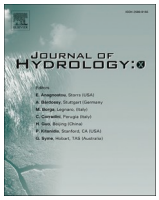
Journal of Hydrology X 15 (2022) 100123

Contents lists available at ScienceDirect

Journal of Hydrology X

journal homepage: www.sciencedirect.com/journal/journal-of-hydrology-x



Research papers

Intercomparison of Sentinel-2 and modelled snow cover maps in a high-elevation Alpine catchment

Florentin Hofmeister [a,*], Leonardo F. Arias-Rodriguez [a], Valentina Premier [b], Carlo Marin [b], Claudia Notarnicola [b], Markus Disse [a], Gabriele Chiogna [a,c]

[a] TUM School of Engineering and Design, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany
[b] Institute for Earth Observation, Eurac Research, Viale Druso 1, 39100 Bolzano, Italy
[c] Institute of Geography, University of Innsbruck, Innrain 52, 6020 Innsbruck, Austria

ARTICLE INFO

Keywords:
Snow cover
Snow redistribution
High-elevation Alpine catchment
Sentinel-2
Physically based snow modelling

ABSTRACT

Modelling runoff generation in high-elevation Alpine catchments requires detailed knowledge on the spatio-temporal distribution of snow storage. With Sentinel-2 MultiSpectral Instrument (MSI), it is possible to map snow cover with a high temporal and spatial resolution. In contrast to the coarse MODIS data, Sentinel-2 MSI enables the investigation of small-scale differences in snow cover duration in complex terrains due to gravitational redistribution (slope), energy balance and wind-driven redistribution (aspect). In this study, we describe the generation of high-resolution spatial and temporal snow cover data sets from Sentinel-2 images for a high-elevation Alpine catchment and discuss how the data contribute to our understanding of the spatio-temporal snow cover distribution. The quality of snow and cloud detection is evaluated against in-situ snow observations and against other snow and cloud products. The main problem was in the false detection of snow in the presence of clouds and in topographically shaded areas. We then seek to explore the potential of the generated high-resolution snow cover maps in calibrating the gravitational snow redistribution module of a physically based snow model, especially for an area with a very data-scarce point snow observation network. Generally, the calibrated snow model is able to simulate both the mean snow cover duration with a high F1 accuracy score of $>$ 0.9 and the fractional snow-covered area with a correlation coefficient of 0.98. The snow model is also able to reproduce spatio-temporal variability in snow cover duration due to surface energy balance dynamics, wind and gravitational redistribution.

1. Introduction

Alpine snow cover and its subsequent melt can dominate local to regional climate and hydrology in both high-latitude areas (Gascoin et al., 2019) and the world's mountainous regions (Dozier and Painter, 2004). In the European Alps, snow is the major driver of Alpine hydrology, storing water during the winter season and releasing it in the spring and summer, with impacts on water supply, agriculture and hydropower production (Matiu et al., 2021). Moreover, ongoing climate change in the Alps affects the abundance of snow (Matiu et al., 2021). Finally, snow accumulation and melting processes also depend on topographical parameters such as slope and aspect, which are linked to incoming solar radiation, gravitational transport and wind redistribution (López-Moreno and Stähli, 2008; Grünewald and Lehning, 2011; López-Moreno et al., 2014; Grünewald et al., 2014; Pedersen et al., 2016; Gurung et al., 2017; Mott et al., 2018; Saydi and Ding, 2020; Vionnet et al., 2021). Therefore, to calibrate and validate models capable of properly capturing snow dynamics in such a complex environment requires accurate observed snow cover maps.

The standard source of information regarding snow cover is a network of automatic and manual ground-based meteorological stations that perform daily or sub-daily observations, mostly of snow depth (Romanov et al., 2000). However, in-situ snow observations are generally insufficient for characterizing the high spatial variability of the snow pack in mountainous regions (Gascoin et al., 2019) and are often only suitable for making qualitative comparisons with hydrological model results (Tuo et al., 2018). Moreover, field measurements are sparse in both space and time and they are subject to several sources of

* Corresponding author.
E-mail addresses: florentin.hofmeister@tum.de (F. Hofmeister), leonardo.arias@tum.de (L.F. Arias-Rodriguez), Valentina.Premier@eurac.edu (V. Premier), Carlo.Marin@eurac.edu (C. Marin), claudia.notarnicola@eurac.edu (C. Notarnicola), markus.disse@tum.de (M. Disse), gabriele.chiogna@tum.de (G. Chiogna).

https://doi.org/10.1016/j.hydroa.2022.100123
Received 12 October 2021; Received in revised form 29 January 2022; Accepted 3 February 2022
Available online 9 February 2022

error (Rasmussen et al., 2012; Marcolini et al., 2019). Furthermore, only few snow observation stations operate at altitudes above 2000 m, which complicates the investigation of altitudinal gradients (Matiu et al., 2019). Despite these limitations, snow depth ground observations are very valuable when it comes to evaluate snow detection quality using remote sensing products (e.g. Foppa et al., 2005; Gascoin et al., 2020; Barrou Dumont et al., 2021).

Since the launch of the first Landsat satellite in the 1970 s, remote sensing has become a key tool for mapping snow cover and revealing snow properties at multiple spatial and temporal resolutions (Romanov et al., 2000; Dozier and Painter, 2004). Snow can be characterized by many variables, such as the snow-covered area (SCA), fractional area (fSCA), albedo, liquid water content, snow depth or snow water equivalent (Frei et al., 2012). In remote sensing applications, snow cover is mainly investigated using SCA and fSCA products on the scale of entire mountain ranges like the Alps or Pyrenees (Gascoin et al., 2019). Remote sensing products supply information about snow-covered areas across elevations, but they have problems with correct snow detection in cloudy conditions. Consequently, when it comes to improving a snow detection algorithm, the main challenge is to reduce misclassification of cloud as snow (Gascoin et al., 2019). Another difficulty is in the detection of snow in dense forest areas, where the ground is obstructed by the canopy. This is particularly the case with evergreen conifer forests in Alpine regions (Di Marco et al. 2020). Besides the influence of forests, topographical features such as steep, shady slopes can impact the detection of snow by optical remote sensing products (Gascoin et al., 2019).

Moreover, there is a trade-off between spatial resolution and swath width (i.e. the acquisition strip of a satellite), which influences the observation frequency (Dozier and Painter, 2004). Despite these limitations, spatial snow information derived from satellite data has been used for decades for various purposes, including hydrology and water resource management (Rango et al., 1996; Hall et al., 2012; Brown et al., 2014; Kääb et al., 2016; Fassnacht et al., 2017). The application of MODIS data to derive snow cover information on a daily or twice-daily basis, depending on the geographical location, is very widespread (Matiu et al., 2019), but their spatial resolution of 500 m is too coarse for hydrological applications in mountain regions, where snow cover properties can vary on scales of 10 m to 100 m (Blöschl, 1999; Gascoin et al., 2019; Vionnet et al., 2021). In fact, although Gurung et al. (2017) used snow cover maps made with MODIS to investigate the effect of topography on the SCA in very large basins (>30,000 km^2), Bouamri et al. (2021) demonstrated that MODIS is not capable of capturing the spatial heterogeneity of snow cover induced by solar radiation, because it does not capture spatial variability below 500 m. This limits the usage of MODIS snow cover maps for calibrating and evaluating spatially distributed snow models (Bouamri et al., 2021).

High-resolution (30 m) snow cover maps can be generated from Landsat images, but the low temporal revisit time of the Landsat mission (16 days) is a significant limitation to snow cover monitoring, and, moreover, data availability can be considerably reduced by cloudiness. The launch of the second Copernicus Sentinel-2 satellite in 2017 has made it possible to map the extent of snow cover at a 20 m resolution, with a revisit time of 5 days (Gascoin et al., 2019). The high-resolution spatial and temporal data collected by Sentinel-2 MultiSpectral Instrument (MSI) are important for enabling detailed investigations of snow cover and snow redistribution processes on the catchment scale (Foppa et al., 2005).

Sentinel-2 data can therefore not only be used to evaluate the parametrizations of snow melt and redistribution in physically based models, but also to actually act or contribute to the objective function used for model calibration (Mott et al., 2010). In general, the challenge with physically based snow models - even without considering snow redistribution processes - is that with increasing catchment size and simultaneous high grid discretization, computation times increase significantly when, for example, multi-layer snow models are applied for long periods, i.e. climate change studies. The computational effort increases even more when mechanistic snow redistribution approaches in physically based models are applied at the catchment scale (Thornton et al., 2021). Hence, snow redistribution approaches are mainly based on empirical assumptions in hydrological models and have different levels of complexity. A research gap thus exists between the development of process-based snow redistribution models, i.e., based on a mechanistic snow redistribution routine (Freudiger et al., 2017), and their experimental validation (Warscher et al., 2013; Schöber et al., 2014; Frey and Holzmann, 2015). Recently, Vionnet et al. (2021) used the Canadian Hydrological Model (CHM) to simulate small-scale variabilities in snow accumulation caused by gravitational redistributions (avalanches) and blowing-snow transport (saltation and suspension), and among others factors. By driving the CHM with spatial high-resolution (50 m) wind fields, it was possible to account for the influence of topographical features due to wind speed and direction. The simulated snow accumulation was evaluated using high-resolution airborne light detection and ranging (lidar) snow depth data and snow persistence indexes derived from remotely sensed imagery. Although the CHM was able to simulate the small-scale variability of snow accumulation, there is still the need for optimization snowdrift-permitting models for large scale application (≥1000 km^2), in particular the representation of subgrid topographic effects on snow transport (Vionnet et al., 2021). Besides the limitation of physically based and fully distributed snow models on the catchment scale due to their heavy computational requirements (Thornton et al., 2021), detailed model input data (e.g. local wind fields) are needed when wind-driven snow redistribution is simulated with models based on semi-empirical parameterizations of the physics of snow transport (Essery et al., 1999; Durand et al., 2005; Liston et al., 2007; Pomeroy et al., 2007) or even with models resolving the 3D turbulent-diffusion equation (Gauer, 1998; Lehning et al., 2008; Sauter et al., 2013; Schneiderbauer and Prokop, 2011; Vionnet et al., 2014) for blown snow particles in the atmosphere (Mott et al., 2018; Vionnet et al., 2021). Although SCA from satellite products have been used in various studies to calibrate spatially distributed snow models to improve model-internal consistency (Duethmann et al., 2014; Freudiger et al., 2017; Thornton et al., 2021), to the best of our knowledge, satellite data have rarely been used to calibrate snow redistribution routines, which are often only calibrated against observed discharge (Warscher et al., 2013; Frey and Holzmann, 2015). In a recent publication, Thornton et al. (2021) also considered parameter optimization of gravitational redistribution in a novel calibration approach for an energy balance-based snow model including snow cover maps derived from Landsat-8.

In this work, we aim to develop a high-resolution, spatial and temporal method of deriving snow cover and, further, to calibrate and evaluate the WaSiM (Water Flow and Balance Simulation Model) snow module (Schulla, 2021). Hence, the objective is to perform a process-oriented analysis of gravitational snow redistribution and energy balance to accurately estimate the snow cover extent. Therefore, we generated two alternative high-resolution snow cover products from Sentinel-2 images for a high-elevation Alpine catchment with a very sparse snow observation network. The use of two products, allows us to acknowledge the uncertainty affecting also satellite products and to consider it in the calibration of the model. To accommodate cloud coverage, each of the two snow cover products used a different snow and cloud detection algorithm and we quantified the difference between the resulting snow cover properties. Cloud detection enables the inclusion of partially clouded Sentinel-2 images in the snow cover dataset by setting clouded areas to no data. Even with cloud cover, these images may contain valuable information beyond the cloud extent regarding the snow cover evolution during the melting season and during the onset of snow accumulation. Finally, snow cover maps with high spatial and temporal resolutions allow an accurate analysis of the effect of topographical features (elevation, slope and aspect) on the snow cover duration and provide useful information for assessing simulated snow

cover maps from a physically based hydrological model for a period of five years.

The main novelties of this paper consist in the application of two different, i.e. one unsupervised and one supervised, snow and cloud detection algorithms for Sentinel-2 images, their use for hydrological model calibration in a topographically complex region, including an assessment of the differences between the two products and an investigation of the effects of topography (elevation, slope and aspect) on mean snow cover duration.

2. Data and methods

2.1. Research area and in-situ snow observation

The upper Martell valley (Martelltal, Val Martello) is an Alpine valley with a SW-NE orientation located in the upper Adige catchment in South Tyrol (Italy), covering an area of 65 km^2. To evaluate the satellite products, the research area was slightly expanded to the west to include the snow observations collected at the Madritsch station located in the neighbouring Sulden valley. The upper Martell valley has an elevation range of 1840 m a.s.l. to 3760 m a.s.l. with a mean altitude of 2814 m a.s.l. (Sonny, 2017), and the dominant land cover classes are bare rock (41%), sparse vegetation (34%) and glaciers (19%) (Environment European Agency CLC Corine 2018). Coniferous forest covers only 6% of the catchment at elevations lower than 2370 m a.s.l.. Along the ridges, extremely steep slopes with inclinations of>60° (highlighted in red in the overview map Fig. 1 can be found throughout the area. The main river of the upper Martell valley is the Plima, which flows into the Zufritt (Gioveretto) reservoir. The runoff regime is dominated by glaciers and snow melt, with high flows during the spring and summer and low flows during winter (Puspitarini et al., 2020). There are also 107 lakes of proglacial origin in the catchment. Previous hydrological studies investigating the mass balance of the Langenferner glacier (Galos et al., 2017; Galos and Klug, 2015) and the impact of glacier shrinking on hydropower production (Puspitarini et al., 2020) also focused on this study area. In addition, the catastrophic flood event from 1987 was the subject of several vulnerability assessment studies (Totschnig and Fuchs, 2013; Papathoma-Köhle et al., 2015).

There is only a sparse observation network of meteorological stations covering the upper Martell valley. The station at the Langenferner (LA) glacier is the only one located inside the upper Martell valley. At the Zufall snow station, snow surveys are performed manually at weekly intervals during the winter period. Continuous snow depth observations are available from the Madritsch, Zufritt and Rossbaenke stations from various elevations and valleys. The investigation period of this study covers the period with most available station data from October 2014 to September 2020. Table 1 gives detailed information of the in-situ snow observations and meteorological stations and the latter are used as model forcing in the hydrological model WaSiM.

2.2. Methods

Two snow and cloud detection approaches, named TUM (unsupervised algorithm) and Eurac (supervised algorithm), respectively, are applied to identify the snow covered area in the catchment over the study period. These data sets are compared to the hydrological model WaSiM results with different parametrizations of snow redistribution (i.e. default and optimized parameters). Fig. 2 gives a comprehensive overview of the individual processing steps used in the observed and simulated snow cover maps.

2.2.1. TUM approach for Sentinel-2 snow cover map

Covering most of the period June 2015 to October 2020, 439 Sentinel-2 Level-1C images were downloaded with the R toolbox Sen2r

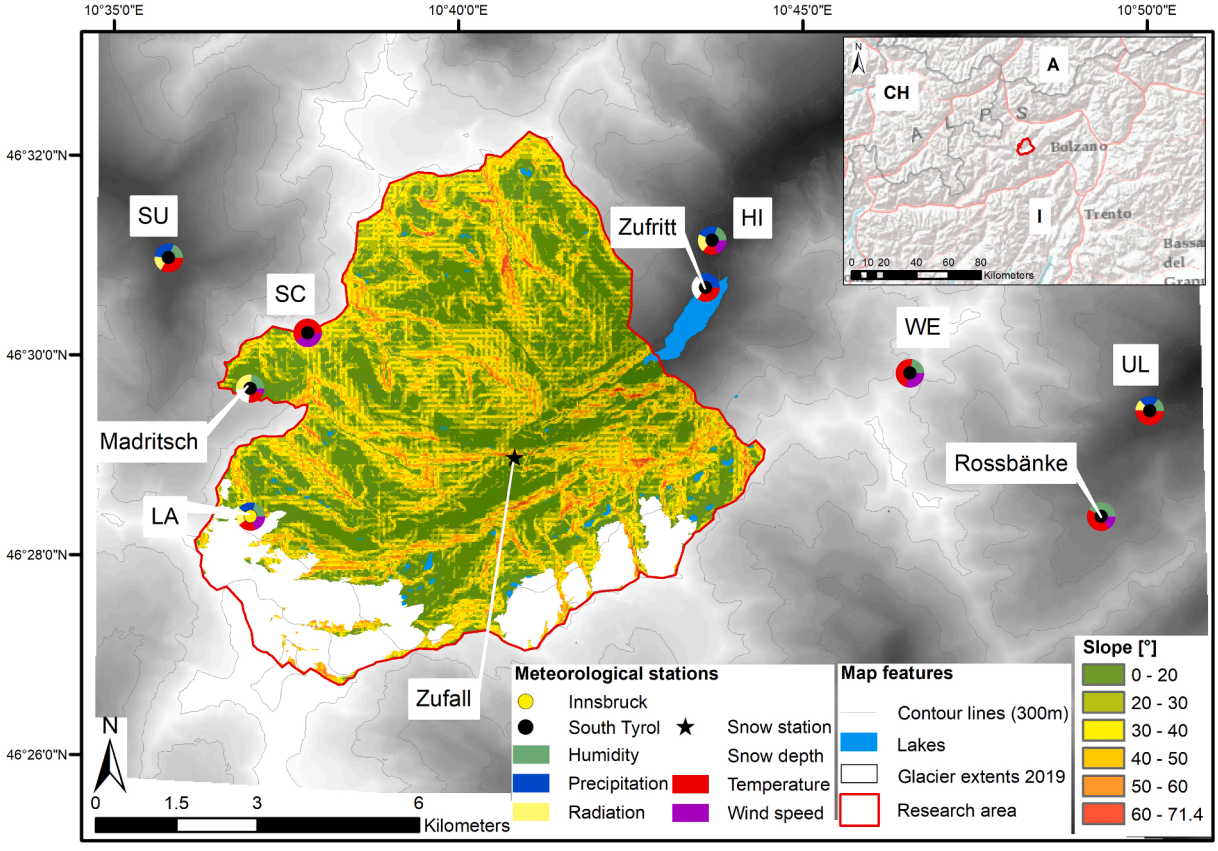


Fig. 1. Map of the research area, the upper Martell valley in South Tyrol (Italy), showing in-situ snow observation stations (full names) and meteorological stations (abbreviations) (Source layer. Esri, Garmin, USGS, NPS, NOAA).

Table 1
Meteorological and snow observation stations. Measured variables: P = Precipitation, T = Temperature, WS = Wind speed, H = Humidity, R = Radiation, SD = Snow depth.

Station	Acronym	Elevation [m a.s.l.]	Latitude	Longitude	Measured variables	Resolution	Temporal coverage	Valley	Provider
Madrisch	–	2825	46.4938	10.6144	P, T, WS, H, R, SD	10-min	2000–2020	Sulden	1
Rossbänke	–	2255	46.469351	10.819436	T, H, WS, SD	10-min	2015–2020	Ulten	1
Zufall	–	2265	46.48129	10.67802	SD	Weekly (manual)	2004–2020	Martell	1
Zufritt	–	1851	46.509063	10.725072	P, T, SD	Daily	1980–2020	Martell	1
Hintermartell	HI	1720	46.5169	10.7269	P, T, WS, H, R	10-min	2009–2020	Martell	1
Langenferner-Felsköpfl	LA	2967	46.47245	10.61391	P, T, WS, H, R	10-min	2012–2020	Martell	2
Schöntaufspitze	SC	3328	46.5029	10.6286	T, WS	10-min	1998–2020	Martell	1
Sulden	SU	1907	46.5159	10.5953	P, T, WS, H, R	10-min	1987–2020	Sulden	1
Ulten Weißbrunn	UL	1900	46.4868	10.8318	P, T, H, R	10-min	1987–2020	Ulten	1
Weißbrunnspitze	WE	3252	46.494	10.774	T, H, WS	10-min	2012–2020	Ulten	1

[1] Autonomous Province of Bozen/Bolzano - South Tyrol
[2] Institute of Atmospheric and Cryospheric Sciences, University of Innsbruck

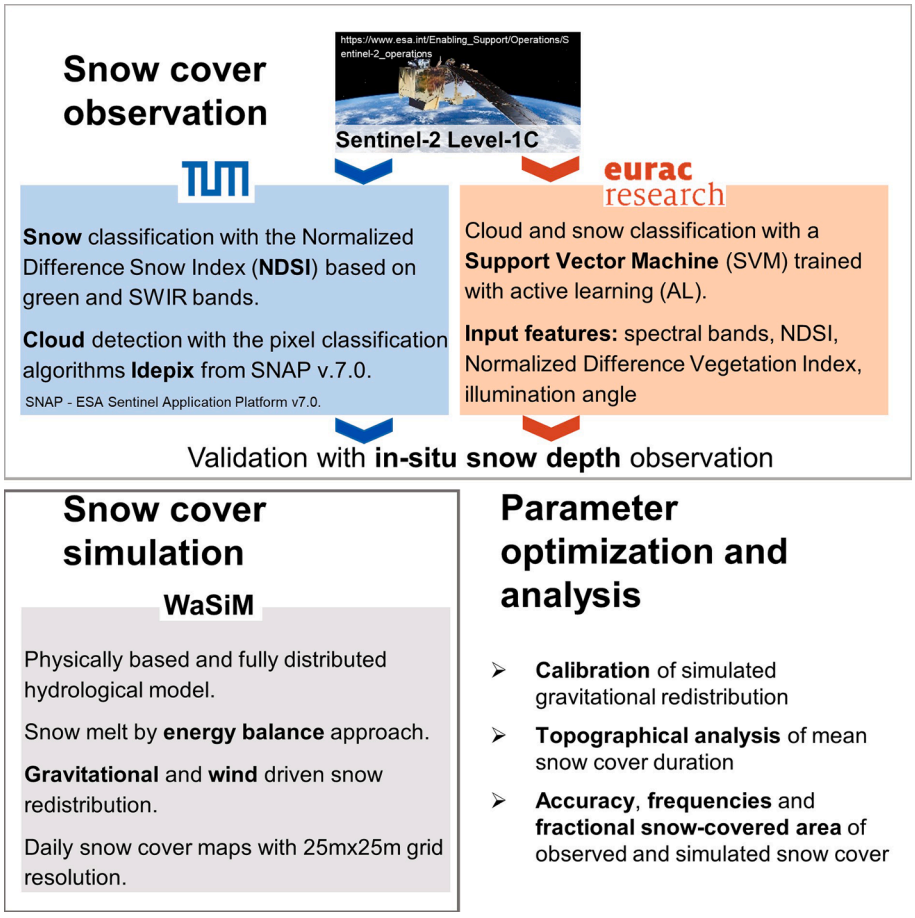


Fig. 2. General workflows for observed and simulated snow cover maps.

(Ranghetti et al., 2020) for the area of interest (AOI) including all seasons of the year. The Level-1C product type provides Top of the Atmosphere reflectance in cartographic geometry projected in UTM/WGS84 (European Space Agency, 2021). The spectral bands, central wavelength and spatial resolution of the Sentinel-2 MultiSpectral Instrument (MSI) are listed in the supplementary material of Hofmeister et al. (2022). Using Eq. (1), the Normalized Difference Snow Index (NDSI) was calculated for each image, after resampling band 11 from 20 m to 10 m resolution (Cimpianu, 2018). The NDSI uses the distinct reflection signature of snow surfaces with very high reflectance in the visible spectrum and strong absorption in the short-wave infrared range (Gascoin et al., 2019).

$$NDSI = \frac{green - SWIR}{green + SWIR} \quad (1)$$

where *green* corresponds to Sentinel-2 band 3 (wavelength: 0.560 μm) and *SWIR* is short-wave infrared represented by band 11 (wavelength: 1.610 μm). An NDSI pixel is classified as snow if the NDSI value is >0.4 (Dozier, 1989). Larger NDSI values are classified as snow-free. The NDSI threshold was the same for all images. Several cloud detection approaches have been developed for Sentinel-2 images, such as Sen2Cor, Idepix, Fmask, MAJA and Sentinel Hub's Cloud Detector (Menekay, 2019). In this study, cloud detection was performed using cloud masks created by the Identification of Pixel Properties algorithms (Idepix) from

the Sentinel Application Platform (SNAP v7.0) (European Space Agency, 2021) at 60 m spatial resolution. Idepix is a single-scene method and therefore suitable for cloud detection over changing terrain conditions, such as snow accumulation and melt. Since it is available as an SNAP plugin, it is very user-friendly, fast and effective at providing cloud masks. Idepix combines cloud masks with sun geometry to search regions of maximum probability for cloud-shadow pixels. In the projected region of potential cloud shadow, the cloud mask is shifted along the illumination path towards the surface reflectance minimum. Idepix can identify different types of pixels, such as semi-transparent clouds, fully opaque clouds, cirrus clouds, cloud shadows and even mountain shadows. This study employs Idepix_Cloud classification, which includes cloud pixels identified with full confidence as well as pixels identified with uncertainty as clouds. Clouded pixels were set to no data.

The snow and cloud maps were masked to the extent of the area of interest and resampled with the nearest neighbour method to the same spatial resolution as that of the hydrological model (25 m). We excluded highly clouded (>80%) images from further analysis, which reduced the total number of images to 366. Following an initial assessment of the cloud maps, we observed that the cloud maps generated with Idepix tended to overestimate cloudiness due to mountain shadows, since low illuminated areas with low spectral reflectance are misclassified as clouds. Therefore, only cloud covered areas for which no snow was detected by the NDSI were set to no data. This enables the snow detection under clouds thanks to semi-transparency in SWIR, which would otherwise be masked by the Idepix cloud mask. In addition, the cloud masks are very conservative, since they are calculated with a coarser resolution (60 m) compared to the snow maps 20 m), which would mask out too many pixels, especially in the peripheral areas of the clouds, for which the snow detection might be correct (Serco Italia, 2017). Water bodies were also set to no data, since their reflectance is often very similar to that of a snow surface (Dozier, 1989).

To test the utility of atmospheric correction over satellite imagery for snow cover detection, a set of atmospherically corrected Sentinel-2 images (September 2018 - October 2019) was analysed and compared against the methods used. The correction was applied using 'Second Simulation of the Satellite Signal in the Solar Spectrum' (6S), as developed by Vermote et al. (1997), which uses Radiative Transfer Models (RTMs) to simulate the passage of solar radiation across the atmosphere. The 6S algorithm is adapted to a Python (Py6S) interface (Wilson, 2012) and was implemented recently for use with the Google Earth Engine (Murphy, 2018) via a Python API and Docker container.

2.2.2. Eurac approach for Sentinel-2 snow cover map

The Eurac snow algorithm also takes the Sentinel-2 Level-1C data as its input. As pre-processing step, the Sentinel-2 images are scaled from digital number (DN) to reflectance values using the quantification value provided in the Sentinel-2 metadata (i.e., equal to 10,000 for all the considered images). Thus, all bands are reprojected and resampled to the final model resolution of 25 m using cubic interpolation. The classification algorithm consists of two steps: i) cloud detection; and ii) snow detection. In both cases, a support vector machine (SVM) classifier is trained with an active learning (AL) procedure. By dividing the procedure into two steps, it is possible to exploit the most representative features of each of the two classification problems. The AL procedure allows us to speed up the learning curve of the classification by asking the user to specify the label of the most uncertain pixels. For cloud detection, when a pixel is classified as cloud or no-cloud, all spectral bands of the Sentinel-2 are considered except for those at 60 m resolution (i.e. bands 1, 9 and 10). Moreover, the Eurac approach uses additional features, such as the normalized difference vegetation index (NDVI), calculated as the normalized difference between near the infrared (NIR) and red bands in addition to the NDSI, which have been shown to introduce benefits into the classification (Tarrio et al., 2020).

For snow detection, there are three possible classes: snow, snow-free and hard shadow. The difference between shadow and hard shadows is defined in terms of energy recorded by the sensor. If the recorded energy is too low to distinguish between snow and snow-free areas, i.e. the signal-to-noise ratio (SNR) of the sensor is too low, we call it hard shadow. This is generally the case when the sun is low on the horizon, which is from approximately mid-November to mid-February at the latitude of the Martell valley, and the terrain is particularly steep. In order to define a threshold under which the recorded energy by the sensor is too low, we selected samples for which the human photo-interpreter cannot distinguish whether the pixel is snow or snow free. Low reflectance situations, generally associated with dark areas e.g., turbid lake, shadow by cloud or terrain, flooded areas, etc., are excluded from snow detection from state of the art and operational algorithms applying thresholds on the visible bands e.g., MODIS snow cover product version 6 (Riggs and Hall, 2015). The shadow detection is performed with a SVM classifier together with the snow detection after masking out the clouds. In addition to the spectral bands and features also used for cloud detection, the classifier includes illumination angle as a further input feature. The illumination angle is calculated from the solar zenith and the solar azimuth angle (Riano et al., 2003) and enables better mitigation of the effect of the differences in the solar illumination during the year.

The SVM has been trained in order to emphasize the presence of snow also in mixed conditions. This is done by assigning the class snow to those pixels whose spectral characteristics can be attributed to the presence of snow by a careful visual inspection. For both SVM models (snow and cloud), we used a radial basis function kernel and performed a model parameters selection according to 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 start with a coarse grid and then given the obtained results is refined around the values of C and gamma that perform the best. The best values are selected by evaluating the mean and standard deviation of the over accuracy calculated in a cross-validation strategy with k folder ($k = 5$). For the cloud detection, we collected 173 training points (97 "cloud free" samples and 76 "cloud" samples). The selected parameters are C equal to 9.885 and gamma equal to 0.896. The number of selected support vectors is 98. The accuracy from the cross validation is 0.82 and the standard deviation is 0.1 by considering the 5-folds. The low accuracy during the training shows the difficulty in the cloud discrimination during winter condition with the spectral bands provided by Sentinel-2 sensor. For the snow detection, we considered three classes of training samples. We collected 177 "snow free" samples, 264 "snow" samples and 138 "shadow" samples, for a total amount of 579 samples. The selected C value is 30.018 and gamma is 0.0003. The number of support vectors was 445 and the accuracy from the cross validation is 0.98 with a standard deviation of 0.1. The samples are collected initially from two scenes and then we adopted an iterative active learning procedure to collect new samples. The active learning procedure ask the user the labels of the most uncertain samples, i.e., the ones with computed probabilities for the assigned class are low. The probability was calculated according to Lin et al. (2007).

It should be mentioned that at the time of the analyses for this publication, the Eurac approach is still in the final stage of the development but it was applied already once in a previous study (Ebner et al., 2021). The Eurac approach was chosen to verify the TUM snow detection results because the supervised algorithm should be more accurate than the unsupervised TUM product. As in the TUM method, water bodies were set to no data.

2.2.3. Snow cover simulation with WaSiM

The snow module of the physically based hydrological model WaSiM (version 10.04.07) (Schulla, 2021) was used as an example for demonstrating the benefits of high-resolution snow cover maps in calibrating simulated snow redistribution. For snow melt simulation, the energy balance approach, including gravitational snow redistribution, was selected, as originally implemented by Warscher et al. (2013). We also

tested the effect of wind-driven snow redistribution (Warscher et al., 2013; Schulla, 2021). Canopy snow interception was not considered in this study because of the small proportion of coniferous forest (6%). The multi-layer snow model was deactivated to keep the computational time reasonable, limit the number of calibration parameters and avoid the need to define multiple soil and snow parameters. The topography-dependent adjustment of radiation and air temperature follows the scheme devised by Oke (2002). The parametrization of the snow accumulation and gravitational snow redistribution (Table 2 originates in part from the WaSiM user manual (Schulla, 2021) and Förster et al. (2018), whereas the correction factor for incoming long-wave radiation (*LWINcorr*) was slightly increased from 1.0 to 1.1 considering the observations available for May and June. The equations for energy balance approach, *Psnow*, *LWINcorr* and *LWOUTcorr* can be found in the WaSiM user manual, which is available in the supplementary material (Hofmeister et al., 2022). To account for solid precipitation undercatch of rain gauges, we used a wind-dependent snow correction factor (*Snowb*) of 0.1 s/m, as also used by Förster et al. (2018) and confirmed by Kochendorfer et al. (2016).

Gravitational snow redistribution is based on a computationally efficient, mass-conserving algorithm that parameterizes the gravitational transport and deposition developed by Gruber (2007) and later implemented in WaSiM by Warscher et al. (2013). Four parameter grids that specify the sliding fractions for each of the cardinal directions (N, E, S, W) were generated using the WaSiM preprocessing tool Tanalys. Total outflow from the gravitational model (*Mout*) is determined by Eq. (2).

$$Mout = \begin{bmatrix} ferosion \times Min - Dgrav, & if\ i \geq ierosion\ and\ SWE > 0 \\ Min - Dgrav, & if\ i \geq ierosion\ and\ SWE = 0 \\ Min - Dgrav, & if\ i < ierosion \end{bmatrix} \quad (2)$$

where *Min* is the inflowing mass from all other neighbouring cells [mm], *SWE* is the snow water equivalent in the current cell [mm], *ferosion* is the erosion factor, which depends on the time step, *i* is the local slope, *ierosion* is the lower inclination limit for snow erosion [°] depending on the spatial resolution and *Dgrav* is the deposition in the current cell [mm]. The current snow deposition in a cell *Dgrav* is defined by Equations (3) and (4).

$$Dgrav = \begin{Bmatrix} Min, & if\ Min < Dmax, grav \\ Dmax, grav, & if\ Min \geq Dmax, grav \end{Bmatrix} \quad (3)$$

Table 2
WaSiM parameters for snow accumulation, gravitational slides, ablation and wind driven redistribution.

Process	WaSiM Parameter	Description	Values
Snow accumulation	T0R	Temperature limit for rain (°C)	0
	Ttrans	½ of temperature-transient zone for rain-snow (°C)	0.5
	Snowb	Wind-dependent snow precipitation correction (s m^{-1})	0.1
Gravitational redistribution	i_{lim}	Maximum deposition slope (°)	55
	D_{lim}	Scaling for maximum deposition (mm)	2
	$i_{erosion}$	Minimum slope for creating slides (°)	50*
	$f_{erosion}$	Fraction of snow pack that forms the slide (–)	0.002*
Snow ablation	LWINcorr	Correction factor for incoming long-wave radiation (–)	1.1
	LWOUTcorr	Correction factor for outgoing long-wave radiation (–)	1.0
Wind redistribution	start azimuth	1st quantile of wind direction (°)	180
	end azimuth	3rd quantile of wind direction (°)	270
	cmin	Minimum correction factor (–)	0.3

* These parameters have been optimized. See Section 3.3 for more information.

$$Dmax, grav = \begin{Bmatrix} \left(1 - \frac{i}{ilim} \times Dlim\right), & if\ i < ilim \\ 0, & if\ \geq ilim \end{Bmatrix} \quad (4)$$

where *Dlim* is the upper deposition mass limit [mm], e.g. the maximum amount of snow that would be deposited on horizontal terrain, and *ilim* is the upper slope limit [°] at which all inflowing masses will be transported to the next downslope cell(s). In this work, we show how *ierosion* and *ferosion* can be calibrated to improve the model's performance using Sentinel-2 snow cover data.

Wind-driven snow redistribution can also be simulated with WaSiM. However, the approach implemented is in fact a kind of snow precipitation correction, by which areas sheltered from the main wind direction receive an increased snow fall and areas exposed to the main wind direction receive a reduced snow fall amount. Which cells are exposed or sheltered is estimated using the directed sky view factor (*SVFdir*). For this study, the main wind direction was set to south west (180° to 270°) which corresponds with the main wind direction of the Langenferner meteorological station (mean wind direction 229° ± 54°) for the snow accumulation period (October to May). The snow precipitation is corrected by Equation (5).

$$Psnow = Psnow + Cwind \times Psnow \quad (5)$$

where *Psnow* is the solid precipitation and *Cwind* is the correction factor, which is determined by Equation (6).

$$Cwind = e \times (Dmax \times (1 - SVFdir) - 1) + cmin \quad (6)$$

where *e* is a linear elevation weighting factor, *SVFdir* is the directed sky view factor, *Dmax* is the maximum possible deposition, and *cmin* is the minimum correction factor for shifting *cwind* to a more or less solid precipitation correction. The parameter *e* ranges from 0 at the lowest elevated pixel to 1 at the highest pixel and linearly scales the amount of snow redistribution (Warscher, 2014). The impact on the mean snow cover duration was tested on different spatial scales by performing different model simulations with and without activated wind-driven snow redistribution.

The time series of all meteorological stations, illustrated in Fig. 1 and listed in Table 1, are spatially interpolated with an elevation-dependent regression (i.e. temperature, wind speed and humidity) and inverse distance weighting (IDW) for precipitation and solar radiation. The simulation period covers six years from 2014/10/01 to 2020/09/30 in hourly time steps. However, the simulated snow maps were only saved as daily means. The selection of the spatial resolution (25 m) was a compromise between the level of detail and the computational demand. Previous hydrological applications of WaSiM used a 50 m spatial and 1 h temporal resolution for Alpine catchments (Kraller et al., 2012; Warscher et al., 2013; Förster et al., 2018) or even 25 m in the most recent application (Thornton et al., 2021). Two different exceedance threshold values (0 mm and 5 mm SWE) for classification as either snow or no snow were tested.

To ensure consistency in the comparison with the observed snow maps, clouded areas (from the TUM product) and water bodies were set to no data in the snow cover maps produced by WaSiM. Since the dynamic glacier model of WaSiM requires a comprehensive calibration, which increases the model complexity, it was not activated, and glaciered areas were set to constantly snow covered in the simulated snow maps. Although glaciers are a key contributor to runoff generation in this area, the focus of this study is to examine the spatio-temporal variability of the snowpack. Glaciered areas were derived for each year from the cloudless snow cover map for August or September taken from the Eurac product, since the TUM product showed an tendency of false snow detection on some very steep (>60°) north-facing slopes, which would partly lead to wrong glacier delineation.

2.2.4. Analysis of observed snow cover maps

The quality of snow/no snow detection is evaluated against the in-situ snow observation by calculating the accuracy score F1 (Equation (7)) based on the confusion matrix in Table 3. The F1 score divides the sum of all true positive (TP) and true negative (TN) matches by the total population (n). Accuracy was only computed at three snow observation stations with continuous snow depth recordings.

2.2.5. Analysis of topographical feature on observed and simulated snow cover

The influence of topographic characteristics, e.g. elevation, slope or aspect, on the observed and simulated mean snow cover duration is analysed such that all grid cells with the same topographical characteristics (e.g. elevation, slope and exposition) are aggregated to mean snow cover duration over the entire period of 6 years following equation (7). The snow cover duration SCD [d] for each pixel of the domain in an hydrological year was computed following Dietz et al. (2012):

$$SCD = \frac{365.25}{N} \sum_{i=1}^{N} (si) \tag{7}$$

where N is the number of days with Sentinel-2 recorded data, beginning with 1 October and ending with 30 September of the next year. On average, the length of a year is 365.25 days considering leap years. *si* refers to the cloud-free daily snow cover data set recorded to values one for snow and zero for snow-free area. Accordingly, we calculated the mean and standard deviation of snow cover duration of all pixels of the 366 snow cover maps with respect to the three topographical feature classes (i.e. aggregation of elevation in ten meter classes, slope and exposition in three degree classes).

2.2.6. Comparison of observed and simulated snow cover maps

A pixel by pixel spatial analysis was performed to evaluate the simulated snow cover maps against the two different observed snow maps at the catchment scale. Two additional performance measures are thus introduced: the F2 score (Equation (8)) and the F3 score (Equation (9)), which are often used to evaluate binary classifications (Aronica et al., 2002; Warscher et al., 2013; Thornton et al., 2021). The range of the F1 and F2 scores is from zero to one while F3 ranges from -∞ to one. All scores are 1 if the simulated snow cover perfectly matches the observed one. The F1 score tends to show the highest performance values, since it takes only true positives and true negatives into account, and these are usually high during winter and summer. F2 and F3 exclude true no snow pixels and are therefore more sensitive to differences in snow cover (Warscher et al., 2013). F3 is even more sensitive to snow extent than F2 due to the subtraction of the false positives in the numerator.

F1 score

$$F1 = \frac{\sum_{i=1}^{n} TP + \sum_{i=1}^{n} TN}{n} \tag{8}$$

F2 score

$$F2 = \frac{\sum_{i=1}^{n} TP}{\sum_{i=1}^{n} TP + \sum_{i=1}^{n} FP + \sum_{i=1}^{n} FN} \tag{9}$$

F3 score

$$F3 = \frac{\sum_{i=1}^{n} TP - \sum_{i=1}^{n} FP}{\sum_{i=1}^{n} TP + \sum_{i=1}^{n} FP + \sum_{i=1}^{n} FN} \tag{10}$$

Table 3
Confusion matrix for evaluating the detected and simulated snow cover.

	Observed snow	Observed no snow
Detected/simulated snow	True positive (TP)	False positive (FP)
Detected/simulated no snow	False negative (FN)	True negative (TN)

For a further spatial comparison of observed and simulated snow products, the fractional snow-covered area (fSCA) is also considered. fSCA is defined as the ratio between the number of pixels classified as snow-covered and the total number of pixels excluding clouds and lakes (Di Marco et al., 2020), as shown in Equation (10).

$$fSCA = \frac{Nsnow}{Ntot - Nclouds} \tag{11}$$

where *Nsnow* is the number of snow cover pixels according to the Sentinel-2 dataset or WaSiM, *Ntot* is the total number of pixels representing the overall catchment area, and *Nclouds* are the pixels classified as cloud and water bodies.

3. Results

3.1. Comparison of cloud detection by TUM and Eurac

In terms of the empirical cumulative cloud detection frequency, Eurac detects a higher proportion of cloud within the range of 15% to 80% areal coverage than the TUM product (Fig. 3a). However, both products are able to detect complete cloud cover and clear sky situations equally well. Besides cirrus (high-altitude clouds) and other clouds, the image processing algorithms also include topographic shadows, which are of particular interest in complex terrains such as the upper Martell valley. The ability to detect topographic shadows is especially important in mid-latitude areas during the winter period, where the solar elevation is very low, typically below 20° (Gascoin et al., 2020). Differences between the two approaches and their limitations in terms of cloud and shadow detection are illustrated in an example for January 27, 2017, when no clouds were present in the AOI (Fig. 3b). However, the low inclination angle of the sun produces extensive shading in the southern and eastern parts of the AOI. Cloud detection by TUM (Fig. 3c) falsely classifies some snow-free ridges or snow-free steep slopes as clouds. On the other hand, the Eurac method (Fig. 3d) detects more shaded areas in the south eastern area. However, both approaches fail to fully detect the topographic shadows on the forested southern valley bottom, which are classified as snow-covered regions.

3.2. In-situ comparisons of snow cover detection by TUM and Eurac

The quality of snow detection was tested against three snow observation stations for various snow depth thresholds that differentiate between snow and no snow. The detection accuracy was calculated by the F1 score (Equation (7)). Snow detection varies not only according to the chosen detection method (TUM or Eurac) but also according to the observation sites (Fig. 4 and supplementary table S.7 in Hofmeister et al. (2022)). The most accurate score (F1 > 0.97) was obtained with both products for snow depth thresholds ranging between 4 cm and 15 cm for the Rossbaenke station, which is surrounded by grassland. Both TUM and Eurac attained a similar detection accuracy at the Madritsch station, however decreasing with increasing snow depth threshold. The highest accuracy of 0.96 (TUM) and 0.97 (Eurac) is reached at a snow depth threshold of 1 cm. Due to the station's relatively high-altitude (2825 m a.s.l.), the land cover of the surrounding area is mainly composed of gravel and debris. Therefore, it can be assumed that there is no interference between the spectral characteristics of the snow cover and those of the vegetation. The main differences between the snow detection products are observed for the lowest snow observation station of Zufritt, which is surrounded by forest and grassland. Whereas the TUM product has a detection accuracy >0.91 for snow depth thresholds between one and ten cm, the detection accuracy of the Eurac product does not exceed 0.83 for the same threshold range. The lowest performance is probably due to the presence of permanently mixed pixels, such as forest and grassland. In fact, the spectral characteristics of these pixels do not show a snow presence >50% (see 2.2.2). In this case, a canopy correction is

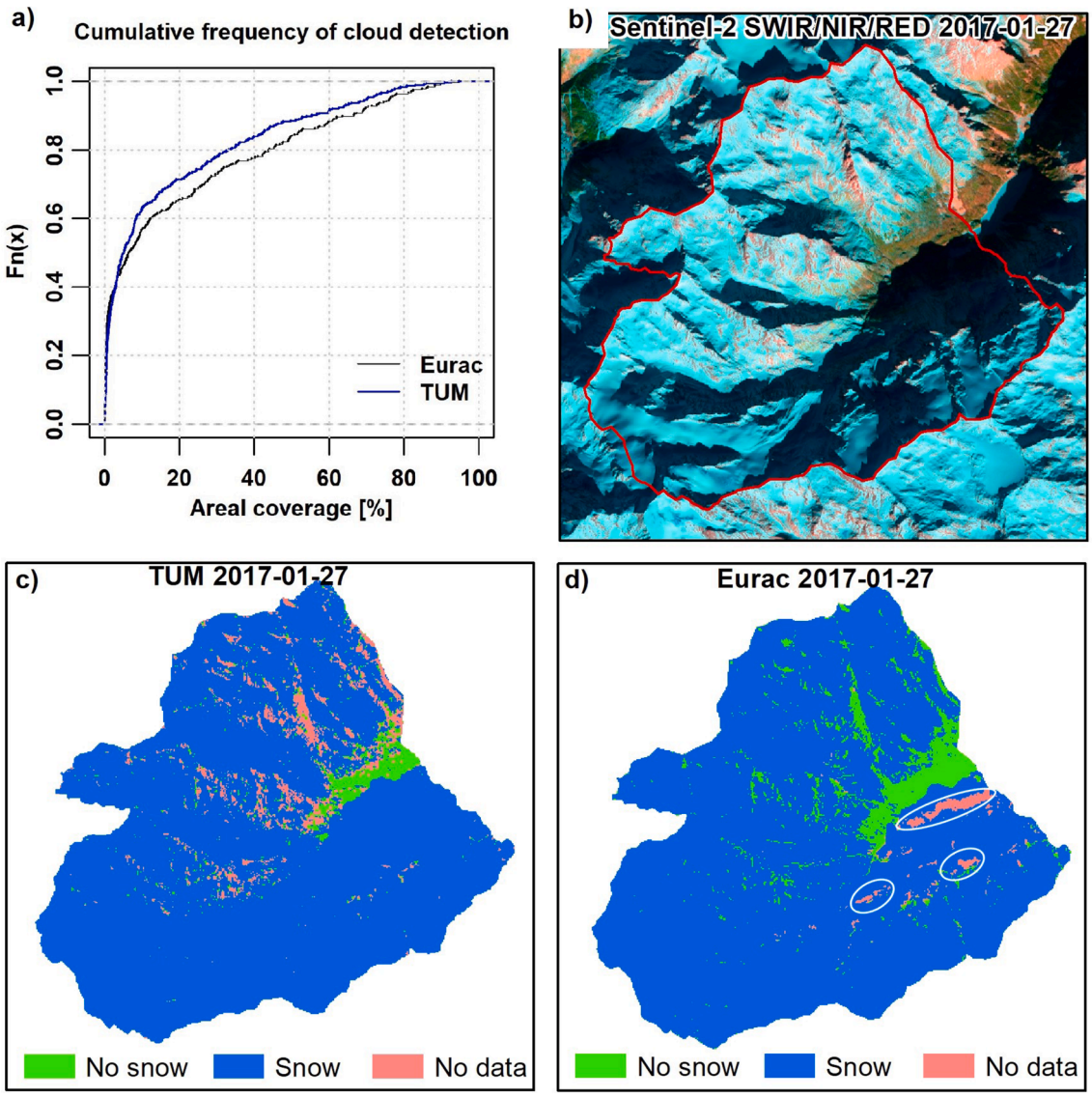


Fig. 3. Empirical cumulative frequency of cloud detection by TUM and Eurac for the whole data set (a), Sentinel-2 false colour composition (SWIR/NIR/RED) in 20 m resolution for one case study on January 27, 2017, showing the AOI boundary (red line) (b) and the cloud and snow detection by TUM (c) and by Eurac (d) in 25 m resolution, for the said date. No data pixels were considered as cloud or shadow by the algorithms. White circles in panel d indicate main differences between TUM and Eurac in detecting topographic shadows.

required to prevent error. The optimum snow detection accuracy of 0.95 for TUM and 0.92 for Eurac based on all stations was found at a 4 cm threshold, showing that, in general, a larger snow quantity, i.e. a closed snow cover with a few centimetres depth (~4 cm), is needed on the ground to enable better identification of snow from Sentinel-2 images.

We also tested two different exceedance thresholds for simulated SWE (0 mm and 5 mm) in order to convert SWE to binary snow cover maps. The commonly used exceedance threshold of 5 mm (Warscher et al., 2013; Schulla, 2021; Thornton et al., 2021) attained a slightly higher F1 prediction accuracy of 0.95, compared to 0.92 for 0 mm at the Madritsch snow observation station. However, since only one snow observation station with continuous snow depth recording is located in the research area, it is not possible to make any general statements about the robustness of the exceedance threshold.

3.3. Calibration and analysis of snow redistribution as modelled by WaSiM

Topographical analysis not only allows the impact of topographic features such as elevation, slope and aspect on the mean snow cover duration to be investigated, but also enables a process-orientated evaluation of the simulation results and calibration of the gravitational redistribution simulated by WaSiM with respect to different slope gradients. A total of four simulation results were analysed with different snow redistributions (Table 4). Simulations A and C were performed with the default WaSiM parametrization of *ierosion* (minimum slope for creating slides) and *ferosion* (fraction of snow pack that forms a slide), while simulations B and C used optimized parameters. For this analysis, all pixel values of all 366 snow cover maps were aggregated by each topographic feature class (elevation, slope and exposure) and the mean and average standard deviation of the snow cover duration (SCD) was estimated for each class (Fig. 5), as explained in section 2.2.5. Although the elevation gradient of the mean snow cover duration and the standard deviation bands (quantified for one mean standard deviation) are very similar for both observation and simulation in the range 2300 to 3500 m a.s.l., distinct differences can be observed in the lower (<2300 m a.s.l.) elevation zones. The Eurac snow maps distinctly show lower mean snow cover duration values than TUM and WaSiM, accompanied by

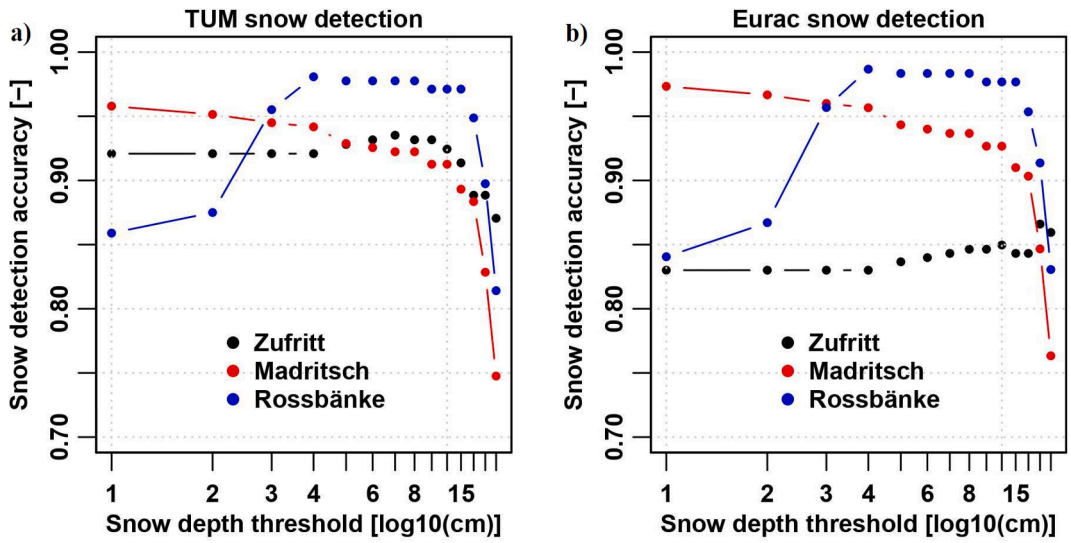


Fig. 4. Snow detection accuracy obtained with TUM (a) and Eurac (b) at three snow observation stations for different snow depth thresholds, shown on a log scale.

Table 4
WaSiM configurations for topographical analysis.

WaSiM Configuration	Simulation A	Simulation B	Simulation C	Simulation D
$i_{erosion}$	45°	50°	45°	50°
$f_{erosion}$	0.007	0.002	0.007	0.002
Wind redistribution	No	No	Yes	Yes

broadening of standard deviation band at the valley bottom (<2300 m a.s.l.) due to the presence of forest and a higher mountain shadow detection (supplementary Figure S.6 in Hofmeister et al. (2022)). Although the TUM product also displays higher variance in the low elevation range, it shows better agreement with the simulated snow cover duration. In the elevation range from 2800 to 3400 m a.s.l., relatively wide standard deviation bands indicate a large spatial heterogeneity in mean SCD. Mean SCD varies only slightly for the highest elevation bands (>3500 m a.s.l.).

The impact of *ierosion* on the mean SCD is visible on the slope feature class in Fig. 5. We can observe that an *ierosion* value of 45° does not reproduce the turning point in the mean SCD, as it is present in both the TUM and Eurac products. Hence, an analysis of the satellite data enables us to determine the value of *ierosion* by applying a grid search approach to find the minimum distance between the peak of mean SCD at slopes between 40° to 50° in observation and model results. The optimum *ierosion* value was found at 50°. The *ferosion* parameter was lowered in 0.001 steps from the default value (0.007) to 0.001 and the best parameter (0.002) determined taking into account the best combination of mean Pearson correlation (0.73) and RMSE (mean SCD of 25.7 [d]) with *ierosion* set to 50° (see supplementary Figure S.8 in Hofmeister et al. (2022)). Wind-driven snow redistribution was activated for the simulation runs C and D.

The small differences in performance of the simulated snow cover duration in the highest elevation zones (>3500 m a.s.l.) can be explained by the steeper topography, in which more snow is redistributed by gravitational slides. The greatest simulated gravitational redistribution occurs with the default parametrization of *ferosion* (Fig. 5a) in this elevation zone. In contrast, the optimized gravitational redistribution (Fig. 5b) shows better agreement with the observed mean SCD. The activation of wind-driven redistribution leads to a distinct reduction in the mean snow cover duration in the range from 2800 to 3500 m a.s.l. (Fig. 5b and d).

The effect of gravitational redistribution is most clearly visible on slope gradients steeper than 30°. With the default parametrization of *ierosion* (45°) and *ferosion* (0.007), WaSiM relocates snow from steep slopes (>45°) to flatter areas, which results in very low snow cover durations for slopes above 45°, accompanied by a sudden change in the standard deviation (Fig. 5a and c). The wind-driven redistribution has barely any detectable effect on the simulated mean snow cover duration on the slope feature class, with the exception of the steepest slopes (>60°), for which a small increase can be observed (Fig. 5b and d).

The final topographical characteristic to be evaluated is aspect (exposure). The mean SCD on areas of different exposure depends mainly on the incoming solar radiation (topographical shadowing) and on wind-driven snow redistribution. The influence of exposure on the mean SCD is clearly visible in Fig. 5, with the highest values (>250 d) on north- and north east-facing slopes and the lowest (<200 d) on south-facing slopes. As with the slopes, the relatively large standard deviation bands indicate a high variability of SCD in areas with similar exposure. Although the mean SCD shape is very similar for the TUM and Eurac products, a systematic offset between both products (overall mean of SCD by TUM 241 d and Eurac 226 d) is present, which is more pronounced with the northern exposure. This can be explained by the differences in cloud and mountain shadow detection by the products. While the Eurac cloud product shows no influence of aspect with a constant overall mean cloud coverage of around 20%, the TUM product detects more clouds on south-facing areas (overall mean cloud coverage of 20%) than on north-facing ones (overall mean cloud coverage of 14%) (supplementary Figure S.6 in Hofmeister et al. (2022)). Activated wind-driven redistribution leads to better agreement between the observed and simulated mean snow cover durations on areas from east to west exposure (Fig. 5d). In the following comparison, we will refer only to the best WaSiM setup with the optimized gravitational snow and activated wind redistribution (configuration D).

3.4. Comparison and accuracy of observed and simulated snow cover products

The cumulative snow and no snow frequency curves show a strong similarity in the snow detection and simulated values (Fig. 6a). The cumulative no snow frequency curves diverge with no snow coverage < 10% in the lower elevation ranges of the AOI. Although the frequency response with snow detection is quite similar, there is a systematic offset between the snow detected by TUM and Eurac from 10% to 90% areal coverage. Eurac tends to detect a lower snow cover probability than

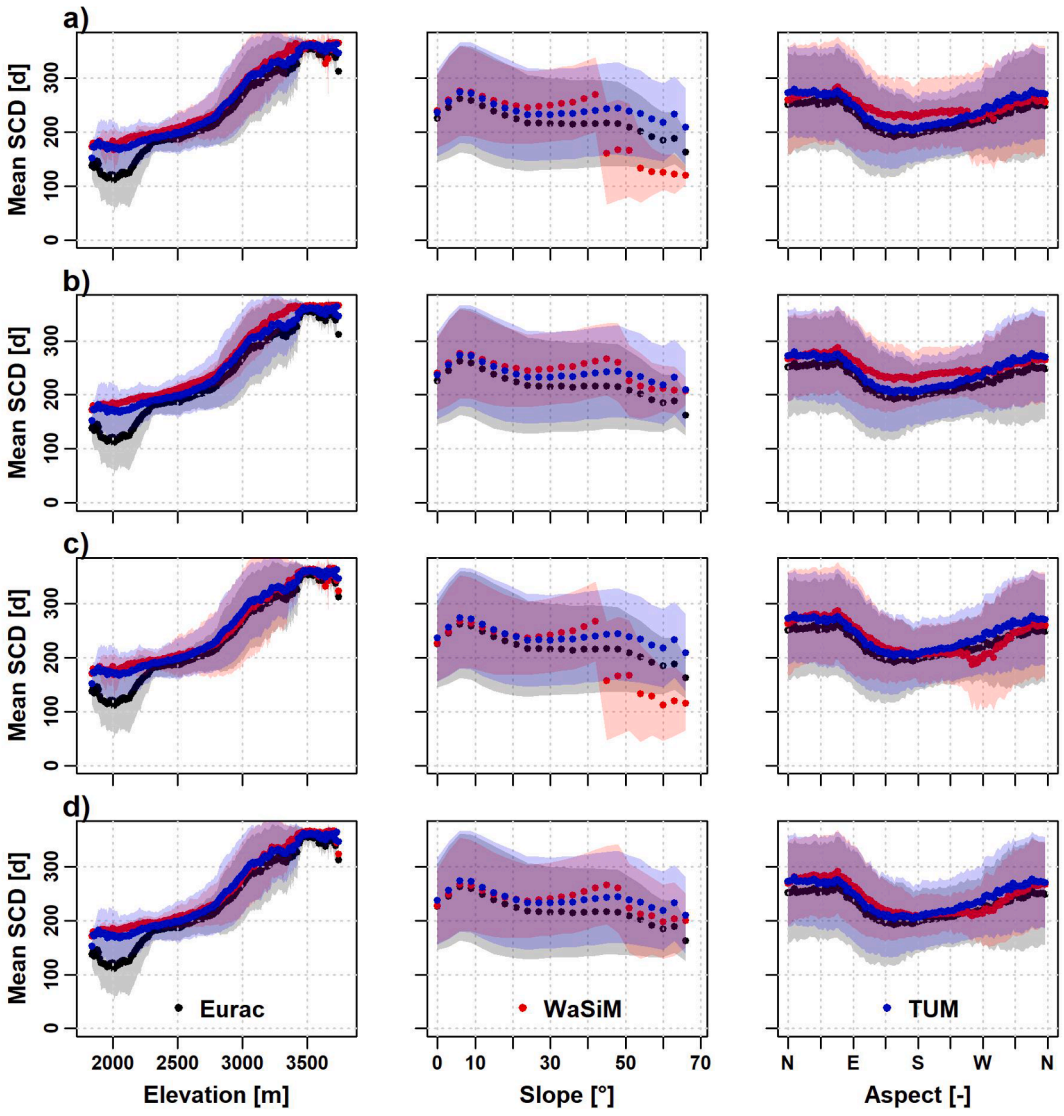


Fig. 5. Topographical analysis of observed and simulated mean snow cover durations with default parametrization of gravitational redistribution and without wind redistribution (a), with optimized parametrization of gravitational redistribution and without wind redistribution (b), with default parametrization of gravitational redistribution and activated wind redistribution (c) and with optimized parametrization of gravitational redistribution and activated wind redistribution (d) over elevation, slope and aspect plus corresponding mean standard deviation.

TUM. This is because the Eurac classification was trained with the majority of mixed pixels being considered as snow-free. The simulated snow cover frequencies follow the same response as the detections, but the model has a slight tendency to overestimate the snow covered area within the 60% to 90% areal coverage range.

Overall accuracy scores for both TUM and Eurac snow detection are very high (F1 = 0.97, F2 = 0.91, F3 = 0.9), only dropping in August to 0.8 for F2 and F3 respectively, as can be seen in Fig. 6b. The overall accuracy scores of TUM-WaSiM (F1 = 0.93, F2 = 0.83, F3 = 0.75) and Eurac-WaSiM (F1 = 0.93, F2 = 0.83, F3 = 0.71) provide further evidence that the model is able to simulate the spatial snow distribution in the research area well. Comparison of the mean scores on a monthly scale reveals the part of the season in which the model has a reduced accuracy in the snow cover distribution simulation. The critical months are the beginning of the accumulation periods (October, with an overall mean fSCA of 45% (Eurac), 57% (TUM) and 56% (WaSiM)) and ablation (June, with an overall mean fSCA of 52% (Eurac), 55% (TUM) and 58% (WaSiM)), although they still have a mean F1 score of above 0.8. The F2 and F3 scores exclude the no snow pixels and accordingly result in lower scores (Warscher et al., 2013).

3.5. Comparison of snow cover frequencies and fractional snow-covered area (fSCA)

Differences between observed and simulated snow cover days (SCDs) at the pixel scale were investigated in more detail for each hydrological year (1 Oct. – 30 Sept.), as illustrated in Fig. 7. In general, the hydrological years 2016 and 2017 had a shorter snow coverage period than the subsequent three years (2018, 2019 and 2020). The lower snow coverage in 2016 and 2017 results in better visual agreement between the two snow observation products. However, the relatively high standard deviations (>74 d) given in Table 5 indicate a greater variance in the snow cover days within the AOI. WaSiM simulates a higher snow cover duration for both years, which also results in a slightly lower standard deviation of SCDs (<73 d). A distinct tendency towards overestimation of the SCDs can be observed, especially for the hydrological year 2016.

For the snow rich years 2018, 2019 and 2020, TUM shows a higher frequency of mean SCDs in the range above 250 d than Eurac. In terms of the observed and simulated SCDs, the WaSiM results are more similar to those of TUM snow detection for the corresponding period than they are

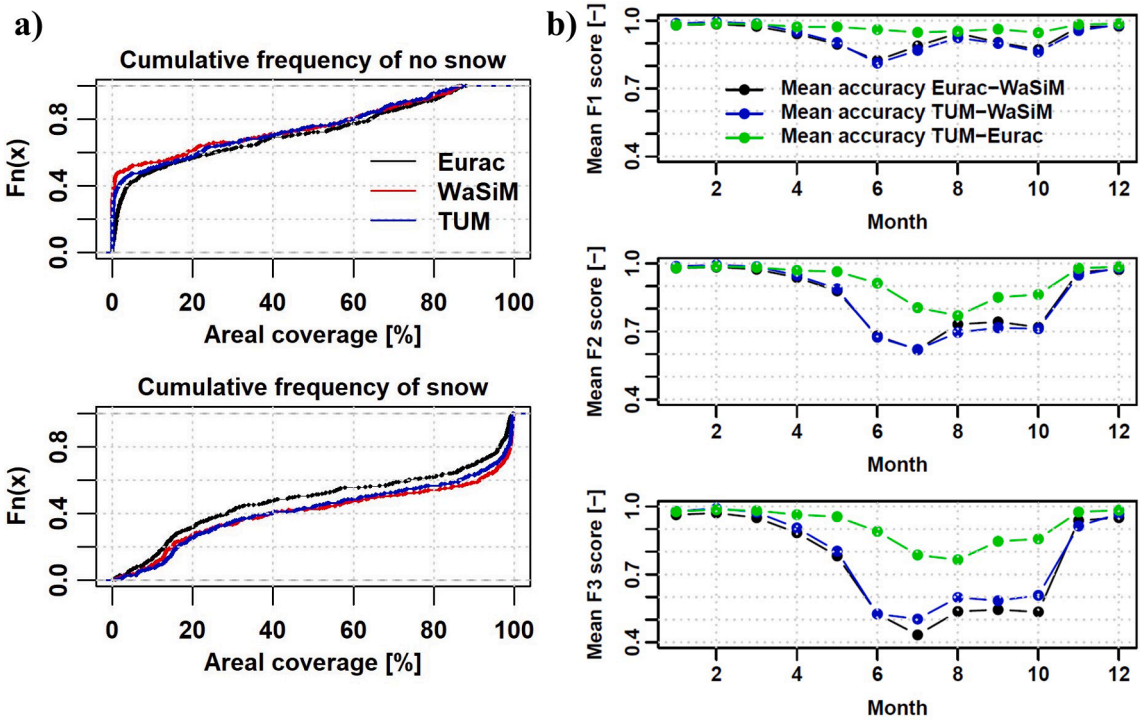


Fig. 6. Empirical cumulative frequencies of no snow and snow (a) and mean monthly accuracy of observed and simulated snow cover areas (b).

to those of Eurac. With increasing mean SCDs, the standard deviation decreases, indicating a lower heterogeneity in the snow cover duration. The overall mean gives the mean and standard deviation of each snow product for the whole investigation period 2016–2020. The overall mean SCDs of 236 d confirms the tendency of Eurac to underestimate the snow cover duration as compared with TUM (251 d) and WaSiM (258 d) and also shows the highest variability in snow cover duration, with a mean standard deviation of 64 d. The standard deviation shows that the snow cover duration varies greatly from year to year and declines in snowy winters with long snow cover durations (such as in 2018/2019). Moreover, the standard deviation of SCDs obtained from WaSiM is systematically lower than the one observed without taking canopy interference into consideration.

Fig. 7 shows the comparison among the frequencies of snow cover days and reveals the tendency of WaSiM to underestimate the snow cover duration in the 360 SCDs bin and to overestimate it in the following bin (>360 SCDs) corresponding to perpetual snow cover. Moreover, WaSiM simulation results in a longer snow cover duration than the two observation products for the classes 260 d and 270 d in the extreme winter 2018/2019.

Given that the frequency of snow cover days does not allow any spatial comparison between the two approaches and the simulated maps, Fig. 8 compares case studies for a hydrological year with (2020) and without (2019) good agreement. Eurac shows differences in snow detection at the valley bottom in forested areas and on steep slopes for both winters (2019 and 2020), these often being set to no data due to the presence of shadows. Moreover, the presence of clouds impacts snow and ice detection on the glaciered areas. The TUM snow cover product mostly has problems with overestimations of snow cover at the bottom of steep north-facing slopes. Topographical features such as ridges and slopes are highly visible in both products, as is interference from vegetation, especially with respect to the year 2020. As snow cover duration increases, topographical features become less visible, as is the case for 2019. However, Eurac also underestimates the snow cover days on the forested valley bottom in the snow rich year 2019 when compared with TUM. The hydrological model WaSiM is well able to simulate the spatial distribution of snow cover days for both seasons, but overestimates the snow cover duration at the valley bottom and to some extent at the highest elevations (>3000 m a.s.l.) for the snow rich year 2019. The main topographical features (ridges and slopes) are also represented by the model. Activated wind redistribution leads to a longer snow cover duration on north east-facing slopes, which are also partially present in the observation products. In contrast, WaSiM underestimates the snow cover duration in west-facing areas due to wind-driven snow erosion, as is the case for the Madritsch snow station and the surrounding area. Canopy snow interception was not considered in the model and is therefore not discernible in the valley bottom.

The fractional snow-covered area (fSCA) of the two observation products and the simulation results were evaluated for a more quantitative comparison (Fig. 9). The fSCA shows strong seasonality in all products. Although the simulated snow coverage given by WaSiM is in line with that observed by Eurac, there are larger fSCA in the ablation period (June and July) of the years 2016 (+14.2%), 2017 (+3.2%) and 2020 (+8.6%) but also for new snow events in August and September of the years 2017 (+8.5%) and to some extent 2018 (+5.5%). This pattern is mainly present when comparing WaSiM with TUM for some dates of the years 2016 and 2017. However, these differences are not as pronounced as for the Eurac product. As an overall mean, WaSiM overestimates the Eurac fSCA by 5.1% and slightly underestimates the TUM fSCA by 0.6% during the ablation period (June and July). The overall Pearson correlation of fractional snow-covered areas is 0.98 for Eurac and WaSiM and 0.98 for TUM and WaSiM.

4. Discussion

4.1. Cloud detection by TUM and Eurac

In this work, we tested the extent to which a simple snow and cloud detection approach (TUM) performs in a complex terrain and determined what limitations might occur. Previous evaluations have found that Idepix overclassifies possible cloud shadow pixels (Alvera-Azcárate et al., 2021). Although Idepix has been tested for a wide selection of regions, it still has limitations and weaknesses in cloud detection that have not been completely resolved. The main difficulties are in its ability

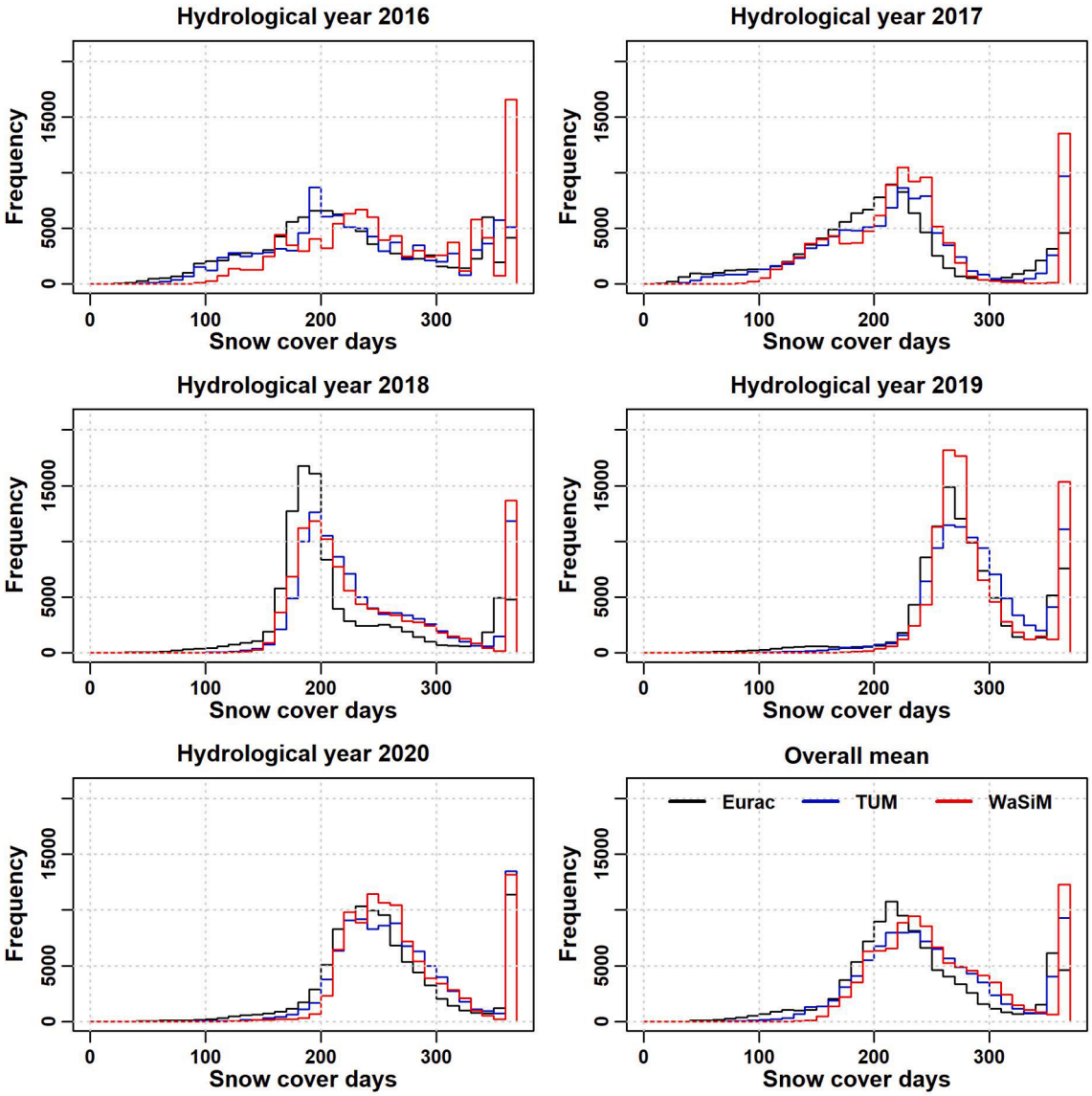


Fig. 7. Frequencies of observed and simulated snow cover days (SCDs) in bins of 10 SCDs for each hydrological year and as average means.

Table 5
List of means and standard deviations (SD) of snow cover days (SCDs) for the different snow products and for each hydrological year.

Hydrological year	Mean SCDs Eurac	Mean SCDs TUM	Mean SCDs WaSiM	SD SCDs Eurac	SD SCDs TUM	SD SCDs WaSiM
2016	221.16	233.20	261.94	77.04	75.99	72.20
2017	206.89	222.55	230.27	74.10	74.33	66.39
2018	218.20	241.58	240.49	63.57	61.59	63.73
2019	275.96	288.79	289.06	50.57	43.62	40.17
2020	256.23	269.78	270.42	55.67	51.44	47.83
Overall mean	235.69	251.18	258.44	64.19	61.40	58.07

to distinguish between cloud and snow/ice and to detect optically very thin clouds (Brockmann Consult GmbH, 2017). In a visual inspection for the years 2015 and 2016 based on Idepix cloud products, we observed that snow was systematically falsely classified as cloud in clear sky situations during the winter due to the presence of topographical shadows (see supplementary Figure S.9 in Hofmeister et al. (2022)), which would lead to an underestimation of snow cover if the erroneously detected clouds are set to no data. Snow detection by NDSI was therefore given a higher confidence rating than cloud detection by Idepix and clouded areas were only set to no data when the binary snow detection value was zero. This resulted in false snow detection for a few pixels on steep shaded slopes with low illumination. This effect is also shown by TUM on the slope feature class (Fig. 5), where the variance in the mean snow cover duration increased for very steep slopes (>60°) when compared to Eurac. The complementation by the Idepix Mountain_Shadow layer might reduce false snow and cloud detection on steep and shaded slopes.

The Eurac algorithm is based on an ML approach that uses multi-temporal training samples. Even if the samples are collected over a large variety of scenes, including different illumination conditions, a single model is used and the quality of the classification still depends on the conditions of the scene. This is mainly due to the limited spectral information provided by Sentinel-2, which is insufficient for solving all ambiguities. Unlike the single-scene cloud detection of Idepix, the Eurac product is able to classify mountain shadows on north-facing slopes, as can be seen in the case study on cloud detection in Fig. 3d. Additionally, it has the highest snow detection confidence, even in steep terrain, as can be concluded from the narrow uncertainty band in Fig. 5. Since it uses an active learning approach, the cloud and snow detection quality relies on the user selecting the appropriate training pixels. It is

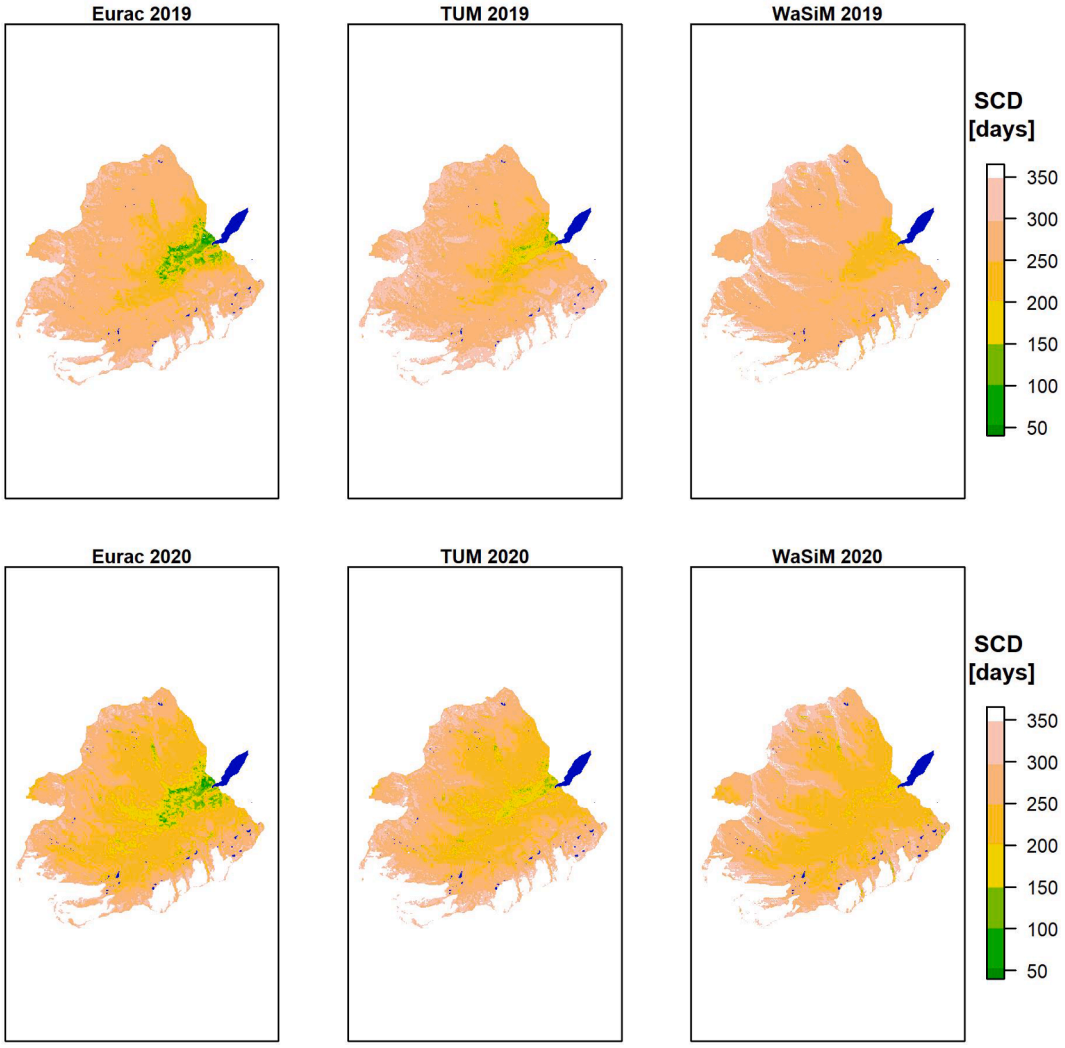


Fig. 8. Observed and simulated snow cover maps for 2019 (above) and 2020 (below).

sometimes difficult to visually assign the correct class to pixels showing mixed characteristics, such as shadowed or forested areas. In these pixels, the classifier usually returns an uncertain probability (i.e. around 50%) of the pixel belonging to the snow class. In general, cloud detection by both Idepix and ML requires deeper investigation over forested areas in Alpine catchments. Moreover, a comparison with other available remote sensing snow cover products (e.g. Gascoin et al., 2019; Di Marco et al., 2020) may be beneficial for the future.

4.2. *Validation of snow detection quality through in-situ snow observation*

The individual selection of a snow depth threshold is recommended when comparing remote sensing and ground snow cover information. In fact, each snow observation site has its own characteristics, such as elevation, aspect, exposure to wind, and land cover, that influence this comparison. Due to the additional interference between vegetation and snow cover detection, validation by in-situ snow observation is more accurate for sites located above the forest line. Moreover, the mixed pixel issue of different spectral signatures can impact the binary snow and no snow information. For small snow depth thresholds (<4 cm), a high snow detection sensitivity can be observed at the Rossbaenke site due to interference of grass cover with the snow (Fig. 4). One way of reducing this error would be to filter the snow depth observations to avoid noise caused by the ground (i.e. grass). Nevertheless, a good level of accuracy can be achieved for lower lying sites that are surrounded by coniferous forest and grassland, as is the case for the TUM snow product. Barrou Dumont et al. (2021) found an optimal threshold value even at 1 cm snow depth by analysing a very large data set of 1764 in-situ snow observation stations. Differences in the snow cover detection accuracy by Sentinel-2 was also noted to be dependent on the land cover type. It was also found that the differences in the accuracy of snow cover detection by Sentinel-2 depend on the type of land cover. In particular, lower accuracies were obtained in closed forests and near water bodies (Barrou Dumont et al., 2021). Nevertheless, systematic false detection, such as in topographic shadows, can only be detected by comparing with another snow and cloud detection method, since most snow observation stations are located in flat terrain and are not shaded.

Comparison between simulated SWE and in-situ snow observation at Madritsch undermined the use of the commonly used SWE threshold of 5 mm to distinguish between snow and no snow. However, this only represents a point comparison at a single station. It might well be that the threshold sensitivity varies from site to site.

In a small case study (September 2018 - October 2019), we tested the utility and effect of atmospheric correction on snow detection with Sentinel-2. From this evaluation, atmospherically (AC) and non-atmospherically (nAC) corrected results mostly differ on lower elevated areas. Using the AC dataset, snow detection accuracy increased for the Madritsch observation station (from 0.85 to 0.94 at a 4 cm threshold) but decreased for the lower lying Zufritt station (from 0.78 to 0.74 at a 4 cm threshold). The full Figure is available in the

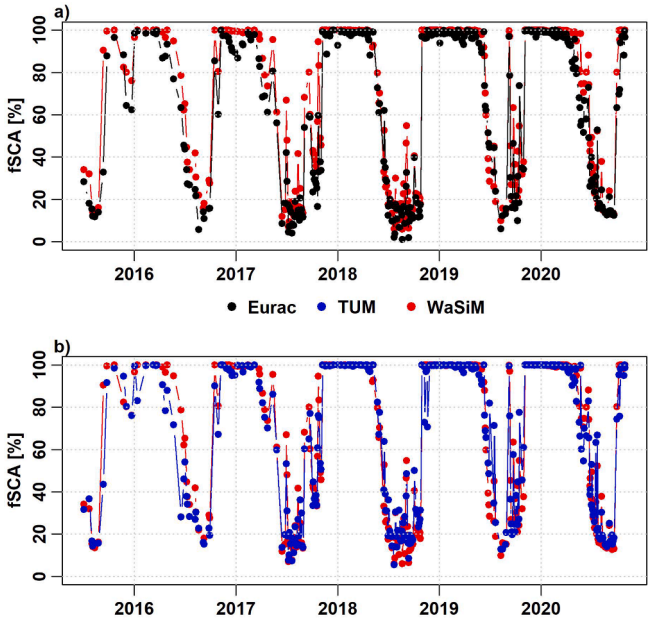


Fig. 9. Fractional snow-covered area (fSCA) for each image and snow product, comparing Eurac and WaSiM (a) and TUM and WaSiM (b).

supplementary material (S.10 in Hofmeister et al. (2022)). Major differences can be seen in the topographical analysis, where the use of the AC dataset led to underestimation of snow at lower valley locations. In this elevation range, AC shows a stronger similarity to the Eurac product (supplementary Figure S.11 in Hofmeister et al. (2022)). In contrast, AC overestimates the snow cover duration in the 2500 to 3500 m elevation range. Moreover, the performance in southern to north western areas no longer agrees with Eurac.

For this analysis, AC is useful for improving snow detection on the point scale, as is the case for the Madritsch station, but leads to underestimation of the snow cover duration at lower elevations (<2500 m) with weaker illumination and in areas with a southern to western aspect, such as the Zufritt station. Comparing the gains and losses in accuracy for both stations, we conclude that it is important to critically evaluate the area studied in order to decide whether to use an AC. The application of these and other AC algorithms may significantly improve accuracy, although, in some instances, its use does not appear to offer sufficiently significant improvements to justify the larger time requirements and higher computational demand for detecting snow cover.

4.3. Topographical analysis and optimized WaSiM setup

The topographical analysis revealed differences between the snow cover observations from Sentinel-2 and simulation with the standard WaSiM setup for snow redistribution (Fig. 5). Both snow observation products and the snow model are able to reproduce the increase in mean snow cover duration with increasing altitude. Larger differences can be observed in particular for lower lying areas (<2200 m a.s.l.). The significant underestimation of snow cover duration by Eurac is due to its difficulties in detecting snow in forested areas compared to TUM and WaSiM. The large standard deviation of the observation products undermine the large variance in snow detection for the corresponding elevation zone, as one elevation zone can include pixels with different slopes, aspects and land cover.

The overestimation of gravitational redistribution can be reduced by optimizing the erosion factor (*ferosion*) (Fig. 5a and b) for the corresponding elevation zone. The simulation of gravitational redistribution is particularly important for avoiding so called “snow towers” at the peaks and ridges. Since the *ferosion* parameter depends on the model time step, it must be recalibrated if, for instance, the model is run with a daily time step. In general, changes in the erosion factor within the range 0.001 to 0.007 with an hourly model time step significantly affected the model results on slope gradients steeper than 30°. Besides the erosion factor, the *ierosion* parameter, which determines the threshold for commencing gravitational redistribution was adjusted from 45° to 50° (Fig. 5a and b). There is no change in the slight overestimation of the simulated mean snow cover duration on slopes from 40° to 50°, even after parameter optimization. It is therefore assumed that with larger snow accumulations, the snowmelt is delayed by the surface energy balance model. The activation of wind-driven snow redistribution impacts mainly the mean snow cover duration on steep slopes with inclines of over 60° due to the intense snow accumulation in this class (Fig. 5d). This shows the limitation of a global erosion factor that is constant in space and time.

Moving to a larger scale, the deficits on the slope feature class are levelled out when comparing the mean snow cover duration against aspect. Several previous studies have determined the effect of aspect on snow cover duration. South-facing slopes receive more solar radiation, which leads to more rapid snow melt and a shorter snow cover duration than on north-facing slopes in the northern hemisphere (López-Moreno and Stähli, 2008; Bouamri et al., 2021). The two observation products mainly differ for north and north west exposed areas, where TUM tends to detect more snow. This is partly due to false positives in shaded areas and lower cloud detection in north-facing areas (supplementary Figure S.6 in Hofmeister et al. (2022)). Although WaSiM was able to adjust the radiation and temperature with respect to the topography, it was determined that the mean snow cover duration was overestimated for south-facing areas (Fig. 5a and b). This overestimation can be reduced by taking into consideration the wind-driven redistribution with a constant south-westerly wind direction (Fig. 5d). Generally, north-west to north-east located areas indicate the largest spatial variability of mean SCD. Though the simulated mean snow cover duration improved on the catchment scale, this simple approach is unable to consider minor or major differences in wind fields and speeds. Comparing the dispersion of standard deviation of all topographical features, it can be seen that the features with the largest spatial variability are slope and aspect.

4.4. Accuracy and frequencies of observed and simulated snow cover and fSCA

Slight differences in the cumulative frequencies of no snow detection occurred mainly for the lower elevation zones, where TUM detects more snow areas than Eurac. It may be that Eurac underestimates snow cover in forest areas (Fig. 4b) and TUM falsely detects snow in areas with low illumination. Since the canopy snow interception model of WaSiM was not activated, the model tends to overestimate mean snow cover durations when compared to observation products at the same elevations (Fig. 5). Besides the systematic offset in cumulative snow detection frequencies between TUM and Eurac from 10% to 90% areal coverage, both approaches agree very well on the presence or absence of full snow coverage (Fig. 6a). The simulated snow covered area approaches the Eurac product for low snow coverage since the glacier extents originate from this product. WaSiM tends to overestimate snow coverages of between 50% and 95% when compared to the observation products. It follows from the distinct decrease in all three accuracy scores in Fig. 6b that WaSiM has difficulties in simulating the snow cover recession in the snow melt period (from May to July). The delayed snow cover recession results in an overestimation of the snow cover for these months, which is also evident from an analysis of the fractional snow-covered area (Fig. 9). This delay can have different reasons. For instance, the longwave incoming and outgoing parameters (*LWINcorr* and *LWOUTcorr*) need a fine tuning based on the snow cover recession during the ablation period (May to July). Consideration of the snow-covered area in a multi-objective calibration approach would improve the model performance

on the catchment scale. In addition, the use of a multi-layer snow model could enhance the simulation of snowmelt by additionally considering heat transfer in the snow pack, especially for snow rich winters (e.g. 2018/2019). However, the multi-layer snow model requires additional calibration parameters to characterize generally unknown soil properties, as well as a new calibration, as a compact snow pack without a layered snow model reacts quite differently to warming or cooling than a layered snow pack (Schulla, 2021).

5. Conclusion

This study presented the potential and limitations using Sentinel-2 images for observing snow cover with a high temporal and spatial resolution in a complex mountainous terrain. We also highlighted the additional benefit of using observation data in the process-oriented calibration of a physically based snow model. To increase the available data, Sentinel-2 images with up to 80% cloud cover were also considered, as they can contain valuable spatial information regarding snow distribution. This requires comprehensive cloud detection to avoid false detection of either snow or no snow. Cloud detection was performed using the Idepix pixel classification algorithm in SNAP for the TUM product. We tested the snow and cloud detection quality against in-situ snow observation on the point scale and also against Eurac, a second snow and cloud product that is based on a supervised algorithm. Both observation products generally attained a very high overall F1 accuracy score (>0.9) with respect to the in-situ station data. The advantages of using two different approaches are that it makes it possible both to perform an intercomparison and validation of the respective results and to identify the limitations of the two approaches. Moreover, the use of two products enables us to estimate the range of uncertainty in snow mapping with high-resolution optical remote sensing data. Although both observation products provide consistent estimates of the mean snow cover above the tree line (>2300 m a.s.l.), distinct differences were observed for the slope feature class. The Idepix tool in particular tends to make false detections of snow and cloud on steep north-facing slopes (>60°) and to falsely classify snow as cloud in some clear sky situations in winter. In contrast, the active learning approach of Eurac underestimates snow cover in evergreen forest areas unless explicit training pixels are selected. The limitation of optical remote sensing products for snow detection under the canopy is a generally well known problem, which is why only the snow cover of high-altitude regions is usually analysed (Gascoin et al., 2019). Future research will be on assessing fractional snow cover in forested areas using optical remote sensing products (Gascoin et al., 2020).

Unlike the coarse MODIS data (Bouamri et al., 2021), Sentinel-2 enables the investigation of small-scale differences in snow cover duration in complex terrains due to gravitational redistribution (slope), energy balance and wind-driven redistribution (aspect). This makes it possible to calibrate and validate the physically based snow model of WaSiM in a process-oriented manner on different scales (point, slope and catchment scale) with high spatial resolution (25 m). Depending on the scale and elevation, each snow detection approach has its advantages; for instance, TUM performs better at the forested valley bottom, while Eurac is more reliable at detecting snow on very steep slopes (>60°). Nevertheless, the two snow detection approaches should be validated with other observational data (e.g. LiDAR) and compared to other algorithms to further investigate the robustness of these approaches.

The spatial comparison of observed and simulated snow cover durations revealed limited WaSiM capabilities on steep slopes (>25°) with the default parametrization, which cloud be partially overcome by calibrating the gravitational redistribution. Moreover, it was shown that the mean snow cover duration computed with WaSiM at different exposures is not only dependent on the incoming solar radiation but also on the wind-driven snow redistribution. The optimized WaSiM model was able to simulate both the mean snow cover duration with a high F1 accuracy score of > 0.9 and the fractional snow-covered area with a very high correlation coefficient of 0.98. Although hydrological analysis would greatly profit from spatially distributed and highly resolved in time information about SWE, the approach that we propose allows us to constrain two important model parameters for the WaSiM model by considering only snow cover information. Our work therefore provides a robust methodology with which it is possible to collect spatio-temporal snow cover information to calibrate empirical gravitational snow redistribution models, and hence enables multi-objective calibration and validation for hydrological model applications in high-elevation Alpine catchments for further studies. We can hypothesis that calibrated WaSiM model results will also provide more accurate SWE estimation and consequently discharge predictions. Such hypothesis should be tested in a multi-objective optimization framework and under the consideration of further observational data, such as SWE and discharge time series. Moreover, snow cover model results can find application beyond the field of hydrology and are of interest for ecological studies (Qi et al., 2021; Wan et al., 2014) as well as for the evaluation of the sustainability of winter tourism (Ebner et al., 2021). Although we did not include glaciers in this study, it is possible to derive further cryospheric products from Sentinel-2 (e.g. glacier extents, snow line on glacier and glacier albedo) to calibrate and validate also glacier modules in hydrological models.

CRediT authorship contribution statement

Florentin Hofmeister: Conceptualization, Methodology, Software, Formal analysis, Data curation, Visualization, Writing – original draft. **Leonardo F. Arias-Rodriguez:** Methodology, Software, Data curation, Formal analysis. **Valentina Premier:** . **Carlo Marin:** Software, Data curation, Formal analysis. **Claudia Notarnicola:** Conceptualization, Writing – review & editing. **Markus Disse:** Conceptualization, Writing – review & editing, Funding acquisition. **Gabriele Chiogna:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Thulasi Vishwanath for processing the snow cover maps, Marco Borga for his advice and critical discussion, and the Autonomous Province of Bozen/Bolzano - South Tyrol and the Institute of Atmospheric and Cryospheric Sciences - University of Innsbruck for providing the meteorological station data. The authors GC, MD and FH acknowledge funding from the DFG (Deutsche Forschungsgemeinschaft) Research Group (FOR2793/1) “Sensitivity of High Alpine Geosystems to Climate Change since 1850” (SEHAG) (Grant DI639/5-1 and CH981/3-1).

We want to thank three anonymous reviewers for their helpful comments and suggestions on the manuscript.

References

Alvera-Azcárate, A., Van der Zande, D., Barth, A., Cardoso dos Santos, J.F., Troupin, C., Beckers, J.-M., 2021. Detection of shadows in high spatial resolution ocean satellite data using DINEOF. Remote Sens. Environ. 253, 112229. https://doi.org/10.1016/j.rse.2020.112229.

Aronica, G., Bates, P.D., Horritt, M.S., 2002. Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. Hydrol. Process. 16 (10), 2001–2016. https://doi.org/10.1002/hyp.398.

Barrou Dumont, Z., Gascoin, S., Hagolle, O., Ablain, M., Jugier, R., Salgues, G., Marti, F., Dupuis, A., Dumont, M., Morin, S., 2021. Brief communication: evaluation of the snow cover detection in the Copernicus High Resolution Snow & Ice Monitoring Service. Cryosphere 15, 4975–4980. https://doi.org/10.5194/tc-15-4975-2021.

Blöschl, G., 1999. Scaling issues in snow hydrology. Hydrol. Process. 13, 2149–2175. https://doi.org/10.1002/(SICI)1099-1085(199910)13:14/15<2149::AID-HYP847>3.0.CO;2-8.

Bouamri, H., Kinnard, C., Boudhar, A., Gascoin, S., Hanich, L., Chehbouni, A., 2021. MODIS does not capture the spatial heterogeneity of snow cover induced by solar radiation. Front. Earth Sci. 9, 640250 https://doi.org/10.3389/feart.2021.640250.

Brockmann Consult GmbH, 2017. 2. The SNAP Cawa TCWV and CTP Processing System — CAWA Software User Manual [WWW Document]. URL https://snap-cawa.readthedocs.io/en/latest/cawa_processing_system.html#the-idepix-pixel-classification-module (accessed 6.2.21).

Brown, M.E., Racoviteanu, A.E., Tarboton, D.G., Gupta, A.S., Nigro, J., Policelli, F., Habib, S., Tokay, M., Shrestha, M.S., Bajracharya, S., Hummel, P., Gray, M., Duda, P., Zaitchik, B., Mahat, V., Artan, G., Tokar, S., 2014. An integrated modeling system for estimating glacier and snow melt driven streamflow from remote sensing and earth system data products in the Himalayas. J. Hydrol. 519, 1859–1869. https://doi.org/10.1016/j.jhydrol.2014.09.050.

Cimpianu, C., 2018. Delineate snow using Normalized Difference Snow Index, Sentinel 2 and QGIS. GISCourse.com. URL https://www.giscourse.com/delineate-snow-using-normalized-difference-snow-index-sentinel-2-and-qgis/.

Di Marco, N., Righetti, M., Avesani, D., Zaramella, M., Notarnicola, C., Borga, M., 2020. Comparison of MODIS and model-derived snow-covered areas: impact of land use and solar illumination conditions. Geosciences 10, 134. https://doi.org/10.3390/geosciences10040134.

Dietz, A.J., Wohner, C., Kuenzer, C., 2012. European snow cover characteristics between 2000 and 2011 derived from improved MODIS daily snow cover products. Remote Sens. 4, 2432–2454. https://doi.org/10.3390/rs4082432.

Dozier, J., 1989. Spectral signature of alpine snow cover from the landsat thematic mapper. Remote Sens. Environ. 28, 9–22. https://doi.org/10.1016/0034-4257(89)90101-6.

Dozier, J., Painter, T.H., 2004. Multispectral and hyperspectral remote sensing of alpine snow properties. Annu. Rev. Earth Planet. Sci. 32 (1), 465–494. https://doi.org/10.1146/annurev.earth.32.101802.120404.

Duethmann, D., Peters, J., Blume, T., Vorogushyn, S., Güntner, A., 2014. The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia. Water Resour. Res. 50 (3), 2002–2021. https://doi.org/10.1002/2013WR014382.

Durand, Y., Guyomarc'h, G., Mérindol, L., Corripio, J.G., 2005. Improvement of a numerical snow drift model and field validation. Cold Reg. Sci. Technol. 43 (1-2), 93–103. https://doi.org/10.1016/j.coldregions.2005.05.008.

Ebner, P.P., Koch, F., Premier, V., Marin, C., Hanzer, F., Carmagnola, C.M., François, H., Günther, D., Monti, F., Hargoaa, O., Strasser, U., Morin, S., Lehning, M., 2021. Evaluating a prediction system for snow management. Cryosphere 15, 3949–3973. https://doi.org/10.5194/tc-15-3949-2021.

Essery, R., Li, L., Pomeroy, J., 1999. A distributed model of blowing snow over complex terrain. Hydrol. Process. 13 (14-15), 2423–2438. https://doi.org/10.1002/(SICI)1099-1085(199910)13:14/15<2423::AID-HYP853>3.0.CO;2-U.

European Space Agency, 2021. Sentinel-2 - Missions - Sentinel Online - Sentinel Online [WWW Document]. URL https://sentinel.esa.int/web/sentinel/missions/sentinel-2 (accessed 7.20.21). European Space Agency, 2019. Sentinel Application Platform (SNAP).

Fassnacht, S.R., López-Moreno, J.I., Ma, C., Weber, A.N., Pfohl, A.K.D., Kampf, S.K., Kappas, M., 2017. Spatio-temporal snowmelt variability across the headwaters of the Southern Rocky Mountains. Front. Earth Sci. 11 (3), 505–514. https://doi.org/10.1007/s11707-017-0641-4.

Foppa, N., Stoffel, A., Meister, R., 2005. Snow depth mapping in the Alps: merging of in situ and remotely-sensed data. EARSeL eProceedings 4.

Förster, K., Garvelmann, J., Meißl, G., Strasser, U., 2018. Modelling forest snow processes with a new version of WaSiM. Hydrol. Sci. J. 63 (10), 1540–1557. https://doi.org/10.1080/02626667.2018.1518626.

Frei, A., Tedesco, M., Lee, S., Foster, J., Hall, D.K., Kelly, R., Robinson, D.A., 2012. A review of global satellite-derived snow products. Adv. Space Res. 50 (8), 1007–1029. https://doi.org/10.1016/j.asr.2011.12.021.

Freudiger, D., Kohn, I., Seibert, J., Stahl, K., Weiler, M., 2017. Snow redistribution for the hydrological modeling of alpine catchments: snow redistribution for hydrological modeling. Wiley Interdiscip. Rev. Water 4 (5), e1232. https://doi.org/10.1002/wat2.1232.

Frey, S., Holzmann, H., 2015. A conceptual, distributed snow redistribution model. Hydrol. Earth Syst. Sci. 19 (11), 4517–4530. https://doi.org/10.5194/hess-19-4517-201510.5194/hess-19-4517-2015-supplement.

Galos, S.P., Klug, R., 2015. Recent glacier changes and related contribution potential to river discharge in the Vinschgau / Val Venosta, Italian Alps. Geogr. Fis. E Din. Quat. 143–154 https://doi.org/10.4461/GFDQ.2015.38.13.

Galos, S.P., Klug, C., Maussion, F., Covi, F., Nicholson, L., Rieg, L., Gurgiser, W., Mölg, T., Kaser, G., 2017. Reanalysis of a 10-year record (2004–2013) of seasonal mass balances at Langenferner/Vedretta Lunga, Ortler Alps, Italy. Cryosphere 11 (3), 1417–1439. https://doi.org/10.5194/tc-11-1417-2017.

Gascoin, S., Grizonnet, M., Bouchet, M., Salgues, G., Hagolle, O., 2019. Theia Snow collection: high-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data. Earth Syst. Sci. Data 11, 493–514. https://doi.org/10.5194/essd-11-493-2019.

Gascoin, S., Barrou Dumont, Z., Deschamps-Berger, C., Marti, F., Salgues, G., López-Moreno, J.I., Revuelto, J., Michon, T., Schattan, P., Hagolle, O., 2020. Estimating fractional snow cover in open terrain from Sentinel-2 using the normalized difference snow index. Remote Sens. 12, 2904. https://doi.org/10.3390/rs12182904.

Gauer, P., 1998. Blowing and drifting snow in Alpine terrain: numerical simulation and related field measurements. Ann. Glaciol. 26, 174–178. https://doi.org/10.3189/1998AoG26-1-174-178.

Gruber, S., 2007. A mass-conserving fast algorithm to parameterize gravitational transport and deposition using digital elevation models: mass-conserving transport and deposition on dems. Water Resour. Res. 43 (6) https://doi.org/10.1029/2006WR004868.

Grünewald, T., Bühler, Y., Lehning, M., 2014. Elevation dependency of mountain snow depth. Cryosphere 8, 2381–2394. https://doi.org/10.5194/tc-8-2381-2014.

Grünewald, T., Lehning, M., 2011. Altitudinal dependency of snow amounts in two small alpine catchments: can catchment-wide snow amounts be estimated via single snow or precipitation stations? Ann. Glaciol. 52 (58), 153–158. https://doi.org/10.3189/172756411797252248.

Gurung, D.R., Maharjan, S.B., Shrestha, A.B., Shrestha, M.S., Bajracharya, S.R., Murthy, M.S.R., 2017. Climate and topographic controls on snow cover dynamics in the Hindu Kush Himalaya. Int. J. Climatol. 37 (10), 3873–3882. https://doi.org/10.1002/joc.4961.

Hall, D.K., Foster, J.L., DiGirolamo, N.E., Riggs, G.A., 2012. Snow cover, snowmelt timing and stream power in the Wind River Range, Wyoming. Geomorphology 137 (1), 87–93. https://doi.org/10.1016/j.geomorph.2010.11.011.

Kääb, A., Winsvold, S., Altena, B., Nuth, C., Nagler, T., Wuite, J., 2016. Glacier remote sensing using Sentinel-2. part I: radiometric and geometric performance, and application to ice velocity. Remote Sens. 8, 598. https://doi.org/10.3390/rs8070598.

Kochendorfer, J., Rasmussen, R., Wolff, M., Baker, B., Hall, M.E., Meyers, T., Landolt, S., Jachcik, A., Isaksen, K., Brækkan, R., Leeper, R., 2016. The Quantification and Correction of Wind-Induced Precipitation Measurement Errors (preprint). Hydrometeorology/Instruments and observation techniques. doi:10.5194/hess-2016-415.

Kraller, G., Warscher, M., Kunstmann, H., Vogl, S., Marke, T., Strasser, U., 2012. Water balance estimation in high Alpine terrain by combining distributed modeling and a neural network approach (Berchtesgaden Alps, Germany). Hydrol. Earth Syst. Sci. 16, 1969–1990. https://doi.org/10.5194/hess-16-1969-2012.

Lehning, M., Löwe, H., Ryser, M., Raderschall, N., 2008. Inhomogeneous precipitation distribution and snow transport in steep terrain: snow drift and inhomogeneous precipitation. Water Resour. Res. 44 https://doi.org/10.1029/2007WR006545.

Lin, H.-T., Lin, C.-J., Weng, R.C., 2007. A note on Platt's probabilistic outputs for support vector machines. Mach. Learn. 68, 267–276. https://doi.org/10.1007/s10994-007-5018-6.

Liston, G.E., Haehnel, R.B., Sturm, M., Hiemstra, C.A., Berezovskaya, S., Tabler, R.D., 2007. Simulating complex snow distributions in windy environments using SnowTran-3D. J. Glaciol. 53, 241–256. https://doi.org/10.3189/172756507782202865.

López-Moreno, J.I., Revuelto, J., Gilaberte, M., Morán-Tejeda, E., Pons, M., Jover, E., Esteban, P., García, C., Pomeroy, J.W., 2014. The effect of slope aspect on the response of snowpack to climate warming in the Pyrenees. Theor. Appl. Climatol. 117, 207–219. https://doi.org/10.1007/s00704-013-0991-0.

López-Moreno, J.I., Stähli, M., 2008. Statistical analysis of the snow cover variability in a subalpine watershed: assessing the role of topography and forest interactions. J. Hydrol. 348, 379–394. https://doi.org/10.1016/j.jhydrol.2007.10.018.

Marcolini, G., Koch, R., Chimani, B., Schöner, W., Bellin, A., Disse, M., Chiogna, G., 2019. Evaluation of homogenization methods for seasonal snow depth data in the Austrian Alps, 1930–2010. Int. J. Climatol. 39, 4514–4530. https://doi.org/10.1002/joc.6095.

Matiu, M., Crespi, A., Bertoldi, G., Carmagnola, C.M., Marty, C., Morin, S., Schöner, W., Cat Berro, D., Chiogna, G., De Gregorio, L., Kotlarski, S., Majone, B., Resch, G., Terzago, S., Valt, M., Beozzo, W., Cianfarra, P., Gouttevin, I., Marcolini, G., Notarnicola, C., Petitta, M., Scherrer, S.C., Strasser, U., Winkler, M., Zebisch, M., Cicogna, A., Cremonini, R., Debernardi, A., Faletto, M., Gaddo, M., Giovannini, L., Mercalli, L., Soubeyroux, J.-M., Sušnik, A., Trenti, A., Urbani, S., Weilguni, V., 2021. Observed snow depth trends in the European Alps: 1971 to 2019. The Cryosphere 15, 1343–1382. https://doi.org/10.5194/tc-15-1343-2021.

Matiu, M., Jacob, A., Notarnicola, C., 2019. Daily MODIS snow cover maps for the European Alps from 2002 onwards at 250m horizontal resolution along with a nearly cloud-free version. doi:10.5281/ZENODO.3566703.

Menekay, D., 2019. Accuracy Assessment Experiments of Cloud Masking Programs (Sen2Cor, Fmask, Idepix) | by Deniz Menekay | Medium [WWW Document]. URL https://medium.com/@denizmenekay/accuracy-assessment-experiments-of-cloud-masking-programs-sen2cor-fmask-idepix-8c1b44fae739 (accessed 6.2.21).

Mott, R., Schirmer, M., Bavay, M., Grünewald, T., Lehning, M., 2010. Understanding snow-transport processes shaping the mountain snow-cover. Cryosphere 4, 545–559. https://doi.org/10.5194/tc-4-545-2010.

Mott, R., Vionnet, V., Grünewald, T., 2018. The seasonal snow cover dynamics: review on wind-driven coupling processes. Front. Earth Sci. 6, 197. https://doi.org/10.3389/feart.2018.00197.

Murphy, S., 2018. Atmospheric Correction of Sentinel 2 Imagery in Google Earth Engine Using Py6S.

Oke, T.R., 2002. Boundary Layer Climates. Routledge.

Papathoma-Köhle, M., Zischg, A., Fuchs, S., Glade, T., Keiler, M., 2015. Loss estimation for landslides in mountain areas – an integrated toolbox for vulnerability assessment and damage documentation. Environ. Model. Softw. 63, 156–169. https://doi.org/10.1016/j.envsoft.2014.10.003.

Pedersen, S.H., Tamstorf, M.P., Abermann, J., Westergaard-Nielsen, A., Lund, M., Skov, K., Sigsgaard, C., Mylius, M.R., Hansen, B.U., Liston, G.E., Schmidt, N.M., 2016. Spatiotemporal characteristics of seasonal snow cover in Northeast Greenland

from in situ observations. Arct. Antarct. Alp. Res. 48, 653–671. https://doi.org/10.1657/AAAR0016-028.
Pomeroy, J.W., Gray, D.M., Brown, T., Hedstrom, N.R., Quinton, W.L., Granger, R.J., Carey, S.K., 2007. The cold regions hydrological model: a platform for basing process representation and model structure on physical evidence. Hydrol. Process. 21, 2650–2667. https://doi.org/10.1002/hyp.6787.
Puspitarini, H.D., François, B., Zaramella, M., Brown, C., Borga, M., 2020. The impact of glacier shrinkage on energy production from hydropower-solar complementarity in alpine river basins. Sci. Total Environ. 719, 137488 https://doi.org/10.1016/j.scitotenv.2020.137488.
Qi, Y., Wang, H., Ma, X., Zhang, J., Yang, R., 2021. Relationship between vegetation phenology and snow cover changes during 2001–2018 in the Qilian Mountains. Ecol. Indic. 133, 108351 https://doi.org/10.1016/j.ecolind.2021.108351.
Ranghetti, L., Boschetti, M., Nutini, F., Busetto, L., 2020. "sen2r": An R toolbox for automatically downloading and preprocessing Sentinel-2 satellite data. Comput. Geosci. 139, 104473 https://doi.org/10.1016/j.cageo.2020.104473.
Rango, A., Wergin, W.P., Erbe, E.F., 1996. Snow crystal imaging using scanning electron microscopy: II. Metamorphosed snow. Hydrol. Sci. J. 41, 235–250. https://doi.org/10.1080/02626669609491495.
Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A.P., Black, J., Thériault, J.M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K., Gutmann, E., 2012. How well are we measuring snow: the NOAA/FAA/NCAR winter precipitation test bed. Bull. Am. Meteorol. Soc. 93, 811–829. https://doi.org/10.1175/BAMS-D-11-00052.1.
Riano, D., Chuvieco, E., Salas, J., Aguado, I., 2003. Assessment of different topographic corrections in landsat-TM data for mapping vegetation types (2003). IEEE Trans. Geosci. Remote Sens. 41, 1056–1061. https://doi.org/10.1109/TGRS.2003.811693.
Riggs, G.A., Hall, D.K., 2015. MODIS Snow Products Collection 6 User Guide.
Romanov, P., Gutman, G., Csiszar, I., 2000. Automated monitoring of snow cover over north america with multispectral satellite data. J. Appl. Meteorol. 39, 1866–1880. https://doi.org/10.1175/1520-0450(2000)039<1866:AMOSCO>2.0.CO;2.
Sauter, T., Möller, M., Finkelnburg, R., Grabiec, M., Scherer, D., Schneider, C., 2013. Snowdrift modelling for the Vestfonna ice cap, north-eastern Svalbard. Cryosphere 7, 1287–1301. https://doi.org/10.5194/tc-7-1287-2013.
Saydi, M., Ding, J., 2020. Impacts of topographic factors on regional snow cover characteristics. Water Sci. Eng. 13, 171–180. https://doi.org/10.1016/j.wse.2020.09.002.
Schneiderbauer, S., Prokop, A., 2011. The atmospheric snow-transport model: SnowDrift3D. J. Glaciol. 57, 526–542. https://doi.org/10.3189/002214311796905677.
Schöber, J., Schneider, K., Helfricht, K., Schattan, P., Achleitner, S., Schöberl, F., Kirnbauer, R., 2014. Snow cover characteristics in a glacierized catchment in the Tyrolean Alps - improved spatially distributed modelling by usage of Lidar data. J. Hydrol. 519, 3492–3510. https://doi.org/10.1016/j.jhydrol.2013.12.054.
Schulla, J., 2021. Model Description WaSiM (Water balance Simulation Model).
Serco Italia SPA, 2017. Snow Cover Mapping with Sentinel-2 (SNAP).
Sonny, 2017. Digitale LiDAR-Geländemodelle von Italien, San Marino, Vatikanstadt, Malta | Digital LiDAR-Terrain Models of Italy, San Marino, Vatican City, Malta - Open Data Portal Austria [WWW Document]. URL http://data.opendataportal.at/dataset/dtm-italy (accessed 10.12.21).
Tarrio, K., Tang, X., Masek, J.G., Claverie, M., Ju, J., Qiu, S., Zhu, Z., Woodcock, C.E., 2020. Comparison of cloud detection algorithms for Sentinel-2 imagery. Sci. Remote Sens. 2, 100010 https://doi.org/10.1016/j.srs.2020.100010.
Thornton, J.M., Brauchli, T., Mariethoz, G., Brunner, P., 2021. Efficient multi-objective calibration and uncertainty analysis of distributed snow simulations in rugged alpine terrain. J. Hydrol. 598, 126241 https://doi.org/10.1016/j.jhydrol.2021.126241.
Totschnig, R., Fuchs, S., 2013. Mountain torrents: quantifying vulnerability and assessing uncertainties. Eng. Geol. 155, 31–44. https://doi.org/10.1016/j.enggeo.2012.12.019.
Tuo, Y., Marcolini, G., Disse, M., Chiogna, G., 2018. A multi-objective approach to improve SWAT model calibration in alpine catchments. J. Hydrol. 559, 347–360. https://doi.org/10.1016/j.jhydrol.2018.02.055.
Vermote, E.F., Tanre, D., Deuze, J.L., Herman, M., Morcette, J.-J., 1997. Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: an overview. IEEE Trans. Geosci. Remote Sens. 35, 675–686. https://doi.org/10.1109/36.581987.
Vionnet, V., Martin, E., Masson, V., Guyomarc'h, G., Naaim-Bouvet, F., Prokop, A., Durand, Y., Lac, C., 2014. Simulation of wind-induced snow transport and sublimation in alpine terrain using a fully coupled snowpack/atmosphere model. The Cryosphere 8, 395–415. doi:10.5194/tc-8-395-2014.
Vionnet, V., Marsh, C.B., Menounos, B., Gascoin, S., Wayand, N.E., Shea, J., Mukherjee, K., Pomeroy, J.W., 2021. Multi-scale snowdrift-permitting modelling of mountain snowpack. Cryosphere 15, 743–769. https://doi.org/10.5194/tc-15-743-2021.
Wan, Y., Gao, Q., Li, Y., Qin, X., Ganjurjav, Zhang, W., Ma, X., Liu, S., 2014. Change of Snow Cover and Its Impact on Alpine Vegetation in the Source Regions of Large Rivers on the Qinghai-Tibetan Plateau, China. Arct. Antarct. Alp. Res. 46, 632–644. doi:10.1657/1938-4246-46.3.632.
Warscher, M., 2014. Performance of Complex Snow Cover Descriptions in a Distributed Hydrological Model System and Simulation of Future Snow Cover and Discharge Characteristics: A Case Study for the High Alpine Terrain of the Berchtesgaden Alps. University Augsburg, Augsburg.
Warscher, M., Strasser, U., Kraller, G., Marke, T., Franz, H., Kunstmann, H., 2013. Performance of complex snow cover descriptions in a distributed hydrological model system: a case study for the high Alpine terrain of the Berchtesgaden Alps. Water Resour. Res. 49, 2619–2637. https://doi.org/10.1002/wrcr.20219.
Wilson, R.T., 2012. Py6S: A Python interface to the 6S radiative transfer model. Comput. Geosci. 51, 166–171. https://doi.org/10.1016/j.cageo.2012.08.002.