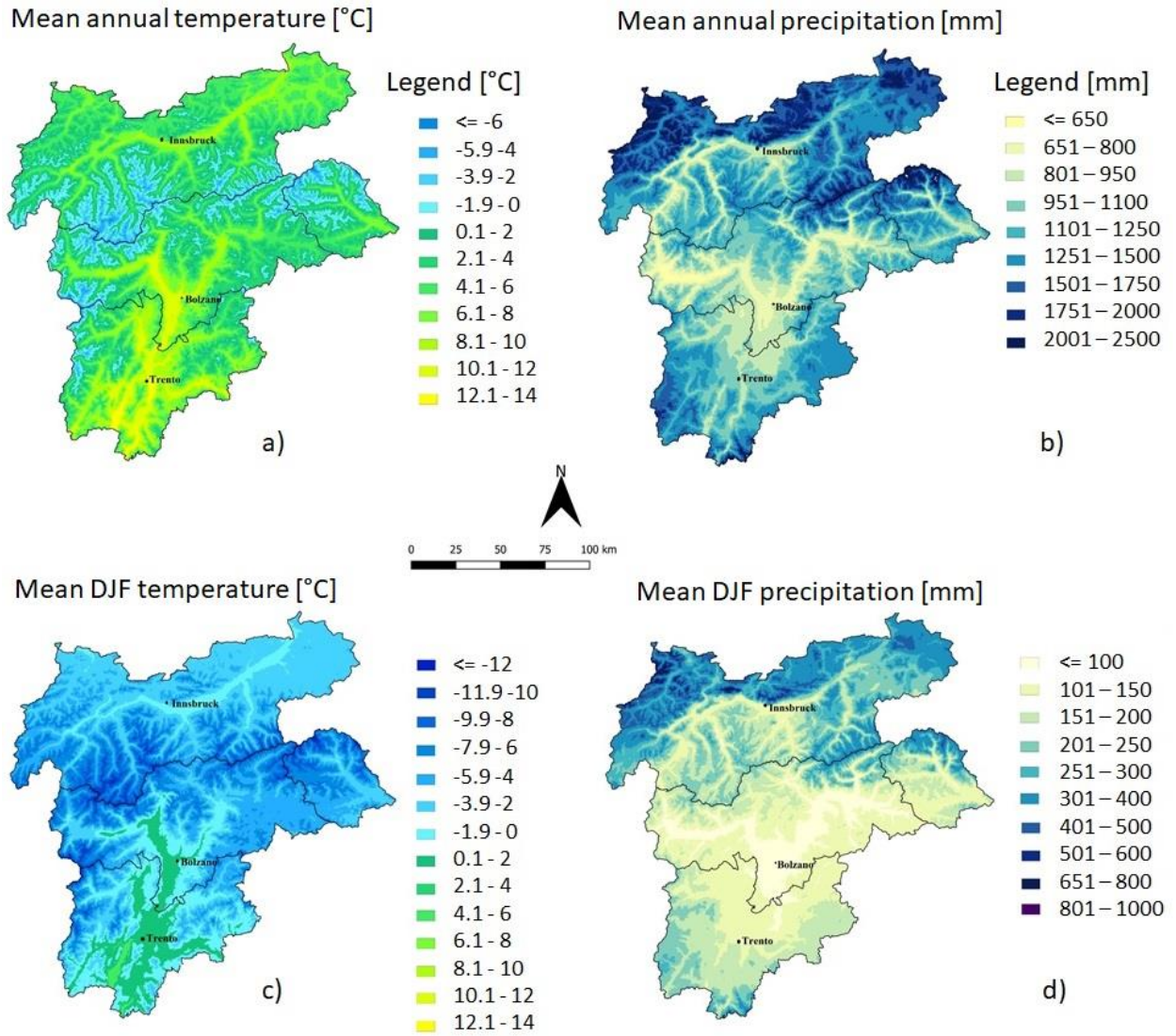

Figure 1.8 Figure represents the mean values computed over the period 1981-2010 of the following parameters: a): Mean annual air temperature; b): Mean annual precipitation; c) Mean DJF air temperature; d) Mean DJF precipitation. Source: the book "Il clima del Tirolo - Alto Adige – Bellunese", output of the Interreg project 3PCLIM, http://www.clima-alpino.eu/).

The winter season with the relative snowfalls starts, on average, earlier in the northern slopes than in southern ones. At an altitude of 2000 m, for example, in the north, the first snowfalls are to be expected in the first half of September, while on the southern slopes at the beginning of October. Vice versa, at these altitudes, the last snowpack melts, on average, around the summer solstice in the northern Alps and already at the beginning of June on the southern alpine side. Regarding lower altitudes, at an altitude of 1000 m a.s.l., the northern Alps are covered by a snowpack mainly

between the end of October and the second half of April, while on the southern alpine side the first snowfalls occur in mid-November and the last snow melts in early April.

1.6 Thesis Objectives and Contributions

The limitations of existing approaches, described in previous sections, and the lack of continuous and spatially homogeneous distribution of snow measurements (both manual and automatic) have pointed out the importance of further improvements in the estimation and monitoring of the heterogeneous distribution of snow cover and of its properties. The high complexity and the non-linearity of snow parameters retrieval problem require the development and the use of advanced methods.

Traditional approaches exploit the strengths of different sources (theoretical model simulations, remote sensing images and ground measurements) through data assimilation techniques [31], [32], [33]. However, a class of effective regression methods, which has been successfully introduced in the field of geo/bio-physical variable estimation in the last decades as an alternative to data assimilation techniques, is represented by non-linear machine learning techniques.

Due to advanced learning strategies, machine learning techniques can learn and approximate even complex non-linear systems, exploiting the information contained in a set of reference samples. These techniques have also the advantage of not requiring any assumptions a priori about the data distribution. Due to this property, the retrieval process can integrate data coming from different sources with poorly defined (or unknown) probability density functions but that are correlated to the target variable. Machine learning methods have shown their versatility in different contexts by using optical and radar data, by fusing remotely sensed data with ground data, as well as by exploiting data derived from theoretical model simulations. These approaches have been also compared to other parametric approaches (such as iterative or Bayesian approaches), indicating that in most of the cases, machine learning methods outperformed these latest ones [34].

In this context, the main objective of this thesis is the development of new methods based on machine learning techniques for improving the estimation of two crucial snow parameters: the *snow coverage* and the *SWE*. The study area is the alpine area that includes Trentino - Alto Adige region (in the north-eastern part of Italy) and Tyrol region in (in south Austria). To achieve this, satellite products together with ground measures and data derived from a snow model and two

coupled electromagnetic models are jointly exploited in a data fusion approach through the use of machine learning techniques.

The following specific objectives of the thesis have been identified:

- To develop a novel data fusion technique for improving the estimation of snow coverage over the study area. The innovative aspect is the joint exploitation of remotely sensed data and snow model simulations, differing from traditional techniques where remote sensing is mainly used for model tuning or in data assimilation approaches. In this thesis a decision-level fusion process is implemented. The approach first retrieves the snow cover maps and their quality measures separately from the two different sources. Then, the two maps are fused to obtain a new and enhanced product that overcomes the aforementioned limitations and takes advantage of both the specific properties of remote sensing data and of the snow model simulations.
- To develop a new concept to improve the distributed estimation of SWE derived by the snow model (the AMUNDSEN model is used in the thesis), by exploiting both topographic parameters and auxiliary products from optical remote sensing data to correct the model with respect to the ground measurements. The novelty of this proposed method is the approach based on the error estimation with an adaptive k-NN algorithm for improving the SWE derived from model simulations.
- To assess the contribution of the information derived from the COSMO-SkyMed X-band SAR for SWE retrieval. To this end, a Support Vector Regressor (SVR) has been trained on a dataset consisting of simulated backscattering coefficients and then tested on available satellite data. In both cases, the ground measurements derived SWE values are used as reference dataset. The simulated values of backscattering have been obtained through the use of two electromagnetic models (Dense Medium Radiative Transfer theory, DMRT [35], coupled with the Advanced Integral Equation Method, AIEM [36]) and by exploiting as models inputs the snow parameters collected from ground measurements. The novelty of this part is the exploitation of the potential of X-band data together with simulated data and ground measurements for SWE retrieval by means of a machine learning technique.

From a global perspective, the main goal of this thesis is thus to improve the estimation of snow coverage and SWE through the development of *general* methods transferable even in other regions with respect to the study area and applicable to other physically based model.

The main novel contributions related to the above-mentioned objectives are:

- The development of a method that jointly exploit remotely sensed data and physical model simulations, differing from traditional approaches where remote sensing is mainly used for model tuning or in a data assimilation context. In this thesis, a decision-level fusion process is implemented, where the snow cover maps and their quality measures are retrieved separately from the two different sources and then integrated by SVM to exploit their complementarities and to address their uncertainties.
- The development of an advanced estimation technique to derive SWE by jointly using snow model simulations, ground data and auxiliary products based on optical remote sensing. The presented approach, based on the error estimation with an adaptive k-NN algorithm, represents a novel and relevant contribution for the SWE retrieval, and in general in the field of snow hydrology.
- The sensitivity of backscattering at X-band to the snow parameters is still controversial, due to different behaviors depending on the variable snow characteristics. In this thesis, the main novel contribution regards the exploitation of COSMO-SkyMed X-band SAR data by means of a machine learning technique, based on SVR, for the estimation of SWE.

1.7 Thesis Structure

This thesis is organized as follows:

Chapter 1: provides the background about the topic of this work, by introducing the role of snow in Earth's climate system and hydrological cycle as well as the effect of changes in snow cover on human life. Then the state of the art of the existing techniques for snow parameters retrieval is presented and the main objectives of the thesis introduced.

Chapter 2: is dedicated to the first part of the research, i.e. the development of a novel technique for improving the snow cover mapping, by generating a time series of snow maps. The method has been validated by exploiting both high-resolution satellite images and ground measurements of snow depth on which a threshold value has been imposed to obtain binary values of snow presence.

Chapter 3: provides a description of the method adopted for improving the SWE estimation derived by the snow model AMUNDSEN. The snow model, the ground-based and the remote sensing data as well as the methodological workflow are presented. Afterwards, results are

analyzed and compared with ground data and MODIS snow maps in order to validate and verify the potential and limitations of proposed method for SWE retrieving.

Chapter 4: explores the potential of COSMO-SkyMed X-band SAR for SWE retrieval. In this chapter, COSMO-SkyMed mission is presented together with the theoretical model used for simulating backscattering values, in order to increase the size of training dataset. The proposed method tries to catch the relation between X-band backscattering and SWE by means of a machine learning technique, based on the Support Vector Machine (SVM).

Chapter 5: draws the conclusion of this thesis and presents a brief summary on the possible further developments of the research activities.

References

[1] A. Frei and D. A. Robinson, "Northern Hemisphere snow extent: Regional variability 1972–1994," *International Journal of Climatology: A Journal of the Royal Meteorological Society,* vol. 19, no. 14, pp. 1535-1560, 1999.

[2] N. National Snow & Ice Data Center, "National Snow & Ice Data Center (NSIDC)," 20 June 2019. [Online]. Available: https://nsidc.org/cryosphere/snow/climate.html.

[3] P. Groisman, T. Karl and R. Knight, "Observed impact of snow cover on the heat balance and the rise of continental spring temperatures," *Science,* vol. 263, pp. 198-200, 1994.

[4] R. Brown and D. Robinson, "Northern Hemisphere spring snow cover variability and change over 1922–2010 including an assessment of uncertainty," *The Cryosphere,* vol. 5, pp. 219-229, 2011.

[5] S. Peng, S. Piao, P. Ciais, P. Friedlingstein, L. Zhou and T. Wang., "Change in snow phenology and its potential feedback to temperature in the Northern Hemisphere over the last three decades," *Environmental Research Letters,* vol. 8, 2013.

[6] D. Robinson, K. Dewey and R. H. Jr., "Global snow cover monitoring: an update," *Bull. Am. Meteorol. Soc,* vol. 74, pp. 1689-1696, 1993.

[7] A. Frei, M. Tedesco, S. Lee, J. Foster, D. Hall, R. Kelly and R. Robinson, "A review of global satellite-derived snow products," *Adv. Space Res.,* vol. 50, pp. 1007-1029, 2012.

[8] T. Estilow, A. Young and D. Robinson, "A long-term northern hemisphere snow cover extent data record for climate studies and monitoring," *Earth Syst. Sci. Data,* pp. 137-142, 2015.

[9] K. Kunkel, D. Robinson, S. Champion, X. Yin, T. Estilow and R. Frankson, "Trends and extremes in Northern Hemisphere snow characteristics," *Current Climate Change Reports,* vol. 2, pp. 65-73, 2016.

[10] A. Haberkorn, J. Helmert, L. Leppänen, J. López-Moreno and R. Pirazzini, European Snow Booklet - an Inventory of Snow Measurements in Europe, 2019, p. 363.

[11] B. E. Goodison, H. L. Ferguson and G. A. and McKay, "Measurement and Data Analysis," in *Handbook of Snow: Principles, Processes, Management and Use*, The Blackburn Press, 1981, p. 191–274.

[12] E. L. Peck, V. C. Bissell, E. B. Jones and D. L. Burge, "Evaluation of snow water equivalent by airborne measurement of passive terrestrial gamma radiation," *Water Resources Research,* vol. 7, no. 5, pp. 1151-1159, 1971.

[13] D. Robinson and G. Kukla, "Maximum surface albedo of seasonally snow covered lands in the Northern Hemisphere," *Journal of Climate and Applied Meteorology,* vol. 24, p. 402–411, 1985.

[14] N. Grody, "Relationship between snow parameters and microwave satellite measurements: theory compared with AMSU observations from 23 to 150 GHz," *Journal of Geophysical Research,* vol. 113, no. D22108, 2008.

[15] M. Tedesco and E. Kim, "Inter-comparison of electromagnetic models for passive microwave remote sensing of snow," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 44, no. 10, pp. 2654-2666, 2006.

[16] F. Ulaby, R. Moore and A. Fung, Microwave remote sensing, active and passive. Volume II: Radar remote sensing and surface emission theory, Norwood, Massachusetts: Artech House, 1982.

[17] C. Derksen, "The contribution of AMSR-E 18.7 and 10.7 GHz measurements to improved boreal forest snow water equivalent retrievals," *Remote Sensing of Environment ,* vol. 112, pp. 2701-2710, 2008.

[18] C. Derksen, A. Walker, B. Goodison and J. Strapp, "Integrating in situ and multiscale passive microwave data for estimation of subgrid scale snow water equivalent distribution and variability," *IEEE Transactions on Geoscience and Remote Sensing ,* vol. 43, no. 5, pp. 960-972, 2005.

[19] J. Pulliainen and M. Hallikainen, "Retrieval of regional snow water equivalent from space-borne passive microwave observations," *Remote Sensing of Environment ,* vol. 75, no. 1, pp. 76-85, 2001.

[20] D. Clifford, "Global estimates of snow water equivalent from passive microwave instruments: history, challenges and future developments.," *International Journal of Remote Sensing ,* vol. 31, no. 14, p. 3707–3726, 2010.

[21] F. T. Ulaby and W. H. Stiles, "The active and passive microwave response to snow parameters: 2. Water equivalent of dry snow," *Journal of Geophysical Research: Oceans,* vol. 85, no. C2, pp. 1045-1049, 1980.

[22] J. R. Kendra, K. Sarabandi and F. T. Ulaby, "Radar measurements of snow experiment and analysis," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 36, pp. 864-879, 1998.

[23] S. Sun, T. Che, J. Wang, H. Li, X. Hao, Z. Wang and J. Wang, "Estimation and analysis of snow water equivalents based on C-band SAR data and field measurements," *Arctic, antarctic, and alpine research,* vol. 47, no. 2, pp. 313-326, 2015.

[24] V. Conde, G. Nico, P. Mateus, J. Catalão, A. Kontu and M. Gritsevich, "On the estimation of temporal changes of snow water equivalent by spaceborne SAR interferometry: a new application for the Sentinel-1 mission," *Journal of Hydrology and Hydromechanics,* vol. 67, no. 1, pp. 93-100, 2019.

[25] R. Essery, "Snow modelling," in *ECMWF Seminar on Polar Meteorology*, Reading, UK, 4-8 September, 2006.

[26] P. Bartelt and M. Lehning, "A physical SNOWPACK model for the Swiss avalanche warning: Part I: numerical model," *Cold Regions Science and Technology,* vol. 35, no. 3, pp. 123-145, 2002.

[27] E. Brun, P. David, M. Sudul and G. Brunot, "A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting," *ournal of Glaciology,* vol. 38, no. 128, pp. 13-22, 1992.

[28] G. Blöschl and R. Kirnbauer, "Point snowmelt models with differ-ent degrees of complexity – internal processes," *J. Hydrol.,* vol. 129, p. 27–147, 1991.

[29] H. Koivasulo and M. Heikenkeimo, "Surface energy exchange overa boreal snowpack," *Hydrol. Process.,* vol. 13, p. 2395–2408, 1999.

[30] R. Essery, S. Morin, Y. Lejeune and C. B. Ménard, "A comparison of 1701 snow models using observations from an alpine site," *Advances in water resources,* vol. 55, pp. 131-148, 2013.

[31] E. E. Stigter, N. Wanders, T. M. Saloranta, J. M. Shea, M. F. Bierkens and W. W. Immerzeel, "Assimilation of snow cover and snow depth into a snow model to estimate snow water equivalent and snowmelt runoff in a Himalayan catchment," *The Cryosphere,* vol. 11, no. 4, pp. 1647-1664, 2017.

[32] T. M. Saloranta, "Operational snow mapping with simplified data assimilation using the seNorge snow model," *Journal of Hydrology,* vol. 538, pp. 314-325, 2016.

[33] N. Griessinger, J. Seibert, J. Magnusson and T. Jonas, "Assessing the benefit of snow data assimilation for runoff modeling in Alpine catchments," *Hydrology and Earth System Sciences,* vol. 20, no. 9, pp. 3895-3905, 2016.

[34] S. Paloscia, P. Pampaloni, S. Pettinato and E. Santi, " A comparison of algorithms for retrieving soil moisture from ENVISAT/ASAR images," *IEEE Trans. Geosci. Remote Sens.,* vol. 46, pp. 3274-3284, 2008.

[35] L. Tsang, J. Pan, D. Liang, Z. Li, D. W. Cline and Y. Tan, "Modeling active microwave remote sensing of snow using dense media radiative transfer (DMRT) theory with multiple-scattering effects," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 45, no. 4, pp. 990-1004, 2007.

[36] K. S. Chen, W. T. D., L. Tsang, Q. Li, J. Shi and A. K. Fung, "Emission of rough surfaces calculated by the integral equation method with comparison to three-dimensional moment method simulations," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 41, no. 1, pp. 90-101, 2003.

Chapter 2

2. A NOVEL DATA FUSION TECHNIQUE FOR SNOW COVER RETRIEVAL

This chapter[1] presents a novel data fusion technique for improving the snow cover monitoring for a mesoscale Alpine region.

The presented methodological innovation consists in the integration of remote sensing data products and the numerical simulation results by means of a machine learning classifier (Support Vector Machine), capable to extract information from their quality measures. This differs from the existing approaches where remote sensing is only used for model tuning or data assimilation. The technique has been tested to generate a time series of about 1300 snow maps for the period between October 2012 and July 2016.

The results show an average agreement between the fused product and the reference ground data of 96%, compared to 90% of the MODIS data product and 92% of the numerical model simulation. Moreover, one of the most important results is observed from the analysis of snow cover area (SCA) time series, where the fused product seems to overcome the well know underestimation of snow in forest of the MODIS product, by accurately reproducing the SCA peaks of winter season.

[1] This chapter has been published in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing:* De Gregorio, L., Callegari, M., Marin, C., Zebisch, M., Bruzzone, L., Demir, B., Strasser, U., Marke, T., Günther, D., Nadalet, R. and Notarnicola, C. (2019). A Novel Data Fusion Technique for Snow Cover Retrieval. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. DOI: 10.1109/JSTARS.2019.2920676

2.1 Introduction

Snow is a dynamically changing water resource that plays an important role in the hydrological cycle in mountainous areas. The traditional acquisition means for the snow cover distribution and variability are in-situ monitoring stations that provide point observations for their locations. The locations of most of these stations are in easily accessible and valley areas, whereas in the higher Alpine regions there are only few in operation, a notable exception, e.g., being the special observation networks of the national avalanche warning services. Therefore, in mountain regions, where the spatial variability of the snow cover is particularly high, the related hydrological processes are mostly unknown due to the lack of spatially and temporally continuous observations [1]. To fill this gap, remote sensing can make a valuable contribution by providing high spatial and temporal resolution data. Snow cover mapping by multispectral remote sensing images implies some limitations and problems. Sources of misinterpretation can be related to:

Clouds: one of the major problems in snow detection by satellite is the distinction between clouds and snow. Depending on the spectral channels available and cloud type, very bright reflectance of some clouds can make them indistinguishable from snow cover [2].

Forest cover: the reflectance of forested areas can be much lower than the one of non-forested areas, even with a considerable snowpack beneath the trees. The forest cover obscures the snow beneath and hence hides it from the optical sensors. Additionally, tree crowns intercept snow. Due to a higher crown density, conifer trees intercept considerably more snow than leafless deciduous trees and this affects the melting pattern as well as the accumulation pattern. Therefore, it is still a challenge to accurately detect the ground snow in a forested area [3].

Shadow: shadow can be particularly relevant in the winter season on north-facing slopes in dependence of the relative position between the sun and the sensor. Similarly, cloud shadows may complicate the snow detection process [4].

To reduce the effects of the cloud cover, a possible approach is to combine satellite images acquired at different times. In the case of MODIS satellites, Terra and Aqua composite images by Xie et al. [5] show a higher agreement with ground measurements than the daily Terra or Aqua product alone. Xie et al. applied their method to the Colorado Plateau (USA) and northern Xinjiang (China). For the 2003-2004 hydrological year, the daily Terra/Aqua composite images exhibit ~10-15% less annual mean cloud cover and ~1-4% more annual mean snow cover, compared to the

daily Terra or Aqua products. Parajka and Blöschl [6] and Gafurov and Bardoyy [7] added a temporal window to the Terra/Aqua data combination where 1 or to 2 days in the past and 1 or 2 days into the future were analyzed to produce cloud-free classification results.

New methods to improve the detection of snow under forests have been developed in recent years. Vikhamar and Solberg [3] for example applied a linear spectral mixing model for snow, trees and snow-free ground to calculate a fractional snow cover. The model requires a forest cover map and surface area proportions as input; the reflectance values of snow and forest are derived from in situ reflectance measurements. Wang et al [8] introduced the Normalized Difference Forest Snow Index (NDFSI) to distinguish snow-covered from snow-free evergreen coniferous forests. The index is based on the analysis of the spectral signature of both landcover types in the near-infrared (NIR) and shortwave infrared (SWIR) bands.

In mountainous areas, shadows frequently occur on steep slopes when the sun elevation angles are low. Shaded areas generally have lower reflectance than sunny areas. Fahsi et al [9] demonstrated that, due to the effect of topography, satellite image pixels of the same cover type may have different spectral response, whereas pixels of different cover types may have similar spectral characteristics as well, due to the effect of topography. Therefore, many approaches have been proposed to remove, or at least to reduce the effect of topographic shadowing (topographic correction). Shahtahmassebi et al. [10] propose an alternative approach with respect to the conventional technique of cosine correction [11]. They tested two filling functions for estimating the forest areas in mountainous shadows in Landsat images using information about the land cover type of neighboring pixels. The drawback of this technique is the assumption of uniform variability of land cover type throughout the whole image during the interpolation phase. Moreover, this approach meets difficulties in complex landscapes where mixed pixels occur, especially at the forest borders. Another common approach to mitigate shadow effect is the multi-source data fusion. A simple and typical procedure, thereby, is to replace shadowed pixels in an image with the no-shadow pixels of the same area in a corresponding image acquired at different time [12]. Dorren et al. [13] used the multi-source approach by exploiting the digital elevation model (DEM) as additional band, in addition to Landsat TM data, to improve the forest mapping in steep mountainous terrain.

An alternative method for retrieving information about snow characteristic is the application of

distributed, numerical snow models. These models use meteorological observations to simulate the accumulation, storage and melt of a seasonal snow cover. Many types of snow models, suitable for many different application purposes, have been developed, resulting in a wide variety of methodical simulation approaches, from purely empirical to more physically oriented approaches [14]. Even though extensively tested and validated at well-equipped research sites, the complex energy-balance based models can be subject to rather large uncertainties if used in spatially distributed applications. These uncertainties may originate mainly from uncertainties in i) the meteorological input data, ii) snowpack process representations and iii) model parameter sets [15].

Due to the uncertainties in any single data source used to produce a snow cover map (data gaps, nonlinear dynamics or surface heterogeneity that make difficult parameters retrieval, model error and inaccurate processing algorithms), a single "best" remotely-sensed data product or snow cover simulations result to monitor snow cover does not exist [16]. In this context, data fusion methods are a good alternative to overcome these limitations and exploit the strengths of the two different methods. In general, data fusion refers to a formal concept for combining data from different sources in order to provide new products of higher quality (in a broad sense) than the individual input datasets and thus to minimize the difference between true measurements and generated products. The most common use of this approach exploits information derived from spectral reflectivities provided by different terrestrial, airborne or satellite sensors. An example is the work done by Cammalleri et al. [17], who proposed a new approach for evapotranspiration (ET) retrieval. Since satellite-based thermal sensors are characterized by either low spatial resolution and high repeatability or by moderate/high spatial resolution and low frequency, they fused characteristics of both classes of sensors, by exploiting daily MODIS images at 1 km and biweekly Landsat imagery at 30 m, to provide optimal spatiotemporal coverage.

In literature, only few works perform a real data fusion between remote sensing and model products. Painter et al. [18] combined results from the Airborne Snow Observatory (ASO), a coupled scanning lidar system and imaging spectrometer, with a distributed snow model in order to obtain the snow spectral/broadband albedo and the snow water equivalent (SWE). First, spectrometer data have been fused with lidar data and then combined with the snow model simulations in order to obtain a higher-level product, such as the SWE, that is retrieved from the combination of lidar-derived snow height [19, 20] and modeled snow density.

The most common approach for using snow cover simulation together with remote sensing data in a synergistic way is involving the latter in a data assimilation approach or in the calibration phase. Data assimilation techniques have undergone continuous development in the last decades: in weather forecasting, the assimilation of satellite, atmospheric and surface observations into numerical weather prediction (NWP) models has led to an extreme improvement in the forecast skill [21]. The data assimilation techniques have also been developed and implemented in many other applications from hydrology [22, 23] to biogeochemistry [24, 25]. However, dynamical incorporation of remotely sensed data into any model systems is not a trivial task and is computationally expensive. Finger et al. [26] proposed a multiple data set calibration approach to estimate runoff composition using hydrological models with three levels of complexity. The results indicate that all three observational data sets are reproduced adequately by the model, allowing an accurate estimation of the runoff in the three mountain streams.

The objective of this study is to develop a novel fusion approach for snow cover maps generation by using physically based model simulations and remotely sensed products. The fusion aims at improving the snow cover detection in those areas where data sources disagree. As such, we try to overcome the aforementioned limitations of traditional methods and to take advantage of both the specific properties of remote sensing data (such as detailed spatial representation of the estimated parameters), and of the physical basis (independency from atmospheric and shadowing conditions) of the model simulations. The proposed fusion approach is based on a Support Vector Machine (SVM), a machine learning technique which has many important properties relevant for the analysis of remotely sensing data (i.e. high generalization capability, relatively high accuracy, sparsity of the solution and fast processing in the test phase [27]). Moreover, due to the minimization of the structural risk, it is more robust than other pattern recognition techniques in training datasets with a small number of labeled samples.

The innovative aspect of the presented approach is the joint exploitation of remotely sensed data and physical model results, differing from approaches where remote sensing is mainly used for model tuning. In the decision-level fusion process, the snow cover maps and their quality measures are retrieved separately from the two different sources, then they are integrated by SVM to exploit their complementarities and to address their uncertainties.

The final output of this research is a time series of about 1300 fused snow maps obtained by

applying the method to the whole simulation period (October 2012 - July 2016).

The chapter is organized as follows: after introducing the study area and datasets in section 2.2, the method for data fusion is described in section 2.3; results are shown and discussed in section 2.4 and, finally, conclusions on current applicability and indications for future development are drawn in section 2.5.

2.2 Study Area and Dataset

2.2.1 Study Area

The study area of this research is the area including Tyrol (Austria), South Tyrol (Italy) and Trentino (Italy) (Figure 2.1). This area is a good field laboratory because it is well instrumented and this guarantees a high data availability. The climatological conditions are representative for different Alpine zones: precipitation reaches its maximum in the northern and southern prealpine areas (up to 2200 mm/year), whereas the inner region is drier (less than 600 mm/year in the Venosta region) [28]. Permanent snow line ranges between 3200 and 2800 m a.s.l. Most of the rivers in the central and northern part of the region considered have a nivo-glacial regime with maximum discharge during the later summer months, whereas in the southern part of Trentino maximum discharge is usually found during spring or fall with an earlier snowmelt [29]. The region is covered by a dense network of meteorological and snow monitoring stations, operated by the Hydrographical Services of the regional authorities, which provide an excellent validation dataset for the proposed methodologies.

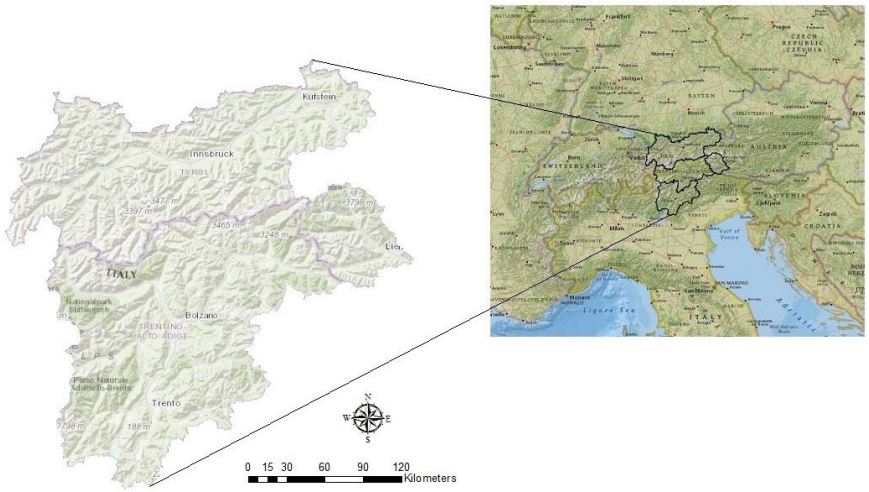

Figure 2.1 Study area: Tyrol (Austria), South Tyrol (Italy) and Trentino (Italy).

2.2.2 Data Description

The fusion method adopted in this study involves the use of snow maps and respective quality measures, originating independently from satellite remote sensing data and from distributed, numerical snow model simulations (Table 2.1).

Table 2.1 The data sources used for the presented SVM fusion approach.

Type	Source		Resolution	Use
Snow cover maps+ quality measure	Remote sensing	EURAC MODIS	Spatial: 250 m Temporal: daily	Input data for data fusion approach
Snow cover maps	Remote sensing	Sentinel 2	Spatial: 10 m (optical bands) Temporal: 12-days	Reference dataset
		Landsat 8	Spatial: 30m Temporal: 16-days	
Snow cover maps+ quality measure	Hydrological Model (AMUNDSEN)		Spatial: 250 m Temporal: daily	Input data for data fusion approach
Ground data (snow height)	Manual measurements		Temporal: weekly/bi-weekly	Dataset for validation

1) Remote Sensing Data: In this study two types of satellite data-derived products have been used:

- MODIS snow maps developed by Eurac Research.
- Sentinel-2 and Landsat-8 RGB images.

The MODIS images, which are freely provided by NASA (http://modis.gsfc.nasa.gov/), have been processed by Eurac Research (Bolzano, Italy) by applying a specific algorithm adapted to mountain areas to obtain snow maps with 250 m spatial resolution [30, 31]. The spatial resolution higher than the standard MODIS product (which has 500 m spatial resolution) can better represent the snow variability in mountainous terrain with very complex topography. The MODIS product derived from the algorithm has been extensively validated by comparison with high resolution SCA maps derived from Landsat 7 ETM+ images, with the NASA standard SCA products MOD10 (MYD10) and with snow height measured by ground stations in selected test sites in Austria, Slovakia, Germany and Italy [31]. Overall accuracies for the different regions between the Eurac SCA product and in situ snow measurements range between 82.4% and 93.7%. The comparison with Landsat shows a mean overall accuracy of around 88.1% in forested areas, whereas in open areas the accuracy reaches 93.6%. The same behavior was found in the comparison with the NASA

product, where the accuracy is 90.2% and decreases to 85.4% in forested areas [32]. In open areas the performances are quite similar, with the advantage that more detailed features are detectable with respect to the 500 m MOD10 (MYD10) maps. All snow maps are provided together with a quality measure, which is based (as explained in the following section relative to the "Data collection") on NDSI. This index, unlike the standard MODIS product, is not used for snow cover area estimation and thus can be used for estimating a quality layer. Further details about the algorithm are explained in [30]. These snow maps, together with the quality measure, have been used as inputs to the fusion process.

The high-resolution images provided by Sentinel-2 and Landsat-8 have been used for extracting reference values used in the training phase of the data fusion classifier. For this purpose, RGB images (with a spatial resolution of 10 and 30 m for Sentinel-2 and Landsat-8, respectively) have been used in a visual interpretation to find suitable reference points. Unlike the MODIS data-derived snow maps that are available daily, Sentinel-2 mission consists of two satellites flying on the same orbit but phased at 180°, which have a revisit frequency of 5 days at the Equator. The temporal resolution of Landsat-8 is instead 16 days. Thus, the selection of dates for extracting the reference points has been constrained by the availability of Sentinel-2 and Landsat-8 images.

2) Snow cover simulations: The evolution of the seasonal snowpack is simulated with the distributed, physically based hydroclimatological model framework "Alpine MUltiscale Numerical Distributed Simulation Engine" (AMUNDSEN) [32]. For every time step and grid cell, a meteorological preprocessor computes all necessary inputs to solve the coupled mass and energy balance of the snowpack and does not require any calibration. The functionality of the model includes sophisticated routines for (i) the regionalization of meteorological input data of various sources [33], (ii) the simulation of short- and longwave radiation including the consideration of shadows and cloudiness [34], (iii) the simulation of the snowpack thermodynamics by means of the factorial snowpack model (FSM) [35] and (iv) the simulation of canopy effects between the trees and snow accumulation on the ground [36, 37]. The presented study focusses on snow cover mapping, hence the model set-up is limited to simulate the snowpack evolution and any processes subsequent to snow melt are neglected (e.g. no simulation of stream flow).

Snow cover simulations are forced with hourly recordings of air temperature, precipitation, global radiation, wind speed and humidity from 325 climate stations in the regions. Furthermore,

AMUNDSEN requires a digital elevation model and maps of land use, soil properties and watershed delineation as inputs in order to distribute input meteorology and parameter sets across the simulation domain.

AMUNDSEN has proven its performance in a variety of applications in most different natural environments [38]. The model ability to predict the seasonal snowpack accumulation and ablation processes in the region was validated at 38 stations with automated snow height recordings. Additionally, 16 stations operated by the hydrographic service of the province Bolzano provide recordings of snow surface temperature, offering the opportunity to validate the mass and energy balance separately. Generally, daily snow height was predicted with acceptable accuracy with a mean Nash-Sutcliffe efficiency (NSE) of 0.68 (ranging from 0.25 to 0.96). However, especially at stations prone to significantly lateral fluxes of blowing snow the observed snow height dynamic could not be reproduced accurately. We explain this primarily by the precipitation under catch corrections, which are well performing at most stations in the region but fail under such extreme conditions. Surface temperature observations could be reproduced with a mean NSE of 0.88, indicating that the model is well capable of solving the energy balance of the snowpack.

3) Ground Data: The ground measurements of snow height, used for validating our results, are collected through different procedures depending on the region. In South Tyrol, measurement campaigns are carried out every day, at 7 a.m., during the whole winter season (from October to May) by private citizens appointed by the public administration. The objective of the survey is to characterize the day from a nivometeorological point of view through a series of significant parameters that can be extracted with simple and fast procedures. The measurements concern parameters such as: snowpack height, fresh snow height, air temperature, weather conditions, cloudiness, visibility and wind activity at high altitude. In particular, for the snow height measurement, a snow measurement stick is inserted vertically into the snowpack until the bottom of the stick rests on the ground; the total height of the snowpack is read on the graduated scale, at the surface of the snowpack.

The measurement sites should be representative of the surrounding area from a nivological point of view, i.e. with regular snow deposition and snowpack evolution by avoiding zones with too fast changes due to the action of wind. The ideal terrain for measurement sites is a flat or slightly sloping terrain ($< 10°$). The data have been provided by the Autonomous Province of Bolzano,

Agency for Civil Protection – Hydrographic Office.

In Trentino, snow height data have been collected from the snow profiles weekly performed by the operators of the Avalanche Office of Province of Trento, alpine guides or avalanche commission members. During the campaign a snow profile is carried out by the operators who analyze and extract parameters which help to identify weaknesses and processes in the snowpack for an avalanche risk evaluation. The extracted parameters are snow height, snow density, grain size and shape for each snowpack layer and air temperature. The data have been provided by Autonomous Province of Trento – Risk prevention service, Forecasting and planning office. The homogeneity of ground measurements is guaranteed by the common protocol used for the snow parameters acquisition, i.e. the AINEVA protocol, used regularly from both South Tyrol and Trentino operators. Moreover, the observation provided by the Province of Bolzano are performed by observers trained and paid by the Province, able to accurately measure snow parameters. Finally, regarding Tyrol, data have been collected from some automatic nivometeorological stations and provided by the Hydrographic Service of Tyrol. Figure 2.2 shows the measurement sites location in the test area.

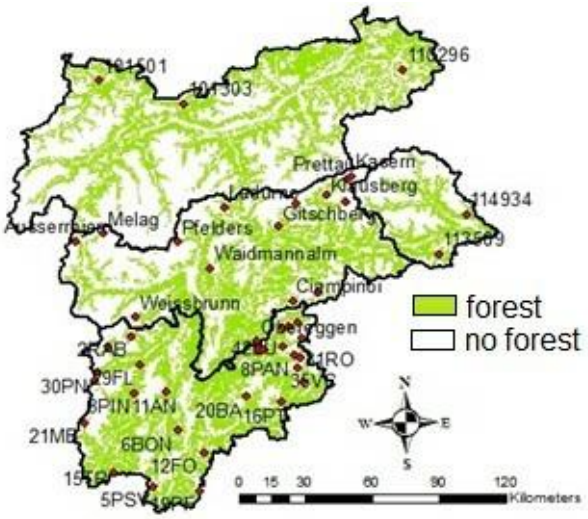

Figure 2.2 Location of measurement sites in the test region. In Tyrol the measurement sites are indicated with numbers, in South Tyrol with the name of location and in Trentino with alphanumeric codes. For reproducibility of the analysis, the stations names have been simply reported as provided by the Province databases.

2.3 Method

The aim of the proposed fusion approach is to improve the snow cover mapping in the areas

where remote sensing product and the simulation results disagree, by taking advantage of both the specific properties of remote sensing data (such as detailed spatial representation of the estimated parameters) and the characteristics typical of physical model (such as solid physical basis and good generalization capabilities). The satellite data-derived snow maps and the model-simulated snow distribution are considered as independent data sets, with individual, spatially varying accuracy. Hereafter, the snow maps derived from the satellite data and those derived from the model simulations will be called MODIS and AMUNDSEN products, respectively.

In the next sections the method for quality measure computation will be explained and later the fusion strategy is presented.

2.3.1 Computation of Quality Measures

The first step of the proposed method consists in the calculation of the quality measures of the two snow cover maps, provided by remote sensing and snow model simulations, respectively. The techniques for the calculation of these quality measures, which are later used as input together with the snow maps for the classifier, are explained in the following sections.

A. Quality Measure for the MODIS Product

The quality measure for the MODIS product is based on NDSI (Normalized Difference Snow Index). It is computed only for the two classes of interest, i.e. snow and no snow, whereas for all the other classes it is not considered. NDSI is an index related to the presence of snow in a pixel and is based on the different reflectivity values of the surface between a band in the visible and one in the short-wavelength infrared (or near-infrared) parts of the spectrum. Since snow is highly reflective in the visible bands and highly absorptive in the short-wavelength infrared (or near-infrared), this index allows a good distinction between snow and clouds, most of which have a high reflectivity in both sections of the spectrum.

$$NDSI = \frac{VIS - SWIR}{VIS + SWIR} \qquad \rightarrow \quad -1 \leq NDSI \leq 1 \tag{2.1}$$

One of the main differences between the Eurac and NASA algorithms in the detection of snow is the use of NDSI index (bands for this index are at 500 m). The NASA algorithm adopts a combined use of NDVI and NDSI, which improves the snow detection in forested areas. Vice versa, the Eurac algorithm uses only the NDVI and B1 (the red band) to preserve the resolution of 250 m. This allows us to use the NDSI for assessing the quality of the snow classification in each

pixel. For the snow and no snow classes, the quality measure (U) can vary between 0 (low quality) and 1 (high quality) and is computed as follows:

$$\begin{cases} U = \frac{1+NDSI}{2} \ \ for \ \ the \ \ class \ \ snow \\ U = \frac{1-NDSI}{2} \ \ for \ \ the \ \ class \ \ no \ \ snow \end{cases} \tag{2.2}$$

B. Quality Measure for the AMUNDSEN Product

Snow maps derived from physically based model simulations comprise a large number of state variables for the snowpack in each pixel. First, however, we only use the binary information of snow presence (i.e., whether snow is present in a certain pixel, or not) for the processing of the snow maps. These are derived from simulated snow water equivalent (SWE):

$$\begin{cases} if \quad SWE < th \rightarrow x = 0(snowfree) \\ if \quad SWE > th \rightarrow x = 1(snowcovered) \end{cases} \tag{2.3}$$

Where $th = 5mm$ is a threshold that accounts for the scale discrepancy between a point location and the pixel dimension. The resulting map is a binary image with values being 0 (no snow) or 1 (snow). The quality measure for the AMUNDSEN product is computed in two different ways for snow-covered and snow-free pixels. The quality measure for snow covered pixels is very simplistic and merely links the uncertainty information of the pixels to the magnitude of predicted SWE value. The assumption behind this approach is that, due to the cumulative nature of the snowpack, large errors (in total snow mass) are needed for a misclassification of snow-covered pixels when a deep snowpack is predicted, whereas smaller errors in snow mass suffice for a misclassification when a shallow snowpack is predicted. The certainty of the classification is assumed to increase with increasing snow mass in a hyperbolic manner. For deep snowpack far enough away from the snow cover threshold, an increase in snow mass is assumed to not further increase the certainty of the snow cover classification. Starting from these definitions, the quality measure for snow-covered pixels is calculated considering that a higher quality in the snow map is associated with a larger snow mass and, thus, a larger SWE value:

$$U = -tanh\left(\frac{SWE}{SWE_t}\right) + 1 \tag{2.4}$$

with SWE_t = 100 mm and SWE the snow water equivalent value of the pixel.

For snow-free pixels, the quality of the snow cover classification is assumed to increase over time until a threshold is reached: the higher the number of previous snow-free days, the higher is the probability that the pixel is snow-free. This quality approximation relies on the time distance to the predicted melt out of a cell. Errors in the simulation of accumulation and ablation processes will translate to error in melt-out timing. With an increasing time distance to the predicted melt-out, larger model errors would be required for a misclassification. In order to maintain a reasonable scaling of the quality measure, the growth of the certainty is limited. Otherwise the certainty of a no-snow classification in autumn would be unrealistic high compared to one just after the snow ablation in spring. The respective quality measure hence is:

$$U = -tanh\left(\frac{\Delta t}{\Delta t_t}\right) + 1 \tag{2.5}$$

with Δt_t= 10 days and Δt the number of no-snow days.

The quality measures as defined here can be considered as proxy quantities of the model accuracies for the detection of snow and snow-free pixels.

2.3.2 Data Fusion Strategy

The fusion strategy involves the disagreement points through the use of SVM technique and exploiting as input features the snow maps from MODIS and from model simulations, as well as the relative quality measures. The procedure is summarized in Figure 2.3. It includes three phases:

- *Data collection*: MODIS and AMUNDSEN snow maps, together with their quality measures are prepared to be then used as inputs to the data fusion process. The MODIS snow maps considered are of binary type (snow/no-snow), with other classes (clouds, water and no-data) masked. Simultaneously, high-resolution RGB images from Sentinel-2 and Landsat-8 acquired during the period October 2012 - July 2016 are selected and collected.
- *Data selection and SVM training*: the input data and the corresponding reference data has been selected for the estimation of the SVM model parameters during the training phase. Since the dataset for performance evaluation in the testing phase shall be independent, two datasets (one for training and one for testing the classifier) have been collected: the first step was the selection of some dates in different periods of the year, in order to consider the seasonal variability of snow coverage. Then, on these randomly selected dates, the pixels locations have been selected and extracted from the snow maps. The corresponding reference dataset with the true labels has been

extracted through a visual interpretation of S2 and L8 images and the whole data set created. This resulting dataset was then randomly divided into a training dataset (80%, about 720 points) and a test dataset (20%, about 180 points).

- *Maps generation and performance evaluation*: finally, the classifier has been tested on the independent dataset (test dataset) to evaluate the performance.

2.3.3 Support Vector Machine Approach

SVMs are supervised learning models for classification and regression procedures. They can address both linear and non-linear relations and work well for many practical applications. SVMs have been proved to have a higher classification accuracy than other widely used pattern recognition techniques, such as the maximum likelihood and the multilayer perceptron neural network classifiers [27]. Moreover, SVMs appear to be especially advantageous when only few training samples are available [27].

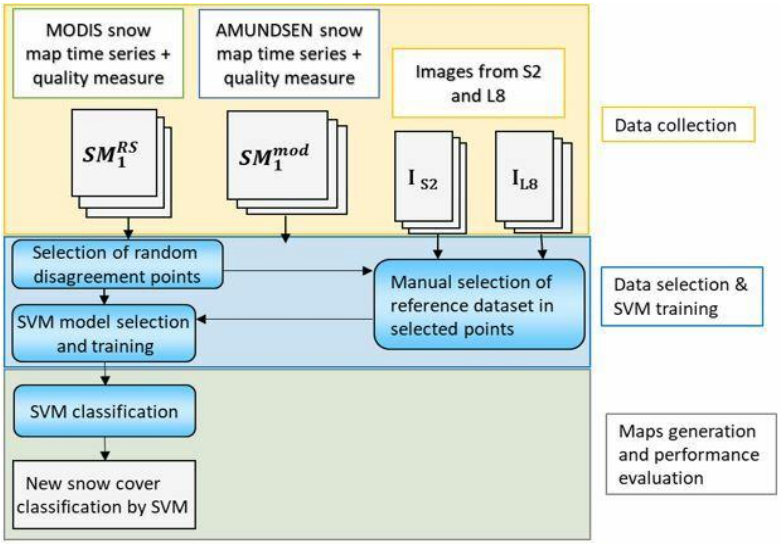

Figure 2.3 Data fusion strategy flowchart.

An important property of SVM models is that they do not require the knowledge of the statistical distributions of classes to carry out the classification, as they exploit the concept of margin maximization [27]. The growing interest in SVMs is mainly related to a) the higher effectiveness with respect to traditional classifiers, resulting in high classification accuracies and very good generalization capabilities; b) the relatively low effort required for architecture design (only few control parameters); c) applicability to linearly constrained quadratic optimization problems.

These described properties, together with a strong ability to deal with remotely sensed data [27], make SVM the suitable approach to address the presented classification problem. Further technical details on SVM mathematical formulation can be found in [27].

2.3.4 Validation Strategy

The validation of the data fusion method has been conducted at two different levels: the first one exploits the data from high-resolution remotely sensed images, whereas the second considers the ground data collected by measurement sites located throughout the test area.

In order to compare snow height ground measurements with the binary maps (snow or no-snow) obtained from the fusion approach, as well as those derived by both MODIS and AMUNDSEN, a threshold needs to be selected for discriminating between snow and no-snow. Two different threshold values on snow height, i.e. 5 cm and 10 cm, were tested to assess the impact of this choice on results. As shown by Thyrel et al. [39], the agreement with ground data seems to improve for lower snow height threshold. Other snow height threshold values have been used in literature, such as 1 cm [40], 2.54 cm [41] and [42], 2 cm [31]. The final choice of setting 5 cm as threshold is related to the observation that a shallow snow layer can rapidly melt during the day and might thus not reveal the actual snow status at the time of the satellite acquisition. Vice versa, a too low threshold may not be representative of the surrounding area.

The validation with ground data has been carried out by using points that are, except some cases, snow-covered, since the snow height measurements are performed in winter season and manual observations for snow-free conditions are lacking.

2.4 Results and Discussion

In this section, the validation of the proposed method and the results derived from the analysis of time series are presented. Hereafter, in the validation with ground data, the term “points” indicates the single ground data at a specific date and in a specific observation site.

2.4.1 Validation with High-Resolution Images

By validating the fusion data method on the test dataset (180 points), the overall accuracy reaches 89%, with respect to 40% (MODIS) and 60% (AMUNDSEN). Figure 2.4 shows the confusion matrices and some statistical indices for the test points. The overall accuracy (OA, in %) is defined as the sum of snow/snow agreement and no-snow/no-snow agreement divided by the total number

of observations available. The True Positive Rate (TPR) indicates the percentage of snow samples that are correctly identified and the True Negative Rate (TNR) represents the proportion of no-snow points that are correctly identified.

MODIS	No_snow	Snow
No_snow	1	38
Snow	69	71

a) OA = 0.40
TPR = 0.65
TNR = 0.01

AMUNDSEN	No_snow	Snow
No_snow	63	64
Snow	7	45

b) OA = 0.60
TPR = 0.41
TNR = 0.90

Fused product	No_snow	Snow
No_snow	61	11
Snow	9	98

c) OA = 0.89
TPR = 0.90
TNR = 0.87

Figure 2.4 Confusion matrix relative to a) MODIS product; b) AMUNDSEN product; c) fused product. OA= Overall Accuracy; TPR= True Positive Rate; TNR= True Negative Rate.

From the confusion matrices, it results that, on average, the MODIS product tends to overestimate the snow coverage while the AMUNDSEN product seems to underestimate it. The new fused product balances these behaviors by improving the overall accuracy with the high-resolution images.

Table 2.2 shows the agreement with reference dataset divided per area, as fraction of total points matching with the selected points in high-resolution images. The three products, i.e. MODIS, AMUNDSEN and fused products, are reported separately for the three areas. The column "Points per area" indicates the total number of points selected for South Tyrol, Trentino and Tyrol, respectively. Results indicate that AMUNDSEN shows a higher agreement with selected reference points in South Tyrol with respect to the other areas; vice versa, MODIS seems to perform better in Tyrol than in South Tyrol and in Trentino.

Table 2.2 Agreement (%) for each region (South Tyrol, Trentino and Tyrol) with validation points extracted from high-resolution images.

Area	MODIS	AMUNDSEN	FUSED	Points per area
SOUTH TYROL	% of agreement	% of agreement	% of agreement	
	39%	69%	88%	60
TRENTINO	% of agreement	% of agreement	% of agreement	
	35%	56%	90%	60
TYROL	% of agreement	% of agreement	% of agreement	
	47%	55%	88%	59

2.4.2 Validation with Ground Data

For the second type of validation analysis we compare the data fusion product to ground data. The general results are shown in Figure 2.5, where the confusion matrices for the three products are reported. Fig. 5 shows the same behavior found in Table 2.2: MODIS has the best performances in Tyrol, whereas AMUNDSEN accuracy is higher for South Tyrolean territory.

By using a threshold value of 5 cm, the agreement percentages for each observation are calculated for all three snow cover maps (Table 2.3). The column "Points per station" indicates the number of points available for each station. The agreement percentages have been evaluated by considering the number of points, which can vary considerably among the observation sites. This could lead to different performances: in most of Trentino site, for example, the number of available measurements may be very low (in the worst case only 2 measurements are available) so the percentages may also be very high. If the measurement site is located in a pixel classified as "cloud" in MODIS product, it is excluded from the analysis.

SOUTH TYROL

MODIS	No_snow	Snow
No_snow	70	230
Snow	19	49

$O.A. = 0.32$
$T.P.R. = 0.18$
$T.N.R. = 0.79$

AMUNDSEN	No_snow	Snow
No_snow	19	49
Snow	70	230

$O.A. = 0.68$
$T.P.R. = 0.82$
$T.N.R. = 0.21$

Fused_product	No_snow	Snow
No_snow	39	43
Snow	50	236

$O.A. = 0.75$
$T.P.R. = 0.85$
$T.N.R. = 0.44$

TRENTINO

MODIS	No_snow	Snow
No_snow	0	20
Snow	0	20

$O.A. = 0.50$
$T.P.R. = 0.5$
$T.N.R. = /$

AMUNDSEN	No_snow	Snow
No_snow	0	20
Snow	0	20

$O.A. = 0.50$
$T.P.R. = 0.5$
$T.N.R. = /$

Fused_product	No_snow	Snow
No_snow	0	0
Snow	0	40

$O.A. = 1$
$T.P.R. = 1$
$T.N.R. = /$

TYROL

MODIS	No_snow	Snow
No_snow	19	22
Snow	28	64

$O.A. = 0.62$
$T.P.R. = 0.74$
$T.N.R. = 0.40$

AMUNDSEN	No_snow	Snow
No_snow	28	64
Snow	19	22

$O.A. = 0.38$
$T.P.R. = 0.26$
$T.N.R. = 0.60$

Fused_product	No_snow	Snow
No_snow	30	20
Snow	17	66

$O.A. = 0.72$
$T.P.R. = 0.77$
$T.N.R. = 0.64$

Figure 2.5 Validation with ground data. Confusion matrix relative to a) MODIS product; b) AMUNDSEN product; c) fused product. OA= Overall Accuracy; TPR= True Positive Rate; TNR= True Negative Rate.

The average agreement between the fused product and observations is 96% with respect to 90% (MODIS) and to 92% (AMUNDSEN) (Table 2.3). In this type of validation, ground data can involve both pixels where MODIS and AMUNDSEN disagree and where they agree. Since the fusion process is applied only on disagreement pixels, in order to assess how the method works, some statistics have been computed by considering only these points, i.e. where AMUNDSEN indicates “snow” and MODIS “no-snow” (or vice versa).

Figure 2.6 presents the results about the validation in South Tyrol: for each measurement site the two columns indicate on the left the AMUNDSEN and MODIS behavior and on the right the fused product behavior. White bars represent the number of total available points (dates) for each measurement site; light and dark blue respectively show the samples where MODIS is wrong and AMUNDSEN is correct with respect to the ground data and vice versa. Above these two bars, the cyan bars indicate the points where both model and satellite data are wrong. The remaining points above the cyan bars are the points where AMUNDSEN and MODIS agree and give correct classification. In these points, as well as in the points of cyan bars, the fusion does not work because the model and the satellite products agree. For each measuring site, the sum of pink bars gives an idea of the improvement provided by the presented approach with respect to the single sources (MODIS and the AMUNDSEN snow maps), by showing the number of disagreement points that are correctly classified after fusion approach.

By averaging the results on all the stations, one can observe that 76% of the disagreement points are correctly classified by the SVM classifier. Moreover, in about 73% of these correctly classified points, the fused product follows AMUNDSEN, while in the remaining 27% it coincides with MODIS. This behavior could be explained by considering that about 68% of considered disagreement points in South Tyrol correspond to measurement sites located in pixels classified as forest. This means that these sites are probably located near the forest and are representative of such type of land cover. The well-known problem of MODIS in detecting snow in these areas could lead the fusion method to be more confident with AMUNDSEN product. Moreover, approximately 38% of the remaining disagreement points in open areas correspond to north-facing sites. In this case the lower reliability of MODIS product could be ascribed to the underestimation in low-light conditions which frequently happen during wintertime as reported in [31].

Table 2.3 Agreement (%) for each measuring station in South Tyrol, Trentino and Tyrol. In Trentino, only 12 out of 28 stations have at least one point of disagreement between MODIS and AMUNDSEN. In the table, only these stations are mentioned, by omitting all those where AMUNDSEN and MODIS are always in agreement.

Station name	MODIS	Model	FUSED	Points per station
SOUTH TYROL	% of agreement	% of agreement	% of agreement	
Ausserrojen	96%	94%	95%	182
Ciampinoi	98%	99%	99%	241
Gitschberg	97%	86%	99%	127
Kasern	77%	89%	88%	191
Klausberg	49%	99%	87%	168
Ladurns	88%	100%	94%	172
Melag	97%	93%	97%	212
Obereggen	90%	94%	97%	230
Pfelders	96%	98%	98%	178
Piz la lla	74%	98%	96%	106
Prettau	79%	80%	97%	98
Rein in Taufers	84%	82%	85%	207
Stausee Neves	83%	83%	83%	201
Waidmannalm	92%	95%	92%	132
Weissbrunn	93%	97%	96%	287
TRENTINO				
10PM	94%	100%	100%	18
11AN	80%	100%	100%	5
12FO	100%	83%	100%	12
14PO	78%	100%	100%	9
16PT	100%	92%	100%	13
18SB	100%	33%	100%	12
19PF	67%	93%	100%	15
20BA	67%	100%	100%	6
29FL	73%	100%	100%	30
30PN	100%	83%	100%	42
5PSV	100%	50%	100%	2
8PAN	50%	100%	100%	2
TYROL				
St.101303	67%	66%	67%	312
St.101501	79%	77%	80%	352
St.110296	88%	90%	92%	329
St.113589	75%	65%	75%	411
St.114934	93%	92%	93%	322
	Average agreement	Average agreement	Average agreement	Total points
	90%	92%	96%	4625

The same procedure has been applied to data collected in Trentino and in Tyrol. For the data from the Trentino region, in 16 out of 28 measurement sites MODIS and AMUNDSEN are always in agreement and both accurately reproduce the ground observations. These sites are not shown in the histogram. As shown in Figure 2.7, the points where MODIS and AMUNDSEN provide different results are correctly classified by the fusion procedure. Moreover, the points where MODIS and AMUNDSEN provide the same results are all correctly classified with respect to the

observations. Hence, for the Trentino area and the period considered, the fused product is able to correct all errors present in the two snow cover maps. In particular, this total agreement of fused product with ground data is symmetrically distributed between the two sources of information: in 50% of cases the fused product matches the MODIS product and in the remaining 50% it coincides with AMUNDSEN.

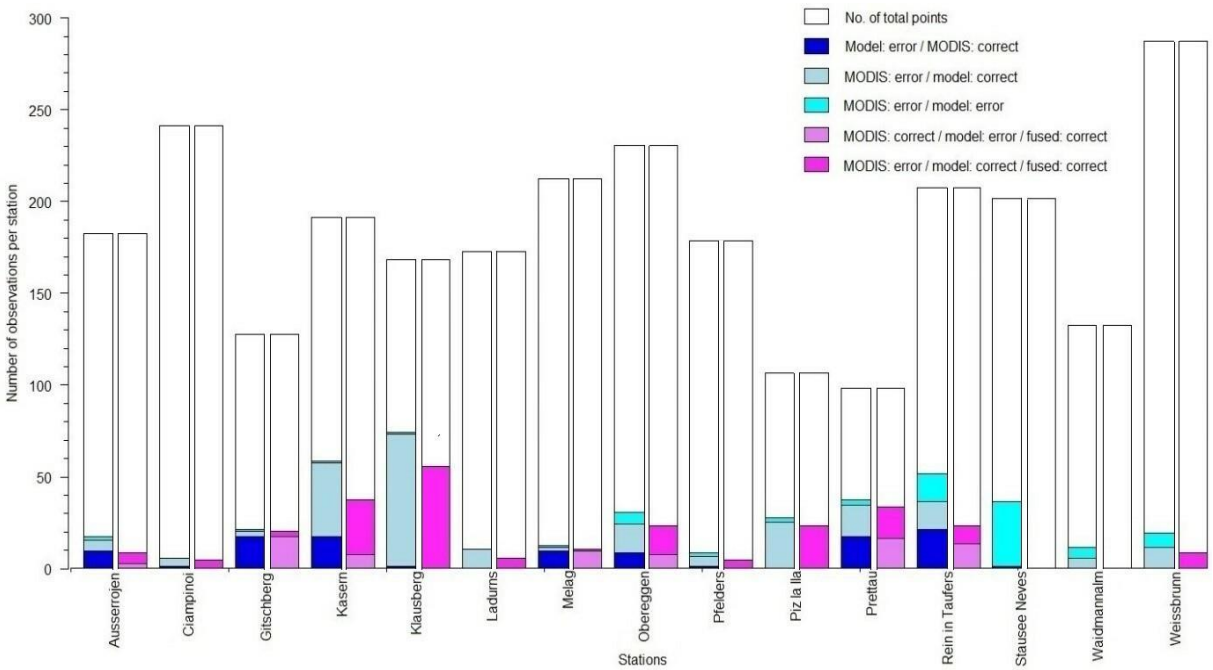

Figure 2.6 Validation with ground data from measurement sites in South Tyrol: white bars indicate the total points for each station; dark blue marks points where MODIS is correct and AMUNDSEN is wrong with respect to ground data; light blue shows points where MODIS is wrong and AMUNDSEN is correct; cyan bars represent points where both MODIS and AMUNDSEN are wrong; light pink are points where the fused product is correct and equal to MODIS; dark pink are points where the fused product is correct and equal to AMUNDSEN.

For the Tyrol region, 70% of the disagreement points have been correctly classified by the SVM classifier (Figure 2.8). In about 74% of these correctly classified points, the fused product matches the MODIS product, whereas in the remaining 26% of cases it agrees with the modelled value. In this case, the behavior of fused product seems to be opposite to the one found in South Tyrol: fusion method seems to be more confident with MODIS product than with AMUNDSEN one. This could be ascribed to a lower number of disagreement points in forested areas. Unlike what happens in South Tyrol, in fact, in Tyrol most of the selected disagreement points (about 58%) is associated to measurements sites located in pixels classified as open areas.

In order to understand if the differences with ground data are due to an underestimation or an overestimation of snow, the histogram in Figure 2.9 shows, by considering all stations, the number of times where each snow cover map disagrees with the ground data, per month. From the histogram, the MODIS product underestimates the presence of snow in most cases, especially in December and January (value of the yellow bars). This behavior is in line with accuracy variation reported for standard MODIS product by NASA [42] and [43].

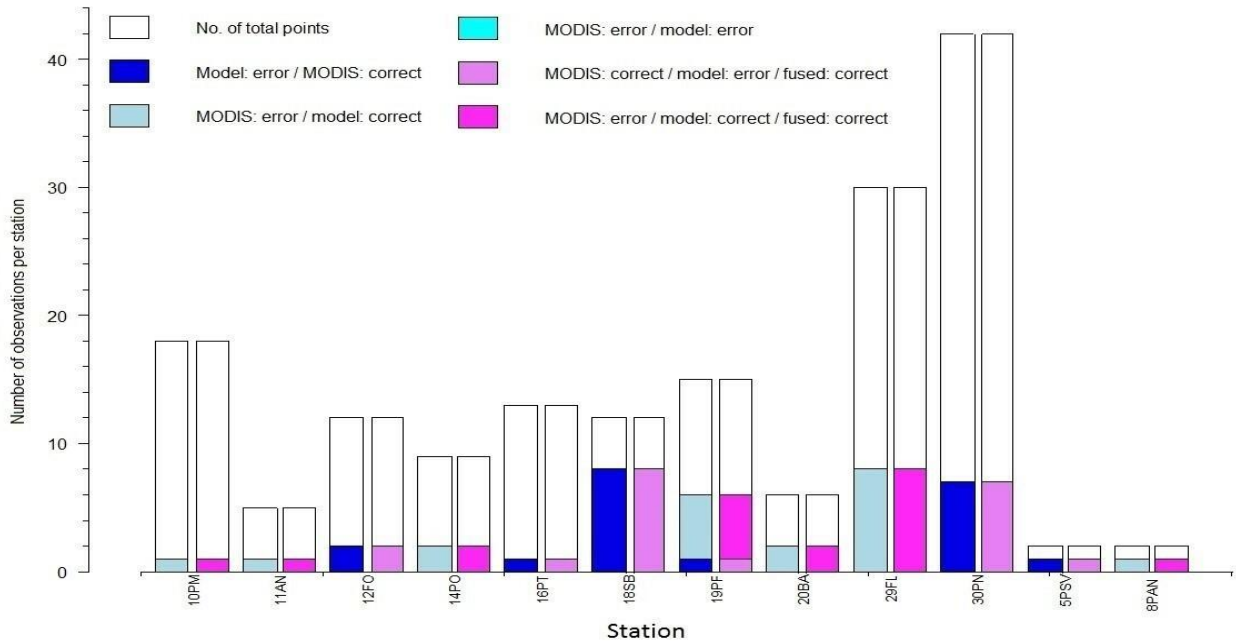

Figure 2.7 Validation with ground data from measurement sites in Trentino.

In the fused product, the smallest errors occur in February when the amount of snow is large. AMUNDSEN seems to produce the largest errors in spring, due to the accumulative nature of the errors in the computation of the accumulation, redistribution and melt processes [15]. This effect might lead to a higher uncertainty in snow detection in this period.

The decision fusion classifier has been applied to about 1300 maps in the considered simulation period to generate the resulting time series of fused snow maps. The accuracy of snow detection from satellite data is, in general, significantly higher in open areas than in forested areas. Indeed, trees increase the complexity of the scene by masking the snow on the ground and altering the radiance measured by the MODIS satellite [3, 8]. Since elevation also strongly affects quantity and distribution patterns of precipitation and snow, we analyzed the snow cover area (SCA) for

different land use (i.e. forest and open areas) and elevation bands. Figure 2.10 shows the SCA (i.e. the total number of snow-covered pixels divided by the total number of snow-covered and snow-free pixels) behavior in time of the three snow products for the entire period (October 2012 - July 2016).

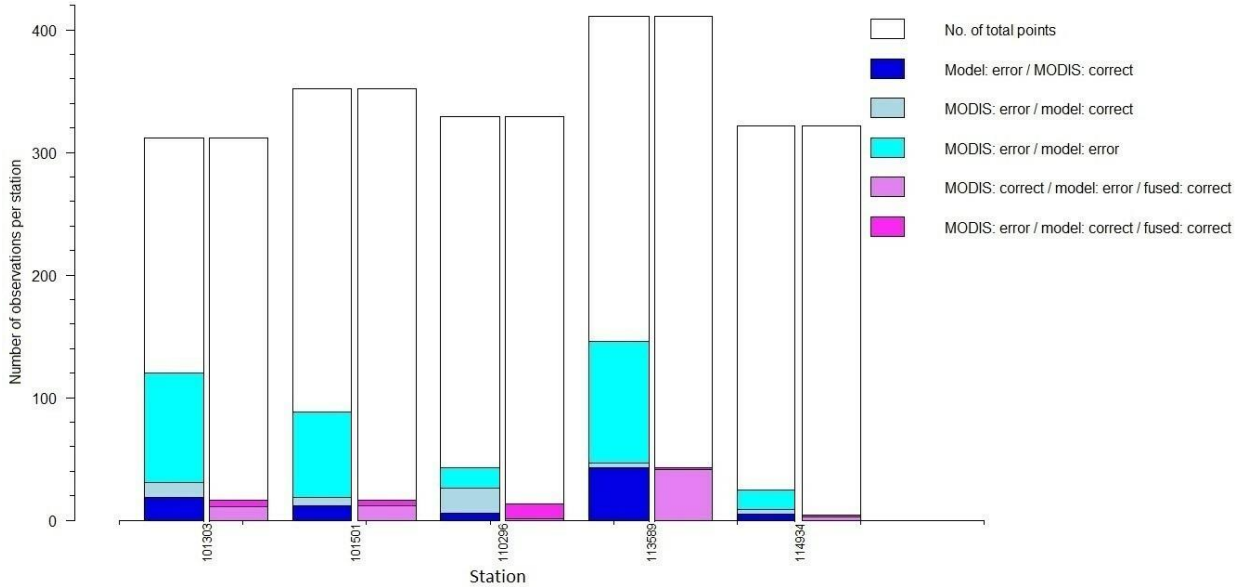

Figure 2.8 Validation with ground data from measurement sites in Tyrol.

The underestimation of snow in forest as found in the MODIS product seems to be solved in the fused product, which follows the accurate simulation of the forest snow cover in AMUNDSEN: in forest, for all elevation bands, the fused product accurately reproduces the SCA peaks of winter season, also when there is a sharp underestimation in the MODIS product. In open areas, the behavior of AMUNDSEN and MODIS products is similar and the fused snow cover maps well reproduce the seasonal variability of the winter peaks and summer minima.

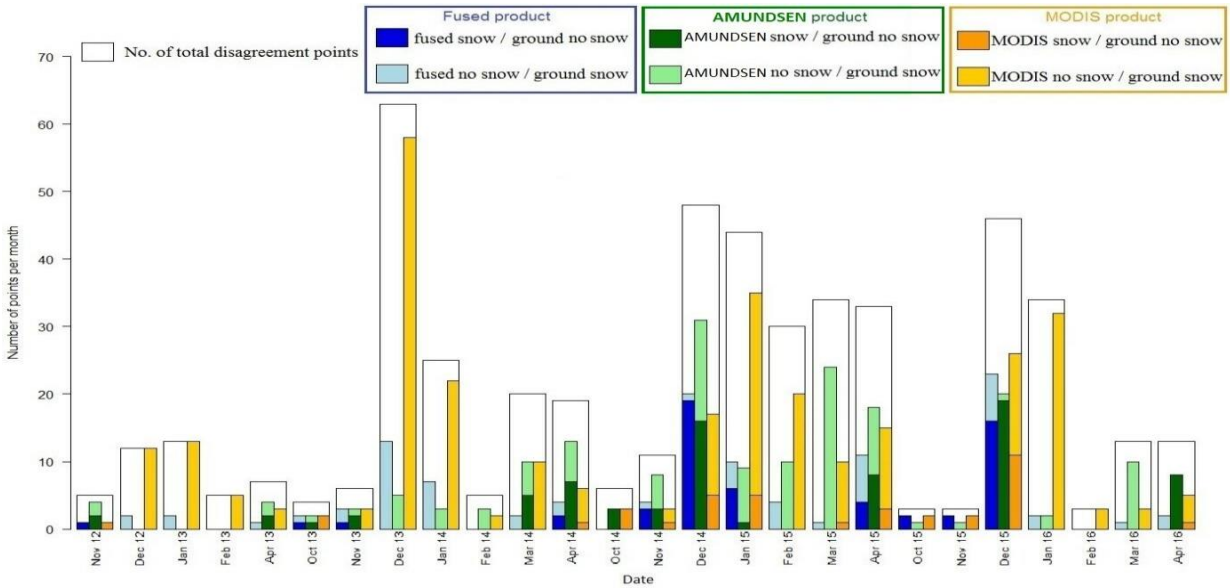

Figure 2.9 Monthly difference between snow products and observations.

2.4.3 Cloud Effect Correction

A further improvement of the final data fusion snow maps can be achieved by a cloud correction approach applied to the regions where the MODIS snow maps are incomplete, due to the cloud presence. Hence, the final product consists of a map having the pixel value obtained by the fusion method in those pixels where two original snow cover maps (MODIS and AMUNDSEN) disagree and the AMUNDSEN pixel value where MODIS indicates “cloud”.

Figure 2.11 shows two examples of snow maps at the end (on April 17th, 2014) of the winter season 2013-2014 and at the start (on November 23th, 2014) of winter season 2014-2015. The right figures show the images with clouds, whereas the left ones show the corrected images, as above explained. The colors highlight the different behaviors of the fused product: green and white represent the pixels where AMUNDSEN and MODIS agree and, therefore, where the data fusion approach is not applied and consequently has the same value of the two single sources; the dark and light blue are the pixels where the fused snow map has the same value of the MODIS map;

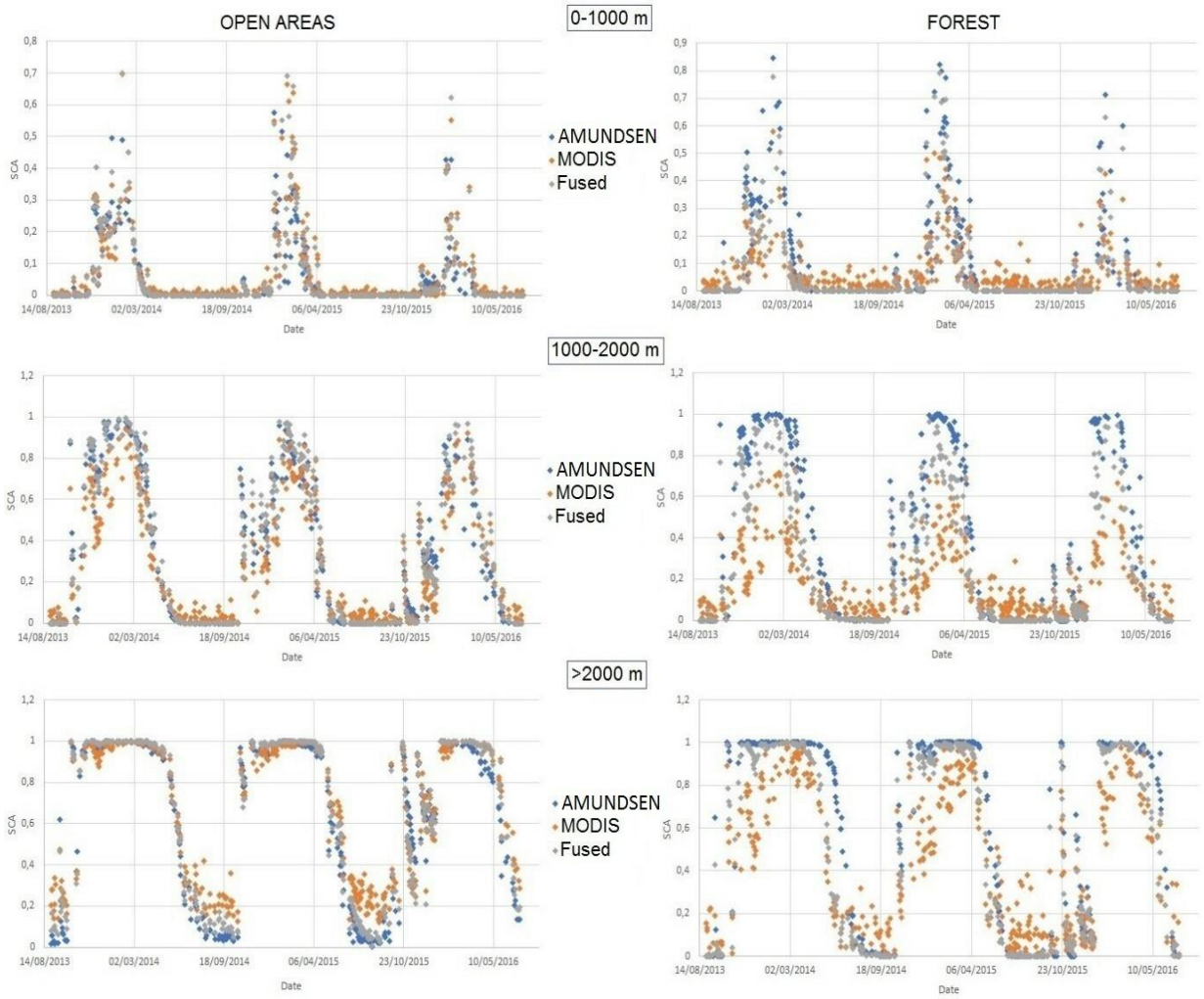

Figure 2.10 Snow cover area (SCA) behavior in time for AMUNDSEN (blue), MODIS (orange) and the fused (grey) products, respectively. The analysis is carried out in open (left) and forested (right) areas for different elevation bands. The red circles in forested area highlight the winter snow SCA peaks where the MODIS underestimation is more evident.

finally, the dark and light pink indicate the pixel where the fused snow map follows the behavior of the AMUNDSEN simulation. In the winter image, most of the pixels classified as snow by AMUNDSEN as well as by the fused product (dark pink) are located on the northern exposure. This behavior may be ascribed to the MODIS underestimation in low-light conditions which frequently happen during wintertime as reported in [31]. The cyan color indicates that the fused product follows the MODIS product behavior in detecting the snow absence. Most of these areas are located in forest: as highlighted in Figure 2.9, in forested area the fused product results follow the AMUNDSEN behavior because of the well-known limitation of optical satellites to detect snow under the canopy. In this context, it is worthwhile mentioning that snow detection in forest

is very complex and it depends on many factors such as the location of the forest (north/south), the density of the forest, the type of the forest (broadleaf or conifer). It is found that normally MODIS product tends to underestimate the snow cover in forested areas [44, 45]. At the same time, at the beginning and end of the season, it can be supposed that AMUNDSEN model may simulate low values of SWE in these transient periods, so that SVM classifier can give in some cases, as shown in Figure 2.10, the priority to the MODIS product. This behavior highlights the importance in the selection of the feature to be used in the data fusion approach, both the inputs and the related quality measures. These measures shall provide both an evaluation of the quality of the inputs and try as well to cover the different spatial and temporal variability, which the snow has in mountain areas. As a future step, different quality measures will be evaluated in order to understand their impact on the final products and how they can tackle the heterogeneity of snow cover in complex terrain.

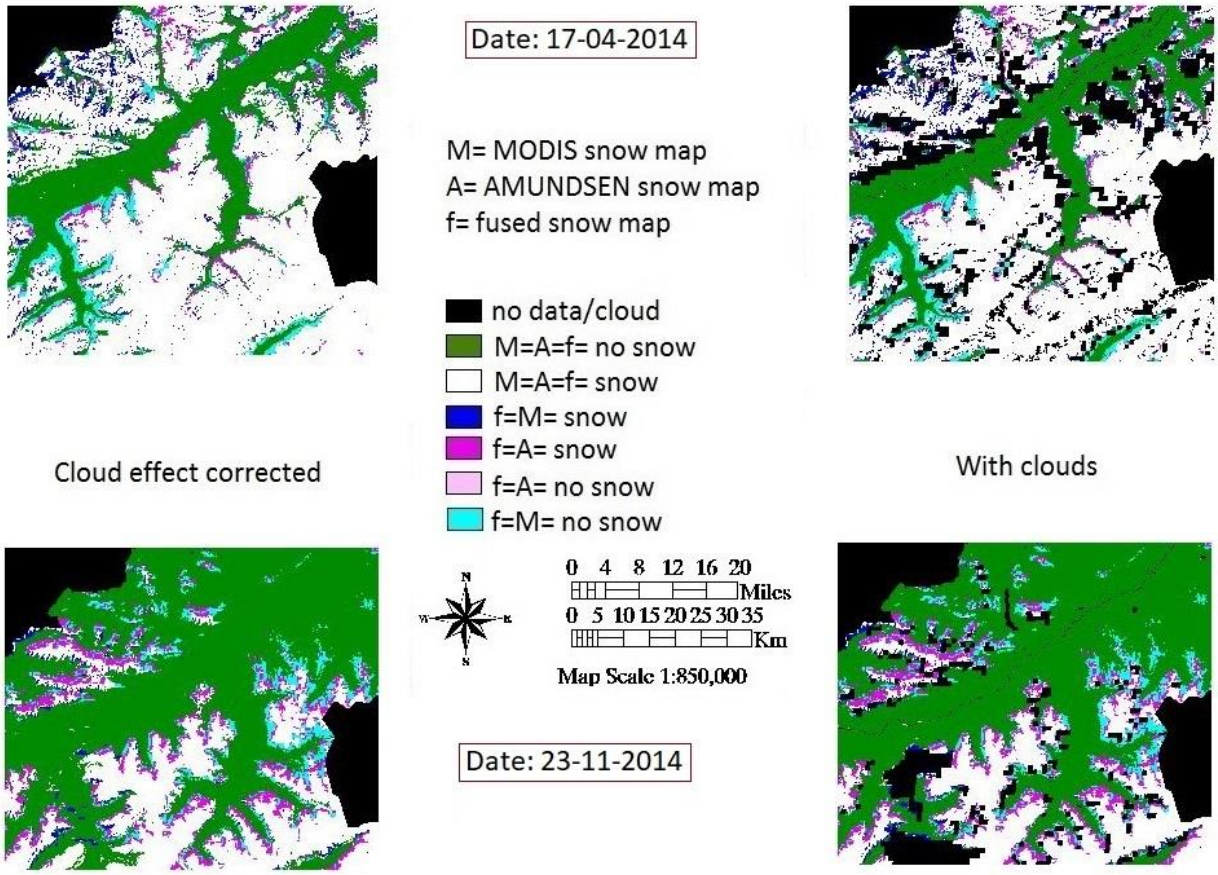

Figure 2.11 Example of fused product, with clouds (right) and with cloud effect correction (left). The colors show the different behavior of the fused product.

2.5 Conclusion and Outlook

In this chaper we present a method to overcome the limitations of existing remote sensing and modelling techniques for snow cover mapping. The data fusion approach developed takes advantage of the specific properties of the remote sensing data (such as independency from meteorological observations and a spatialized representation of snow cover) and those of a physical snow model (such as solid physical basis and the independency from cloud coverage). The objective of the data fusion is solving the ambiguity of disagreement points, i.e. those pixels where the snow model indicates snow presence and the satellite product snow absence (or vice versa).

The agreement points (where both MODIS and AMUNDSEN say “snow” or “no snow”) cannot be improved from this fusion approach, but an analysis on these situations has been carried out in order to understand the behavior of products with respect to the ground data. In South Tyrol and Trentino less than 1% of agreement pixels is wrongly classified and they occur especially in situation of shallow snowpack (less than 5 cm, which is the threshold imposed on the ground snow height values to obtain binary values) when MODIS and AMUNDSEN say snow and the ground measurement registers no snow.

The case of Tyrol is slightly different because the ground data come from automatic nivometeorological stations. The snow height measurements are continuous and are also collected in the summer period. However, measurements in summer period are critical: the grass grows at stations and is measured by the ultrasonic sensor by causing a supposed increase in the height of the snowpack. If it snows in summer, the grass is flattened. A snowfall causes, thus, a sudden drop in the measured height. Most of wrong classification of agreement pixels in Tyrol area occur in summer period, when measurements from automatic stations in summer are not reliable. Some cases occur at the beginning of the winter season with first snowfalls and shallow snowpack.

The fusion is carried out by means of an SVM, a pattern recognition technique often adopted in the field of bio-physical parameter retrieval for its capability to handle complex and non-linear problems and to manage different kinds of inputs. The results show that the presented data fusion method is able to produce a more accurate snow cover map than could be provided by remote sensing or snow modelling alone. The validation of the fused snow cover product was performed by using ground data derived from measurements carried out in open sites in the test region, resulting in a very good agreement. The average accuracy of 90% (MODIS) and 92%

(AMUNDSEN) is increased to 96% in the fused product. In future work we will extend the validation to forested sites.

Moreover, it is worth saying that the analysis of the uncertainties shows that there are very few cases in which they are similar (difference between AMUNDSEN and MODIS uncertainties lower than 10%). These cases represent approximately the 3% and usually occur in December and in April. SVR behavior in these cases strongly depend on period of the year: in spring season, in average, SVR seems to be more confident with MODIS product even in the cases when AMUNDSEN uncertainty is slightly lower than in MODIS. Vice versa, when the case of similar uncertainties occurs in December, the SVR follows the AMUNDSEN behavior. This confirms that SVR approach is more than a simple classification based on uncertainties (choice of lower uncertainty) and that probably the regressor catches a seasonal trend from the training dataset that leads it to choose the most reliable product also depending on the period and not only on uncertainty values.

A further improvement was carried out by applying a cloud clearing that makes use of the snow model result in areas that are classified as "cloud" in the MODIS product. This procedure allows to obtain a final snow cover map with coverage on the entire area. As further development, in addition to the two snow cover maps of satellite and model origin and their quality measures, other input features can be tested for the fusion procedure, e.g. the sun incident angle (to account for different illumination conditions) or the percentage of forest coverage in the pixel (to account the quality of remote sensing product that is, as mentioned, affected by this parameter).

These results are promising if compared to what already exists in the literature: Parajka and Blöschl [6] presented a method for improving the existing MODIS daily snow products by reducing cloud coverage. They improve the combined Aqua and Terra snow cover product by using first a spatial filter and then a temporal filter for reducing the cloud covered pixels. Their approach allows a reduction in cloud coverage of more than 95%, with an overall annual accuracy of more than 92%, based on a comparison with ground snow height measurements. The same overall accuracy evaluation was applied by Pu et al. [46] who tested the MODIS 8-day composite snow product against ground snow height data on the Tibet Plateau. They reported an average of 90% overall accuracy in the period 2000–2003. Şorman et al. [47] compared daily snow cover maps obtained from MODIS images with ground observations in mountainous terrain of Turkey

for the winter season of 2002–2003 and 2003–2004 during the accumulation and ablation periods of snow. The comparison shows good agreement with overall accuracies in between 62 to 82 % considering a 2-day shift during cloudy days.

Results obtained in this work encourage further research on the development of a general method being able to provide improved snow cover maps, transferable even in other regions or to exploit this fusion method for retrieving other snow parameters, such as snow water equivalent (SWE). Satellite products at high-medium resolution cannot deliver such variable and can contribute only with some auxiliary data. The objective of the fusion, in this case, will be the improvement of the reliability of physical model product by exploiting remotely sensed products as proxy information.

References

[1] P. Singh e V. Singh, Snow and Glacier Hydrology, Dortrecht, Boston: Kluwer Academic Publishers, 2001, p. 741.

[2] D. Hall e G.A. Riggs, «Accuracy assessment of the MODIS snow products,» *Hydrological Processes: An International Journal,* vol. 21, n. 12, pp. 1534-1547, 2007.

[3] D. Vikhamar e R. Solberg, «Subpixel mapping of snow cover in forests by optical remote sensing,» *Remote Sensing of Environment,* vol. 84, n. 1, pp. 69-82, 2003.

[4] P. Giles, «Remote sensing and cast shadows in mountainous terrain,» *Photogrammetric engineering and remote sensing,* vol. 67, n. 7, pp. 833-840, 2001.

[5] H. Xie, X. Wang e T. Liang, «Development and assessment of combined Terra and Aqua MODIS snow cover products in Colorado Plateau, USA and Northern Xinjiang, China,» *Journal of Applied Remote Sensing,* vol. 3, pp. 1-15, 2009.

[6] J. Parajka e G. Blöschl, «Spatio-temporal combination of MODIS images –potential for snow cover mapping,» *Water Resources Research,* vol. 44, pp. 1-13, 2008.

[7] A. Gafurov e A. Bardoyy, «Cloud removal methodology from MODIS snow cover product,» *Hydrology and Earth System Sciences,* vol. 13, pp. 1361-1373, 2009.

[8] X. Wang, J. Wang, Z. Jiang, H. Li e X. Hao, «An Effective Method for Snow-Cover Mapping of Dense Coniferous Forests in the Upper Heihe River Basin Using Landsat Operational Land Imager Data,» *Remote Sensing,* vol. 7, n. 12, pp. 17246-17257, 2015.

[9] A. Fahsi, T. Tsegaye, W. Tadesse e T. Coleman, «Incorporation of digital elevation models with Landsat-TM data to improve land cover classification accuracy,» *Forest Ecology and Management,* vol. 128, n. 1, pp. 57-64, 2000.

[10] A. Shahtahmassebi, K. Wang, Z. Shen, J. Deng, W. Zhu, N. Han, F. Lin e N. Moore, «Evaluation on the two filling functions for the recovery of forest information in mountainous shadows on Landsat ETM+ Image,» *Journal of Mountain Science,* vol. 8, n. 3, pp. 414-426, 2011.

[11] D. Civco, «Topographic normalization of Landsat Thematic Mapper digital imagery,» *Photogrammetric Engineering and Remote Sensing,* vol. 55, n. 9, pp. 1303-1309, 1989.

[12] P. Dare, «Shadow analysis in high-resolution satellite imagery of urban areas,» *Photogrammetric Engineering & Remote Sensing,* vol. 71, n. 2, pp. 169-177, 2005.

[13] L. Dorren, B. Maier e A. Seijmonsbergen, «Improved Landsat-based forest mapping in steep mountainous terrain using object-based classification,» *Forest Ecology and Management,* vol. 183, n. 1, pp. 31-45, 2003.

[14] R. Essery, S. Morin, Y. Lejeune e C. Bauduin-Menard, «A comparison of 1701 snow models using observations from an alpine site,» *Advances in Water Resources,* vol. 55, pp. 131-148, 2013.

[15] M. Zappa, F. Pos, U. Strasser, P. Warmerdam e J. Gurtz, «Seasonal water balance of an Alpine catchment as evaluated by different methods for spatially distributed snowmelt modelling,» *Nordic Hydrol.,* vol. 34, n. 3, pp. 179-202, 2003.

[16] L. Renzullo, D. Barrett, A. Marks, M. Hill, J. Guerschman, Q. Mu e S. Running, «Multi-sensor model-data fusion for estimation of hydrologic and energy flux parameters,» *Remote Sensing of Environment,* vol. 112, n. 4, pp. 1306-1319, 2008.

[17] C. Cammalleri, M. Anderson, F. Gao, C. Hain e W. Kustas, «A data fusion approach for mapping daily evapotranspiration at field scale,» *Water Resources Research,* vol. 49, n. 8, pp. 4672-4686, 2013.

[18] T. Painter, D. Berisford, J. Boardman, K. Bormann, J. Deems, F. Gehrke, A. Hedrick, M. Joyce, R. Laidlaw, D. Marks, C. Mattmann, B. McGurk, P. Ramirez, M. Richardson, S. Skiles, F. Seidel e A. Winstral, «The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo,» *Remote Sensing of Environment,* vol. 184, pp. 139-152, 2016.

[19] A. Prokop, «Assessing the applicability of terrestrial laser scanning for spatial snow depth measurements,» *Cold Reg. Sci. Technol.,* vol. 54, pp. 155-163, 2008.

[20] J. Deems, P. Gadomski, D. Vellone, R. Evanczyk, A. LeWinter, K. Birkeland e D. Finnegan, «Mapping starting zone snow depth with a ground-based lidar to assist avalanche control and forecasting,» *Cold Reg. Sci. Technol.,* vol. 120, pp. 197-204, 2015.

[21] A. Simmons e A. Hollingsworth, «Some aspects of the improvement in skill of numerical weather prediction,» *Quarterly Journal of the Royal Meteorological Society,* vol. 128, pp. 647-677, 2002.

[22] G. Heathman, P. Starks, L. Ahuja e T. Jackson, «Assimilation of surface soil moisture to estimate profile soil water content,» *Journal of Hydrology,* vol. 279, pp. 1-17, 2003.

[23] D. MacKay, S. Samanta, R. Nemani e L. Band, «Multi-objective parameter estimation for simulating canopy transpiration in forested watersheds,» *Journal of Hydrology,* vol. 277, pp. 230-247, 2003.

[24] P. Rayner, M. Scholze, W. Knorr, T. Kaminski, R. Giering e H. Widmann, «Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system,» *Global Biogeochemical Cycles,* vol. 19, 2005.

[25] W. Sacks, D. Schimel, R. Monson e B. Braswell, «Model data synthesis of diurnal and seasonal CO2 fluxes at Niwot Ridge, Colorado,» *Global Change Biology,* vol. 12, pp. 240-259, 2006.

[26] D. Finger, M. Vis, M. Huss e J. Seibert, «The value of multiple data set calibration versus model complexity for improving the performance of hydrological models in mountain catchments,» *Water Resources Research,* vol. 51, n. 4, pp. 1939-1958, 2015.

[27] F. Melgani e L. Bruzzone, «Classification of hyperspectral remote sensing images with support vector machines,» *IEEE Transactions on geoscience and remote sensing,* vol. 42, n. 8, pp. 1778-1790, 2004.

[28] F. Isotta, C. Frei, V. Weilguni, M. P. Tadić, P. Lassègues, B. Rudolf, V. Pavan, C. Cacciamani, G. Antolini, S. Ratto, M. Munari, S. Micheletti, V. Bonati, C. Lussana, C. Ronchi, E. Panettieri, G. Marigo e G. Vertačnik, «The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data,» *International Journal of Climatology,* vol. 34, n. 5, pp. 1657-1675, 2013.

[29] G. Zolezzi, A. Bellin, M. Bruno, B. Maiolini e A. Siviglia, «Assessing hydrological alterations at multiple temporal scales: Adige River, Italy,» *Water Resources Research,* vol. 45, pp. 1-15, 2009.

[30] C. Notarnicola, M. Duguay, N. Moelg, T. Schellenberger, A. Tetzlaff, R. Monsorno, A. Costa, C. Steurer e M. Zebisch, «Snow cover maps from MODIS images at 250 m resolution, Part 1: Algorithm description,» *Remote Sensing,* vol. 5, n. 1, pp. 110-126, 2013.

[31] C. Notarnicola, M. Duguay, N. Moelg, T. Schellenberger, A. Tetzlaff, R. Monsorno, A. Costa, C. Steurer e M. Zebisch, «Snow cover maps from MODIS images at 250 m resolution, Part 2: Validation,» *Remote Sensing,* vol. 5, n. 4, pp. 1568-1587, 2013.

[32] U. Strasser, «Die Modellierung der Gebirgsschneedecke im Nationalpark Berchtesgaden. Modelling of the mountain snow cover in the Berchtesgaden National Park,» *Berchtesgaden National Park research report,* vol. 55.

[33] U. Strasser, J. Corripio, B. Brock, F. Pellicciotti, P. Burlando e M. Funk, «Spatial and Temporal Variability of Meteorological Variables at Haut Glacier d'Arolla (Switzerland) During the Ablation Season 2001: Measurements and Simulations,» *J. Geophys. Res.,* vol. 109.

[34] J. Corripio, «Vectorial algebra algorithms for calculating terrain parameters from DEMs and solar radiation modelling in mountainous terrain,» *International Journal of Geographical Information Science,* vol. 17, n. 1, pp. 1-23, 2003.

[35] R. Essery, «A factorial snowpack model (FSM 1.0),» *Geoscientific Model Development,* vol. 8, n. 12, pp. 3867-3876, 2015.

[36] U. Strasser, M. Bernhardt, M. Weber, G. Liston e W. Mauser, «Is snow sublimation important in the alpine water balance?,» *The Cryosphere,* vol. 2, pp. 53-66, 2008.

[37] G. Liston e K. Elder, «A Distributed Snow–Evolution Modeling System (SnowModel),» *J. Hydromet.,* vol. 7, n. 2, pp. 217-234, 2006.

[38] F. Hanzer, K. Helfricht, T. Marke e U. Strasser, «Multi-level spatiotemporal validation of snow/ice mass balance and runoff modeling in glacierized catchments,» *The Cryosphere,* vol. 10, pp. 1859-1881, 2016.

[39] G. Thirel, C. Notarnicola, M. Kalas, T. Schellenberger, A. Tetzlaff, M. Duguay, N. Mölg, P. Burek e A. d. Roo, «Assessing the quality of a real-time Snow Cover Area product for hydrological applications,» *Remote Sensing of Environment,* vol. 127, pp. 271-287, 2012.

[40] E. Maurer, J. Rhoads, R. Dubayah e D. Lettenmaier, «Evaluation of the snow-covered area data product from MODIS,» *Hydrological Processes,* vol. 17, n. 1, pp. 59-71, 2003.

[41] A. Tekeli, Z. Akyürek, A. Şorman, A. Şensoy e A. Şorman, «Using MODIS snow cover maps in modeling snowmelt runoff process in the eastern part of Turkey,» *Remote Sensing of Environment,* vol. 97, n. 2, pp. 216-230, 2005.

[42] A. Simic, R. Fernandes, R. Brown, P. Romanov e W. Park, «Validation of VEGETATION, MODIS, and GOES+ SSM/I snow-cover products over Canada based on surface snow depth observations,» *Hydrological Processes,* vol. 18, n. 6, pp. 1089-1104, 2004.

[43] J. Parajka e G. Blöschl, «Validation of MODIS snow cover images over Austria,» *Hydrology and Earth System Sciences Discussions,* vol. 3, n. 4, pp. 1569-1601, 2006.

[44] D. Vikhamar e R. Solberg, «Snow-cover mapping in forests by constrained linear spectral unmixing of MODIS data,» *Remote Sensing of Environment,* vol. 88, n. 3, pp. 309-323, 2003.

[45] R. Solberg, D. Hiltbrunner, J. Koskinen, T. Guneriussen, K. Rautiainen e M. Hallikainen, «Snow algorithms and products – Review and recommendations for research and development,» *NR Report* , vol. 924, 1997.

[46] Z. Pu, L. Xu e V. V. Salomonson, «MODIS/Terra observed seasonal variations of snow cover over the Tibetan Plateau,» *Geophysical Research Letters,* vol. 34, n. 6, 2007.

[47] A. Ü. Şorman, Z. Akyürek, A. Şensoy, A. A. Şorman e A. E. Tekeli, «Commentary on comparison of MODIS snow cover and albedo products with ground observations over the mountainous terrain of Turkey,» *Hydrology and Earth System Sciences,* vol. 11, n. 4, pp. 1353-1360, 2007.

Chapter 3

3. IMPROVING SWE ESTIMATION BY FUSION OF SNOW MODEL WITH TOPOGRAPHIC AND REMOTELY SENSED DATA

This chapter[2] presents a new concept to derive the snow water equivalent (SWE) based on the joint use of snow model (AMUNDSEN) simulation, ground data, and auxiliary products derived from remote sensing. The main objective is to characterize the spatial-temporal distribution of the model-derived SWE deviation with respect to the real SWE values derived from ground measurements. This deviation is due to the intrinsic uncertainty of any theoretical model, related to the approximations in the analytical formulation. The method, based on the k-NN algorithm, computes the deviation for some labeled samples, i.e. samples for which ground measurements are available, in order to characterize and model the deviations associated to unlabeled samples (no ground measurements available), by assuming that the deviations of samples vary depending on the location within the feature space. Obtained results indicate an improved performance with respect to AMUNDSEN model, by decreasing the RMSE and the MAE with ground data, on average, from 154 to 75 mm and from 99 to 45 mm, respectively. Furthermore, the slope of regression line between estimated SWE and ground reference samples reaches 0.9 from 0.6 of AMUNDSEN simulations, by reducing the data spread and the number of outliers.

[2] This chapter has been published in *Remote Sensing*:
De Gregorio, L.; Günther, D.; Callegari, M.; Strasser, U.; Zebisch, M.; Bruzzone, L.; Notarnicola, C. Improving SWE Estimation by Fusion of Snow Models with Topographic and Remotely Sensed Data. *Remote Sens.* **2019**, *11*, 2033. http://dx.doi.org/10.3390/rs11172033

3.1 Introduction

Melt water from snow and glaciers plays a key role in the hydrological cycle by contributing to the river flow and water resources in many parts of the world. It is estimated that about one-sixth of the world's population depends on snow- and ice-melt for the supply with drinking water [1]. Therefore, for hydrological assessments in these regions, knowledge about the spatial and temporal distribution of the snow water equivalent (SWE) is of uttermost importance. SWE is defined as the amount of water contained within the snowpack: it can be thought as the depth of water that would theoretically result if the entire snowpack would melt instantaneously [2]. Where available, point ground measurements of SWE remain the main direct information about the snow mass. However, given the large spatial heterogeneity of snow they may not be representative of large areas. A spatialized estimation of SWE in mountain areas, which are typically complex terrains with high topographic heterogeneity, is currently one of the most important challenges of snow hydrology [3]. An improved knowledge of the spatial distribution of SWE and its evolution over time would allow a better management of mountain water resources for drinking water supply, agriculture and hydropower, as well as for flood protection.

In literature, several approaches to the estimation of the spatial distribution of SWE exist. One of the most common methods is the interpolation of SWE ground measurements, constrained by remotely sensed maps of the snow extent. If enough ground measurements with a good spatial distribution are available, this approach may produce accurate SWE results [4]. Two different types of snow extent products derived from satellite exist: fractional and binary snow cover maps. Fassnacht et al. [4] and Molotch et al. [5] use of the fractional product that provides information about the percentage (from 0% to 100%) of snow coverage for each pixel. Elder et al. [6], instead, utilize binary mapping techniques with a set of thresholds to determine whether a pixel is snow-covered or not. A common statistical technique for spatial interpolation is based on binary regression trees that have been successfully applied to obtain interpolated SWE values from ground observations [6, 7]. However, numerous studies show that individual point observations of SWE are not necessarily representative of the surrounding area [8, 9, 10], thus limiting the feasibility of this approach.

Several statistical models have been developed to spatially interpolate the point-based snow information, e.g. multivariate linear regression can relate physiographic variables, historical SWE

data and snow-covered area imagery to the observed SWE. The accuracy of this simple method can be better than those of more complex techniques such as inverse-distance weighting [11]. Because of their accuracy and ability to preserve patterns from observations [12], nearest-neighbor approaches are an alternative methodical approach for spatio-temporal modeling biophysical parameters. However, in literature, only few studies exist based on the use of k-NN algorithms for modeling snow parameters. Among them, Zheng et al. [12] developed an approach to estimate SWE through the interpolation of spatially representative point measurements using a k-NN algorithm and historical spatial SWE data. Schneider et al. [13] estimated the relationships between SWE, snow covered area and topography to extend the Airborne Snow Observatory (ASO) dataset. In their analysis, they also used a nearest neighbor approach for resampling fractional snow-covered area maps. Another common approach for retrieving spatially distributed SWE is the reconstruction based on both remotely sensed snow cover maps and the estimation of snowmelt. The main idea, developed by Martinec and Rango [14], is to identify the date of snow disappearance for each pixel starting from Landsat snow cover maps; then, through a backward calculation of the melt rate, the accumulated SWE for each day back to the last significant snowfall is reconstructed. The sources of uncertainty for this approach are mainly related to the melt model structure and its meteorological forcing. Moreover, the main disadvantage of this approach is that it works properly only in areas with distinct accumulation and ablation periods. Furthermore, it operates retroactively only after snow disappearance, and hence does not enable the application for streamflow forecasting. Bair et al. [15] validated two different SWE reconstruction methods with the NASA ASO data in the upper Tuolumne River Basin in California's Sierra Nevada. The first approach uses an energy balance model to calculate snowmelt, integrating different remotely sensed products like daily MODIS fractional snow-covered area and albedo; it also considers ephemeral snow (i.e., snow that rapidly appears and disappears). The second reconstruction model implements a net radiation restricted degree-day approach [16]. The first method results, on average, more accurate than the second one in the SWE reconstruction, by showing no bias (0%) and a low mean absolute error (26%). Other successful examples of reconstructed SWE for basins in Sierra Nevada are shown by Girotto et al. [17], Guan et al. [18] and Rittger et al. [19].

An accurate estimation of SWE from remotely sensed images represents a longstanding challenge. Satellite data in the visible bands may provide information about the presence or absence of snow cover [20] but require cloud-free conditions. However, no indication on the total

amount of the snow mass can be derived. Passive microwave (PM) instruments are able to estimate the brightness temperature naturally emitted from the Earth and can be used to estimate SWE. When snow covers the ground, microwave radiation transmitted through the snowpack is absorbed and scattered by snow grains by decreasing the measured radiation. A deeper snowpack includes a larger number of snow grains, which are the main responsible for signal attenuation. This inverse relationship between snow depth and temperature brightness is the basis of SWE retrieval from PM measurements [21]. Vuyovich and Jacobs [21] compared snow hydrology model results to remotely sensed data to determine if passive microwave estimates of SWE can be used to characterize the snowpack and estimate runoff from snowmelt in the Helmand River, in Afghanistan. Mizukami and Perica [22] tried to identify SWE retrieval algorithms feasible for large-scale operational applications. In their study, Vuyovich et al. [23] compared the daily AMSR-E and SSM/I SWE products over nine winter seasons with spatially distributed model output of the SNOw Data Assimilation System (SNODAS) at watershed scale (25 km of spatial resolution) for 2100 watersheds in the United States. Results show large areas where the passive microwave SWE products are highly correlated to the SNODAS data, except in heavily forested areas and regions with a deep snowpack, where passive microwave SWE is significantly underestimated with respect to SNODAS. The best correlation is associated with basins in which maximum annual SWE value is lower than 200 mm and forest fraction is less than 20%. Forest cover has been proven to be one of the most relevant sources of uncertainty in SWE retrieval with PM sensors by acting as a mask for the snowpack microwave emission [24, 25]. Moreover, snow metamorphism affects the snowpack microwave emission by changing the crystal sizes, caused by temperature and water vapor gradients [26, 27]. Finally, SWE estimation from PM sensors suffers from several issues related to the coarse spatial resolution of the sensors (~ 25 km): in mountain regions, indeed, the spatial variability of snow cover and snow properties over a 25-km grid is large due to topographic influences. In the last decades, scientists have also extensively investigated the potential of Synthetic Aperture Radar (SAR) data for deriving SWE. Sun et al. [111] used microwave scattering models to analyze the C-band SAR scattering characteristics of snow-covered areas and estimated the distribution of the SWE using SAR data and snow cover data measured in the field. Conde et al. [112] presented a methodology for mapping the temporal variation of SWE through the SAR Interferometry technique and Sentinel-1 data.

Information about snow state variables can also be obtained from hydrological models. Many of the existing snowpack models are based on the same physical principles and solve the surface energy balance problem of a snowpack [30]. The main difference among these models is related to the way they represent physical processes in the snowpack such as absorption of incoming radiation, advection and convection, and how they represent the internal structure of the snowpack. In a cross-comparison with 33 models, Rutter et al. [30] found that the correlation of models' performance between years is always stronger at the open sites than in the forest, suggesting that models are more robust at open sites. The increasing complexity of snow-cover models demands high-quality forcing data. However, meteorological forcing data as provided by weather station recordings or atmospheric simulations suffer from several errors such as those induced by inaccuracy of the measurement, the regionalization scheme or boundary conditions. The process representations in deterministic, physically based snow models (which simulate physical processes in the snowpack) are an abstraction of reality, and hence inherently introduce uncertainty through simplification and the choice of parameter values. For fully distributed snow models, the spatial resolution is a compromise between computational feasibility and adequacy in mirroring the spatial scale of physical processes. Especially if the resolution (i.e. cell size) is much larger than the processes considered in the model, this choice is associated with uncertainty.

On the basis of this analysis, the main objective of this work is to generate a spatialized product of SWE over an Alpine area composed of Tyrol, South Tyrol and Trentino (Euregio region), by overcoming the aforementioned problems of hydrological models related to intrinsic uncertainty of the forcing data and correcting the spatial-temporal distribution of SWE as simulated by the snow model AMUNDSEN. The correction is performed using a specific k-NN algorithm and exploiting ground measurement-derived SWE data. The innovative aspect of our work is the joint use of snow model simulations, ground data, auxiliary products based on remote sensing and an advanced estimation technique to derive SWE. In this way our approach differs from traditional data assimilation techniques.

The chapter is organized as follows: section 3.2 introduces the study area and, after a description of the dataset, the method for SWE retrieval is presented in the last part of the section; results are then shown and discussed in Section 3.3 and, finally, conclusions and future perspectives are drawn in Section 3.5.

3.2 Materials and Methods

3.2.1 Study area

The considered study area is the Alpine region that includes Tyrol (Austria), and South Tyrol and Trentino (North-East Italy, Figure 3.1). Most of the rivers in the central and northern part of the considered region have a nivo-glacial regime with maximum discharge during the late summer months, whereas in the southern part of Trentino maximum discharge usually occurs during spring with an earlier snowmelt [31]. The area is covered by a relatively dense network of measurements sites (Figure 3.1), where snow profiles are periodically collected by the operators of the Avalanche Offices of the Provinces of Trento and of Bolzano (for Trentino and South Tyrol, respectively) and by the Hydrographic Service and the Zentralanstalt für Meteorologie und Geodynamik (ZAMG) for the Tyrol region.

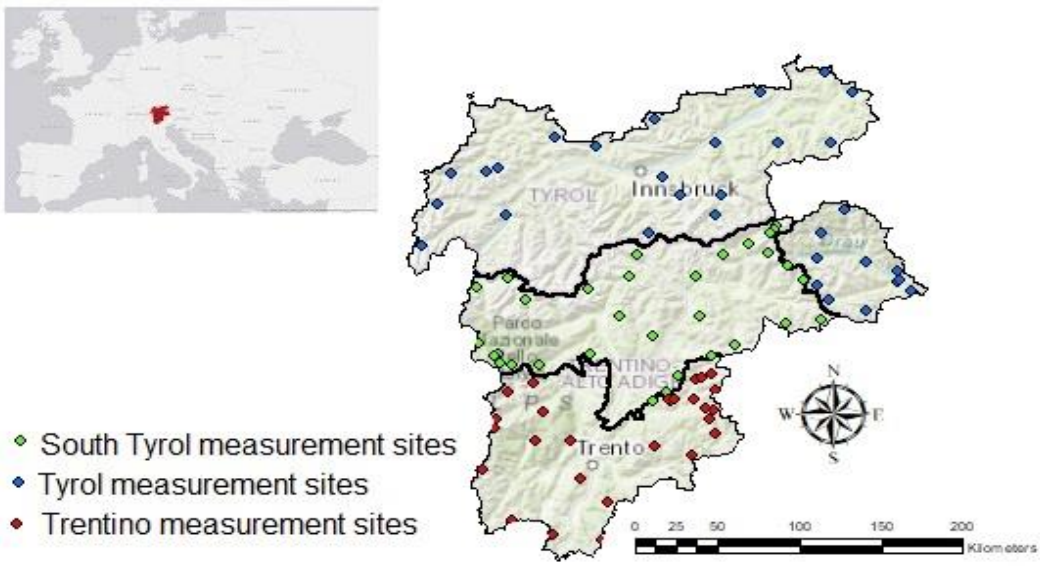

Figure 3.1 Study area: Tyrol (Austria), South Tyrol (Italy) and Trentino (Italy) and location of the measurement sites.

3.2.2 Data description

This section describes the input (features) and the target variables used in the proposed method. The same features have been selected for the three regions. Table 3.1 summarizes the features selected for implementing the k-NN algorithm and, below, shows the number of SWE ground samples available for each region. The following subsections will describe the single features used in the analyses.

Table 3.1 Features selected for implementing the k-NN algorithm (above); Number of SWE ground samples available for each region (below).

Feature name	Feature description
Altitude	Measurement site altitude
Geographic coordinates	Measurement site latitude and longitude
Forest coverage	Percentage of pixel (containing measurement site) covered by forest
Slope	Pixel slope
Aspect	Pixel aspect
Day of the year (DOY)	DOY rescaled with respect to the start of hydrological year (the day number 1 is the 1st of October).
SWE value from AMUNDSEN	SWE value from AMUNDSEN corresponding to location and date of each ground measurement.
SWE climatology	Calculated as average SWE value (from AMUNDSEN) on the other years (10 years).
SWE quality measure (CV)	Coefficient of Variation (standard deviation divided by the mean) on ensemble SWE simulations.
Land Surface Temperature (LST)	Mean daily LST calculated on previous month and number of days in previous month having positive temperatures.
Region	**# of SWE ground samples**
South Tyrol	1270
Tyrol	1467
Trentino	605

A. The AMUNDSEN model and its uncertainty

SWE is simulated using the distributed, physically based snow model AMUNDSEN ("Alpine MUltiscale Numerical Distributed Simulation ENgine") [32]. The regionalization and approximation of measured and unmeasured meteorological forcing and the inclusion of snow-canopy interactions are performed by the meteorological preprocessor of the model. Then the

coupled mass- and energy-balance is solved at every raster cell by means of the energy balance scheme of the integrated 1-D Factorial Snowpack Model (FSM) [33]. AMUNDSEN has proven its performance in a variety of applications in different natural environments [34]. In our application, the model has been validated at 38 stations with automated snow depth recordings. Additionally, 16 stations operated by the Hydrographic Service of the Province of Bolzano provide recordings of the snow surface temperature to validate the mass and energy balance separately. Daily snow height was predicted with a mean Nash-Sutcliffe efficiency (NSE) of 0.68 (ranging from 0.25 to 0.96).

In this work three snowpack variables provided by AMUNDSEN are used as features for implementing the k-NN algorithm: (i) SWE, corresponding to location and date of respective ground measurements, (ii) the associated uncertainty value, and (iii) a "SWE climatology" parameter. The latter is the average of the SWE values at the point and for the date corresponding to the ground measurement calculated for the other years. The uncertainty associated to the AMUNDSEN SWE simulation is based on ensemble simulation comparisons. Such ensemble simulations are a common way for assessing the uncertainty of model output. In many disciplines, such as hydrology, meteorology and cryospheric sciences, ensemble simulations have demonstrated their potential in improving the robustness of forecasts [35] and assimilation schemes [36]. In this study we follow a multi-model approach to generate an ensemble and include as many sources of uncertainty as possible. However, given the large extension of our study site the resulting computational costs need to be considered. In order to resolve critical snow-related processes such as snow redistribution and absorption of incoming shortwave radiation, hourly simulations are carried out with a spatial resolution of 250x250m. A maximum of 96 ensemble members were considered feasible, parallelized on a 96-core cluster. In order to reduce the number of the ensemble members while still enabling a certain amount of dispersion, just the most sensitive model configurations, i.e. those that explain most of the output variance, are accounted for.

An uncertainty and sensitivity analysis of FSM at one station in the study region identified the albedo formulations as well as the liquid water transport scheme inside the snowpack as the origin of the highest explanatory power for the performance variance [37]. Errors in precipitation sums and the approximation of the precipitation phase together with errors in air temperature and the radiative forcing are responsible for most uncertainties from a forcing data perspective. We reproduced the spread of a larger ensemble by a manual selection, result of a point-scale sensitivity

analysis aimed at identifying the most important uncertainty sources (input data, model structures and parameters choice) to explain the variance of the model performance. The selection is based on the findings of the Guenther et al. study [37]. However, in order to reduce the number of the ensemble members (in this study limited to 96 for computational reasons) while still representing the uncertainty for spatial distributed simulations, we just perturbed some of the most sensitive model settings. Particularly, we considered the following sources of uncertainty:

- *Precipitation phase*: The wet-bulb temperature (Tw), obtained through an iterative solution of the psychometric equation, has shown to improve predictions of snow and rainfall transition [38]. Lower and upper wet-bulb temperature limits, between which mixed snow and rainfall events are possible, are set. Precipitation undercatch and errors in elevation gradient and lateral redistribution: uncertainties associated with these factors are lumped together and their influence is approximated with two different precipitation correction factors.
- *Longwave irradiance*: Incoming longwave radiation is sparsely measured in the study area. Therefore, this input variable is derived from recordings of shortwave irradiance, air temperature (Ta), water-vapor content (ea) and the subsequent computation of atmospheric transmissivity and surface temperatures of surrounding slopes [39]. We utilize two different formulations of the clear-sky emissivity (εcs) estimation for a rough uncertainty estimation of this factor.
- *Snow albedo (α_S)*: in FSM two different albedo evolution representations are implemented. The prognostic option decreases albedo as snow ages over a timescale factor τa (with different values for cold and melting snow, respectively) towards a minimum (α_{min}), and increases albedo according to the amount of fresh snowfall (S_f) relative to a required snowfall amount to refresh the albedo (S_α) to its maximum (α_{max}). The second option predicts albedo as a function of surface temperature (T_S) in relation to the melting temperature (T_m). We employ both albedo options with two sets of parameters each for minimum and maximum albedo.
- *Snowpack hydraulics:* liquid water in snow layers is parameterized by a simple bucket model, where the maximum amount of liquid water (W_{max}) that a snow layer i can contain is dependent on the porosity (φ_i), the snow layer depth (h_i) and the irreducible liquid water content (W_{irr}). In the ensemble we apply this scheme with three different values of W_{irr}. Setting W_{irr} to 0 corresponds to switching off this option.

The combination of all presented model options and parameter sets results in an ensemble with 96 members.

B. Topographic and auxiliary parameters

This section refers to all those parameters that do not vary in time and that are used as features in the k-NN algorithm for SWE derivation. Topographic parameters can be used as proxies for the meteorological drivers, such as precipitation or wind for sublimation and redistribution or solar radiation (and temperature) for snowmelt. In addition, vegetation, and in particular the presence and density of a canopy, affects local meteorological conditions [40]. Several works aim at understanding the relationship between snowpack distribution and properties, and topographic variables. With the purpose of producing SWE maps, Erxleben et al. [41] considered elevation, slope, aspect, and forest coverage. Since elevation and SWE are known to be highly correlated [4], Fassnacht et al. [40] examined the relation between SWE and other topographic parameters, including location, canopy density, slope and aspect. In this study, the following parameters have been included for the estimation of SWE:

- Geographic coordinates (latitude and longitude)
- Altitude
- Slope and aspect
- Forest coverage as percentage (from 0% = no forest coverage to 100% = fully forested)
- Day number in the hydrological year (day number 1 is the 1st of October)

The day of the year has been included as a parameter in order to take into account the correlation between the AMUNDSEN performance and the period of the year. This correlation is due to the cumulative nature of the SWE, leading to a propagation of the deviation in time.

C. Satellite products

SWE is the amount of water that results from the melt of a snowpack with given depth and density. The latter can vary considerably: new snow generally has the lowest density of about 100 kg m3, and it can increase due to metamorphism to about 350–400 kg m3 for dry old snow and up to 500 kg m3 for wet old snow. The velocity at which the metamorphism takes place varies depending on the ambient conditions. As a general rule, the higher the temperature and the greater the temperature difference between the inner layers and the surface, the more rapidly the snow

structure changes [42]. Since snow temperature is generally close to 0°C near the ground, an estimation of snow surface temperature gives an idea of what stage of metamorphism is going on and therefore what kind of grains are present in the snowpack. Snow surface temperature can therefore be a proxy for snowpack conditions and hence be useful for SWE estimation.

In this study we exploit the MODIS product MOD11A1, i.e. the Land Surface Temperature (LST) images at 1-km spatial resolution. Collection 6 (C6) has been validated for Stage 2 via a series of field campaigns conducted in 2000-2007, and for more locations and time periods through radiance-based validation studies [43]. Further technical information can be retrieved in [44]. MOD11 can be downloaded from the NASA website [45]. LST product has a considerable dependency on surface material, vegetation cover, and topography and this makes validation results obtained for a single station alone never globally representative. Over surfaces with a heterogeneous land cover or with large topographic differences, satellite LST data are exposed to larger variations than over more homogeneous regions [46]. For this reason, Martin et al. [46], in their analysis, evaluated the accuracy of the LST data sets obtained from several sensors (AATSR, GOES, MODIS, and SEVIRI) by exploiting multiple years of in situ data from globally distributed stations representing various land cover types and topographies, including mountainous areas. An important reason for differences between satellite and in situ LST data is the upscaling of in situ data, because satellite measurements usually cover considerably larger areas than in situ point measurements, which may result in a lack of representativeness. The representativeness of the surrounding environment is very much dependent on the land cover and topography of each station, and therefore each station has to be examined individually [46]. In the Table Mountain station, authors found that the median accuracy, i.e. the satellite LST minus the station LST, of the MODIS product for the study years (2003-2012) is within $\pm 1\,K$ and by considering all measurement stations within $\pm 2\,K$.In particular, in this work, two LST-derived products have been used as features for implementing the k-NN algorithm: the mean LST calculated for the last 30 days with respect to each measurement acquisition date and the number of days, during these last 30 days, in which the temperature was positive. Both products have been chosen to broadly characterize different snowpack conditions. The mean surface temperature is used as a proxy for indicating the general condition of the snowpack, as mentioned by Oesch et al. [47] who proved the feasibility of snow surface temperature product derived from the NOAA-AVHRR sensor for monitoring snowmelt processes in snow covered pixels. The surface temperature, indeed, cannot

only be used for calibrating and calculating snow surface energy budget models, it is also possible to monitor the snow melting process itself. Furthermore, Colombo et al. [48], in their study on thermal inertia for monitoring snowmelt processes, remark the importance of accurate surface temperature measurements to infer snow density, especially during melting period. Because of the cloudiness, in this work daily product of LST has not been used and different mean values calculated over different time windows (10-15-30 days) have been tested in order to evaluate the product with larger sensitivity to the SWE retrieval. Moreover, in addition to the temporal resolution, also the spatial resolution of LST product (1km) could affect the sensitivity because that spatial scale may not be able to capture the snowpack variations. The basic idea is therefore that the mean value calculated over the last 30 days is the parameter that best captures the spatial and temporal variation of the snowpack, also considering the uncertainty of the satellite product. The number of positive temperature days, instead, can be used as a measure for "counting melting events", since mid-winter melt events could be correlated to the model SWE error, as explained in the model uncertainty description. The underlying hypothesis for the use of these parameters is that the AMUNDSEN behavior could be different for different snowpack conditions (e.g. the relative model error may be smaller for cold snowpacks than for snowpacks near melting conditions; model error is larger for repeated mid-winter melt events, etc.).

D. Ground data

The ground measurements of SWE, used partly in the training phase as target and partly to validate the proposed strategy, are collected through manual measurements performed by the foresters and operators of the Avalanche Office of the Provinces of Bolzano and Trento for South Tyrol and Trentino, and by the Hydrographic Service or the Zentralanstalt für Meteorologie und Geodynamik (ZAMG) for the Tyrol region. Measurement campaigns were carried out about every 2 weeks (South Tyrol and Trentino) and every week in Tyrol, or individually after significant snow and weather events (e.g., heavy snowfall, sudden and significant temperature change or wind activity) during the period of snow coverage. The main objective of the snow profile observations is the investigation of the physical and mechanical characteristics of the different layers of the snowpack, to identify weak layers and a potential instability. Regarding the choice of the measurement sites, these have to be safe and mostly representative for the slope of interest. Measurements were supposed to be preferably carried out for slopes with an inclination close to

or slightly less than 30°. Care was taken to select locations with mostly undisturbed snowpack. During the surveys, several physical parameters of the snowpack were measured by stratigraphic analysis, including the density of the different layers and the depth of the snowpack. The average density (ρ_s) and depth of the snowpack (HS) allow an estimation of the snow water equivalent by means of the following formula:

$$SWE\ (mm) = HS * \rho_s \tag{3.1}$$

In Trentino and South Tyrol the manual estimations of SWE are performed according to the AINEVA protocol [49]. In Tyrol, operators use a similar protocol, based on snow pit and manual measurements of snow depth and density from which SWE is derived. It is worth noting that the manual ground measurements can be affected by transcription errors (by the operator), measurement errors (not reached the bottom of the snowpack and thus wrong estimation of snow depth) or errors in the metadata (e.g. coordinates) or measurement units. Moreover, the manual observations can have significant limitations in consistency, continuity, spatial and temporal resolution and time and manpower consumption. Nevertheless, this type of data represents the most reliable estimate of the true SWE available for the study area and will therefore be used as ground truth in this study.

3.2.3 Proposed Method

In this section, the method used for SWE retrieval and the basic concepts of the adopted k-NN algorithm will be introduced. The proposed approach aims to overcome the errors inherent in the results from any snow modelling. Accordingly, the SWE values resulting from the AMUNDSEN simulations (SWEA) can be affected by uncertainties compared to the SWE derived from ground measurements (SWEg). The i-th SWE real value can be written as the sum of the estimation provided by AMUNDSEN and a deviation term δi:

$$SWE_i^g = SWE_i^A + \delta_i \tag{3.2}$$

The deviation is defined only for the samples where ground real values are available, hereafter called labeled samples. The characterization of deviation for unlabeled samples (no ground value available) is crucial for generating the new improved SWE product. Thus, the aim of our approach is to characterize the distribution of the model deviation in an automatically identified feature

space using the ground observation, and then to estimate the final SWE value for unlabeled samples.

A feature-selection technique based on a genetic algorithm (GA) and a proper cost function has been used for each region (i.e. Tyrol, South Tryol and Trentino) of the study area, in order to assess which variables are more relevant for the estimation of the deviation term (target variable). The procedure adopts the approach presented in [50] and is shown in Figure 3.2.

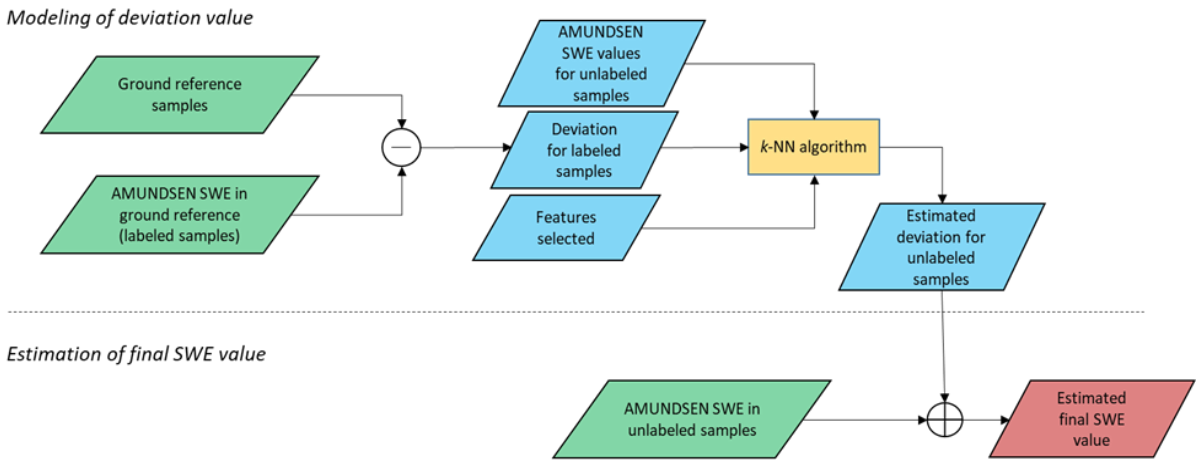

Figure 3.2 Flow chart of the proposed method for SWE retrieval.

A. Modeling of deviation value

This phase aims at computing the deviation values for unlabeled samples starting from the training dataset. For doing this, first the deviations for labeled samples are computed by calculating the difference between the AMUNDSEN SWE values and the respective ground samples. Then the deviation distribution is characterized in the feature space (consisting of the variables reported in Table 1). We adopted the Local Deviation Bias (LDB) strategy, which was tested to have better performance and describe the deviations more accurately with respect to the Global Deviation Bias (GDB) strategy [50]. LDB approach assumes that the AMUNDSEN model can provide different accuracies depending on the sample location in the feature space. In other words, the deviation locally changes in the space of the features and its value for an unlabeled sample is related to that of training samples located in the same portion of the feature space. The estimation of the deviations for the unlabeled samples is performed through the k-NN algorithm: for each unlabeled sample, the k-nearest labeled samples having the smallest distance in the feature space are

identified and the deviation for the unlabeled sample is then calculated as the average deviation value of the k-nearest labeled samples.

The application of the k-NN algorithm to our study can be schematized as follows: given x_i labeled samples of training dataset with $i = 1, \dots, M$, the output variable is represented by the deviation (between modelled and observed SWE), which is defined for each unlabeled sample x_j as the average deviation value of the k-nearest labeled samples in the feature space:

$$\delta(x_j) = \frac{\sum_{i=1}^{M} \delta(x_i) W(x_j, x_i)}{\sum_{i=1}^{M} W(x_j, x_i)} \tag{3.3}$$

where $W(x_j, x_i)$ is 0 or 1 depending on whether x_i is among the k-NN's of the unlabeled sample x_j or not. This means that $W(x_j, x_i) = 1$ if x_i is one of the k-NN's of x_j, and $W(x_j, x_i) = 0$ otherwise. An important question in this approach is how to select an optimal value of parameter k. In this study, we use the well-known rule of setting k as the square root of the half of the total number of reference samples [51].

B. Estimation of final SWE value

Once the deviations for all unlabeled samples ($\widehat{\delta_\iota}$) are calculated, the final corrected SWE values ($SWE_i^A\_corr$) are obtained by adding them to the respective AMUNDSEN SWE value:

$$SWE_i^A\_corr = SWE_i^A + \widehat{\delta_\iota} \tag{3.4}$$

In other words, the estimate of SWE from AMUNDSEN simulations is corrected by the use of the deviation. The deviations differ from each other depending on the sample location in the feature space.

C. Validation strategy

The above explained method has been applied for each region in the study area (Tyrol, South Tyrol and Trentino) separately as well as for the whole dataset, which includes all three regions. The method has been firstly validated by exploiting the ground data and then, once applied the method overall the study area, the generated SWE maps have been compared with binary MODIS snow maps. In the following, the two steps of validation and comparison are described.

Validation with ground data

For each region and the whole dataset, the following procedure has been applied. The dataset has been divided into two independent datasets: the learning (70%) and test (30%) ones (Figure 3.3).

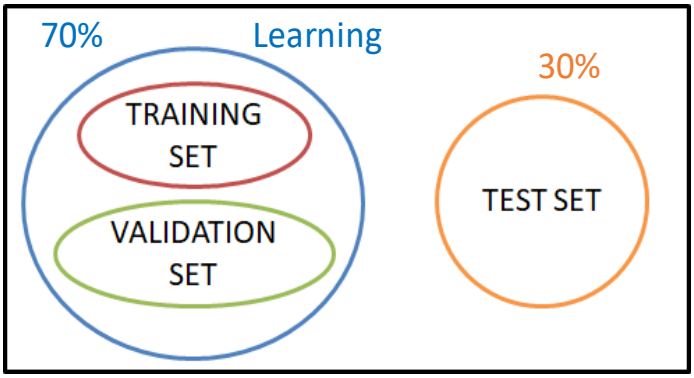

Figure 3.3 Separation of dataset for generating and testing the k-NN algorithm.

The 70% of learning dataset is used for generating the algorithm and is composed by a training and a validation set, used by applying a repeated 10-fold cross validation for 10 times. To ensure independence between datasets, as the deviation is time-correlated for each measurement point, the folds have been selected such that no points in the validation dataset are present in the training dataset even with a different time. This means that each time the algorithm uses 9 folds composed by certain measurement sites points as training dataset and the remaining one fold, which includes different measurement sites points, as validation dataset. Once the algorithm has been implemented, it has been tested on an independent test dataset, which include different measurement sites points with respect to the learning dataset, in order to evaluate the performances. The SWE values obtained have been compared with ground samples through the computation of some statistic metrics in order to evaluating the improvement achieved with the proposed method with respect to the AMUNDSEN simulations. The statistical metrics are: the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), the determination coefficient (R^2) and the bias. These metrics have been computes for both the training (as mean value of the repeated 10-fold cross validation results) and the test datasets in order to verify that the performance of the two datasets were consistent and without overfitting phenomena. Moreover, for test dataset, a scatterplot graph between estimated and ground samples together with the relative intercept and slope values has been reported.

Comparison with snow cover maps

The comparison with snow cover maps involves the information derived from the MODIS snow cover maps developed by Eurac Research, having a spatial resolution of 250 m [52, 53]. In order to evaluate the agreement between the SWE maps from the AMUNDSEN simulations and the proposed method and the MODIS snow maps, a pixel-based analysis was performed. The SWE values of the maps are therefore converted into binary values. To this purpose, different values ranging between 20 mm and 50 mm [54] were tested and an acceptable SWE threshold was found to be equal to 50 mm, value for which it is very likely that a pixel is classified as snow-covered in the MODIS product. In this way the agreement between pixels in the SWE maps and those in the snow cover maps has been computed, by analyzing separately the two class (snow/no snow) and two different altitude belts.

3.3 Results and Discussion

In this section, we present analyses and results obtained with the proposed method. In Section 3.3.1 we show the preliminary analyses relative to the AMUNDSEN SWE simulations. Then, in the Sections 3.3.2-3.4 we present the results obtained by the application of the proposed method.

3.3.1 Analysis of AMUNDSEN SWE simulations

The analysis of the AMUNDSEN simulations helps to understand how the model results vary with respect to the period of the year, the altitude and the different regions included in the study area. This analysis will guide the identification of the training data samples that are representative of the area under study. Figure 3.4 shows the temporal evolution of the deviations between modelled and observed SWE for labeled samples. The main evidence, observed for all years, is the temporal increase of the spread in the deviations due to the cumulative nature in the SWE variable, so the deviation propagates in time. Table 3.2 shows the number of points for each year and the relative mean percentage error (MPE), calculated as the ratio between the deviation and the corresponding observed SWE value. The percentage error is a relative error and expresses how large the absolute error (namely deviation) is, compared to the total amount of the measured SWE. The lower maximum value of SWE observed in the hydrological year 2005-2006 is due to lower values of snow depth recorded in this year with respect to the other studied years. It is useful for comparing samples having differing size. In our case, SWE derived from ground measurements

(hereafter also called “ground SWE”) can range from few mm up to around 1450 mm, as reported in the last column of Table 3.2.

The analysis of the simulated SWE with respect to the altitude for the first study year 2005/2006 is shown in Figure 3.5 (the other years show similar behavior). By comparing the temporal evolution for altitudes lower than 1000 m and higher or equal to 1000 m, a different behavior can be observed: the distribution of deviation values for lower altitudes seems to be asymmetrical with respect to zero: the simulated SWE is higher (negative values) than the observed one (Figure 3.5a). This asymmetrical deviation distribution at low altitudes could be due to several reasons such as an error in estimation of the precipitation phase or gradient in the model [32, 33], or the non-representativity of the observation sites at low altitudes.

Another factor to be considered is the thickness of the snowpack. At locations where the snowpack is shallower (typically at lower altitudes) and therefore with low SWE values, absolute underestimation cannot be high, since the SWE value is limited by a prediction of 0 mm. On the other side, there is no such limitation for the overestimation. This asymmetry in the deviation distribution does not appear at higher altitudes, where the snowpack is generally thicker. In this case the main evidence is the increasing temporal spread, as shown in Figure 3.5b.

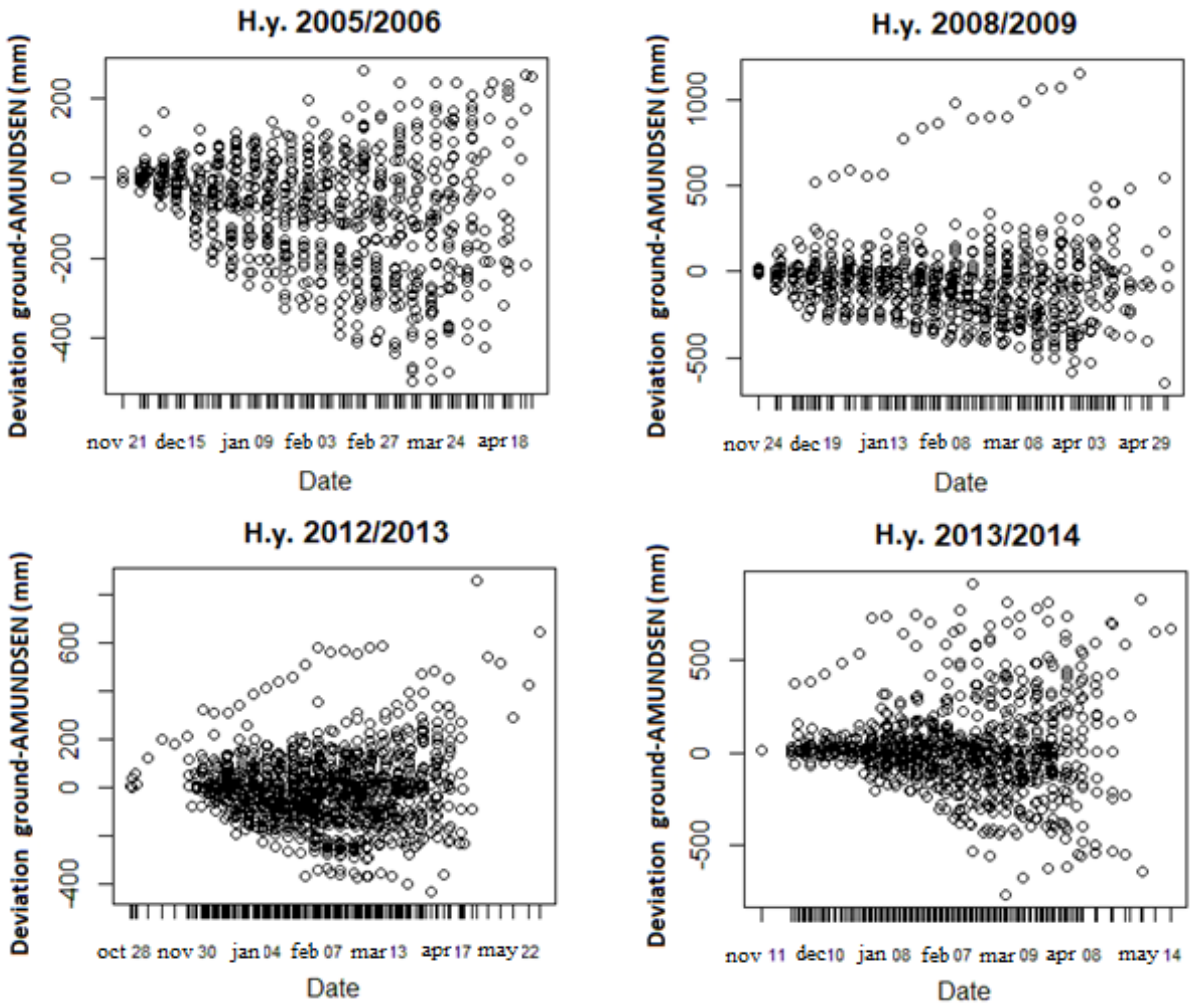

Figure 3.4 Temporal evolution of the deviation of AMUNDSEN model from ground measurements derived SWE for the 4 years analyzed.

Table 3.2 Number of samples and Mean Percentage Error (MPE) for each considered study year.

Hydrological year	Number of samples	Mean percentage error	Min - Max values of observed SWE (mm)
2005-2006	760	23.5%	4-512
2008-2009	708	18.6%	15-1446
2012-2013	1017	9.6%	8-1264
2013-2014	856	5.2%	5-997

Finally, an analysis per region was performed. Figure 3.6 shows the deviation of the AMUNDSEN simulations from the ground values of local observations for Tyrol, South Tyrol and Trentino. In both graphs, the main remark is about the AMUNDSEN behavior for Tyrol area. As for low altitudes, the deviations are asymmetric (again, by showing an overestimation of SWE by the snow model). Also in this case, this behavior could be ascribed to the measurement sites altitude. About 42% of measurement sites in Tyrol are located below 1000 m, while in Trentino and South Tyrol altitudes are always above 1000 m (in Trentino) and 1500 m (in South Tyrol). These preliminary analyses suggest different model performance depending on the period of the year and on the region of the study area. To evaluate the proposed method, we tested it on three different datasets, one per region, as well as on the entire dataset in order to identify differences in the performances that depend on the regional sampling.

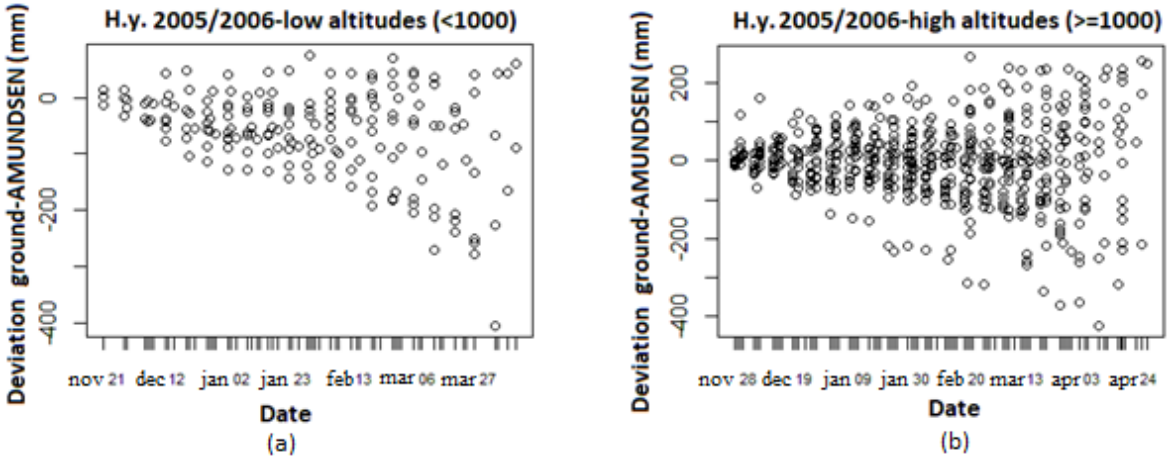

Figure 3.5 Deviation evolution with respect to the altitude, for low altitudes, <1000 meters (a) and for higher altitudes, >=1000 meters (b).

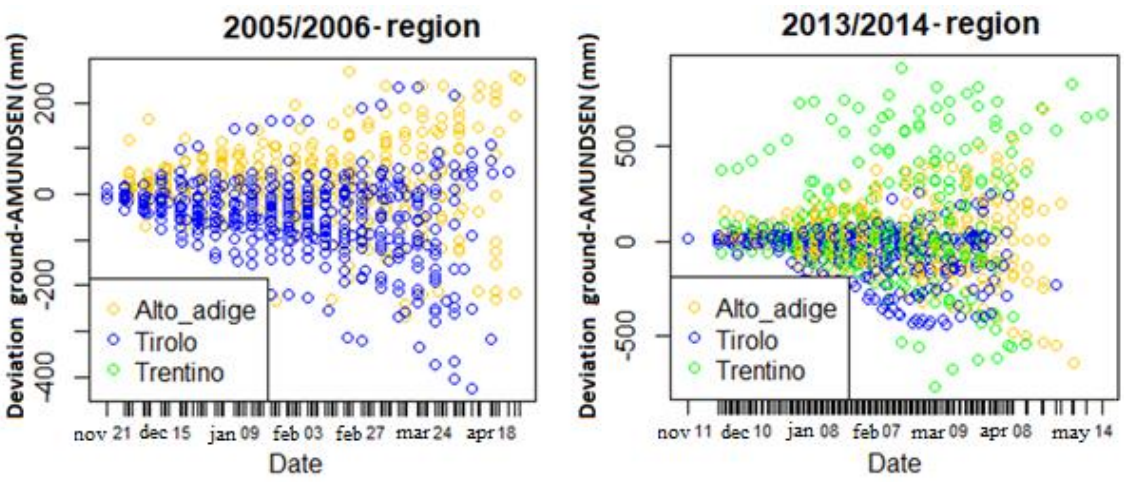

Figure 3.6 Analysis per region. For 2005/2006 no data from Trentino are available.

3.3.2 Results: South Tyrol dataset

For South Tyrol, 1270 observations are available. The k-NN algorithm was implemented by using the 70% of the sample, i.e. a sub-dataset of almost 900 samples. The target variable is the deviation, i.e. the difference between the AMUNDSEN simulation and the ground measurements derived SWE, and the feature space includes all variables indicated in Table 3.1. The resulting algorithm was then applied to the remaining 380 samples (test dataset) in order to evaluate the performance on a new and independent dataset. Once errors values are obtained, they are added to the corresponding simulated SWE in order to estimate its corrected value. Table 3.3 shows the performance in the estimation of SWE on both the training and the test data of the proposed method and the AMUNDSEN simulations. The k-NN algorithm seems to halve both RMSE and MAE compared to the modelled SWE. However, the statistical metrics used are no relative errors and should be contextualized with respect to the range of respective absolute measured SWE, which in this case can reach very high values (up to 1450 mm).

Figure 3.7 shows the comparison of scatterplots between observed SWE reference samples versus AMUNDSEN simulations (Figure 3.7a) and with the proposed method (Figure 3.7b). The absolute improvement of the SWE estimation is higher for higher observed values. Higher SWE values typically occur in the later season where the difference between the AMUNDSEN model results and the observations is larger.

Table 3.3 Estimation performance obtained with AMUNDSEN simulation and with proposed method on the test dataset of South Tyrol.

Region	Dataset	Estimation method	RMSE (mm)	MAE (mm)	R^2	Bias
South Tyrol	Training	AMUNDSEN	166.3	111.2	0.4	-37.5
		Proposed method	77.2	49.5	0.8	1.4
	Test	AMUNDSEN	167.7	109.7	0.4	-33.8
		Proposed method	80.9	53.4	0.8	2.5

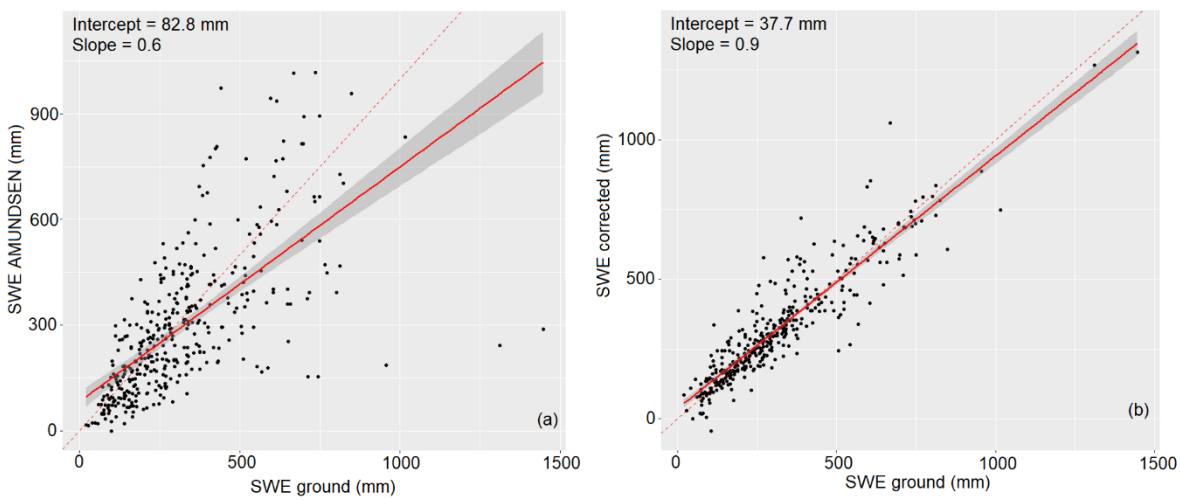

Figure 3.7 Regression scatterplot for South Tyrol dataset: simulated SWE vs observation (a) and corrected SWE vs observation (b). The dashed line represents the 1:1 line.

3.3.3 Results: Tyrol dataset

The analysis of the Tyrol dataset involves around 1470 observations. 70% of them (around 1030 samples) are used for implementing the algorithm with the same validation approach used for South Tyrol. Results relative to the remaining 30% of data (around 440 samples, test dataset) are shown together with the training dataset in Table 3.4. Also in this case, the proposed method

provides a more accurate estimation in terms of MAE, RMSE and bias compared to the model simulation.

Figure 3.8 shows the scatterplots of estimated versus observed SWE values. Similar to the previous case of South Tyrol, the main result is that the proposed method reduces the difference between the two sources of SWE by increasing the slope of the regression line up to 0.9 and reducing the intercept value to 10 mm.

Table 3.4 Estimation performance obtained with AMUNDSEN simulation and with proposed method on the test dataset relative to Tyrol.

Region	Dataset	Estimation method	RMSE (mm)	MAE (mm)	R^2	Bias
Tyrol	Training	AMUNDSEN	89.4	65.1	0.6	34.8
		Proposed method	39.4	26.2	0.9	0.7
	Test	AMUNDSEN	88.4	64.9	0.6	32.6
		Proposed method	44.8	28.1	0.8	1.3

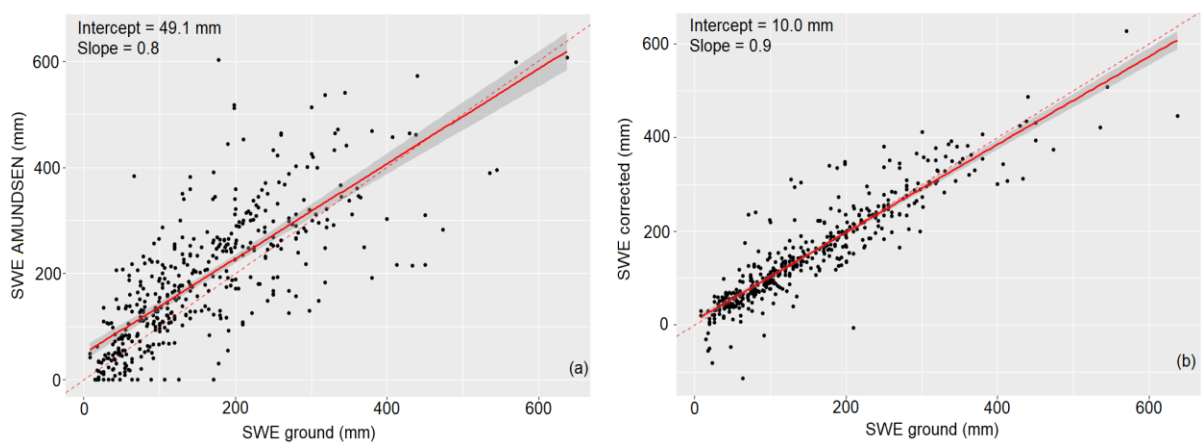

Figure 3.8 Regression scatterplot for Tyrol dataset: simulated SWE vs observation (a) and corrected SWE vs observation (b). The dashed line represents the 1:1 line.

3.3.4 Results: Trentino dataset

The third data set involves around 600 labeled observations. Also in this case, results are tested on 30% of the samples, i.e. around 180 data points. Table 3.5 reports the obtained values of MAE

and RMSE together with the R-squared and the bias. Also in this case, the assumption that the deviation is varying depending on the sample location in the feature space leads to an improvement in the SWE estimation. The high RMSE of 240.7 mm for the AMUNDSEN simulations is probably due to the presence of numerous outliers and the small number of test points. Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors, by resulting impacted by the presence of outliers. Figure 3.9 shows the proposed method based on the k-NN algorithm reduces the data spread and increases the slope of the regression line up to 0.9, while the RMSE sharply decreases to 102.8 mm.

Table 3.5 Estimation performance obtained with both the AMUNDSEN simulation and the proposed method on the test dataset relative to Trentino.

Region	Dataset	Estimation method	RMSE (mm)	MAE (mm)	R^2	Bias
Trentino	Training	AMUNDSEN	251.4	170.8	0.3	-82.6
		Proposed method	101.5	63.2	0.8	12.9
	Test	AMUNDSEN	240.7	162.7	0.3	-91.0
		Proposed method	102.8	65.3	0.8	18.6

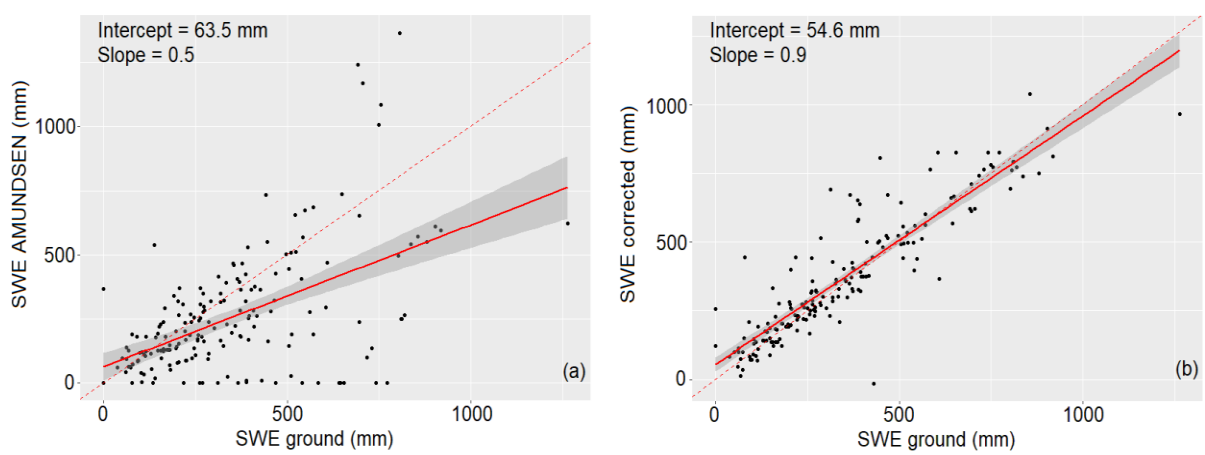

Figure 3.9 Regression scatterplot for Trentino dataset: simulated SWE vs observation (a) and corrected SWE vs observation (b). The dashed line represents the 1:1 line.

3.3.5 Results: the whole dataset

The last analysis was conducted by using the whole dataset available, i.e. around 3300 observations, including the 4 years and the entire study area. 30% of the samples (i.e. around 1000 samples) were used for evaluating the performances of the proposed method. Table 3.6 reports the statistical metrics for the SWE estimations obtained with both the AMUNDSEN simulations and the proposed method. The performances for the whole data set are approximately equal to the mean performances achieved over the three regions separately. Figure 3.10 shows the scatterplots of simulated versus observed SWE, as well as a comparison of the proposed method results to the observations. The scatterplots confirm the results derived by quantitative analysis given in Table 3.6, pointing out an increase of the slope value and a corresponding decrease in the value of the square error RMSE for the proposed method.

Table 3.6 Estimation performance obtained with both the AMUNDSEN simulation and the proposed method on the test dataset relative to entire study area.

Region	Dataset	Estimation method	RMSE (mm)	MAE (mm)	R^2	Bias
Euregio	Training	AMUNDSEN	163.2	103.4	0.4	-14.1
		Proposed method	70.0	41.6	0.8	1.3
	Test	AMUNDSEN	153.7	98.7	0.4	-12.2
		Proposed method	75.3	45.0	0.8	-1.7

Performances were then evaluated by analyzing different periods of the year and different altitudes. The test dataset was composed of around 600 measures from the winter period (i.e. November to February), and 400 points from spring (March to May). Around 200 of the test points are located below 1000 m, and the remaining 800 above 1000 m altitude. This disparity in test sample distribution with elevation is due to the fact that only 13.6% of the observation sites are located below 1000 m. Table 3.7 shows the RMSE and MAE in relation to the seasonal periods and altitude bands. As already mentioned in section 3.3.1, the cumulative nature of SWE leads to a temporal increase of the deviations between the simulations and the results of the proposed

method: from a value of RMSE of 124.8 mm in the winter period to a value of 200.8 mm in springtime. At low altitudes the uncertainty in the AMUNDSEN results is smaller than for high altitudes. This is probably due to the absolute nature of both RMSE and MAE and to the shallower snowpack at lower altitudes. This implies low SWE values and therefore lower absolute errors than for higher altitudes.

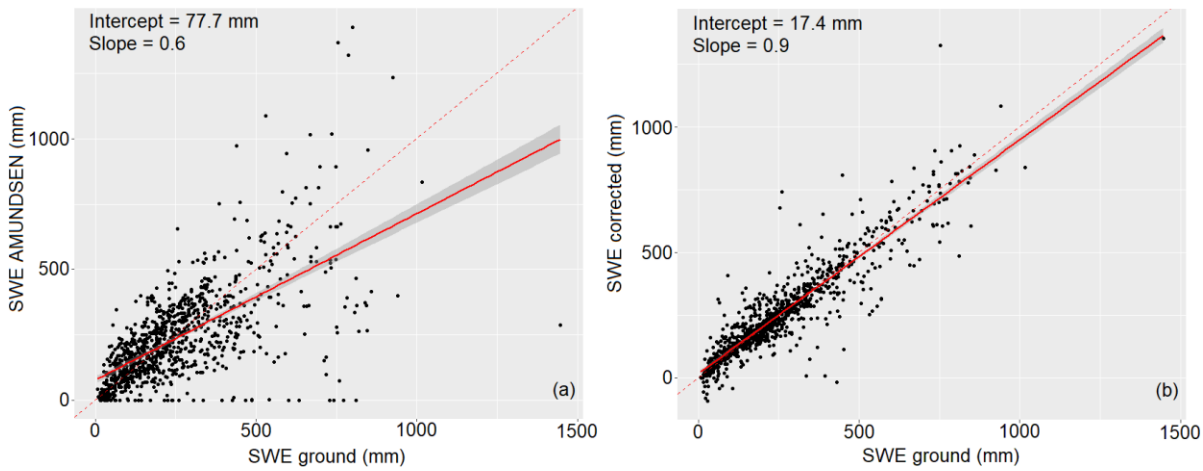

Figure 3.10 Regression scatterplot for the whole dataset: simulated SWE vs observation (a) and corrected SWE vs observation (b). The dashed line represents the 1:1 line.

3.4 SWE maps

Previous analyses provide the basis to create SWE maps for the entire study area. It was shown that applying the proposed method to the whole dataset results in a performance similar to the mean performance of the individual data sets. Furthermore, implementing a single algorithm for the whole study region reduces the computational cost significantly. For this reason, the generation of corrected SWE maps is based on the application of the proposed technique trained on the whole dataset. The resulting algorithm from the training procedure is then applied to the spatially distributed simulations of the Euregio region in order to generate a SWE map time series. Figure 3.11 and Figure 3.12 show two examples of SWE maps obtained with the proposed method, compared to AMUNDSEN simulations and the MODIS snow cover maps developed by Eurac Research.

The map in Figure 3.11 refers to an end-of-season situation (7 March 2014), while the maps in Figure 3.12 refer to a begin of the season (29 November 2013). In both cases, the proposed method shows lower SWE values compared to the AMUNDSEN simulations, especially for higher

altitudes (more than 2000 m) where the difference between AMUNDSEN simulations and the SWE values estimated by the proposed method reach values up to 67 mm.

Table 3.7 Performance obtained with both the AMUNDSEN simulations and the proposed method by dividing the test dataset in two seasonal periods and two altitude bands.

Period	Estimation method	RMSE (mm)	MAE (mm)
Winter	AMUNDSEN	124.8	77.9
	Proposed method	64.5	37.0
Spring	AMUNDSEN	200.8	142.0
	Proposed method	93.8	61.4
Altitude	**Estimation method**	**RMSE (mm)**	**MAE (mm)**
Low altitudes (<=1000m)	AMUNDSEN	89.3	68.1
	Proposed method	55.5	32.7
High altitudes (>1000m)	AMUNDSEN	163.0	104.5
	Proposed method	78.4	47.3

At the begin of the season, differences between the model and the proposed method results are more evident with respect to those of the end of the season, especially in the southern and in the northern part of the study area. The lower SWE values as evident in the map derived with the proposed method, in the southern part lead to an improved matching with the snow cover map derived by MODIS by better capturing the snow free areas.

Table 3. 8 shows the pixel-based agreement in percentage between the SWE maps and the MODIS product. We can confirm the behavior found in Figure 3.11 and Figure 3.12, i.e. that the proposed method, in both cases, improves the estimation of snow-free areas, but shows lower values in the snow-covered areas, generally located at higher altitudes for the dates analyzed. An improvement could be achieved by integrating the dataset with more high-altitude points (in this study, only 15% is located above 2000 m) in order to provide more training data to the algorithm.

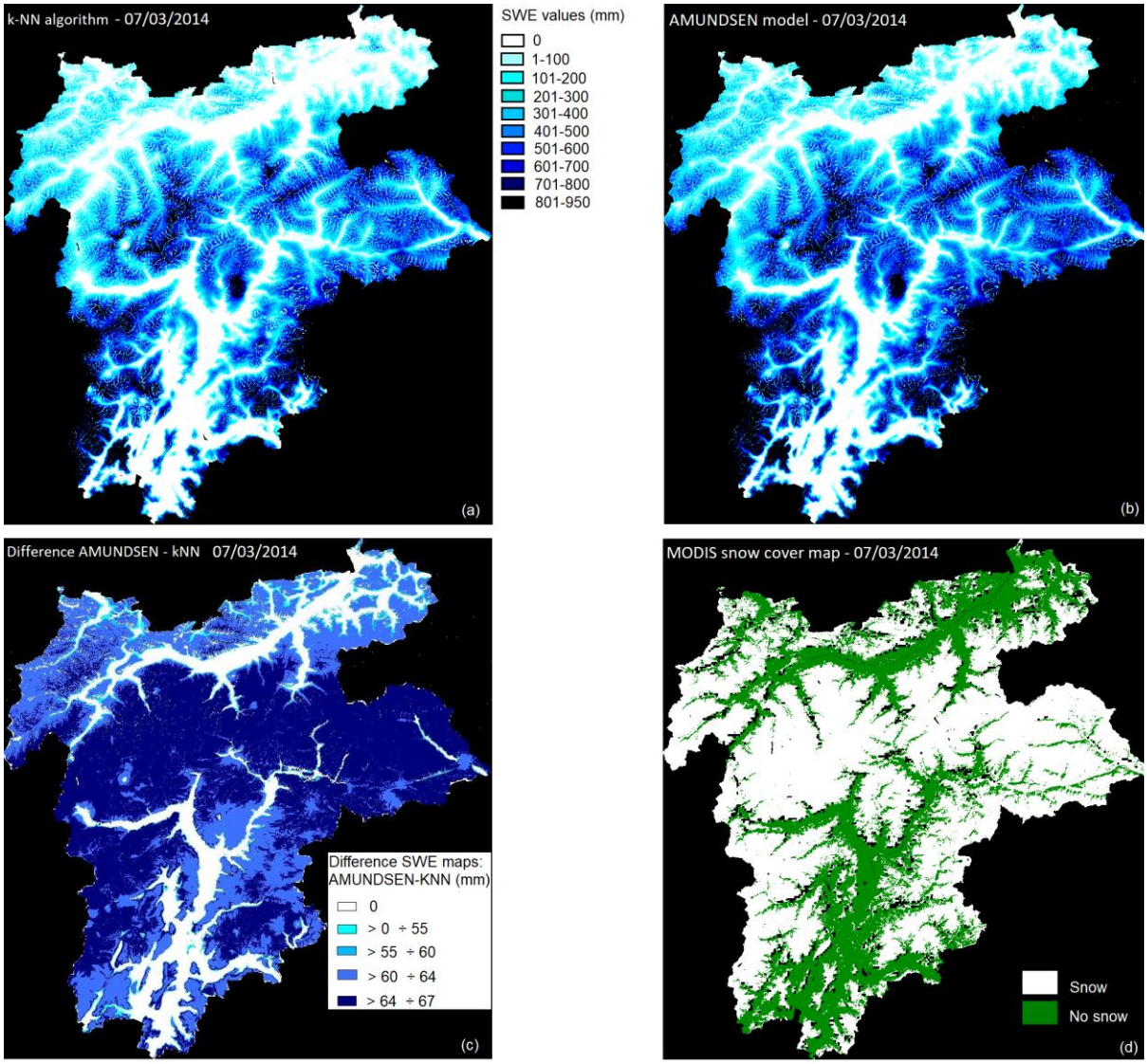

Figure 3.11 March 7th, 2014: (a) SWE map generated by the proposed method and (b) by the AMUNDSEN simulations. (c) is the difference between the two products and (d) the snow cover map product by MODIS.

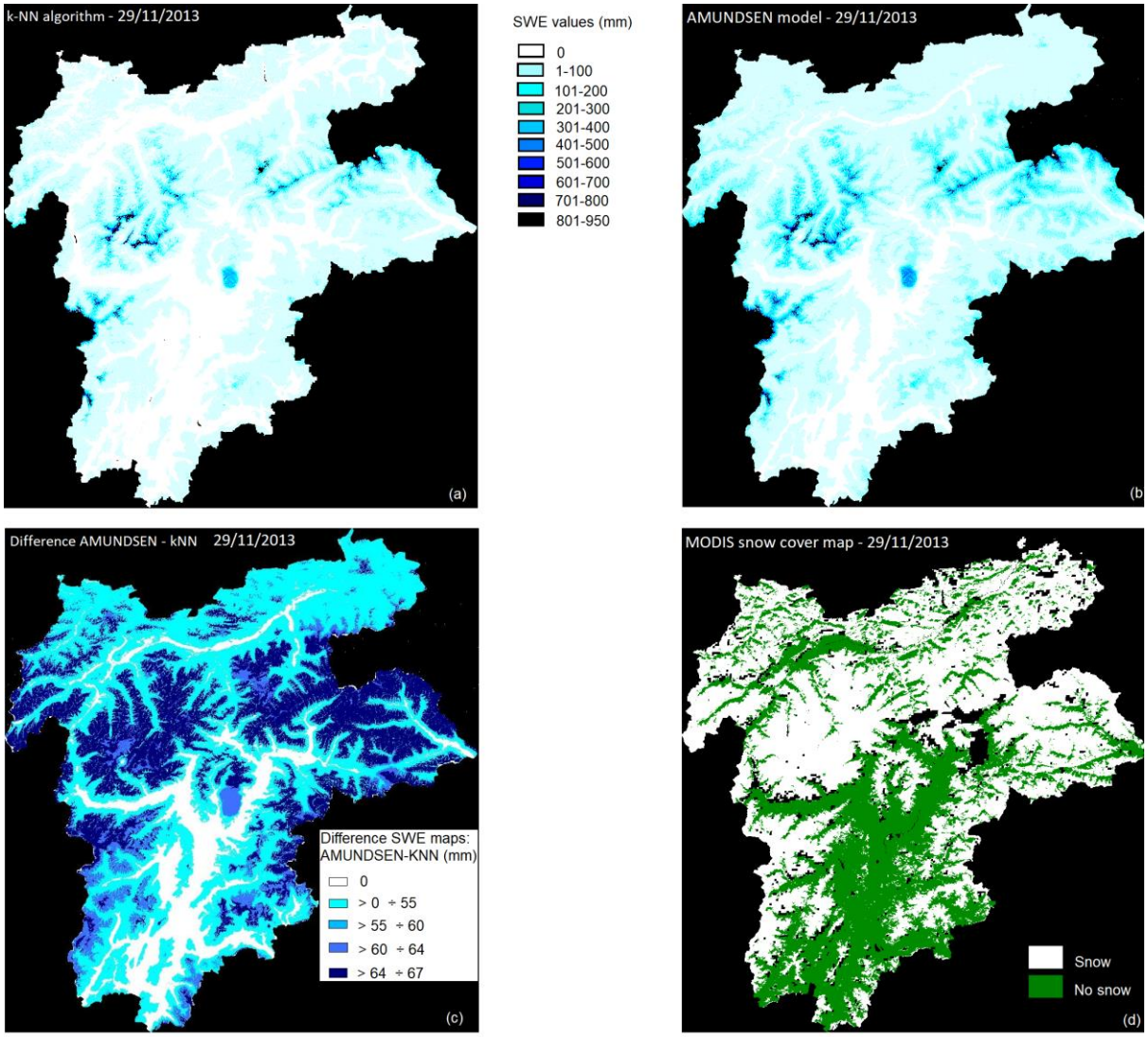

Figure 3.12 November 29th, 2013: (a) SWE map generated by the proposed method and (b) by the AMUNDSEN simulations. (c) is the difference between the two products and (d) the snow cover map product by MODIS.

3.5 Conclusion

In this chapter a new concept to improve the distributed estimation of snow water equivalent (SWE) is presented. The proposed method exploits a physically based model (AMUNDSEN), field observations, some topographic and auxiliary parameters and products from optical remote sensing for creating a time series of SWE maps for a region including Tyrol, South Tyrol and Trentino (Euregio area). Available ground reference samples are used for characterizing deviations of the snow model simulations affected, as any theoretical model, by uncertainties from approximations in the analytical formulation with respect to the observation. The hypothesis is that such deviations

are varying depending on their location in the feature space. This behavior can be characterized by exploiting the properties of a specific k-Nearest Neighbor (k-NN) estimator, based on a "feature similarity" principle, to predict values of any new data point. Once the deviation is computed, it is added to the modelled SWE in order to obtain a corrected value.

Obtained results are promising with a significant improvement of performance: the new method in our data decreased, on average, the RMSE and the MAE from 154 to 75 mm and from 99 to 45 mm, respectively compared to the AMUNDSEN simulations. Furthermore, the slope of the regression line between estimated SWE and ground observations increases from 0.6 to 0.9 by reducing the data spread and the number of outliers.

Table 3. 8 Agreement between MODIS snow maps and SWE maps estimated with AMUNDSEN and with proposed method.

	Estimation method	Agreement with MODIS **no snow**		Agreement with MODIS **snow**	
29/11/2013	AMUNDSEN	86%	34% (<=1000m)	59%	0% (<=1000m)
			52% (>1000m)		59% (>1000m)
	Proposed method (k-NN)	98%	34% (<=1000m)	29%	0% (<=1000m)
			64% (>1000m)		29% (>1000m)

	Estimation method	Agreement with MODIS **no snow**		Agreement with MODIS **snow**	
07/03/2014	AMUNDSEN	53%	44% (<=1000m)	97%	1% (<=1000m)
			9% (>1000m)		96% (>1000m)
	Proposed method (k-NN)	61%	46% (<=1000m)	93%	1% (<=1000m)
			15% (>1000m)		92% (>1000m)

In the approach presented in this study, two aspects are critical: the feature selection and the amount of observation samples. In this work, the feature selection in this work was performed

through a genetic algorithm, by considering several variables supposed to be related to SWE computation. Different products from optical remote sensing were included in the feature selection, such as snow cover duration, snow cover fraction, different reflectance bands and the land surface temperature. The latter was found to be the only product relevant in our analysis. In particular, we exploited the mean surface temperature and the number of positive-temperature days, both computed on the last 30 days with respect to the date of ground acquisition. Certainly, many other parameters from remote sensing could be tested, such as products from radar sensors that are sensitive to the water presence in the snowpack [55]. A deeper and more extensive feature selection could for sure improve the results obtained. Regarding the amount of ground observations, an improvement to the proposed approach could be achieved by increasing the dataset variability in the feature space. This could be done by acquiring, for example, ground measurements that are more differentiated in the feature space, such as different altitudes or different percentage of forest cover or slope.

We can conclude that the proposed approach effectively handles the variability of deviations between simulations and observations in the feature space and can be applied to other study areas and to other physically based snow models.

References

[1] R. Hock, G. Rees, M. W. Williams e E. Ramirez, «Contribution from glaciers and snow cover to runoff from mountains in different climates. Hydrological Processes: An International Journal,» vol. 20, n. 10, pp. 2089-2090, 2006.

[2] A. Haberkorn, «European Snow Booklet – an Inventory of Snow Measurements in Europe,» EnviDat., 2019.

[3] K. A. Dressler, G. H. Leavesley, R. C. Bales e S. R. Fassnacht, «Evaluation of gridded snow water equivalent and satellite snow cover products for mountain basins in a hydrologic model,» Hydrological Processes: An International Journal, vol. 20, n. 4, pp. 673-688, 2006.

[4] S. R. Fassnacht, K. A. Dressler e R. C. Bales, «Snow water equivalent interpolation for the Colorado River Basin from snow telemetry (SNOTEL) data,» Water Resources Research, vol. 39, n. 8, 2003.

[5] N. P. Molotch, S. R. Fassnacht, R. C. Bales e S. R. Helfrich, «Estimating the distribution of snow water equivalent and snow extent beneath cloud cover in the Salt–Verde River basin, Arizona,» Hydrological Processes, vol. 18, n. 9, pp. 1595-1611, 2004.

[6] K. Elder, W. Rosenthal e R. E. Davis, «Estimating the spatial distribution of snow water equivalence in a montane watershed,» Hydrological Processes, vol. 12, n. 10-11, pp. 1793-1808, 1998.

[7] N. P. Molotch e R. C. Bales, «SNOTEL representativeness in the Rio Grande headwaters on the basis of physiographics and remotely sensed snow cover persistence,» Hydrological Processes: An International Journal, vol. 20, n. 4, pp. 723-739, 2006.

[8] N. N. Neumann, C. Derksen, C. Smith e B. Goodison, «Characterizing local scale snow cover using point measurements during the winter season,» Atmosphere-Ocean, vol. 44, n. 3, pp. 257-269, 2006.

[9] R. Rice e R. C. Bales, «Embedded-sensor network design for snow cover measurements around snow pillow and snow course sites in the Sierra Nevada of California,» Water resources research, vol. 46, n. 3, 2010.

[10] J. I. López-Moreno, S. R. Fassnacht, S. Beguería e J. B. P. Latron, «Variability of snow depth at the plot scale: implications for mean depth estimation and sampling strategies,» The Cryosphere, vol. 5, n. 3, pp. 617-629, 2011.

[11] D. Schneider e N. P. Molotch, «Real-time estimation of snow water equivalent in the Upper Colorado River Basin using MODIS-based SWE Reconstructions and SNOTEL data,» Water Resources Research, vol. 52, n. 10, pp. 7892-7910, 2016.

[12] Z. Zheng, N. P. Molotch, C. A. Oroza, M. H. Conklin e R. C. Bales, «Spatial snow water equivalent estimation for mountainous areas using wireless-sensor networks and remote-sensing products,» Remote sensing of environment, vol. 215, pp. 44-56, 2018.

[13] D. Schneider, N. Molotch e J. Deems, «Estimating relationships between snow water equivalent, snow covered area, and topography to extend the Airborne Snow Observatory dataset,» The Cryosphere Discuss., 2017.

[14] J. Martinec e A. Rango, «Areal distribution of snow water equivalent evaluated by snow cover monitoring,» Water Resources Research, vol. 17, n. 5, pp. 1480-1488, 1981.

[15] E. H. Bair, K. Rittger, R. E. Davis, T. H. Painter e J. Dozier, «Validating reconstruction of snow water equivalent in California's Sierra Nevada using measurements from the NASA Airborne Snow Observatory,» Water Resources Research, vol. 52, n. 11, pp. 8437-8460, 2016.

[16] K. A. Y. E. Brubaker, A. Rango e W. Kustas, «Incorporating radiation inputs into the snowmelt runoff model,» Hydrological processes, vol. 10, n. 10, pp. 1329-1343, 1996.

[17] M. Girotto, S. A. Margulis e M. Durand, «Probabilistic SWE reanalysis as a generalization of deterministic SWE reconstruction techniques,» Hydrological processes, vol. 28, n. 12, pp. 3875-3895, 2014.

[18] B. Guan, N. P. Molotch, D. E. Waliser, S. M. Jepsen, T. H. Painter e J. Dozier, «Snow water equivalent in the Sierra Nevada: Blending snow sensor observations with snowmelt model simulations,» Water Resources Research, vol. 49, n. 8, pp. 5029-2046, 2013.

[19] K. Rittger, E. H. Bair, A. Kahl e J. Dozier, «Spatial estimates of snow water equivalent from reconstruction,» Advances in water resources, vol. 94, pp. 345-363, 2016.

[20] T. H. Painter, K. Rittger, C. McKenzie, P. Slaughter, R. E. Davis e J. Dozier, «Retrieval of subpixel snow covered area, grain size, and albedo from MODIS,» Remote Sensing of Environment, vol. 113, n. 4, pp. 868-879, 2009.

[21] C. Vuyovich e J. M. Jacobs, «Snowpack and runoff generation using AMSR-E passive microwave observations in the Upper Helmand Watershed, Afghanistan,» Remote sensing of environment, vol. 115, n. 12, pp. 3313-3321, 2011.

[22] N. Mizukami e S. Perica, «Towards improved snow water equivalent retrieval algorithms for satellite passive microwave data over the mountainous basins of western USA,» Hydrological Processes, vol. 26, n. 13, pp. 1991-2002, 2012.

[23] C. M. Vuyovich, J. M. Jacobs e S. F. Daly, «Comparison of passive microwave and modeled estimates of total watershed SWE in the continental United States,» Water Resources Research, vol. 50, n. 11, pp. 9088-9102, 2014.

[24] A. T. C. Chang, J. L. Foster e D. K. Hall, «Effects of forest on the snow parameters derived from microwave measurements during the BOREAS winter field campaign,» Hydrological Processes, vol. 10, n. 12, pp. 1565-1574, 1996.

[25] A. W. Nolin, «Recent advances in remote sensing of seasonal snow,» Journal of Glaciology, vol. 56, n. 200, pp. 1141-1150, 2010.

[26] E. G. Josberger e N. M. Mognard, «A passive microwave snow depth algorithm with a proxy for snow metamorphism,» Hydrological Processes, vol. 16, n. 8, pp. 1557-1568, 2002.

[27] M. Durand, E. J. Kim, S. A. Margulis e N. P. Molotch, «A first-order characterization of errors from neglecting stratigraphy in forward and inverse passive microwave modeling of snow,» IEEE Geoscience and Remote Sensing Letters, vol. 8, n. 4, pp. 730-734, 2011.

[28] S. C. T. W. J. L. H. H. X. W. Z. &. W. J. Sun, «Estimation and analysis of snow water equivalents based on C-band SAR data and field measurements,» Arctic, antarctic, and alpine research, vol. 47, n. 2, pp. 313-326, 2015.

[29] V. Conde, G. Nico, P. Mateus, J. Catalão, A. Kontu e M. Gritsevich, «On the estimation of temporal changes of snow water equivalent by spaceborne SAR interferometry: a new

application for the Sentinel-1 mission,» Journal of Hydrology and Hydromechanics, vol. 67, n. 1, pp. 93-100, 2019.

[30] N. Rutter, R. Essery, J. Pomeroy, N. Altimir, K. Andreadis, I. Baker, A. Barr, P. Bartlett e e. al., «Evaluation of forest snow processes models (SnowMIP2),» Journal of Geophysical Research: Atmospheres, vol. 114, n. D6, 2009.

[31] G. Zolezzi, A. Bellin, M. Bruno, B. Maiolini e A. Siviglia, «Assessing hydrological alterations at multiple temporal scales: Adige River, Italy,» Water Resources Research, vol. 45, p. 1–15, 2009.

[32] U. Strasser, «Die Modellierung der Gebirgsschneedecke im Nationalpark Berchtesgaden. Modelling of the mountain snow cover in the Berchtesgaden National Park,» Berchtesgaden National Park research report, vol. 55, 2008.

[33] R. Essery, «A Factorial Snowpack Model (FSM 1.0),» Geoscientific Model Development, vol. 8, n. 12, pp. 3867-3876, 2015.

[34] F. Hanzer, K. Helfricht, T. Marke e U. Strasser, «Multi-level spatiotemporal validation of snow/ice mass balance and runoff modeling in glacierized catchments,» The Cryosphere, vol. 10, pp. 1859-1881, 2016.

[35] M. Renner, M. G. F. Werner, S. Rademacher e E. Sprokkereef, «Verification of ensemble flow forecasts for the River Rhine,» Journal of Hydrology, vol. 376, n. 3-4, pp. 463-475, 2009.

[36] L. Charrois, E. Cosme, M. Dumont, M. Lafaysse, S. Morin, Q. Libois e G. Picard, «On the assimilation of optical reflectances and snow depth observations into a detailed snow-pack model,» The Cryosphere, pp. 1021-1038, 2016.

[37] D. Günther, T. Marke, R. Essery e U. Strasser, «Uncertainties in Snowpack Simulations-Assessing the Impact of Model Structure, Parameter Choice and Forcing Data Error on Point-Scale Energy-Balance Snow Model Performance,» Water Resources Research, vol. 55, n. 4, pp. 2779-2800, 2019.

[38] D. Marks, A. Winstral, M. Reba, J. Pomeroy e M. Kumar, «An evaluation of methods for determining during-storm precipitation phase and the rain/snow transition elevation at the surface in a mountain basin,» Advances in Water Resources, vol. 55, pp. 98-110, 2013.

[39] U. Strasser, J. Corripio, B. Brock, F. Pellicciotti, P. Burlando e M. Funk, «Spatial and Temporal Variability of Meteorological Variables at Haut Glacier d'Arolla (Switzerland) During the Ablation Season 2001: Measurements and Simulations,» J. Geophys. Res., vol. 109, 2004.

[40] S. R. Fassnacht, K. A. Dressler, D. M. Hultstrand, R. C. Bales e G. Patterson, «Temporal inconsistencies in coarse-scale snow water equivalent patterns: Colorado River Basin snow telemetry-topography regressions,» PirineoRevista de Ecología de Montaña, vol. 167, pp. 165-185, 2012.

[41] J. Erxleben, K. Elder e R. Davis, «Comparison of spatial interpolation methods for estimating snow distribution in the Colorado Rocky Mountains,» Hydrological Processes, vol. 16, n. 18, pp. 3627-3649, 2002.

[42] «Snow metamorphism,» [Online]. Available: https://www.slf.ch/en/snow/snow-as-a-material/snow-metamorphism.html. [Consultato il giorno 6 May 2019].

[43] «Status for: Land Surface Temperature and Emissivity (MOD11),» [Online]. Available: https://landval.gsfc.nasa.gov/ProductStatus.php?ProductID=MOD11. [Consultato il giorno 6 May 2019].

[44] «Collection 6 - MODIS Land Surface Temperature Products - Users' Guide,» [Online]. Available: https://icess.eri.ucsb.edu/modis/LstUsrGuide/MODIS_LST_products_Users_guide_Collection-6.pdf. [Consultato il giorno 6 May 2019].

[45] «Earthdata,» [Online]. Available: https://earthdata.nasa.gov/ . [Consultato il giorno 6 May 2019].

[46] M. A. G. D. P. A. C. G. F. M. C. J. &. R. J. J. Martin, «Comprehensive in situ validation of five satellite land surface temperature data sets over multiple stations and years,» Remote Sensing, vol. 11, n. 5, p. 479, 2019.

[47] D. W. S. &. H. A. Oesch, «Snow surface temperature from AVHRR as a proxy for snowmelt in the Alps,» in EARSeL-LISSIG-Workshop Observing our Cryosphere from Space, Bern, 2002.

[48] R. Colombo, R. Garzonio, B. Di Mauro, M. Dumont, F. Tuzet, S. Cogliati e E. Cremonese, «Introducing thermal inertia for monitoring snowmelt processes with remote sensing,» Geophysical Research Letters, vol. 46, n. 8, pp. 4308-4319, 2019.

[49] «AINEVA - Modelli AINEVA per i rilievi,» [Online]. Available: https://www.aineva.it/modelli-aineva-per-i-rilievi/. [Consultato il giorno 14 May 2019].

[50] D. Castelletti, L. Pasolli, L. Bruzzone, C. Notarnicola e B. Demir, «A novel hybrid method for the correction of the theoretical model inversion in bio/geophysical parameter estimation,» IEEE Transactions on Geoscience and Remote Sensing, vol. 54, n. 8, pp. 4764-4774, 2016.

[51] K. Mardia, J. Kent e J. Bibby, Multivariate analysis, New York: USA:Academic, 1979.

[52] C. Notarnicola, M. Duguay, N. Moelg, T. Schellenberger, A. Tetzlaff, R. Monsorno, A. Costa, C. Steurer e M. Zebisch, «Snow cover maps from MODIS images at 250 m resolution, Part 1: Algorithm description,» Remote Sensing, vol. 5, n. 1, pp. 110-126, 2013.

[53] C. Notarnicola, M. Duguay, N. Moelg, T. Schellenberger, A. Tetzlaff, R. Monsorno, A. Costa, C. Steurer e M. Zebisch, «Snow cover maps from MODIS images at 250 m resolution, Part 2: Validation,» Remote Sensing, vol. 5, n. 4, pp. 1568-1587, 2013.

[54] S. Gascoin, O. Hagolle, M. Huc, L. Jarlan, J. F. Dejoux, C. Szczypta, R. Marti e R. Sánchez, «A snow cover climatology for the Pyrenees from MODIS snow products,» Hydrology and Earth System Sciences, vol. 19, n. 5, pp. 2337-2351, 2015.

[55] J. Lemmetyinen, C. Derksen, H. Rott, G. Macelloni, J. King, M. Schneebeli, A. Wiesmann, L. Leppänen, A. Kontu e J. Pulliainen, «Retrieval of Effective Correlation Length and Snow Water Equivalent from Radar and Passive Microwave Measurements,» Remote Sensing, vol. 10, n. 2, p. 170, 2018.

Chapter 4

4. SWE RETRIEVAL BY EXPLOITING COSMO-SKYMED X-BAND SAR IMAGERY AND GROUND DATA THROUGH A MACHINE LEARNING APPROACH

The main objective of this chapter[3] is to estimate Snow Water Equivalent (SWE) by jointly exploiting the information derived from X-band Synthetic Aperture Radar (SAR) imagery acquired by the Italian Space Agency COSMO-SkyMed satellite constellation in StripMap HIMAGE mode and manual SWE ground measurements. The idea is to verify the sensitivity of the backscattering coefficient at X-band to the SWE and, by means of a Support Vector Regression (SVR) algorithm, to estimate the SWE for the South Tyrol region, north-eastern Italy. The regressor is trained by exploiting about 1,000 simulated backscattering coefficients corresponding to different snowpack conditions, obtained with a theoretical model based on the Dense Media Radiative Transfer theory - Quasi-crystalline approximation Mie scattering of Sticky spheres (DMRT-QMS). Then, the performance is evaluated on the backscattering values derived from COSMO-SkyMed satellite images and using the corresponding ground measurements of SWE as references. The results show a correlation coefficient equal to 0.6, a bias of 10.5 mm and a RMSE of 51.8 mm between estimated SWE values and ground measurements. The limited performance could be related to the DMRT-QMS theoretical model used for the simulations that results to be very sensitive to snow grain size and may have generated a training dataset only partially representative of satellite derived backscattering coefficients used for testing the algorithm.

[3] This chapter has been presented at SPIE Remote Sensing conference 2019:
L. De Gregorio, F. Cigna, G. Cuozzo, A. Jacob, S. Paloscia, S. Pettinato, E. Santi, D. Tapete, C. Notarnicola, L. Bruzzone (**2019**): SWE retrieval by exploiting COSMO-SkyMed X-band SAR imagery and ground data through a machine learning approach. SPIE Remote Sensing conference, 9-12 September 2019, Strasbourg, France.

4.1 Introduction

Snow coverage plays a crucial role in mountainous regions, especially in snowmelt-driven watersheds where solid precipitation is the main water input of the hydrological cycle [1]. Snowfall accumulates as snowpack throughout the cold season by storing large volumes of water, and gradually melts releasing water during the warm season [1]. The parameter that characterizes the hydrological importance of snow cover is the snow water equivalent (SWE). An accurate estimation of the spatial and temporal distribution of SWE in mountain environments is still a relevant challenge for the scientific community.

In the last decades, the estimation of snow-related parameters by means of satellite images has become increasingly popular. The large-scale monitoring of the Earth's surface from space-borne sensors improves the limited availability and representativeness of traditional point-wise in situ measurements, thus allowing a better spatialization of land surface parameters. Optical sensors are suitable for snow cover extent estimation, but to obtain volumetric information about snowpack, microwave sensors are needed. The penetration of microwaves in the snowpack depends on the related frequency and the snow conditions, such as water content and grain size [2] [3]. Several studies demonstrated the potential use of microwave sensors for the estimation of snow physical parameters [2] [3] [4] [5] [6] [7] [8]. From the literature, current methods for retrieving SWE from space rely on passive microwave sensors [4] [5]. The exploitation of passive microwave observations from space for snow cover properties detection is appealing due to the availability of a long time series of daily observations with near global coverage, extending back almost 40 years [5]. However, the use these sensors for snow properties detection is limited by the poor spatial resolution that implies difficulty in considering mixed pixel effects over heterogeneous landscapes, such as in mountainous areas. The use of Synthetic Aperture Radar (SAR) at suitable frequencies has been suggested as a potential observation method to overcome the coarse resolution of passive microwave sensors.

Most studies concerning SAR for snow properties retrieval rely on wet snow detection or snow depth estimation, only few of them are focused on SWE estimation. Shi and Dozier [2] developed semi-empirical models for characterizing the snow–ground interaction terms, the relationships between the ground surface backscattering components, and the snowpack extinction properties at C-band and X-band. With these relationships, snow depth and optical equivalent grain size can be estimated from SIR-C/X-SAR measurements. Nagler and Rott [3] developed an algorithm for

mapping wet snow in mountainous terrain using repeat pass SAR images. Guneriussen et al [6] analyzed RADARSAT SAR images to determine the optimum image modes for snow monitoring. They found a large contrast between a wet snow cover and bare ground for high incidence angle data. This large contrast is supported by surface scattering model results by assuming wet snow has a smoother surface than bare ground. Moreover, temporal analysis shows that the backscattering coefficient of dry snow is 2–3 dB lower than for bare ground. Pettinato et al. [7] suggested a new approach to the retrieval of snow-covered areas with C-band SAR, where the snow cover fraction is retrieved by using a gradual transition between snow-free and snow-covered conditions. Sun et al. [9] used microwave scattering models to analyze the C-band SAR scattering characteristics of snow-covered areas and estimated the distribution of the SWE by exploiting SAR data and snow cover data measured in the field. Pettinato et al. [8] conducted a study on the sensitivity of X-band backscattering data from the Italian Space Agency (ASI)'s COSMO-SkyMed SAR sensors to snow characteristics. They found that X-band data contribute to SWE retrieval, provided that the snowpack is characterized by a snow depth of about 60-70 cm, i.e. SWE > 100-150 cm, and with relatively large snow grain size.

In this perspective, this work addresses SWE retrieval in the whole region of South Tyrol, in north-eastern Italy, by exploiting the information derived from both the X-band SAR imagery acquired by the COSMO-SkyMed constellation in StripMap HIMAGE mode at 3 m ground resolution and the manual ground measurements. The SWE ground data are derived from manual snow profiles achieved by expert operators and provided by the Hydrographic Office of the Autonomous Province of Bolzano. The SWE retrieval has been performed by means of a machine learning technique, namely Support Vector Regression (SVR), by exploiting as input features the backscattering values and the relative incidence angle at each ground measurements location. In order to increase the number of samples included in training dataset, further backscattering values have been simulated by using the DMRT-QMS model, an implementation of the Dense Media Radiative Transfer (DMRT) theory, based on the Quasi-Crystalline Approximation (QCA) of Mie scattering of densely packed Sticky spheres [10].

4.2 Study area and dataset

4.2.1 Study area

The considered study area is the Alpine region of South Tyrol (~7,400 km²), in north-eastern Italy (Figure 4.1). The area is almost entirely mountainous and the altitude ranges between 200 and

3905 meters a. s. l. The wide altitude range implies a great variation in snow condition and snow cover duration in the period between November and May. Three main topographical landscapes can be distinguished in South Tyrol: the alpine ridge, with the highest mountains in the north of the region; the Adige valley, which separates at the top the Alps (to the West) from the Dolomites (to the East); and the Pre-Alps and Lake Garda (to the West) from the Venetian pre-Alps (to the East). The mountain ridges act as a natural obstacle on which larger-scale weather system can be deflected or modified. These landscape characteristics influence the spatial and temporal variability of the seasonal snow cover.

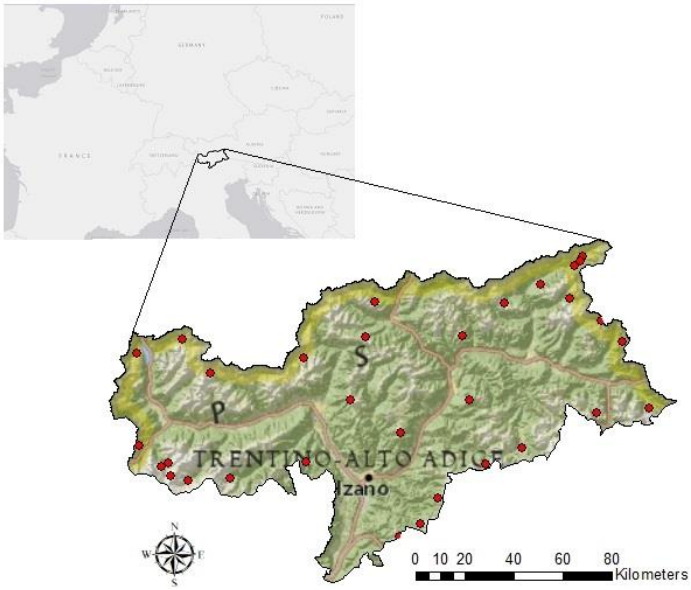

Figure 4.1 Study area: South Tyrol, Italy. Red points identify the snow measurement sites maintained by the Autonomous Province of Bolzano.

4.2.2 Data

This section describes all data involved in the study, including ground data and SAR backscattering values derived both from COSMO-SkyMed imagery and theoretical model simulations.

A. Ground measurements

In this work, the ground measurements of SWE are used as reference samples for SWE retrieval. SWE values are collected through manual measurements performed by the foresters and operators of the Avalanche Office of the Provinces of Bolzano every 2 weeks. During field campaigns, operators collect several physical parameters of the snowpack by means of stratigraphic analysis (Figure 4.2), including the density of the different layers and the depth of the snowpack. Through

the average density (ρ_s) and the depth of the snowpack (*SD*), it is possible to estimate the SWE with the following formula:

$$SWE\ (mm) = HS \cdot \rho_s \tag{4.1}$$

The sites where snow measurements are performed have to be safe and mostly representative for the surrounding environment. Measurements are supposed to be preferably carried out in flat areas or with a slope of maximum 30°. Since the presence of liquid water in the snowpack is a limiting factor in the SWE estimation at X-band (this is due to the increase of absorption and therefore the reduction of signal penetration [11]), in this study 45 ground measurements-derived values of SWE have been collected when the snowpack is supposed to be dry, i.e. during the winter months of January and February for the years 2013, 2014 and 2015. The measurements have been chosen by selecting the dates corresponding to the COSMO-SkyMed acquisitions in the study period.

Figure 4.2 SWE retrieval in a snow pit with a cylinder. Source: Institute for Snow and Avalanche Research (SLF) Davos, Switzerland.

B. COSMO-SkyMed data

The COSMO-SkyMed mission consists of a constellation of four sun-synchronous, near-polar and low-Earth orbiting midsize satellites, each equipped with a multimode high-resolution SAR operating at the X-band (9.6 GHz frequency; 3.1 cm wavelength). The SAR instruments can be operated using different beam modes which include: Spotlight (mode 2 and mode 1, with mode 1 for defence use only); StripMap (HIMAGE and PingPong); and ScanSAR (Wide Region or Huge Region) [12]. In this work, we exploited images acquired in HIMAGE mode, i.e. wide field, single

polarization imaging mode with 40 km swath and 3 m ground resolution. Further details are available on the COSMO-SkyMed SAR Products Handbook [13].

Table 4.1 shows the 14 images used in the study period and their characteristics, including orbit direction, polarization, time and date of acquisition. Each image covers only a part of the study area, therefore only the measurement sites included in that portion of study area and for which ground data are available at that date have been used. The satellite images have been pre-processed by applying radiometric calibration, multi-looking and speckle filtering for speckle reduction, and finally terrain correction for geocoding the images by correcting SAR geometric distortions. The final product has a spatial resolution of 20 meters. For each image, at the available measurement sites, the radar backscattering value and the local incidence angle have been extracted.

Table 4.1 COSMO-SkyMed StripMap HIMAGE mode HH-polarized images used for extracting the backscattering values.

ID	Satellite_name	Acquisition_start	Acquisition_stop	Orbit_direction
630884	COSMO-SkyMed-1	2015-02-05 at 04:42:12 UTC	2015-02-05 at 04:44:01 UTC	ASCENDING
227599	COSMO-SkyMed-2	2013-01-17 at 04:52:20 UTC	2013-01-17 at 04:54:01 UTC	ASCENDING
234317	COSMO-SkyMed-2	2013-02-07 at 04:46:06 UTC	2013-02-07 at 04:47:51 UTC	ASCENDING
343844	COSMO-SkyMed-2	2014-01-09 at 04:43:55 UTC	2014-01-09 at 04:45:40 UTC	ASCENDING
359414	COSMO-SkyMed-2	2014-02-26 at 04:43:39 UTC	2014-02-26 at 04:45:24 UTC	ASCENDING
439357	COSMO-SkyMed-3	2015-01-29 at 04:43:32 UTC	2015-01-29 at 04:44:02 UTC	ASCENDING
444756	COSMO-SkyMed-3	2015-02-25 at 04:48:39 UTC	2015-02-25 at 04:50:02 UTC	ASCENDING
239292	COSMO-SkyMed-4	2013-02-22 at 04:52:08 UTC	2013-02-22 at 04:53:52 UTC	ASCENDING
441790	COSMO-SkyMed-4	2015-02-12 at 04:48:25 UTC	2015-02-12 at 04:50:07 UTC	ASCENDING
233787	COSMO-SkyMed-1	2013-02-07 at 17:20:12 UTC	2013-02-07 at 17:21:29 UTC	DESCENDING
231128	COSMO-SkyMed-2	2013-01-30 at 17:20:09 UTC	2013-01-30 at 17:21:35 UTC	DESCENDING
440580	COSMO-SkyMed-2	2015-02-05 at 17:16:19 UTC	2015-02-05 at 17:17:44 UTC	DESCENDING
232124	COSMO-SkyMed-3	2013-01-31 at 17:20:02 UTC	2013-01-31 at 17:21:31 UTC	DESCENDING
227438	COSMO-SkyMed-4	2013-01-18 at 17:20:07 UTC	2013-01-18 at 17:21:37 UTC	DESCENDING

C. Model data

Further backscattering coefficients, in addition to those extracted from the satellite images, have been generated by means of a theoretical model. The model is an implementation of the Dense Media Radiative Transfer (DMRT) theory, applying the scattering model of QCA (Quasi-Cristalline Approximation) Mie of densely packed Sticky spheres (DMRT-QMS) [10]. The DMRT describes the scattering in a medium with particle fractional volume >10% (independent scattering is not valid). The model has been run for a single snow layer of identical scatterers, with a smooth air-snow interface and a rough snow-ground interface (Figure 4.3) with the objective of simulating further backscattering values, in addition to those extracted from the satellite images, with the aim to increase the dataset. Two snow parameters derived from ground measurements (snow depth, *SD*, and snow density, ρ_s) are used as inputs for model simulations. The grain diameter (d_g) and stickiness (s) values, which are not available from ground measurements and used as input parameters for the model, have been set by performing the simulations in correspondence of available ground data and COSMO-SkyMed images, from which a measured backscattering value has been derived. In this way, it has been possible to select the best values by minimizing the difference between modeled and measured backscattering values as a function of these parameters: the grain diameter has been ranged between 0.5 and 3 mm, and the stickiness between 0.1 and 0.4 [14]. The mean difference between measured and corresponding modeled backscattering values is 0.9 dB.

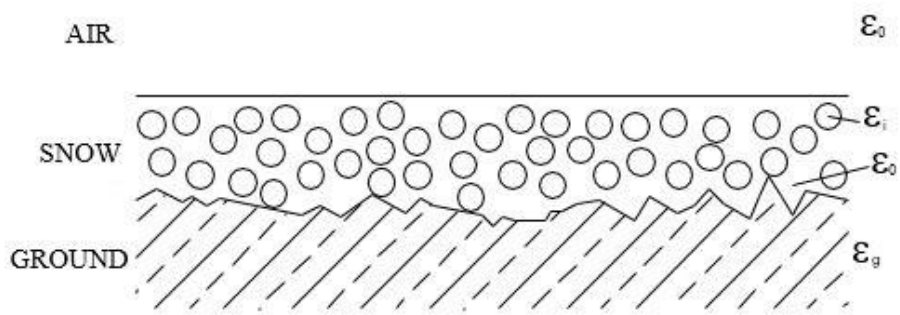

Figure 4.3 Scheme of a single layer approximation snowpack. The ε indicate dielectric constant of the medium: ε_0 is relative to the air, ε_i to the ice particles and ε_g to the ground.

4.3 Methodology

The idea of the proposed approach is to verify the sensitivity of the radar backscattering at X-band to the snow characteristics for SWE retrieval through the use of an SVR technique. The procedure is shown in the flow chart of Figure 4.4 and summarized as follows:

- After the first phase of model setting described above, the training dataset for SVR has been created by simulating backscattering coefficients. The model has been applied to about 1,300 ground measurements of SD and ρ_s in order to simulate the corresponding backscattering value. The ground measurements used for the simulations have been collected in the same dry-snow period, i.e. the months of January and February, and in the same sites of those corresponding to the satellite images and used for setting the model parameters. Therefore, it is reasonable to assume that the conditions of the snowpack are similar in all measurements. The model simulations have been iterated by randomly varying the grain diameter and the stickiness in the range obtained from the simulations in the setting phase of the model. Through this procedure, we obtained a set of backscattering coefficients for each input vector of snow parameters.
- The simulations outputs, together with the incidence angle for each measurement site derived from the COSMO-SkyMed images, have been exploited for training the SVR by using as reference the corresponding values of SWE from ground measurements.
- Subsequently, the SVR has been tested on observed data of backscattering (from COSMO-SkyMed images) and of SWE (from the ground measurements corresponding to the COSMO-SkyMed images) in order to evaluate the performance on an independent dataset.

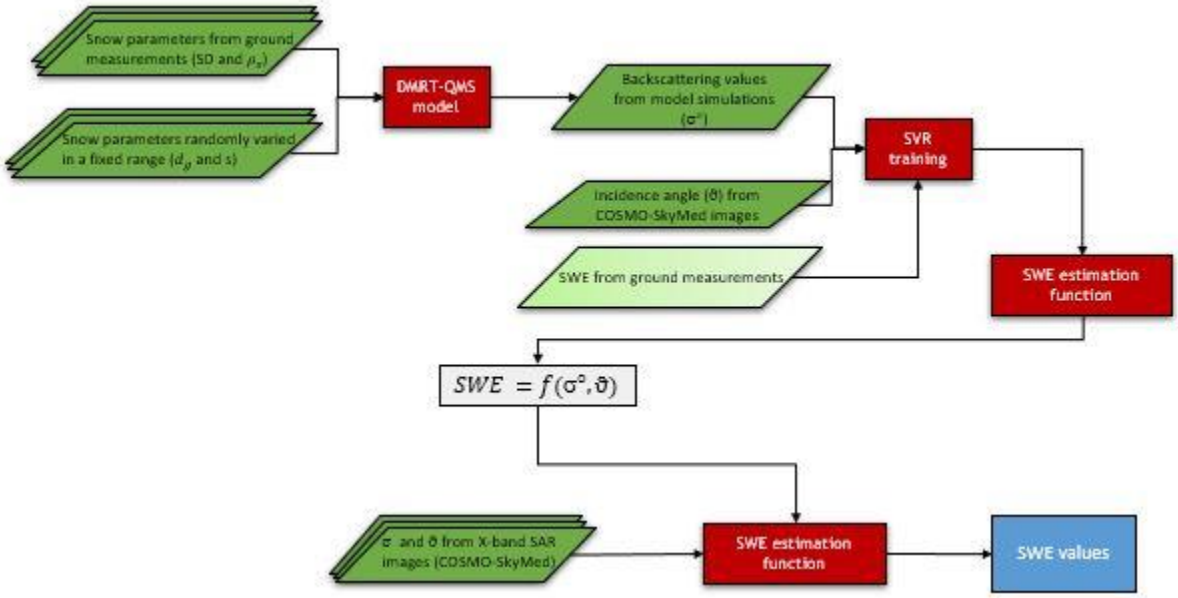

Figure 4.4 Flow chart of the procedure for SWE retrieval. σ^0 indicates the backscattering value derived from the model simulations, while σ and ϑ are the backscattering value and the incidence angle extracted by the satellite images, respectively. SVR and DMRT-QMS stand for Support Vector Regression and Dense Media Radiative Transfer theory - Quasi-crystalline approximation Mie scattering of Sticky spheres, respectively.

4.3.1 Support Vector Regression

The Support Vector Regression (SVR) is a machine learning technique developed from statistical learning theory that in the last decades has found numerous applications for the retrieval of biophysical parameters. It is a supervised non-parametric learning model used for regression (SVR) problems. Therefore, no assumptions on the underlying data distribution are necessary for the application of this technique.

The application of this technique for biophysical parameters retrieval has already been successfully tested in other works: Bruzzone and Melgani [15] presented a novel approach to the estimation of biophysical parameters from remote sensing images based on a multiple estimator system; Camps-Valls et al. [16] proposed a robust ε-Huber SVR technique for the estimation of biophysical parameters extracted from remotely sensed data; Xiao et al. [17] developed a snow-depth retrieval algorithm based SVR technique using passive microwave remote sensing data and other auxiliary data; Pasolli et al. [18] present an experimental analysis of the application of the ε-insensitive support vector regression (SVR) technique to soil moisture content estimation from remotely sensed data at field/basin scale.

4.4 Experimental Results

4.4.1 Sensitivity analysis

The sensitivity of X-band SAR to snow parameters has already been demonstrated in [8] and [14]. In this work, a sensitivity analysis has been conducted to verify the validity of this assumption for our study area. In order to have more backscattering data from COSMO-SkyMed images available, ground measurements performed on dates close (± 5 days) to the date of satellite acquisition has also been considered. The analysis has been performed by investigating the sensitivity for ascending and descending orbits separately because of the different acquisition times (i.e. at dawn and dusk; see Table 4.1) that may correspond to different snowpack conditions. The results are shown in Figure 4.5, where two behaviors can be pointed out: in the ascending mode the satellite acquisitions occur in early morning, i.e. when the snowpack is drier and the liquid water content in the snowpack is low or absent; this leads to a better correlation between SWE values and backscattering coefficients. Viceversa, in descending mode, the acquisition occurs in late afternoon, i.e. when the snowpack is potentially wetter due to surface snow melting. In this case, the scatterplot in Figure 4.5b shows lower correlation, due to the already mentioned limitation in the SWE estimation at X-band when the snowpack contains liquid water.

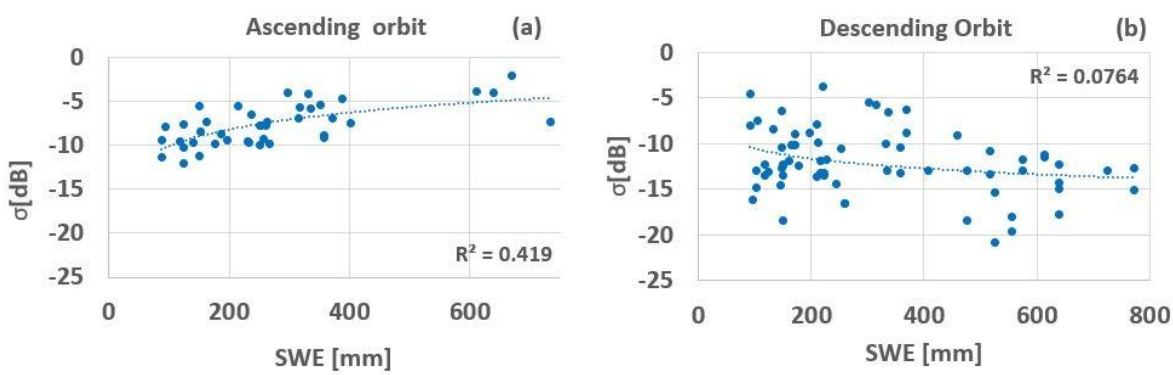

Figure 4.5 Sensitivity analysis: backscattering coefficients (σ) derived from COSMO-SkyMed StripMap HIMAGE HH-polarized images versus the corresponding SWE values derived from ground measurements for (a) ascending orbit and (b) descending orbit.

4.4.2 Performance evaluation

The SVR has been applied to a dataset of almost 1,000 simulated backscattering coefficients, selected from the initial 1,300, by considering only the samples with SWE values in the range of those of the test dataset, i.e. the dataset where backscattering coefficients are extracted from COSMO-SkyMed images. 70% of the dataset, i.e. about 700 samples, has been used to train the

regressor and the remaining 30% for a test on a different dataset to evaluate the generalization capability of the algorithm.

Table 4.2 reports different statistical metrics for both the training and the test datasets. The root mean square error (RMSE), the mean absolute error (MAE), the Pearson correlation coefficient (*R*) and the bias value are calculated between reference ground values and estimated SWE values. From the analysis of the results, we observe that the performance is consistent between the two datasets. This means that the regressor has good generalization capabilities. However, although the bias has low values in both cases, the Pearson correlation coefficients are 0.5 for the training dataset, and 0.6 for the test one, which are relatively low values.

Table 4.2 Performance of the SVR retrieval algorithm on the training and test datasets.

Dataset	RMSE (mm)	MAE (mm)	R	Bias (mm)
Training	61.2	49.1	0.5	0.8
Test	65.5	53.9	0.6	3.8

Subsequently, once the algorithm has been trained and validated, it has been tested using as input data the backscattering coefficients and the incidence angles extracted from COSMO-SkyMed images available, and as reference dataset the corresponding ground measurements derived SWE values. The performance is shown in Figure 4.6 and the results are comparable with those indicated in Table 4.2. The scatterplot shows the results of the test conducted on the backscattering coefficients extracted from the ascending mode images only that, as confirmed by the sensitivity analysis, are found to be more sensitive to the SWE parameter. The filter on the orbit direction allows improving the performance compared to what is obtained by considering all the images, i.e. ascending and descending mode together. Indeed, the MAE decreases from 46.2 to 42.2 mm, the *R* coefficient increases from 0.5 to 0.6, the slope of the regression line increases from 0.4 to 0.5, and the intercept decreases from 167.9 mm to 143.2 mm. The positive bias value indicates that, on average, the SVR overestimates the SWE values compared to the true values from ground measurements.

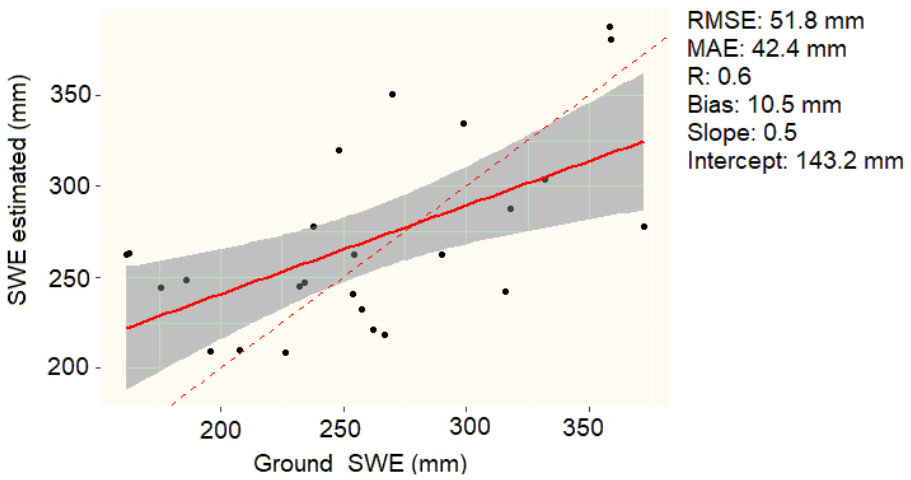

Figure 4.6 SWE estimated by the SVR regressor applied to the COSMO-SkyMed satellite data.

This limited performance could be related to two different factors. On one side, the X-band images analyzed in this chapter, are in HH polarization. This polarization is not the optimal choice for the sensitivity to snow parameters. To increase the sensitivity, dual polarization data should be considered, such as images in VV/VH polarization. The need for such data is being addressed at the site of Val Senales by a dedicated acquisition campaign with COSMO-SkyMed in StripMap PingPong mode that has been launched in mid 2019. The availability of PingPong data for this site will allow assessing the improvement in the algorithm performance that can be achieved using dual-polarization X-band data.

On the other side, the theoretical model used for simulating the backscattering coefficients can be considered a first order approximation of the interaction between the SAR signal and the snowpack. The snow parameters used as inputs for implementing the model are partially derived from ground measurements (snow depth and snow density) and partially randomly generated (snow grain size and stickiness) by varying the values in a range defined starting from ground measurements corresponding to the COSMO-SkyMed images and used for testing the SVR. However, the computation of the backscattering coefficient by the theoretical model is very sensitive to these parameters and different grain radii or stickiness values can lead to very different backscattering values. This means that the simulated dataset used for training the algorithm could be only partially representative of the dataset derived from satellite images and used for testing the algorithm.

In a the next step of our work we plan to consider the multilayer snowpack structure, by extracting the real snow grain diameters for each layer and using them for implementing the model and generating more realistic and representative backscattering coefficients to train the SVR.

4.5 Conclusions

In this work, the sensitivity of COSMO-SkyMed X-band SAR has been verified and exploited for the retrieval of SWE in the alpine region of South Tyrol, in north-eastern Italy, by means of an SVR approach. The results of the sensitivity analysis showed a better correlation between SWE and backscattering signal at X-band in ascending orbit mode with respect to those in descending orbit mode. This behavior is confirmed by the performance analysis where the error obtained by SVR in SWE estimation by exploiting ascending mode images information only is lower than that obtained if both ascending and descending modes are considered. However, the results derived by the comparison between estimated and true SWE values are not completely satisfactory, by showing a correlation coefficient of only 0.6, a bias value of 10.5 mm and a slope and intercept value of the regression line of 0.5 and 143.2 mm, respectively.

There are two main possible reasons of this performance: i) the use of X-band data in HH polarization and ii) the limited representativeness of the training dataset. Regarding the latter, the simplified modeling of the single-layer snowpack used for simulating the backscattering coefficients of the training dataset, can be improved through a more detailed description of snow characteristics and its microstructure. Moreover, the assumption at the basis of this work is that the snowpack is dry: the selected ground data, and the corresponding satellite images, have been collected in a typical dry period for alpine regions, i.e. in January and February. However, this assumption should be supported by manual measurements that confirm the dry conditions of the snowpack.

References

[1] K. Ishida, A. Ercan, T. Trinh, M. L. Kavvas, N. Ohara, K. Carr and M. L. Anderson, "Analysis of future climate change impacts on snow distribution over mountainous watersheds in Northern California by means of a physically-based snow distribution model," *Science of the total environment,* vol. 645, pp. 1065-1082, 2018.

[2] J. Shi and J. Dozier, "Estimation of snow water equivalence using SIR-C/X-SAR. I. Inferring snow density and subsurface properties," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 38, no. 6, pp. 2465-2474, 2000.

[3] T. Nagler and H. Rott, "Retrieval of wet snow by means of multitemporal SAR data," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 38, no. 2, pp. 754-765, 2000.

[4] J. Pan, M. T. Durand, V. J. B. J. and D. Liu, "Application of a Markov Chain Monte Carlo algorithm for snow water equivalent retrieval from passive microwave measurements," *Remote sensing of environment,* vol. 192, pp. 150-165, 2017.

[5] J. Lemmetyinen, C. Derksen, H. Rott, G. Macelloni, J. King, M. Schneebeli, A. Wiesmann, L. Leppänen, A. Kontu and J. Pulliainen, "Retrieval of effective correlation length and snow water equivalent from radar and passive microwave measurements," *Remote Sensing,* vol. 10, no. 2, p. 170, 2018.

[6] T. Guneriussen, H. Johnsen and I. Lauknes, " Snow cover mapping capabilities using RADARSAT standard mode data”," *Can. Remote. Sens.,* vol. 27, no. 2, p. 109–117 , 2001.

[7] S. Pettinato, E. Malnes and J. Haarpaintner, "Snow cover maps with satellite borne SAR: A new approach in harmony with fractional optical SCA retrieval algorithms," in *Proc. IEEE IGARSS*, Denver, CO, 31 jul- 4 Aug 2006.

[8] S. Pettinato, E. Santi, M. Brogioni, S. Paloscia, E. Palchetti and C. Xiong, "The potential of COSMO-SkyMed SAR images in monitoring snow cover characteristics," *IEEE Geoscience and Remote Sensing Letters,* vol. 10, no. 1, pp. 9-13, 2012.

[9] S. Sun, T. Che, J. Wang, H. Li, X. Hao, Z. Wang and J. Wang, "Estimation and analysis of snow water equivalents based on C-band SAR data and field measurements," *Arctic, antarctic, and alpine research,* vol. 47, no. 2, pp. 313-326, 2015.

[10] L. Tsang, J. Pan, D. Liang, Z. Li, D. W. Cline and Y. Tan, "Modeling active microwave remote sensing of snow using dense media radiative transfer (DMRT) theory with multiple-scattering effects," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 45, no. 4, pp. 990-1004, 2007.

[11] M. Tedesco, E. J. Kim, A. W. England, R. D. De Roo and J. P. Hardy, "Brightness temperatures of snow melting/refreezing cycles: Observations and modeling using a multilayer dense medium theory-based model," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 44, no. 12, pp. 3563-3573, 2006.

[12] F. Caltagirone, A. Capuzi, A. Coletta, G. F. De Luca, E. Scorzafava, R. Leonardi, S. Rivola, S. Fagioli, G. Angino, M. L'Abbate, M. Piemontese, E. Zampolini Faustini, A. Torre, C. De Libero and P. Esposito, "The COSMO-SkyMed dual use earth observation program: Development, qualification, and results of the commissioning of the overall constellation.," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 7, no. 7, pp. 2754-2762.

[13] "https://earth.esa.int/c/document_library/get_file?uuid=3b4bdce5-e75b-49fa-be10-113d22c86b74&groupId=10174," [Online]. [Accessed 07 08 2019].

[14] S. Paloscia, S. Pettinato, E. Santi and M. Valt, "COSMO-SkyMed image investigation of snow features in alpine environment," *Sensors,* vol. 17, no. 1, 2017.

[15] L. Bruzzone, 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, January 2005, pp. 159-174

[16] G. Camps-Valls, L. Bruzzone, J.L. Rojo-Alvarez and F. Melgani,, " Robust Support Vector Regression for Biophysical Variable Estimation from Remotely Sensed Images ", *IEEE Geoscience and Remote Sensing Letters*, Vol. 3, No. 3, pp. 339-343, 2006.

[17] X. Xiao, T. Zhang, X. Zhong, W. Shao and X. Li, "Support vector regression snow-depth retrieval algorithm using passive microwave remote sensing data," *Remote sensing of environment,* vol. 210, pp. 48-64, 2018.

[18] L. Pasolli, C. Notarnicola, L. Bruzzone, “Estimating Soil Moisture with the Support Vector Regression Technique,” *IEEE Geoscience and Remote Sensing Letters*, Vol. 8, No. 6, pp. 1080-1084, November 2011.

Chapter 5

5. CONCLUSION

5.1 Thesis Summary and General Conclusions

Although snow only occurs in certain parts of the world, it has strong effects on regional weather patterns. By studying snow distribution, snowpack characteristics and changes over time, scientists can help to improve weather forecast modeling and to learn more about the interaction between snow and local weather. Moreover, scientists also study snow cover to understand the effect of snow cover extent changes on climate and water supplies around the world.

In this context, the main objective of this thesis is presenting novel methods for the accurate estimation of two snow parameters based on the use of model-based simulations and remote sensing data and ground measurement data as reference. In particular, the focus of this work is the improvement of the snow cover extent and the SWE over the alpine Euregio area, consisting of Tyrol region (Austria), South Tyrol region and Trentino region (Italy). The proposed methods represent a relevant contribution to the snow hydrology field and the automatic estimation of snow parameters.

The automatic methods based on machine learning techniques presented in this work are capable of taking advantage from the specific properties of the different data sources, by providing a more comprehensive representation of the snowpack properties and overcoming the limits of existing approaches for snow parameters retrieval, highlighted from previous studies.

Chapter 2 presents a method for generating a snow cover extent product, starting from the snow cover maps derived from both snow model simulations and remotely sensed data, able to solve the ambiguity of the disagreement points, i.e. those points where the two data sources disagree. From this chapter it is clear the importance of combining different data sources by means of a supervised non-parametric learning model, in our case SVM. Indeed, the results demonstrate that the proposed approach benefits from the specific properties of the remote sensing data (such as independency

from meteorological observations) to compensate the weaknesses of the physical model and at the same time takes advantage from the latter (e.g., for its solid physical basis and the independency from cloud coverage) overcoming the well-known limitations of optical remote sensing. The new snow cover product has been validated by using ground data derived from manual measurements, by showing a very good agreement. The average accuracy of 90% (MODIS) and 92% (AMUNDSEN, the snow model used in this thesis) is increased to 96% in the fused product. These results encourage further research for a generalization of the proposed method, by testing it even in other regions and by exploiting other snow models. In this chapter, the properties of SVMs have been exploited. Their ability to deal with both linear and non-linear problems make them very suitable and flexible in many practical applications. Moreover, SVMs have been proved to have a higher classification accuracy than other widely used pattern recognition techniques, such as the maximum likelihood and the multilayer perceptron neural network classifiers (ref [27] chapter 2) and they also work well when only few training samples are available. Finally, an important property of SVM models is that no prior knowledge in terms of statistical distribution of the dataset is required, as they are based on the concept of margin maximization.

Chapter 3 introduces a new method to characterize and correct the deviations of a hydrological model (AMUNDSEN, in our case) from ground reference data in order to generate an improved SWE product with respect to the one derived from the model simulations. The hypothesis is that such deviations can be characterized by analyzing their behavior in the feature space, since they vary depending on the portion of the feature space. To this purpose, the correction strategy, inspired to a k-Nearest Neighbor approach, has been successfully used after an accurate feature selection by means of a genetic algorithm. The results obtained with the proposed approach are promising by showing a significant improvement of performance: on average, the RMSE and the MAE decrease from 154 to 75 mm and from 99 to 45 mm respectively, if compared to the AMUNDSEN simulations. Furthermore, the slope of the regression line between estimated SWE and ground observations increases from 0.6 to 0.9 by reducing the data spread and the number of outliers. In this work, the feature selection involved several products derived by optical remote sensing that have been tested to select those with most relevant auxiliary information for the estimation of model deviation. Only two land surface temperature derived products resulted significative by providing a marginal contribution (improvement of about 3% in the performance). These results stimulate further research in feature selection, by exploiting other satellite products,

such as higher resolution optical data or radar data. Nonetheless, experimental results obtained show an improvement in SWE estimation with respect to the simulated product, by confirming the ability of the proposed approach to handle and modeling the variability of deviations between simulations and observations in the feature space.

In this chapter the k-NN algorithm has been exploited. This algorithm is of the simplest and most popular methods for the estimation of statistical variables. Indeed, the only two parameters to tune are the distance metric and k. Moreover, k-NN algorithm a non-parametric model and this means that no prior assumption on the data are needed. To classify the new data, the k-NN algorithm reads through whole dataset to find out k-nearest neighbors. Finally, this method has already been successfully applied to biophysical parameter retrieval by Castelletti et al. (ref. 50 chapter 3).

The genetic algorithm (GA) used for the feature selection is an adaptive search technique which have demonstrated substantial improvement over a variety of random and local search methods [1]. This is accomplished by their ability to exploit accumulating information about an initially unknown search space in order to bias subsequent search into promising subspaces. GAs derive their name from the fact that they are based on models of genetic change in a population of individuals. These models consist of three basic elements: a) a Darwinian notion of "fitness," which governs the extent to which an individual can influence future generations; b) a "mating operator," which produces offspring for the next generation; and c) "genetic operators," which determine the genetic makeup of offspring from the genetic material of the parents [1].

The conclusion drawn from chapter 3 have led to the development of the method for SWE retrieval described in chapter 4. In this chapter the sensitivity of the backscattering coefficient at X-band to the SWE has been analyzed and then exploited for generating an algorithm being able to estimate this snow parameter. To this purpose, we used the X-band SAR imagery acquired by the Italian Space Agency COSMO-SkyMed constellation in StripMap HIMAGE mode together with the manual ground measurements of snow water equivalent (SWE) and theoretical model simulations. In this case, the proposed retrieval method is based on a machine learning technique (SVR) that exploits simulated backscattering coefficients for the training phase and then is tested with satellite derived data. Ground data are used as reference dataset in both training and test phases. The results derived from the sensitivity analysis show a correlation between backscattering coefficients extracted from the satellite acquisitions in ascending mode and the ground measurements of SWE,

are partially confirmed by the subsequent performances analysis of proposed method. Indeed, the results derived from the comparison between estimated and measured SWE values are not completely satisfactory and show a poor correlation and low accuracy. The experimental results obtained in this chapter encourage further analysis to improve the simulations used in the SVR training, as well as the use of more detailed information about the wetness status of the snowpack aimed at confirming the assumption of dry snow.

In this chapter, the regressive problem of SWE estimation has been addressed with SVR. Although less popular than SVM, SVR has been proven to be an effective tool in real-value function estimation. One of the main advantages of SVR is that its computational complexity does not depend on the dimensionality of the input space. Additionally, it has excellent generalization capability, with high prediction accuracy [2].

In general, the choice of methods exploited in this thesis is mainly related to the data availability and to the parameter to estimate. A well-known advantage of the SVM techniques 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 to 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 to the target variable. This ability makes these techniques very suitable in the case of snow cover area estimation, where the snow cover maps derived from two different sources, i.e. remote sensing and snow model, are exploited together with their quality measures, in order to generate an enhanced product in a decision-level fusion approach. In the case of SWE retrieval with COSMO-SkyMed images information, the ability of SVR in approximating even complex non-linear systems through the information contained in a set of reference samples is exploited. Finally, the case of SWE retrieval by using the AMUNDSEN product (cap 3) is a different situation because in this case the idea is correcting an existing product with respect to the ground reference. The error estimation has been addressed by exploiting the simplicity and the already tested capability of a k-NN algorithm (ref. 50 chapter 3).

The following table, Table 5. 1, summarizes the techniques exploited in this thesis in order to give to the reader a general vision of the machine learning techniques at the basis of the developed algorithms.

Table 5. 1 Methods exploited in the developed approach for snow parameters retrieval

Snow parameter estimated (chapter)	Method(s)	Purpose
Snow cover area (chapter 2)	SVM	To estimate the snow cover area
Snow water equivalent (chapter 3) with optical data	k-NN	To estimate the SWE
	Genetic algorithm	For feature selection
Snow water equivalent (chapter 4) with radar data	SVR	To estimate the SWE
	Theoretical model (DMRT-QMS)	For simulating the backscattering coefficients used as input in SVR

5.2 Future developments

The research activities shown in this thesis aim at developing methods that can significantly improve the capability of automatically estimating snow parameters by exploiting snow and electromagnetic models as well as remotely sensed data and auxiliary topographic features in a data fusion at decision level approach, based on machine learning techniques. On the basis of the developed methods and the experimental results obtained, some interesting future research lines can be identified.

First, since the main objective of this thesis is to develop *general* methods to improve snow parameters estimation, we aim to test the proposed methods on other study areas and by using other physically based models to confirm the robustness of the developed approaches. Another general issue to be further developed in this research is relative to ground data that have been always used as reference dataset. To make the proposed methods as general as possible, an accurate and deep validation analysis is needed, by including ground data that are representative of as many conditions as possible, e.g. by carrying out ad hoc measurement campaigns in forested areas (where available measurements are scarce) or on non-flat slopes (which are excluded from the snow measurement protocol). Moreover, also phenocam or drone images could be exploited for obtaining information at high spatial resolution of limited areas to integrate the pointwise manual measurements.

Regarding the snow cover extent product, further developments could be devoted to the use of high-resolution optical sensors. The high potential of the new generation of satellites, such as Sentinel missions, that provide open-data products with high spatial resolution, can improve the

snow detection, by catching the wide spatial snow variability in forested areas and in topographically complex terrains, typical of mountainous areas.

By focusing the attention on the method for SWE retrieval with the use of both optical satellite data and topographic parameters, two important aspects should be addressed in future research: the feature selection and the amount ground data. Regarding the former, we aim to further investigate the combination of different satellite products derived from optical sensors for creating informative feature spaces; about the latter, an improvement to the proposed approach could be achieved by increasing the dataset variability in the feature space.

Finally, attention will be devoted to the last part of this work, regarding the SWE estimation by exploiting the backscattering coefficient at X-band. In particular, we aim to improve the knowledge of the snowpack. This improvement would lead to a double benefit: on the one side, a more detailed description of snow microstructure and characteristics, used as input parameters for the model, would lead to a more precise and realistic estimation of the backscattering coefficient, used as training dataset for the SVR. On the other side, the knowledge of the wetness status of the snowpack, would be useful to verify and confirm the basic hypothesis of the proposed approach regarding the dry condition snowpack. Moreover, in our study, we only use HH polarization images. The use of dual polarization data to increase the sensitivity to SWE would represent an interesting future research direction. Another interesting aspect to study is the use of the data in L-band frequency, which, with interferometric approaches, have shown significant relationship to SWE variations. In this view, the use of the new SAOCOM satellite can be of high interest.

References

[1] K. De Jong, «Learning with genetic algorithms: An overview,» *Machine learning,* vol. 3, n. 2-3, pp. 121-138, 1988.

[2] M. Awad e R. Khanna, «Support vector regression,» in *Efficient Learning Machines*, Berkeley, CA, Apress, 2015, pp. 67-80.

6. Appendix A

In this appendix a more detailed description of the AMUNDSEN snow model is done in order to give a general view of the model used in this work to reader who is not familiar with it. Firstly, a general description of the model will be provided and then version used in this work, replacing all the snowpack thermodynamics with a different model with respect to the standard one, will be explained. More technical and analytic details of the model can be found in [1].

The modular, physically based, distributed modeling system AMUNDSEN (Alpine MUltiscale Numerical Distributed Simulation ENgine) has been designed to specifically address the requirements of snow modeling in mountain regions under climate change conditions and has already been extensively validated in several Alpine sites [1] [2] [3] [4] [5]. It provides distribute time series of snow process variables employing a wide range of interpolation, parametrization and simulation procedures. The input parameters needed for the model implementation are [1]: a DEM of the investigation domain with a spatial resolution typically of tens to hundreds of meters (the relatively high resolution is necessary for adequately capturing the small-scale processes shaping the snow cover in complex terrain); hourly to 3-hourly recordings of the meteorological variables, such as air temperature, relative humidity, precipitation, global radiation, and wind speed; various other spatial input fields, such as land cover, soil, catchment boundaries, canopy height) in order to run specific submodules (canopy module, evapotranspiration, runoff); finally, several derived topographic parameters (slope, aspect, sky-view factor, openness) can either be preprocessed or calculated during runtime.

For an accurate simulation of snow cover heterogeneity and dynamics in mountain environment, both adequate model algorithm and meteorological input variables are required. Indeed, the understanding of spatial and temporal variability of meteorological variables is crucial for the description of the interaction between climate and the energy and mass balance of snow cover. This variability, relevant for snow accumulation and ablation, is often extrapolated from the observation available at the station in the study area or, alternatively, by using the simulations from a meteorological model.

Spatial interpolation of meteorological variables

In AMUNDSEN, distributed meteorological variables are computed hourly from stations recording, as follows: firstly, the so-called *trend field* is derived by calculating linear regression of the meteorological observations with elevation; this regression is then applied to the whole area represented by the DEM. Then, the residuals (i.e. the deviation of measurements from the trend field) are spatially interpolated by applying an *Inverse Distance Weighting* (IDW) approach, resulting in a so-called *residual field*, which represents the local deviation of a specific meteorological variable from its value in the trend field. The weights are the inverse distances between any location defined by a DEM pixel and the stations. This algorithm ensures the spatialization of the station observations and can be applied regardless of whether a relation of the meteorological variable with elevation exist. Indeed, the algorithm is applied for each time step by ensuring that for each variable the variation rate is dynamically adjusted according to the observation for that time step [1]. This procedure is applied for the following meteorological variables: temperature, wind speed, shortwave incoming radiation and precipitation. Relative humidity (%) is converted into absolute humidity ($kg \cdot m^3$), then spatial interpolated and re-converted afterwards. It is worth to note that all interpolations within a time step can be limited in the range between the minimum and maximum values of the observation for that time step, to avoid the generation of unrealistic extrapolation values [1].

Radiation terms

In this section, a short description of radiation terms is provided.

Incoming shortwave radiation

All shortwave radiation components are derived from local terrain characteristics and from physical and empirical relations described in the following.

The algorithms compute the effects of shading by the surrounding terrain, the decrease of atmospheric transmittance due to the individual processes of scattering, and multiple reflections between the atmosphere and the ground as well as reflections from surrounding terrain. The latter aspect is very important for high mountain regions, where slope reflections can considerably increase incoming shortwave radiation, in particular if these slopes are covered with snow [6].

Direct and diffuse shortwave radiative fluxes for each grid cell are parameterized using efficient vectorial algebra algorithms, the principles of which are described in detail in [6]. First, potential direct solar radiation for a clear sky is determined for each grid cell for its specific geographical position and date of simulation, considering the different effects influencing incoming radiation, i.e. the hill shading, the transmission losses due to scattering (Rayleigh and aerosol scattering) and absorption (by water vapor, ozone, and other trace gases), the transmission gains due to multiple reflections between the atmosphere and the ground, and reflections from surrounding terrain.

Shading by adjacent terrain is computed by scanning the projection of cells onto a solar illumination plane perpendicular to the sun direction. By checking the projection of a grid cell over this plane, it is determined whether a point is in the sun or in the shade of another cell.

The influence of clouds is accounted for through the use of the cloud factor which is computed as fraction of observed global radiation to the respective simulation result representing the clear sky situation. If more than one observation is available (more meteorological stations) in a single model timestep, then the cloud factor is spatially distributed over the domain in the same way as the meteorological variables [1].

Incoming longwave radiation

The majority of the longwave radiation reaching the surface is emitted from the lowest layers of the atmosphere which are not necessarily correctly represented by measurements of temperature at the 2 m level [7]. Therefore, the longwave radiative flux emitted can be different depending on the location, on the mixture conditions in the boundary layer, on the height of a potential inversion layer and a subsequent different temperature profile.

In AMUNDSEN model incoming longwave radiation is also derived following [6] using parameterizations for the radiation fractions coming from the clear sky, from clouds, and from surrounding slopes. These three phenomena have to be quantified and modeled for improving the final snow cover estimation. Indeed, spatial differences in the incoming longwave radiative flux can be attributed mainly to three phenomena: (i) differences in air temperature and its vertical profile, (ii) differences in the effect of north-facing slopes and (iii) differences in the effect of clouds.

Snowpack thermodynamics

The module relative to the snowpack thermodynamics, i.e. everything about surface energy-balance, cold content or albedo, in this study has been implemented with a different model with respect to the standard one, by exploiting the factorial snowpack model (FSM) implemented by Essery [8]. The open-source FSM solves the coupled mass and energy balance of a snowpack in a control volume of 1-m^2 surface area and height H_s. A maximum of three snow layers can be selected by the user and the total snow depth governs the number and thickness of the snow layers [8]. The model can be run in 32 different configurations of varying complexity by switching on or off independently the five process parameterizations shown in Table A1 [8]. For each process FSM allows a simpler representation (option 0) and a more complex one (option 1). A more analytical and detailed description of the parameterized processes is done in [8].

Table A1 Process parameterizations available in the FSM

Process	Implementation	Option
Absorption of solar radiation	Snow albedo evolution	Function of surface temperature (0) decays with time (1)
Heat conduction in snow	Thermal conductivity	Constant (0) Function of snow density (1)
Compaction of snow	Snow density	Constant (0) Compaction (1)
Transfer of heat from the air to snow	Correction for atmospheric stability	Off (0) On (1)
Transport of liquid water	Snowpack hydraulics	Immediate drainage (0) Bucket model (1)

Regarding the precipitation, two main contributes are considered in the FSM: rainfall and snowfall. The solid mass fluxes at the surface are snowfall (or deposition of wind-blown snow) S_f and sublimation E. Solid mass fluxes between layers are included because redistribution of mass is required by the discretization when the snow depth changes. The liquid mass fluxes into and out of the snow column are rainfall R_f and melt M at the surface and runoff R_b at the base of the snowpack; evaporation of liquid water in the snow is neglected. The relatively short run time and

the rather small set of model parameters make FSM ideal to investigate various model configuration and parameter settings and for this reason it has been used in this work for simulating the snowpack processes within the AMUNDSEN model.

References

[1] U. Strasser, «Modelling of the mountain snow cover in the Berchtesgaden National Park,» Berchtesgaden, 2008.

[2] U. Strasser, J. Corripio, F. Pellicciotti, P. Burlando, B. Brock and M. Funk, «Spatial and temporal variability of meteorological variables at Haut Glacier d'Arolla (Switzerland) during the ablation season 2001: Measurements and simulations,» *J. Geophys. Res.,* vol. 109, n. D03103, 2004.

[3] F. Pellicciotti, B. Brock, U. Strasser, P. Burlando, M. Funk and J. Corripio, «An enhanced temperature-index glacier melt model including the shortwave radiation balance: Development and testing for Haut Glacier d'Arolla, Switzerland,» *J. Glaciol.,* vol. 51, p. 573–587, 2005.

[4] U. Strasser, M. Bernhardt, M. Weber, G. E. Liston and W. Mauser, «Is snow sublimation important in the alpine water balance?,» *The Cryosphere,* vol. 2, pp. 53-66, 2008.

[5] T. Marke, U. Strasser, F. Hanzer, J. Stötter, R. A. I. Wilcke and A. Gobiet, «Scenarios of Future Snow Conditions in Styria (Austrian Alps),» *J. Hydrometeorol.,* vol. 16, p. 261–277, 2015.

[6] J. G. Corripio, «Vectorial algebra algorithms for calculating terrain parameters from DEMs and solar radiation modelling in mountainous terrain,» *International Journal of Geographical Information Science,* vol. 17, n. 1, pp. 1-23, 2003.

[7] A. Ohmura, «Physical basis for the temperature-based melt-index method,» *Journal of applied Meteorology,* vol. 40, n. 4, pp. 753-761, 2001.

[8] R. Essery, «A factorial snowpack model (FSM 1.0),» *Geosci. Model Developm.,* vol. 8, pp. 3867-3876, 2015.

DOCTORAL PROGRAM IN INFORMATION AND COMMUNICATION TECHNOLOGY

Doctoral candidate

Ludovica De Gregorio

Cycle	**Cycle** 31
Thesis	**Title** Development of new data fusion techniques for improving snow parameters estimation
Advisor	**Name Surname (University/Institute):** Lorenzo Bruzzone (University of Trento)
Co-advisor	**Name Surname (University/Institute):** Dr. Claudia Notarnicola (Eurac Research- Institute for Earth Observation)

1. List of publications

[Papers already published or in press - papers under review can be also included but in a separate list]

Journals (already published)

1. **De Gregorio, L.**, Günther, D., Callegari, M., Strasser, U., Zebisch, M., Bruzzone, L. and Notarnicola, C. (**2019**). Improving SWE Estimation by Fusion of Snow Models with Topographic and Remotely Sensed Data. *Remote Sens.,* 11, 2033. http://dx.doi.org/10.3390/rs11172033.

2. **De Gregorio, L.**, Callegari, M., Marin, C., Zebisch, M., Bruzzone, L., Demir, B., Strasser, U., Marke, T., Günther, D., Nadalet, R. and Notarnicola, C. (**2019**). A Novel Data Fusion Technique for Snow Cover Retrieval. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. DOI: 10.1109/JSTARS.2019.2920676.

3. **L. De Gregorio**, M. Callegari, P. Mazzoli, S. Bagli, D. Broccoli, A. Pistocchi, C. Notarnicola (**2018**): Operational River Discharge Forecasting with Support Vector Regression Technique Applied to Alpine Catchments: Results, Advantages, Limits and Lesson Learned, *Water Resource Management*, Vol 32, pp 229-242, doi: https://doi.org/10.1007/s11269-017-1806-3

4. Xie, J., Kneubühler, M., Garonna, I., de Jong, R., Notarnicola, C., **De Gregorio, L.**, & Schaepman, M. E. (**2018**). Relative influence of timing and accumulation of snow on alpine land surface phenology. *Journal of Geophysical Research: Biogeosciences*, 123(2), 561-576. https://doi.org/10.1002/2017JG004099

5. S. Asam, M. Callegari, M. Matiu, G. Fiore, **L. De Gregorio**, A. Jacob, A. Menzel, M. Zebisch, C. Notarnicola (**2018**). Relationship between Spatiotemporal Variations of Climate, Snow Cover and Plant Phenology over the Alps - An Earth Observation-Based Analysis. *Remote Sensing Journal*

(Special Issue Mountain Remote Sensing), Vol. 10 (11), 1757; doi: https://doi.org/10.3390/rs10111757

6. J. Xie, M. Kneubühler, I. Garonna, C. Notarnicola, **L. De Gregorio**, R. De Jong, B. Chimani, M. Schaepman (**2017**): Altitude-dependent influence of snow cover on alpine land surface phenology, *Journal of Geophysical Research: Biogeosciences*, https://doi.org/10.1002/2016JG003728.

Conferences

1. C. Notarnicola, S. Asam, M. Callegari, **L. De Gregorio**, F. Greifeneder, C. Marin, R. Sonnenschein, M. Zebisch (**2016**): Monitoring ecosystems dynamics in alpine environments using Sentinel data, ESA Living Planet Symposium 2016, 9–13 May 2016, Prague, Czech Republic.
2. S. Asam, M. Callegari, **L. De Gregorio**, A. Jacob, C. Notarnicola, M. Zebisch, M. Matiu, A. Menzel, G. Fiore (**2017**): Spatiotemporal variations of alpine climate, snow cover and phenology. 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images, June 27-29 2017, Bruges, Belgium.
3. **L. De Gregorio**, M. Callegari, C. Marin, M. Zebisch, L. Bruzzone, B. Demir, U. Strasser, D. Günther, C. Notarnicola (**2018**): Fusion of remote sensing and hydrological model data for improving snow mapping. SnowHydro - 1st International Conference on Snow Hydrology, 12-15 February 2018, Heidelberg, Germany.
4. M. Callegari, **L. De Gregorio**, F. Greifeneder, C. Marin, M. Petitta, A. Troccoli, M. Zebisch, C. Notarnicola (**2018**): A machine learning approach to exploit snow remote sensing data for run - off prediction. SnowHydro - 1st International Conference on Snow Hydrology, 12-15 February 2018, Heidelberg, Germany. (poster)
5. **L. De Gregorio**, M. Callegari, C. Marin, M. Zebisch, L. Bruzzone, B. Demir, U. Strasser, T. Marke, D. Günther, M. J. Polo, M. J. Pérez-Palazón, C. Notarnicola (**2018**): Integration of a hydroclimatological model and remote sensing products for improving snow cover mapping in mountain areas. Remote Sensing & Hydrology Symposium, 8-10 May 2018, Córdoba, Spain.
6. R. Pimentel, C. Marin, **L. De Gregorio**, M. Callegari, M. J. Pérez-Palazón, C. Notarnicola, M. J. Polo (**2018**): Validating improved-MODIS products from spectral mixture-Landsat snow cover maps in a mountain region in southern Spain. Remote Sensing & Hydrology Symposium, 8-10 May 2018, Córdoba, Spain.
7. **L. De Gregorio**, M. Callegari, C. Marin, M. Zebisch, L. Bruzzone, B. Demir, U. Strasser, D. Günther, T. Marke, C. Notarnicola (**2018**): A novel data fusion technique for snow parameter retrieval. IGARSS, 22-27 July 2018, Valencia, Spain.
8. **L. De Gregorio**, M. Callegari, Carlo Marin, M. Zebisch, L. Bruzzone, B. Demir, U. Strasser, T. Marke, D. Günther, R. Nadalet, C. Notarnicola (**2018**): Improved snow parameters estimation through integration

of simulated and remotely sensed snow cover information. International Snow Science Workshop (ISSW), 7-12 October 2018, Innsbruck, Austria.

9. Pimentel, R., Marín, C., **De Gregorio, L.**, Callegari, M., Pérez-Palazón, M. J., Notarnicola, C., and Polo, M. J. (**2018**): Validating improved-MODIS products from spectral mixture-Landsat snow cover maps in a mountain region in southern Spain, Proc. IAHS, 380, 67-72, https://doi.org/10.5194/piahs-380-67-2018.
10. M. Callegari, G. Bertoldi, C. Brida, F. Bovolo, L. Bruzzone, **L. De Gregorio**, D. Günther, C. Marin, T. Marke, M. Niroumand-Jadidi, I. Podsiadło, P. Rastner, U. Strasser, M. Zebisch, C. Notarnicola (**2019**): Integration of hydrological model outputs with remote sensing products to improve cryosphere process understanding. ESA Living Planet Symposium 2019, 13-17 May 2019, Milan, Italy. (poster)
11. R. Barella, M. Callegari, C. Marin, M. Zebisch, **L. De Gregorio**, R. Dinale, S. Benetton, R. Sailer, S. Galos, C. Klug, L. Nicholson, C. Notarnicola (**2019**): A Novel automatic method for glacier outlines extraction from Sentinel data. ESA Living Planet Symposium 2019, 13-17 May 2019, Milan, Italy. (poster)
12. **De Gregorio, L.**, Callegari, M., Marin, C., Zebisch, M., Bruzzone, L., Demir, B., Strasser, U., Marke, T., Günther, D., & Notarnicola, C. (**2019**): Improved snow parameters estimation through integration of simulated and remotely sensed snow cover information. CRYOMON-SciPro project workshop, 23 May 2019, Bolzano, Italy.
13. **De Gregorio, L.**, Callegari, M., Marin, C., Zebisch, M., Bruzzone, L., Strasser, U., Marke, T., Günther, D., & Notarnicola, C. (**2019**): New fusion method for improving SWE estimation from snow simulation by exploiting satellite products. CRYOMON-SciPro project workshop, 23 May 2019, Bolzano, Italy.
14. S. Pettinato, S. Paloscia, E. Santi, E. Palchetti, C. Notarnicola, M. Callegari, C. Marin, **L. De Gregorio**, D. Tapete, F. Cigna (**2019**): Exploiting the synergy between Sentinel-1 and COSMO-SkyMed data for snow monitoring in alpine areas. IGARSS, July 28 - August 2, 2019, Yokohama, Japain.
15. **L. De Gregorio**, F. Cigna, G. Cuozzo, A. Jacob, S. Paloscia, S. Pettinato, E. Santi, D. Tapete, C. Notarnicola, L. Bruzzone (**2019**): SWE retrieval by exploiting COSMO-SkyMed X-band SAR imagery and ground data through a machine learning approach. SPIE Remote Sensing conference, 9-12 September 2019, Strasbourg, France.

16.

2. Research/study activities

[Describe your research/study activities during the Doctoral programme. It may include (but not limited to):

- Awards
- Research and training periods spent abroad
- Participation in international mobility programs
- Participation in research projects
- Teaching activities

Conference/summer school/internship	from	to	Venue	Country
SnowHydro - 1st International Conference on Snow Hydrology	12/02/2018	15/02/2018	University of Heidelberg	Heidelberg, Germany
IGARSS	22/7/2018	27/7/2018	Feria Valencia-Convention & Exhibition Center	Valencia, Spain
International Snow Science Workshop (ISSW)	7/10/2018	12/10/2018	Congress Innsbruck	Innsbruck, Austria
CRYOMON-SciPro project workshop	23/5/2019	23/5/2019	Eurac Research	Bolzano, Italy

- Collaboration in CRYOMON-SciPro project
- Collaboration in Algoritmi project