

Data Descriptor

Not peer-reviewed version

Two datasets over South Tyrol and Tyrol areas to understand and characterize water resource dynamics in mountain regions

[Ludovica De Gregorio](#)*, [Giovanni Cuzzo](#)*, [Riccardo Barella](#), Francisco Corvalán, Felix Greifeneder, Peter Grosse, [Abraham Mejia-Aguilar](#), [Georg Niedrist](#), Valentina Premier, [Paul Schattan](#), [Alessandro Zandonai](#), and [Claudia Notarnicola](#)

Posted Date: 27 September 2024

doi: 10.20944/preprints202409.2161.v1

Keywords: remote sensing; hydrology; snow cover fraction; soil moisture



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Data Descriptor

Two Datasets over South Tyrol and Tyrol Areas to Understand and Characterize Water Resource Dynamics in Mountain Regions

Ludovica De Gregorio ^{1,*}, Giovanni Cuozzo ^{1,*}, Riccardo Barella ¹, Francisco Corvalán ², Felix Greifeneder ³, Peter Grosse ⁴, Abraham Mejia-Aguilar ¹, Georg Niedrist ¹, Valentina Premier ¹, Paul Schattan ^{5,6}, Alessandro Zandonai ¹ and Claudia Notarnicola ¹

¹ Eurac Research, Viale Druso 1, Bolzano/Bozen, 39100, Italy

² Universidad Nacional de Cuyo, Argentina

³ Chloris Geospatial, 399 Boylston Street, Suite 600, Boston, MA 02116 USA

⁴ Institute of Environmental Science and Geography, University of Potsdam, Karl-Liebknecht-Straße 24/25, 14476 Potsdam, Germany

⁵ Institute of Geography, University of Innsbruck, Innrain 52f 6020 Innsbruck, Austria

⁶ Institute of Hydrology and Water Management (HyWa), BOKU University, Gregor-Mendel-Straße 33, 1180 Wien, Austria

* Correspondence: ludovica.degregorio@eurac.edu; giovanni.cuozzo@eurac.edu

Abstract: In this work, we present two datasets for specific areas located in South Tyrol (Italy) and in Tyrol (Austria) that can be exploited to monitor and understand water resource dynamics in mountain regions. The idea is to provide the reader with information about the different sources of water supply over five defined test areas. The Snow Cover Fraction (SCF) and Soil Moisture Content (SMC) datasets are derived from machine learning algorithms based on remote sensing data. Both SCF and SMC products are characterized by a spatial resolution of 20 m and are provided for the period from October 2020 to May 2023 (SCF) and from October 2019 to September 2022 (SMC) respectively, each covering 3 seasons of interest, winter for SCF and spring-summer for SMC. For SCF maps, the validation with very high-resolution images shows high correlation coefficients of around 0.9. The SMC products were originally produced with an algorithm validated at global scale, but here, to obtain more insight in the specific alpine mountain environment, the values estimated from the maps are compared with ground measurements of automatic stations located at different altitudes and characterized by different aspects in the Val Mazia catchment in South Tyrol (Italy). In this case a MAE between 0.05 and 0.08 and an unbiased RMSE between 0.05 and 0.09 m³·m⁻³ were achieved. The datasets presented can be used as input for hydrological models as well as to hydrologically characterize the study alpine area starting from different sources of information.

Dataset: the dataset is submitted and will be published as a supplement to this paper in the journal Data

Dataset License: CC-BY-4.0

Keywords: remote sensing; hydrology; snow cover fraction; soil moisture

1. Summary

The knowledge on the state of individual components of the hydrological cycle is essential for accurate water resources management providing an important tool for climate change adaptation. Snow, for example, plays a key role in the hydrological cycle and is an important indicator of climate change [1,2]. The large amount of water stored in the snowpack during the winter season is crucial for spring runoff, impacting the downstream agricultural production, which relies on irrigation [3]. Acquiring information on the snowpack, which in turn gives information on the amount of water stored, is therefore crucial to improving knowledge of water availability in mountain basins. In the

same way, the soil moisture content (SMC) represents a crucial state variable in the global cycles of water, energy and carbon and it therefore plays a key role in the study of Earth's climate and weather. However, on one side direct measurement of these geophysical parameters by means of field measurements is not always possible, especially in inaccessible areas with complex topography, typical of mountain catchments. On the other side, although automatic stations (e.g., meteorological stations) can provide accurate observation data for a long time series [4,5], the number of stations in mountainous regions remains low [6], impeding a spatially distributed monitoring of these variables.

The scarcity of ground measurements also affects the accuracy of model-based estimation of geophysical parameters, as these measurements represent the input variables for the models. Especially in remote areas, where the availability of observed meteorological data on the ground is very limited, potential errors can be introduced on a regional scale [7]. Thus, collecting snow and soil moisture observations and modeling the snowpack and the soil moisture content can be extremely challenging due to its high spatial and temporal variability in mountainous areas.

Beyond the more traditional modelling approaches and in situ measurements, the great development of techniques to detect geophysical parameters from remote sensing images has significantly contributed to improve snow and soil moisture mapping and to provide spatially continuous measurement over large and remote areas. Traditionally, snow-cover mapping techniques identify pixels as binary snow cover, i.e., either snow-covered or snow-free. However, the binary snow cover area (hereafter referred to as SCA) classification does not efficiently capture sub-pixels characteristics. The variability of snow at sub-grid level requires, instead, the knowledge of the fractional snow cover and its distribution as accurately as possible to integrate this information in numerical models simulating the hydrological or atmospheric surface energy exchange processes [8].

The Fractional Snow Cover (FSC), that is the snow-covered fraction of the pixel area, naturally provides finer information than the binary SCA, which may be insufficient to characterize the snow distribution in areas where partially snow-covered (mixed) pixels are prevalent [9].

Regarding the SMC, the currently available operational products rely on data from coarse-medium resolution passive or active microwave sensors like SMOS (passive) [10], SMAP (passive) [11], or ASCAT (active) [12].

In this context, this paper aims to present the datasets developed in the framework of the project ACR_Water (Assimilating Cosmic-Ray Neutron and Remote Sensing Data for Improved Water Resource Management), whose objective is to improve the analysis of mountain water resources exploiting in a synergic way remote sensing and cosmic ray neutron sensors data. To overcome the limitations of conventional approaches in mountain regions, a multi-dataset assimilation framework based on complementary information sources is proposed. The project exploits the strength-driven combination of spatial patterns of SCF and top layer SMC with temporally continuous Cosmic-Ray Neutron Sensing (CRNS) observations enabling a new quality of water resource analysis. The datasets presented in this paper involve the time series of SCF and SMC maps: the provided data can be used, for example, in the cryosphere sciences (regarding the SCF data), but also for hydrological or climatological purposes (e.g. hydrological models).

2. Data Description

2.1. Study Area

The study area covers the surroundings of the 5 CRNS exploited in the project, which are located in South Tyrol (Italy) and Tyrol (Austria): Corvara, Weisssee, Leutasch, Dresdner Hütte and Obergurgl. The CRNS probes, that are geographically distributed around the study area, are in regions with different topographical, geological, and meteorological characteristics. The probes are situated in the Easatern Alps within the catchments of the Inn, Adige and Leutascher Ache Rivers. The CRNS locations are shown in **Error! Reference source not found.** where the red lines delimit the study area and separate the North (and East) Tyrol (Austria) from the Southern part (Italy).

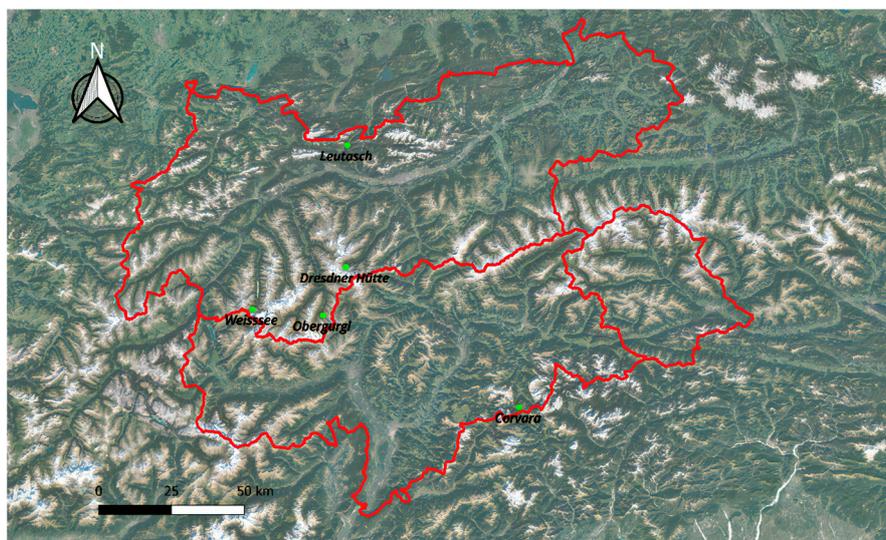


Figure 1. Study area and the five CRNS probes: Leutasch, Dresdner Hütte, Obergurgl and Weissee are located in North Tyrol (Austria), while Corvara is in South Tyrol (Italy).

2.2. Data Description

The SCF and SMC maps are provided as percentage (representing the percentage of snow and the soil moisture content per pixel, respectively) in NetCDF format. The time series cover the period from October 2019 to September 2022 for the SMC product and from October 2020 to May 2023 for the SCF product. The maps have a spatial resolution of 20 meters.

The characteristics of the datasets are reported in **Error! Reference source not found..** The maps were generated for the areas surrounding the CRNSs, whose characteristics are listed in **Error! Reference source not found..**

Table 1. Characteristics of the datasets.

<i>Characteristic</i>	<i>Snow Cover Fraction dataset</i>	<i>Soil Moisture Content dataset</i>
Name	Snow cover fraction	Soil Moisture Content
Data type	Integer (0-100=% pixel snow coverage; 255= No data)	Integer (0-100=% soil moisture content; 0 = No data)
Data format	NetCDF	NetCDF
Projection	WGS 84 / UTM zone 32N	WGS 84 / UTM zone 32N
Spatial coverage	10x10 km surrounding the CRNSs	10x10 km surrounding the CRNSs
Spatial resolution	20 m	20 m
Temporal coverage	2020-2023	2019-2022
Temporal resolution	approximately every 5 days (depending on the cloud coverage)	On average approx. a map every 2-3 days. After 23 rd of December 2021, only half of the maps are available due to the failure of S1-B satellite. Each satellite pass only partially covers the study area

Table 2. CRNS coordinates and altitude.

<i>CRNS name</i>	<i>Latitude</i>	<i>Longitude</i>	<i>Altitude [m]</i>
Corvara	46.544	11.900	1924
Leutasch	47.376	11.162	1111
Weissee	46.873	10.714	2464
Dresdner Hütte	46.997	11.140	2293
Obergurgl	46.849	11.031	2644

3. Methods

3.1. Snow Cover Fraction (SCF) Algorithm

The input data used for snow cover retrieval consists of Sentinel-2 (S2) images. The constellation is comprised of two twin satellites, namely S2-A and S2-B, which together provide a revisit frequency of 5 days in the area of interest. The study area is covered by five tiles: 32TNT, 32TPT, 32TPS, 32TQT, and 32TQS. The S2 Level 1C spectral bands are downloaded from CREODIAS (<https://creodias.eu/>) and appropriately preprocessed in preparation for the core Snow Cover Fraction (SCF) algorithm. Three main steps are applied: i) conversion from digital number (DN) to Top of the Atmosphere (ToA) reflectance values using the quantification value provided in the Sentinel-2 metadata, ii) resampling of all bands to a spatial resolution of 20 m using cubic interpolation and cropping to the area of interest. Here, the raw S2 grid and coordinate reference systems are preserved to minimize changes as much as possible; iii) cloud masking using the S2 cloudless algorithm, which is available at <https://github.com/sentinel-hub/sentinel2-cloud-detector> [13]. Scenes with a cloud coverage greater than 50% are excluded from further processing.

The SCF algorithm retrieval is based on a Support Vector Machine (SVM) classification that is manually trained using an Active Learning (AL) procedure [14].

All spectral bands are utilized as features, along with informative indices such as the Normalized Difference Vegetation Index (NDVI), which is calculated as the normalized difference between the near-infrared (NIR) and red bands, as well as the Normalized Difference Snow Index (NDSI), which is calculated as difference between the green and the shortwave infrared (SWIR). The SVM model utilizes a radial basis function kernel. To determine the model parameters, we employ a grid search strategy to identify the regularization parameter C and the kernel coefficient gamma. The grid is initialized with a user-defined range. The model selection process begins with a coarse grid and based on the obtained results, is refined around the values of C and gamma that performed the best. The best values are selected by evaluating the mean and standard deviation of the accuracy calculated in a cross-validation strategy with k-fold validation (k = 5).

The SVM model is trained exclusively with pure pixels under different illumination conditions i.e., diffuse light, direct light, and shadow. This is accomplished through visual inspection of the spectral signatures of the collected training samples, as well as considering the characteristics of adjacent pixels such as topographical features and vegetation in the surrounding area. The selected training data should represent distinct classes of "snow" and "snow-free" accurately, as they play a crucial role in defining the hyperplane, which is the decision function that separates the two classes. In detail, the trainings that define the hyperplane are the so-called support vectors. Once the hyperplane is defined, we can establish a correlation between the SCF and the distance to the hyperplane. This linear relationship is observed by considering an external SCF reference, specifically a WorldView (WV) image.

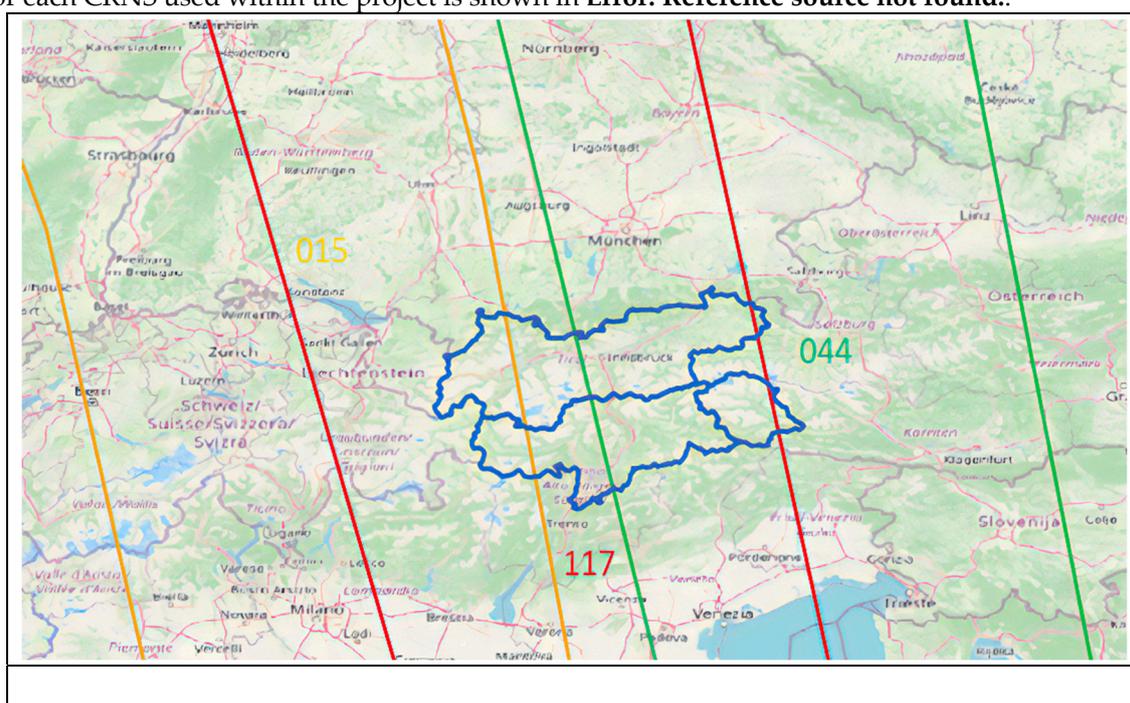
The AL procedure helps accelerate the learning process of the classification by involving the user in collecting training samples iteratively. The training selection was performed ad hoc for each scene and iteratively by the user, visually assessing the results for each scenario. Consequently, different trainings might be selected for different scenes, as well as different model parameters. This approach ensures a scene-specific classification tailored to the final purpose of achieving the highest possible accuracy for the intended product assimilation.

In previous studies [15,16], a similar approach was applied. However, previous studies focused on simple classification, while here we estimate SCF (Snow Cover Fraction). Additionally, an improved version of this algorithm is presented by [17], that is based on the automatic selection of training data. This represents a big advantage when considering large areas and long time-series of data, while it is preferable to collect ad hoc trainings for achieving better performances.

3.2. Soil Moisture Content (SMC) Algorithm

The primary input data for the Soil Moisture Content Estimation consists of Sentinel-1 (S1) images. These satellite data are Synthetic Aperture Radar (SAR) images, which take advantage of providing, after appropriate preprocessing operations, a 2-D backscattered image of the terrain independent from weather conditions and daylight. The joint use of the two satellites S1-A and S1-B made it possible to obtain a repeat cycle of 6 days over the area of interest with SAR Interferometric Wide (IW) swath mode. The dual-polarization (VV + VH) Ground Range Detected products have been used as SAR input data for the soil moisture estimation. However, Copernicus Sentinel-1B encountered an anomaly related to the power supply of the instrument's electronics on 23 December 2021. Since then, the satellite has no longer been able to provide radar data. As a result of this failure, the only available satellite has become S1-A, and consequently, the repeat cycle has increased to 12 days.

Despite that repeat cycle, many orbits partially cover the area of interest of Tyrol and South Tyrol in Ascending or Descending mode, increasing the temporal resolution of the SMC time series using the six relative orbits: 66, 168, 95, 15, 117 and 44 (**Error! Reference source not found.**). The coverage for each CRNS used within the project is shown in **Error! Reference source not found.**



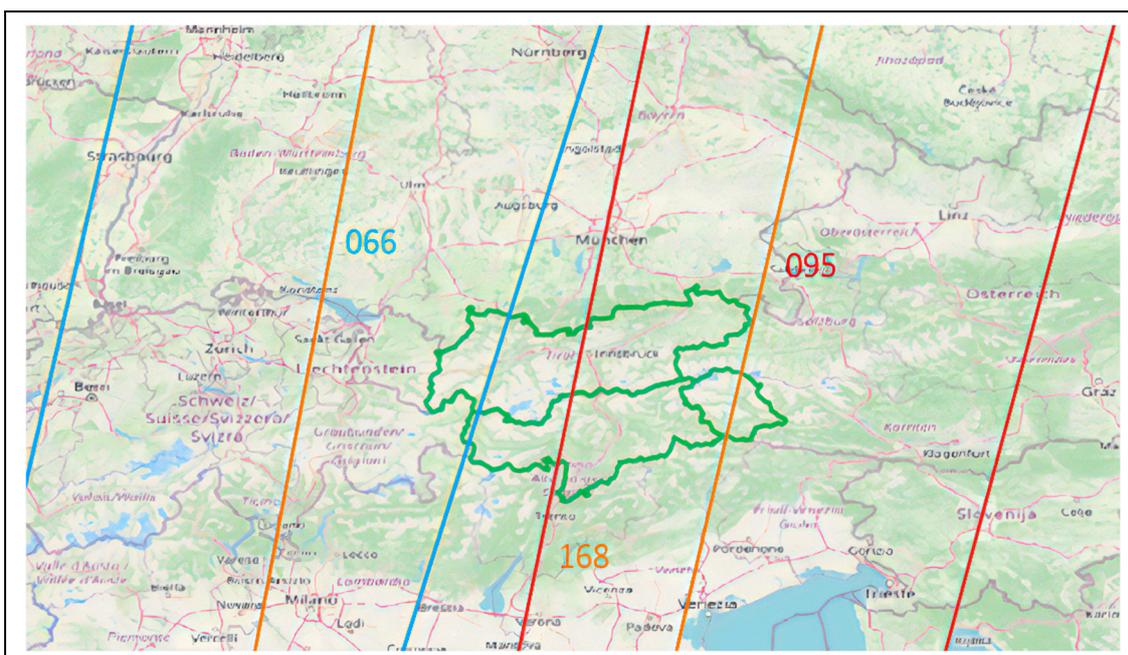


Figure 2. Coverage of the area of interest of the ACR-project with Ascending (relative orbits 015, 117 and 044 in the upper image) and Descending orbits (relative orbits 066, 168 and 095 in the lower image).

Table 3. In the second and third columns are listed the relative orbits of Sentinel-1 in ascending and descending mode respectively covering the sites where the CRNS are installed (first column).

CRNS site	Ascending relative orbit	Descending relative orbit
Leutasch	117	168, 095
Dresdner Hütte	117	168, 095
Weisssee	015, 117	168
Obergurgl	117	168, 095
Corvara	044, 117	168, 095

The algorithm used to estimate the SMC is based on a Machine Learning (ML) approach described in [18], together with the input data utilized and features selected and validated globally. The source code is available online at <https://zenodo.org/records/4552813> and the related documentation is available at: <https://pysmm.readthedocs.io/en/latest/>. The algorithm uses a Gradient Boosted Regression Trees (GBRT). It exploits the server-side processing capabilities of Google Earth Engine (GEE) eliminating the requirements to download or preprocess the input datasets. All the input datasets are available on GEE. They are used as training data to extract features and for masking pixels where the algorithm cannot estimate the SMC (i.e. pixel with snow cover presence, with high vegetation coverage, or in layover shadow).

In particular, the data selected from the 461 automatic stations of the International Soil Moisture Network (<https://ismn.geo.tuwie.n.ac.at>) provide the basis for the large-scale validation of satellite-derived soil moisture products.

The other data collected on GEE are Landsat-8 (shortwave reflectance & thermal radiance) [19], MODIS MOD13Q1 Enhanced Vegetation Index (EVI) [20], soil temperature and snow-water-equivalent from Global Land Data Assimilation System (GLDAS) [21], Copernicus Global Land Cover Layer [22], soil information from OpenLandMap (OLM) [23] and SRTM Aster DEM [24].

The original algorithm was implemented to obtain a spatial resolution of 50 m, but in this work the product was resampled at 20 m using a bilinear interpolation (**Error! Reference source not found.**). In this way, the spatial resolution is consistent with that of the SCF product. The GBRT

algorithm used is a family of tree-based methods. This characteristic entails that it is compatible with different data and scales. Moreover, it has a relatively low computational cost associated with algorithm training and target prediction.

The classes included in the estimation from the Copernicus Global Land Layer are the following: bare/sparse vegetation, cropland, herbaceous vegetation, open forest and shrubs. Moreover, a masking based on several thresholds is implemented. In more detail, stations from ISMN must be available on the same day of the satellite acquisition and considering the limited penetration of C band SAR data, only stations that measure SMC at no more than 5 centimeters are included.

In general, also pixels in layover shadow or foreshortening for S1 images are filtered out, characterized by radiometric saturation, terrain occlusion or clouds in the L8 pixel quality band, with MODIS EVI greater than 0.5 and, regarding GLDAS estimates, pixels below a threshold temperature of 275 K and only SWE values of 0 kg/m² are applied. The algorithm assigns a null value to pixels for which an SMC estimate could not be performed due to masking or because the corresponding satellite acquisition did not cover them.

The SMC dataset made available covers 1 October 2019 to 30 September 2022. This way, three spring-summer seasons are available, like what was delivered for SCF. The algorithm also estimates the SMC in the winter season, although in this case, considering that CRNS stations are located at high altitudes the thresholds set by GLDAS severely limit the number of points at which an estimate is available and, sometimes no estimate is available. In the latter case, the maps were discarded.

The dataset covers an area of 500x500 pixels centered on the 5 CRNS stations to be consistent spatially with the SCF dataset coming from S2 data.

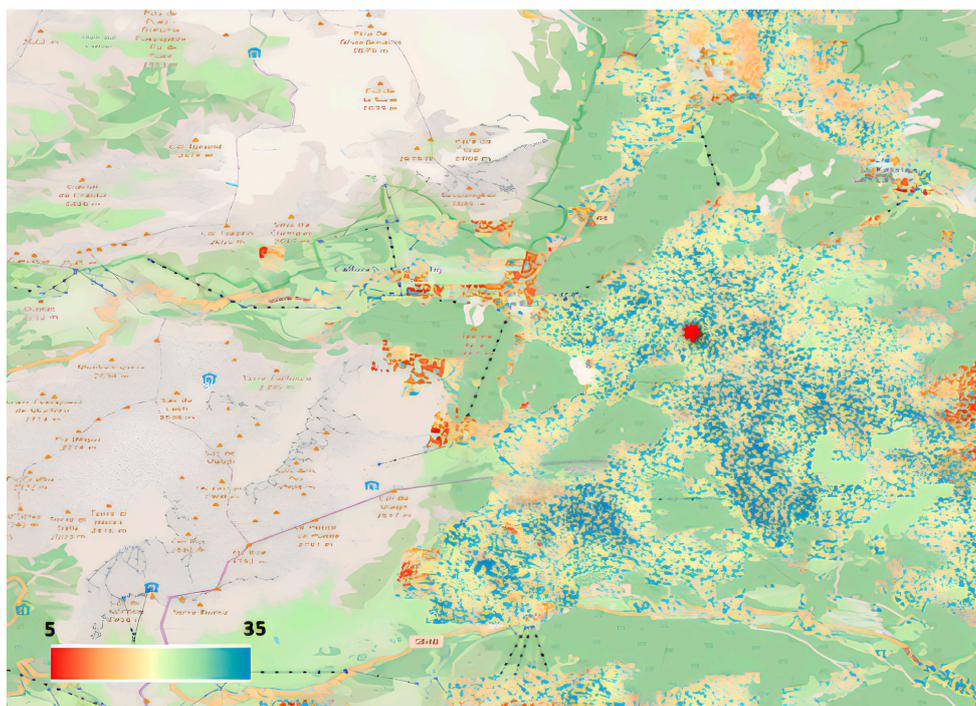


Figure 3. Example of SMC map surrounding the area of Corvara. The red star indicates the point where the Cosmic Ray Neutron Sensor installed during the ACR_Water project.

3.3. Validation SCF

A very high-resolution dataset is exploited to evaluate the quality of the product developed since the lack of ground data does not allow for a systematic validation analysis by considering the different topographical and morphological conditions of the test area, which is predominantly mountainous.

To this purpose, we obtained some WorldView images, that we first prepared through a process of orthorectification and co-registration to make them comparable with S2-derived maps. The collected WorldView images used for verifying the reliability of the developed product are not

always correspondent to the areas covered by the maps here presented (see Livigno and Sonthofen in **Error! Reference source not found.**). Indeed, to obtain a dataset that covers as many scenarios as possible in terms of different topographies as well as different periods of the year or snow coverage, in some case we also exploited images close to the areas of the maps, but not perfectly coincident with them.

Furthermore, a cross-comparison with the Copernicus product is performed to evaluate the developed product with respect to the standard product especially in complex situations as, for example, in case of shadow.

3.3.1. Very High-Resolution Data Preparation

Six WorldView (WV) images are acquired following the criteria of selecting cloud-free scenes, acquisitions possibly coincident or very near to a S2 acquisition and snow cover conditions that might be interesting to test, as variable SCF over the scene (mixed pixels due to vegetation or melting process). The WV images used are both WV2 and WV3.

First, an orthorectification is needed. Very high-resolution (VHR) images often contain geometric distortions due to sensor orientation, topographic relief, and Earth's curvature. The orthorectification procedure aims to correct these distortions and transform the image into a georeferenced and geometrically accurate representation. These data are provided with an orthorectification kit consisting of the Rational Polynomial Coefficients (RPC). These are the coefficients of a fractional polynomial function, that link the image coordinates with the object space.

After being orthorectified, the image is co-registered and aligned to the S2 grid and reference system, by keeping a final resolution of 1 m.

Once the WV and S2 grid are made comparable, the VW needs to be classified. Analogously to what done for the S2 images, we used also for this case a SVM model trained separately for each VW image by manually collecting the training samples. The procedure is very similar to what described in the previous section, with the main difference that our target is here a binary classification.

Finally, we performed an aggregation of the VW images to 20 m spatial resolution thus obtaining the SCF to be compared with the S2 images.

3.3.2. Cross-Comparison with WorldView

The comparison with WV data is done by analyzing different test sites through different metrics to understand the goodness of the SCF estimation in different areas and variables conditions such as different seasons, topography and snow coverage conditions. **Error! Reference source not found.** shows an example of comparison between the developed product and the reference VHR imaged classified: on the left-side the S2 map while on the right-side WV map.

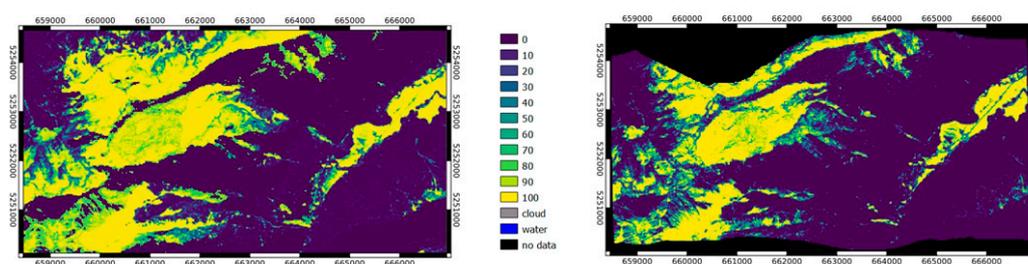


Figure 4. Example of classified SCF maps over the Leutasch area on 27 November 2020: on the left the S2 map and on the right the WV classified image.

The following table (**Error! Reference source not found.**) reports the metrics calculated for the different areas, together with the dates of WV and S2 images, respectively, where the comparison with very high-resolution images has been performed.

Table 4. Comparison of SCF maps with VHR images: metrics and results.

	Date S2	Date WV	BIAS	RMSE	Unbiased RMSE	Correlation
Corvara	13/04/2022	12/04/2022	-0,62	17,44	17,43	0,90
Dresdner	27/11/2020	27/11/2020	-2,65	17,51	17,31	0,91
Leutasch	27/11/2020	27/11/2020	3,45	22,38	22,11	0,86
Livigno	23/09/2021	21/09/2021	-16,33	34,19	30,05	0,51
Sonthofen	12/03/2022	14/03/2022	0,45	21,11	21,1	0,87
Weissee	13/01/2022	13/01/2022	1,17	15,46	15,42	0,90

From the **Error! Reference source not found.** we can see that on average the correlation is for all test sites, around 0.9, except for Livigno, where it is 0.5. The reason behind this drop in performance could be related to the different acquisition dates in the two images, S2 and WV. Indeed, as you can see in **Error! Reference source not found.**, before the WV acquisition there was a snowfall event, which, given the period (September 2021), was only an ephemeral snowfall lasting a few days and indeed it is not visible in the image acquired just two days later by S2. On the left side of the **Error! Reference source not found.** the RGB false color (FC) of SCF maps by S2 (up) and WV (down) are shown; on the right side the classification results used for the comparison and the metric estimation.

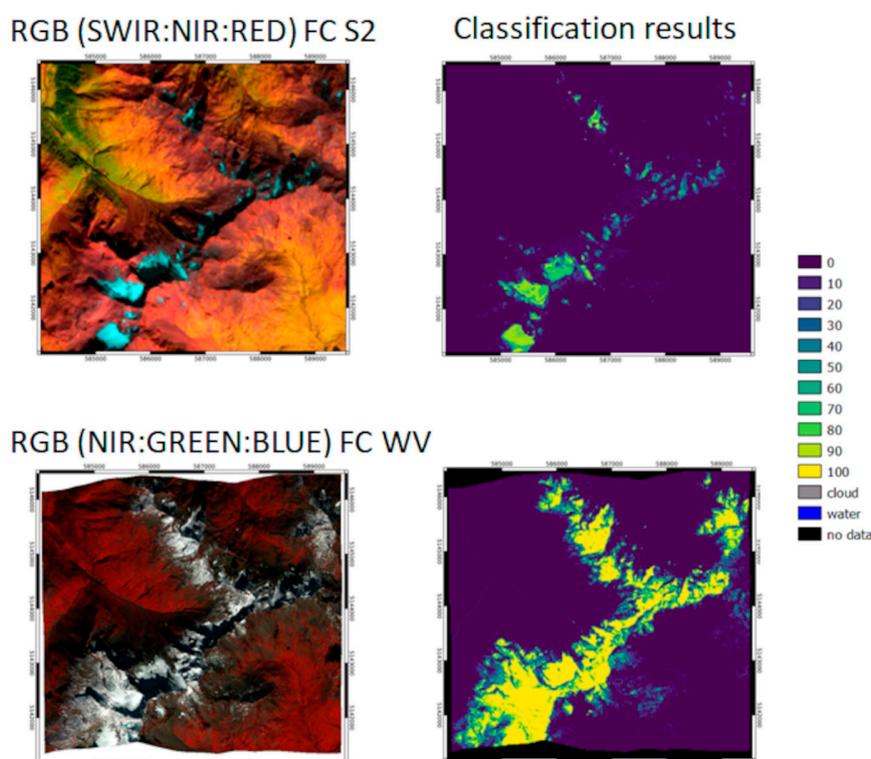


Figure 5. Example of Livigno: comparison between SCF from S2 and WV classification. The different date of acquisition between the two satellites leads to different snow cover due to ephemeral (off-season) snowfall.

3.3.3. Cross-Comparison with Copernicus

To evaluate the snow cover fraction dataset, a cross-comparison with a standard snow product such as the Copernicus high resolution FSC product is performed [25] (Dataset link: <https://sdi.eea.europa.eu/catalogue/copernicus/api/records/3e2b4b7b-a460-41dd-a373-962d032795f3?language=all>) This comparison allows us to evaluate the performance of the developed product with respect to the standard product especially in complex situations as, for example, in case of shadowed areas.

We performed the comparison for the five test sites in the period ranging from 1 October 2020 to 10 May 2023. **Error! Reference source not found.** lists the number of the analyzed scenes for each test site, the period and the extent over which the comparison took place.

Table 5. Comparison between SCF maps derived from S2 and Copernicus product: number of the scenes compared, period and extent.

Test site	Processed S2	Available Copernicus	Matching scenes	Period		Numer of pixel (centered on the CRNS)	Extent
Corvara	105	212	101	01/10/2020	10/05/2023	500x500	717320.0, 5153620.0, 727320.0, 5163620.0
Dresdner	158	295	142	01/10/2020	10/05/2023	500x500	657660.0, 5202080.0, 667660.0, 5212080.0
Leutasch	163	295	162	01/10/2020	10/05/2023	500x500	658180.0, 5244180.0, 668180.0, 5254180.0
Obergurgl	68	327	66	01/01/2022	10/05/2023	500x500	649810.0, 5185346.0, 659810.0, 5195346.0
Weissee	184	329	177	01/10/2020	10/05/2023	500x500	625640.0, 5187460.0, 635640.0, 5197460.0

Starting from the available scenes we computed bias, unbiased RMSE, RMSE and the cross correlation to compare the two products. The results are shown in **Error! Reference source not found.**

Table 6. Results of the intercomparison between SCF product developed and the standard product derived from Copernicus.

Test site	BIAS	RMSE	Unbiased RMSE	Correlation
Corvara	-0,06	16,8	16,05	0,73
Dresdner	5,29	21,89	20,27	0,69
Leutasch	0,76	17,73	18,54	0,61
Obergurgl	0,01	23,74	22,87	0,66
Weissee	3,74	20,69	19,48	0,72

3.4. Validation SMC

The SMC algorithm was already validated by [18] at a global scale. However, in this work, a few alpine automatic monitoring stations measuring soil water content (SWC) in Val Mazia (South Tyrol – Italy) were considered (**Error! Reference source not found.**) for comparisons with SMC maps estimated by the ML algorithm [26]. These stations record meteorological and biophysical variables of the long-term socio-ecological research LT(S)ER site Matschertal / Val di Mazia (<https://browser.lter.eurac.edu/>). The data can be downloaded from this website after registration. This approach was considered because most of the ISMN network used to train the SMC algorithm is not located in alpine or, in general, in mountainous areas like those considered in the project. Consequently, this comparison can be helpful to better understand the behavior of the algorithm in this specific environment. Among many parameters like air temperature, relative humidity, precipitation and wind speed, etc. these climate stations record the SWC at different depths from 2 to 50 cm, averaging 15 samples taken every minute, aggregating them into a single value. Considering the penetration characteristic of the SAR signal, only the measurements at 2 and 5 cm were considered, while the ones at 20, 40 and 50 cm were excluded. For the comparison, the satellite acquisition time was considered, and the ground measurements from one hour before to one hour after that time were further averaged and utilized.

The stations are equipped with Campbell Scientific TDR (Time Domain Reflectometry) sensors, able to measure soil temperature and SWC (range 0 to 52%, accuracy $\pm 3\%$ for Electrical Conductivity ≤ 10 dS/m).

The monitoring stations are located in the western part of South Tyrol (Val Mazia catchment; an inner-alpine dry valley in the Italian Alps) and are shown in **Error! Reference source not found.** The four stations were selected among a network of 24 climate stations considering their peculiar characteristics, especially in terms of elevation, exposure, slope, and landcover type, as observed in the photos in the

Table 7. Climatic stations described in terms of GPS coordinates, elevation, land use and soil characteristics.

Station	Latitude [DD]	Longitude [DD]	Elevation a.s.l. [m]	Aspect [deg]	Slope [°]	Land use	Soil type	Texture
B1	46.661183	10.590244	980	230	12	Irrigated Meadow	Brown earth	Sandy Loam
P2	46.684305	10.585125	1540	230	27	Pasture	Brown earth / Ranker	Loam
B3	46.691694	10.591936	1920	220	11	Pasture	Ranker	Loam
M2	46.711303	10.691794	2210	250	9	Pasture	/	/

. By referring to the period May – October of the three observation years 2020 -2022, different metrics were considered based on what is described in [27–29]. Mean Absolute Error, Bias and Unbiased Root Mean Square Error are reported in **Error! Reference source not found.**. The selection of the study period considers that these months have the lowest probability of snow cover, making it easier to obtain a consistent and continuous series of SMC maps, which is helpful for the time series analysis.



Figure 6. The Val Mazia catchment in South Tyrol defined by the red outline, with the location of the automatic stations considered in this study identified by the blue points.



B1 – 980 m a.s.l.



P2 – 1540 m a.s.l.



B3 – 1920 m a.s.l.



M2 – 2210 m a.s.l.

Figure 7. Automatic stations located in Val Mazia measuring SMC used for comparison with the SMC estimated by the algorithm based on Sentinel-1 images.

Table 7. Climatic stations described in terms of GPS coordinates, elevation, land use and soil characteristics.

Station	Latitude [DD]	Longitude [DD]	Elevation a.s.l. [m]	Aspect [deg]	Slope [°]	Land use	Soil type	Texture
B1	46.661183	10.590244	980	230	12	Irrigated Meadow	Brown earth	Sandy Loam
P2	46.684305	10.585125	1540	230	27	Pasture	Brown earth / Ranker	Loam
B3	46.691694	10.591936	1920	220	11	Pasture	Ranker	Loam
M2	46.711303	10.691794	2210	250	9	Pasture	/	/

Table 8. Performances of the comparison between SMC estimated by the algorithm and measured by the monitoring stations. The suffix _2 and _5 indicate the depth of the station in centimeters.

Station	MAE_2	MAE_5	URMSE_2	URMSE_5	BIAS_2	BIAS_5
B1	0.055	0.047	0.054	0.048	-0.043	-0.035
P2	0.086	0.083	0.091	0.086	0.046	0.050
B3	0.059	0.107	0.071	0.104	0.018	-0.076
M2	0.055	0.050	0.068	0.061	0.006	-0.006

Error! Reference source not found. - Error! Reference source not found. show examples of the trends of the SMC estimated by the algorithm compared to that extracted from the automatic meteo stations at 2 and 5 centimeters. The graphs for different ground stations and years are related to May –October, when the stations are usually not affected by snow, to highlight the temporal trends. The measurements and estimations are acquired on the same day as described above. In 2022, the number of SMC estimates from remote sensing data is reduced as only the S1-A satellite was available then.

The three curves show similarities in the trends although the estimated values present a more compressed dynamic. A possible reason can be related to the temporal filter applied to SAR data to reduce the speckle of the SAR images. The measurements at 2 and 5 centimeters are very similar except for B3 and M2 stations, which are a meadow and a pasture station, respectively, but are characterized by a flat terrain. In this B3 case, where the dynamics are broader, the estimated value is closer to the ground measurement at 2 centimeters, while in M2 case the estimation over the entire analysis period is between the two ground estimations.

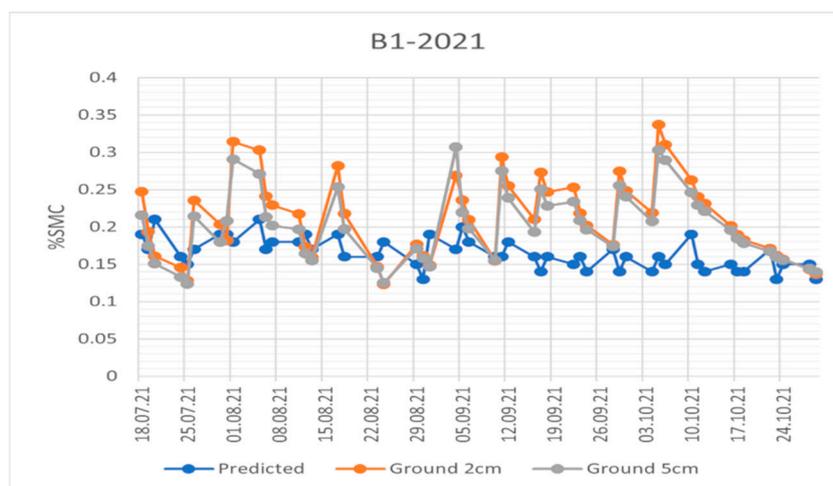


Figure 8. SMC Trend comparison for the automatic meteo station B1 in 2021. The orange line refers to SMC measurements at a depth of 2cm, the grey line to the SMC measurements at 5 cm of depth and the blue line is the SMC estimated by the algorithm.

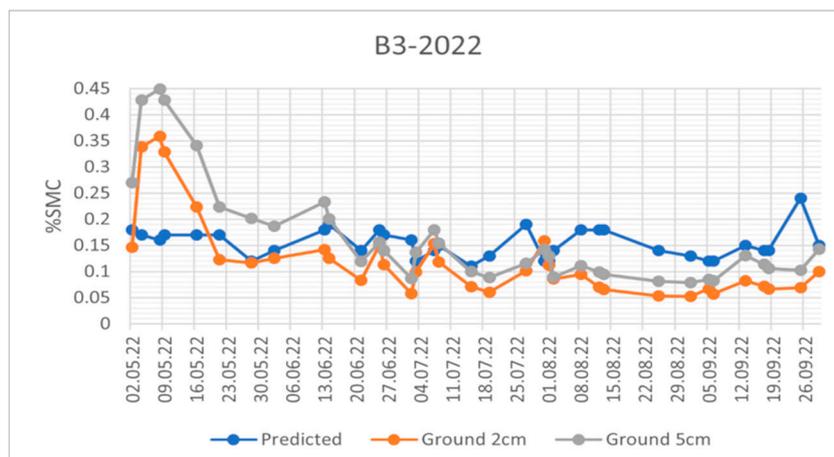


Figure 9. SMC Trend comparison for the automatic meteo station B3 in 2021. The orange line refers to SMC measurements at a depth of 2cm, the grey line to the SMC measurements at 5 cm of depth and the blue line is the SMC estimated by the algorithm.

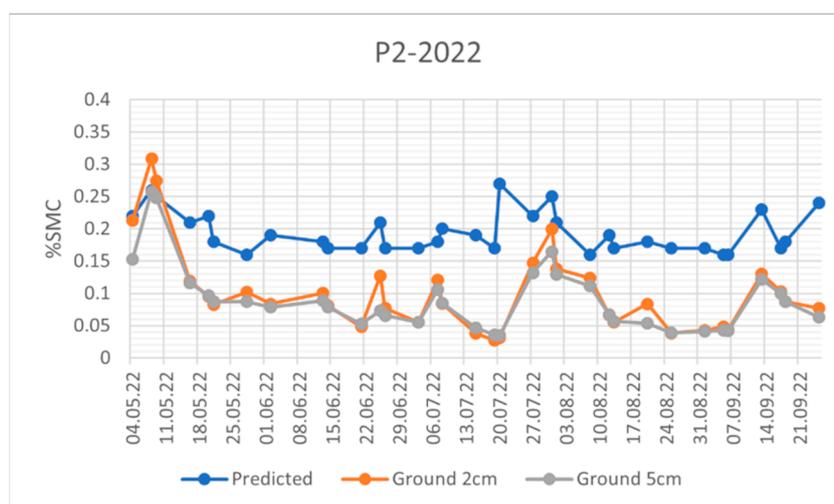


Figure 10. SMC Trend comparison for the automatic meteo station P1 in 2022. The orange line refers to SMC measurements at a depth of 2cm, the grey line to the SMC measurements at 5 cm of depth and the blue line is the SMC estimated by the algorithm.

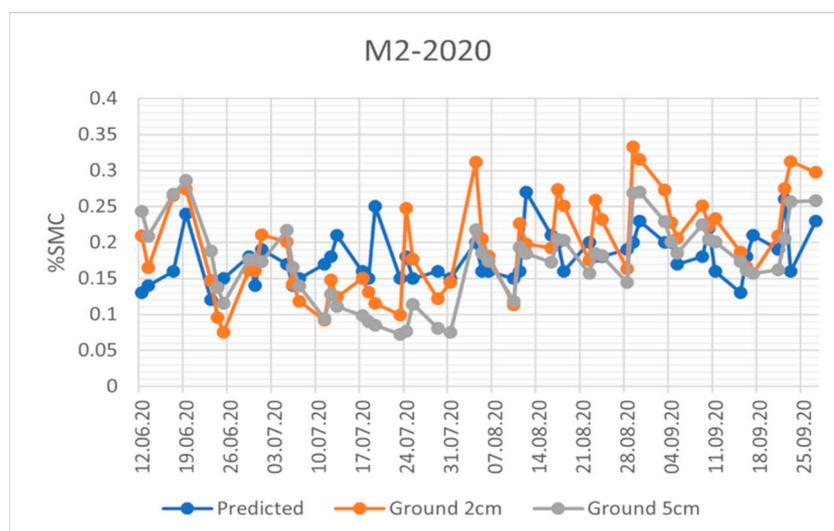


Figure 11. SMC Trend comparison for the automatic meteo station M2 in 2020. The orange line refers to SMC measurements at a depth of 2cm, the grey to the SMC measurements at 5 cm of depth and the blue line is the SMC estimated by the algorithm.

3.4.1. Comparison with CRNS and Point-Scale SMC Data

A further comparison with CRNS and point-scale SMC data illustrates a potential use case for the data at the site of Leutasch. The Leutasch site is part of the COSMOS Europe network [30]. The CRNS based SMC values were obtained using the calibration procedure described in [31]. The data needs to be corrected for changes in incoming cosmic radiation, atmospheric pressure and air humidity. The residual signal is highly sensitive to changes in hydrogen content in soil, vegetation and snow [32]. CRNS represent a large footprint of several hectares and a depth of several decimeters [33]. The signal is thus more representative for the hydrological state of the location than point-scale measurements. Furthermore, a depth-profile of point-scale SMC measurements using SMT100 sensors is available.

The trends of the comparisons obtained are shown in **Error! Reference source not found.** and **Error! Reference source not found.**. Therein, it is visible that CRNS and remote sensing-based SMC estimates show similar temporal soil moisture dynamics. However, due to the larger measurement volume, in particular with regard to the depth of the signal, the CRNS data show different absolute values. The point-scale data of the upper-most SMT100 probe in 5 cm depth, in contrast, gives very similar levels of SMC. Both SMT100 and CRNS sensors have higher dynamic range than the remote sensing-based data. Thus, due to different integration volumes and spatial coverages the combination remote sensing and in-situ based SMC data can give further insights into the hydrologically relevant dynamics. This is particularly true if combining spatially representative CRNS data with regionally available remote sensing products.

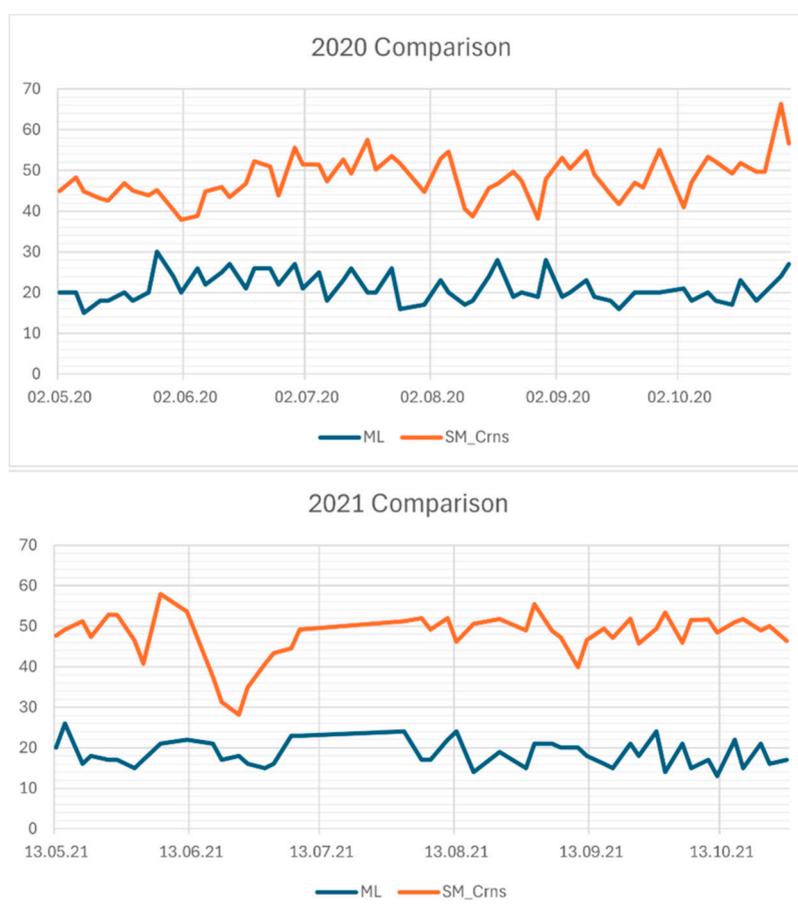


Figure 12. Comparison between SMC estimated from Machine Learning Algorithm (ML, blue line in the graph) based on satellite data and soil moisture estimated from CRNS (SM_Crns, orange line in the graph) on the Leutasch test site. The upper plot represents data for the summer 2020, while the one below shows the data relative to the summer 2021.

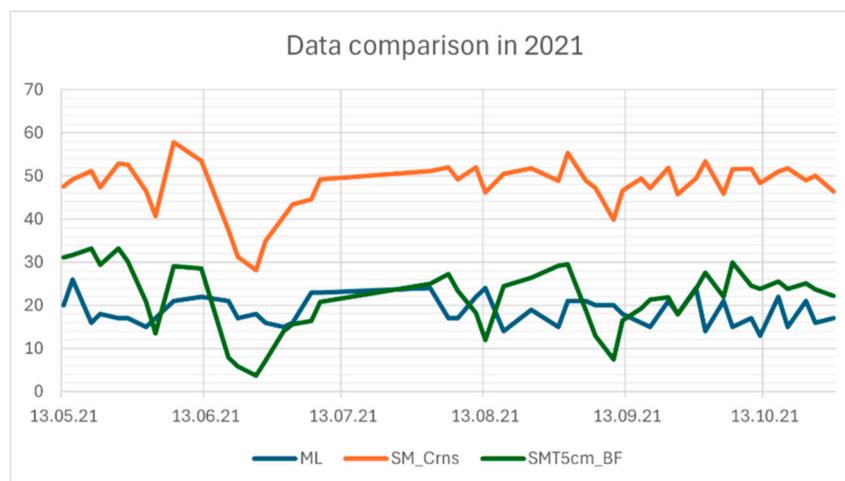


Figure 13. Comparison among SMC estimated from Machine Learning Algorithm (ML, blue line in the graph) based on satellite data soil moisture estimated from CRNS (SM_Crns, orange line in the graph) and raw soil moisture measurements obtained via SMT100 sensors on the Leutasch test site from May to October 2021.

4. User Notes

4.1. Data Access

The two datasets presented in this work are freely accessible online at the website ... and are organized as follows: for each CRNS station, an NCDF file is available for both the SCF and SMC products, providing the entire time series for the study period.

4.2. Example Usage

The developed products can be exploited, for example, as inputs for hydrological models: complementary information sources can indeed be exploited for improving the model output and reducing the uncertainty by overcoming limitations of individual data related for example to an inadequate scale representation or to the uncertainties of the single products as reported in the following section. If relying on one data source only, errors and biases propagate into hydrological modelling [34]. The combined use of remotely sensed datasets together with the CRNSs derived data is of high potential for operational water resource management, allowing for more reliable forecasts in regions with limited accessibility. Another example of usage is the analysis of the temporal behavior of the soil moisture and of the snow cover fraction over a specific point of interest included in the area. This analysis can be useful to study trends and anomalies, for example extreme drought characteristics in terms of onset and duration.

4.3. Cautionary Notes

A few cautionary notes are necessary before using the datasets presented in this paper. The uncertainties related to the proposed products are due to several factors, as explained below, and must be kept in mind when using the data.

Concerning snow maps, in fact, because of the lack and the nature of station data, no comparison has been performed with this type of measurement. The stations provide, when available, punctual measurements of snow height at the station location, and this kind of data are not easily comparable

with a spatialized value derived from the SCF map where a percentage of snow cover for each pixel (20 by 20 meters) is provided. However, an inter-comparison with very high-resolution products (WorldView) and with what is considered as reference product (Copernicus) was performed and the results are encouraging by suggesting a valid and reliable product.

Regarding the SMC product, it should be noted that the model used was obtained starting from the ISMN stations which are mostly located in the USA and, in any case, none of them are in the Alps or in European mountain areas. The consequence is that, also considering URMSE, the error obtained in the tested Alpine sites is higher with respect to $0.04 \text{ m}^3 \cdot \text{m}^{-3}$ obtained in the validation executed at global scale by [18]. Moreover, although the automatic stations in Val Mazia (used for the validation in this paper) cover different kinds of mountainous conditions in the Alps, their locations do not correspond to those of CRNSs probes, where the SMC time series are provided. Nevertheless, the tests carried out show that the data follow a realistic temporal trend. Another limitation is that the algorithm can exclude many points from the SMC estimation because of the masking procedure described in section 0. For example, regarding forested areas, the only type considered are that one classified as “open forest” by the Copernicus Global Land Cover Layer (CGLS-LC100), which has a spatial resolution coarser respect to spatial resolution of the SMC maps.

Finally, both satellite products provided are obtained by exploiting the use of optical sensors (S2 for SCF and L8 for SMC) that are reliable instruments for observing geophysical parameters from space. In particular, the distinctly high reflectance of snow in the visible and its highly absorptive nature in the short-wave infrared wavelengths of the electromagnetic spectrum are captured by multispectral sensors [35]. Nevertheless, optical-based satellite sensors are impacted by atmospheric conditions (e.g., cloud presence, [36]) and by land cover conditions (e.g., forest presence, [37,38]). Clouds coverage represents indeed for parameters estimation with optical sensors: if on one side the optical signal cannot penetrate the cloud coverage, by preventing the parameter detection in that portion of the image, on the other hand, in the case of snow detection, cloud cover can create errors in snow-detection algorithms because of the similar reflectance of cloud and snow at wavelengths utilized for snow identification. Thus, misclassification errors can occur between snow and clouds during cloudy days.

The other challenge for geophysical parameters retrieval from optical sensors are the forested areas because the sensor only detects the viewable soil fraction of a satellite pixel (snow or soil moisture in this case). In case of snow, optical sensors only view snow that is visible beneath leafless deciduous stands (such as aspen), in clearings and forest gaps between coniferous trees (i.e., pine, spruce, and fir) and through thin foliage or needles, or snow that has been intercepted by the forest canopy. Since the study area is in a mountainous environment, this kind of error must be considered by the user who wants to exploit the products in forested areas.

Author Contributions: Conceptualization L.D.G., G.C. and C.N.; methodology L.D.G., G.C., V.P. and C.N.; software G.C., F.G., V.P.; validation G.C. L.D.G. and V.P.; data curation G.N., A.M., P.S. and A.Z.; writing—original draft preparation L.D.G., G.C., V.P. and R.B.; review and editing all authors; supervision C.N. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: This work was conducted within the project ACR Water (Assimilating Cosmic-Ray Neutron and Remote Sensing Data for Improved Water Resource Management), coordinated by the University of Innsbruck and funded by the Autonomous province of Bolzano “Ripartizione Diritto allo Studio, 568 Università e Ricerca Scientifica”. We thank Eurac Research’s long-term socio-ecological research area LT(S)ER IT25 - Matsch/Mazia - Italy, for providing the data, DEIMS.iD: <https://deims.org/11696de6-0ab9-4c94-a06b-7ce40f56c964>. WorldView data were provided by the European Space Agency (ESA), Project Proposal id PP0088762, © DigitalGlobe, Inc. (2023), provided by European Space Imaging, all rights reserved.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Pulliainen, J., Luojus, K., Derksen, C., Mudryk, L., Lemmetyinen, J., Salminen, M., ... & Norberg, J. Patterns and trends of Northern Hemisphere snow mass from 1980 to 2018. *Nature*, **2020**. 581(7808), 294-298.
- Musselman, K. N., Addor, N., Vano, J. A., & Molotch, N. P. Winter melt trends portend widespread declines in snow water resources. *Nature Climate Change*, **2021**. 11(5), 418-424.
- Qin, Y., Abatzoglou, J. T., Siebert, S., Huning, L. S., AghaKouchak, A., Mankin, J. S., ... & Mueller, N. D. Agricultural risks from changing snowmelt. *Nature Climate Change*, **2020**. 10(5), 459-465.
- Ma, N., Yu, K., Zhang, Y., Zhai, J., Zhang, Y., & Zhang, H. Ground observed climatology and trend in snow cover phenology across China with consideration of snow-free breaks. *Climate Dynamics*, **2020**. 55, 2867-2887.
- Matiu, M., Crespi, A., Bertoldi, G., Carmagnola, C. M., Marty, C., Morin, S., ... & Weilguni, V. Observed snow depth trends in the European Alps: 1971 to 2019. *The Cryosphere*, **2021**. 15(3), 1343-1382.
- Lundquist, J., Hughes, M., Gutmann, E., & Kapnick, S. Our skill in modeling mountain rain and snow is bypassing the skill of our observational networks. *Bulletin of the American Meteorological Society*, **2019**. 100(12), 2473-2490.
- Asaoka Y, Kazama S, Sawamoto M. The variation characteristics of snow water resources in a wide area and its geographical and climate dependency. *Journal of Japan Society of Hydrology and Water Resources*, **2002**. 15: 279-289.
- Roesch, A., Wild, M., Gilgen, H., Ohmura, A., & Arugnell, N. C. A new snow cover fraction parameterization for the ECHAM4 GCM. *Climate Dynamics*, **2001**. 17(2), 933 – 946.
- Rittger, K., Painter, T. H., & Dozier, J. Assessment of methods for mapping snow cover from MODIS. *Advances in Water Resources*, **2013**. 51, 367-380.
- Berger, M., Camps, A., Font, J., Kerr, Y., Miller, J., Johannessen, J. A., ... & Attema, E. (2002). Measuring ocean salinity with ESA's SMOS mission—*Advancing the science*, **2002**.
- Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., ... & Van Zyl, J. The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, **2010**. 98(5), 704-716.
- Naeimi, V., Scipal, K., Bartalis, Z., Hasenauer, S., & Wagner, W. An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations. *IEEE Transactions on Geoscience and Remote Sensing*, **2009**. 47(7), 1999-2013.
- Zupanc, A.: Improving cloud detection with machine learning, <https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13> (last access: 9 August 2024), **2017**.
- Tuia, D., Persello, C., & Bruzzone, L. Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE geoscience and remote sensing magazine*, **2016**. 4(2), 41-57.
- Ebner, P. P., Koch, F., Premier, V., Marin, C., Hanzer, F., Carmagnola, C. M., ... & Lehning, M. Evaluating a prediction system for snow management. *The Cryosphere*, **2021**. 15(8), 3949-3973.
- Hofmeister, F., Arias-Rodriguez, L. F., Premier, V., Marin, C., Notarnicola, C., Disse, M., & Chiogna, G. Intercomparison of Sentinel-2 and modelled snow cover maps in a high-elevation Alpine catchment. *Journal of Hydrology*, **2022**. 15, 100123.
- Barella, R., Marin, C., Gianinetto, M., & Notarnicola, C. A novel approach to high resolution snow cover fraction retrieval in mountainous regions. In IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium (pp. 3856-3859), **2022**. IEEE.
- Greifeneder, F., Notarnicola, C., & Wagner, W. A machine learning-based approach for surface soil moisture estimations with google earth engine. *Remote Sensing*, **2021**. 13(11), 2099.
- Chander, G., Markham, B.L., Helder, D.L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens. Environ.* **2009**, 113, 893-903.
- Didan, K. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006, V006 ed.; NASA EOSDIS LP DAAC: Sioux Falls, SD, USA, **2015**.
- Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B.; Radakovich, J.; Bosilovich, M., et al. The Global Land Data Assimilation System. *Bull. Am. Meteorol. Soc.* **2004**, 85, 381-394.
- Buchhorn, M., Smets, B., Bertels, L., Lesiv, M., Tsendbazar, N.-E., Herold, M., Fritz, S. Copernicus Global Land Service: Land Cover 100m: Collection 3: Epoch 2015: Globe. Zenodo **2020**.
- Hengl, T. Soil bulk density (fine earth) 10 x kg/m-cubic at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution. Zenodo **2018**.
- Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., et al. The Shuttle Radar Topography Mission. *Rev. Geophys.* **2007**. 45, 1-33.
- Gascoin, S., Grizonnet, M., Bouchet, M., Salgues, G., & Hagolle, O. Theia Snow collection: High-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data. *Earth System Science Data*, **2019**. 11(2), 493-514.

26. Palma, M., Zandonai, A., Cattani, L., Klotz, J., Genova, G., Brida, C., ... & Della Chiesa, S. Data Browser Matschl Mazia: Web Application to access microclimatic time series of an ecological research site. *Research Ideas and Outcomes*, **2021**. 7, e63748.
27. Yang, G., Guo, P., Li, X., Wan, H., Meng, C., & Wang, B. Assessment with remotely sensed soil moisture products and ground-based observations over three dense networks. *Earth Science Informatics*, **2020**. 13, 663-679.
28. Entekhabi, D., Reichle, R. H., Koster, R. D., & Crow, W. T. Performance metrics for soil moisture retrievals and application requirements. *Journal of Hydrometeorology*, **2010**. 11(3), 832-840.
29. Cui, H., Jiang, L., Du, J., Zhao, S., Wang, G., Lu, Z., & Wang, J. Evaluation and analysis of AMSR-2, SMOS, and SMAP soil moisture products in the Genhe area of China. *Journal of Geophysical Research: Atmospheres*, **2017**. 122(16), 8650-8666.
30. Bogena, H., Schrön, M., Jakobi, J., Ney, P., Zacharias, S., Andreasen, M., ... & Vereecken, H. COSMOS-Europe: A European network of cosmic-ray neutron soil moisture sensors. *Earth System Science Data Discussions*, **2022**. 14, 1125-1151, <https://doi.org/10.5194/essd-14-1125-2022>, 2022.
31. Schrön, M., Köhli, M., Scheiffele, L., Iwema, J., Bogena, H. R., Lv, L., ... & Zacharias, S. Improving calibration and validation of cosmic-ray neutron sensors in the light of spatial sensitivity. *Hydrology and Earth System Sciences*, **2017**. 21(10), 5009-5030.
32. Zreda, M., Shuttleworth, W. J., Zeng, X., Zweck, C., Desilets, D., Franz, T., & Rosolem, R. COSMOS: The cosmic-ray soil moisture observing system. *Hydrology and Earth System Sciences*, **2012**. 16(11), 4079-4099.
33. Köhli, M., Schrön, M., Zreda, M., Schmidt, U., Dietrich, P., & Zacharias, S. Footprint characteristics revised for field-scale soil moisture monitoring with cosmic-ray neutrons. *Water Resources Research*, **2015**. 51(7), 5772-5790.
34. Schattan, P., Schwaizer, G., Schöber, J., & Achleitner, S. The complementary value of cosmic-ray neutron sensing and snow-covered area products for snow hydrological modelling. *Remote Sensing of Environment*, **2020**. 239, 111603.
35. Warren, S. G. Optical properties of snow. *Reviews of Geophysics*, **1982**. 20(1), 67-89.
36. Hall DK, Kelly REJ, Riggs GA, Chang ATC, Foster JL. Assessment of the relative accuracy of hemispheric scale snow-covermaps. *Annals of Glaciology*, **2002b**. 34:24-30
37. Vikhamar D, Solberg R. Snow-cover mapping in forests byconstrained linear spectral unmixing of MODIS data. *Remote Sensing ofEnvironment*, **2003**. 88: 309-323. DOI: 10.1016/j.rse.2003.06.004.
38. Parajka J, Holko L, Kostka Z, Blöschl G. MODIS snow covermapping accuracy in small mountain catchment-comparison betweenopen and forest sites. *Hydrological Earth Systems Sciences Discussion*, **2012a**. 9: 4073-4100. DOI: 10.5194/hessd-9-4073-2012.