

Snow depth estimation at country-scale with high spatial and temporal resolution

Rodrigo Caye Daudt^{a,*}, Hendrik Wulf^b, Elisabeth D. Hafner^{a,c,d}, Yves Bühler^{c,d},
Konrad Schindler^a, Jan Dirk Wegner^{a,e}

^a EcoVision Lab, Photogrammetry and Remote Sensing, ETH Zurich, Switzerland

^b ExoLabs GmbH, Hegibachstrasse 48, Zurich, Switzerland

^c WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland

^d Climate Change, Extremes and Natural Hazards in Alpine Regions Research Center CERC, Davos, Switzerland

^e Institute for Computational Science, University of Zurich, Switzerland

ARTICLE INFO

MSC:

00-01

99-00

Keywords:

Snow depth estimation

Convolutional neural networks

Recurrent neural networks

Data fusion

Probabilistic deep learning

ABSTRACT

Monitoring snow depth is important for applications such as hydrology, energy planning, ecology, and safety evaluation for outdoor winter activities. Most methods able to estimate snow depth for large regions can only do so in a spatial resolution of up to 1 km ground sampling distance (GSD). This limits their usage in high alpine areas, where this resolution fails to capture local snow distribution patterns caused by the pronounced topographical features. In this work we use a recurrent convolutional neural network to estimate snow depth at high spatial resolution (10 m GSD), weekly, and at large scale based on satellite data sources and elevation maps, without the need for measurement stations on the ground. The proposed method achieves unprecedented results for large-scale, high-resolution snow depth mapping. The resulting maps are evaluated over a period of three years against high-fidelity snow depth maps obtained with airborne photogrammetry. Finally, we also produce well-calibrated uncertainty estimates for every individual snow depth estimate via a probabilistic regression framework.

1. Introduction

Water is arguably the most important natural resource for communities and ecosystems worldwide. Monitoring it is essential, especially in the face of challenges due to climate change (Muelchi et al., 2021; Stahl et al., 2022). Earth Observation (EO) holds great potential for large scale, continuous monitoring of such resources and for supporting Sustainable Development Goals (SDGs) (The 17 Goals, 2022), especially when coupled with modern deep machine learning (Zhu et al., 2017; Ma et al., 2019; Persello et al., 2022).

Water stored as snow is crucial for over a billion people worldwide during the warmer months (Lievens et al., 2019). Monitoring snow distribution is essential for planning and managing water resources weeks or months in advance, and becomes more important with increasing climate impacts (Muelchi et al., 2021; Stahl et al., 2022). Research shows that mean snow depth in the European Alps has decreased at a rate of >8% per decade over the past 50 years (Matiu et al., 2021; Rumpf et al., 2022). The melting of snow, glaciers, and permafrost threatens the water supply of many communities and increases the likelihood of landslides and floods (Shukla et al., 2022). Water flows

in high-elevation ecosystems are at high risk of being disrupted due to glacier loss and shrinking snow-cover. With a rapid loss reduction in the world's glaciers (Hugonnet et al., 2021), which is exacerbated by snow albedo feedback (Thackeray et al., 2019), we can expect big changes in the hydrological situations of many communities in the coming decades which will negatively affect water security and livelihoods of millions (Shukla et al., 2022).

Mapping snow distribution and snow depth has therefore been a topic of high scientific and practical interest in the fields of hydrology and remote sensing. Having access to such maps is also important for ecology, especially in alpine ecosystems (Wipf et al., 2009). Snow maps support planning of water resource management, leisure outdoor activities, and promote a better understanding long-term dynamics in the cryosphere (Lievens et al., 2019; Hugonnet et al., 2021). Moreover, they are valuable for studying avalanches (Schweizer et al., 2003; Pérez-Guillén et al., 2021; Bühler et al., 2022), so as to reduce risks and costs. Large-scale snow depth maps, in conjunction with local density estimates, also support the estimation of Snow Water Equivalent (SWE), which is important for many applications in hydrology (Lettenmaier

* Corresponding author.

E-mail address: rcayedaudt@ethz.ch (R.C. Daudt).

<https://doi.org/10.1016/j.isprsjprs.2023.01.017>

Received 23 August 2022; Received in revised form 16 December 2022; Accepted 21 January 2023

0924-2716/© 2023 The Author(s). Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

et al., 2015; Margulis et al., 2019; Jonas et al., 2009; Winkler et al., 2021).

Most previous efforts towards snow depth estimation can be grouped into two categories: large-scale, low spatial resolution estimates (Kelly et al., 2003; Saha et al., 2010; Muñoz-Sabater et al., 2021; Lievens et al., 2019; Information, 2022; Olefs et al., 2020) that cover entire countries or mountain ranges; and small-scale, high spatial resolution estimates (Bühler et al., 2015; Painter et al., 2016; Bühler et al., 2016, 2017; Jacobs et al., 2021; Eberhard et al., 2021) that cover only regions of a few km². The former group varies significantly in resolution, e.g., 1 km (Lievens et al., 2019; Olefs et al., 2020), 9 km (Muñoz-Sabater et al., 2021), 25 km (Kelly et al., 2003), or 47 km (Saha et al., 2010). Such maps are valuable for understanding snow dynamics at planetary scale, as it is currently not possible to map snow depth for the entire cryosphere at high spatial resolution. That being said, they often are of limited use in alpine terrain, because they cannot capture the associated spatial variability. Snow depth in such areas is typically mapped with high-resolution methods that are too costly to scale to large regions. High-resolution, large scale snow depth estimates have been produced in the past with moderate success (Wulf et al., 2020), but such products are rare and rely on extensive ground measurement networks.

Learning-based image analysis has the potential to learn from high-quality snow depth maps and then be used to produce high-resolution, large-scale estimates of snow depth. When applied to globally available, high-resolution open data sources such Sentinel-1 and Sentinel-2 images (Sentinel-1, 2022; Sentinel-2, 2022), the application of such methods is not restricted geographically and can even be done retrospectively.

In this work we present a novel snow depth estimation method that can be used at country-scale to produce spatially dense, weekly snow depth estimates at 10 m GSD with unprecedented accuracy. Such estimates are created using a fully convolutional, recurrent neural network and only data from scalable sources: multispectral optical satellite images, synthetic aperture radar (SAR) images, and a digital elevation map. We use satellite images directly in a deep learning pipeline, while previous work used photogrammetric methods with images of much higher resolution. We train this network with a combination of coarse large-scale snow depth maps and local high-fidelity maps generated by airborne photogrammetry. By employing a probabilistic learning framework we are, moreover, able to complement the snow depth maps with well-calibrated, spatially explicit uncertainty maps. The latter are useful diagnostic information for map users and downstream processes (e.g., estimates of avalanche risk).

The main contributions of this work are:

1. A novel, scalable snow depth estimation algorithm which does not require access to ground station measurements. This independence from in-situ observations and usage of open data make it possible to, in principle, deploy the proposed method in any region of interest and at any scale. In our work paper, we produce weekly maps for the entire country of Switzerland (>40,000 km², respectively 4 · 10⁸ pixels).
2. The ability to additionally produce uncertainty maps for the estimated snow depths. In this way, users can derive confidence margins to inform downstream applications, for instance in hydrology. These uncertainty maps allow one to understand which patterns and locations lead to unreliable snow depth estimates.
3. At a technical level our work is, to the best of our knowledge, the first to combine a convolutional Gated Recurrent Unit (GRU) (Ballas et al., 2016) with the probabilistic regression scheme proposed by Kendall and Gal in Kendall and Gal (2017), thus producing uncertainty estimates that are dense in both space and time.

We compare the proposed method to high-fidelity snow depth surveys acquired with airborne photogrammetry around Davos, Switzerland, and show empirically that it indeed produces snow depth maps with finer details and a higher metric accuracy than existing large-scale methods.

2. Related work

2.1. Snow depth estimation

The first source of snow depth measurements are automatic measurement stations. Manual measurements are also feasible in some situations. Nevertheless, even in areas with a high density of measurement stations interpolation of such data is unable to account for the spatial variability of snow depth (Bühler et al., 2015). Snow maps of this kind nonetheless constitute the state of the art for large-area products, and are made available by snow research institutes (e.g., Information (2022) at 1 km GSD).

Lievens et al. have proposed a physical model to estimate snow depth using Sentinel-1 C-band synthetic aperture radar (SAR) images (Lievens et al., 2019). Focusing on the northern hemisphere, their model uses the backscatter σ^0 to estimate snow depth. Data show that the ratio $\sigma_{\text{VH}}^0/\sigma_{\text{VV}}^0$ strongly correlates with snow depth. This is based on the empirical observations that VV polarized SAR bands are suitable for snow-melt mapping, although they have also observed that C-band SAR has limited sensitivity to changes in dry snow. With that method, snow depth estimates with a resolution of 1 km² GSD can be produced. Olefs et al. have also produced snow depth maps with a GSD of 1 km, using a physically based snow cover model. After producing daily snow depth maps for Austria for the period 1961–2020, they were able to study detailed dynamics of the cryosphere in that region, and its evolution over the last six decades.

High resolution snow depth maps can be created by comparing digital surface maps (DSMs) constructed during summer and winter months. Such DSMs can be generated using a variety of photogrammetric techniques: stereo imaging with high resolution satellite images (Eberhard et al., 2021), structure-from-motion using uncrewed aerial systems (UASs) or airplanes (Bühler et al., 2015, 2016, 2017; Adams et al., 2018; Eker et al., 2019; Eberhard et al., 2021), terrestrial cameras (Eberhard et al., 2021), as well as terrestrial or airborne laser scanning (Eker et al., 2019; Jacobs et al., 2021). Some of the more accurate among these systems achieve root mean squared errors (RMSE) of 0.1–0.3 m, depending on the local topography (Bühler et al., 2016). Experiments by Eberhard et al. (2021) suggest that snow depth maps generated from very high resolution satellite imagery reach accuracies up to 0.5–1.0 m in terms of RMSE. We note that these experiments consider rather small test regions (e.g., 0.10 km² in Jacobs et al. (2021) and 0.12 km² in Adams et al. (2018)), and may not be representative for some of the varied terrain characteristics encountered across a major mountain range.

Wulf et al. (2020) have succeeded in producing large-scale, high resolution snow depth estimates by coupling interpolation of station measurements with features learned from previous high-accuracy snow depth surveys. These features, derived from the DEM, to some degree compensate for the lack of spatial detail in the initial interpolation. That work was the first to successfully produce large-scale (i.e., entire mountain ranges) snow depth maps with GSD of 20 m. That method serves as a strong baseline for current country-wide snow mapping in our study, and is described in more detail in Section 4.2.1. This method will be referred to as TCAM from this point on.

2.2. Satellite image sequence analysis

In recent years, several methods have applied state-of-the-art machine learning approaches for analyzing Earth observation time-series. Previous work has successfully applied random forests (Pelletier et al., 2016, 2019a), temporal convolutional networks (Pelletier et al., 2019b,a), Recurrent Neural Networks (RNNs) (Russwurm and Körner, 2017; Rußwurm and Körner, 2018; Rußwurm et al., 2019; Pelletier et al., 2019a; Garnot et al., 2019; Rußwurm et al., 2020; Turkoglu et al., 2021b,a), neural ODEs (Metzger et al., 2021), and self-attention mechanisms (Rußwurm and Körner, 2020; Garnot et al., 2020; Garnot and Landrieu, 2020, 2021) for a variety of tasks. Such algorithms have achieved excellent results for mapping crops and/or field boundaries, classifying land cover, forecasting of vegetation indices, and more. Among such methods, RNNs are a well-established family of algorithms for various forms of time-series analysis (Russwurm and Körner, 2017; Rußwurm and Körner, 2018; Rußwurm et al., 2019; Pelletier et al., 2019a; Garnot et al., 2019; Rußwurm et al., 2020; Turkoglu et al., 2021b,a). The prevalent architectures are based on Long Short-Term Memory (LSTM) (Russwurm and Körner, 2017; Rußwurm and Körner, 2018; Rußwurm et al., 2019; Rußwurm and Körner, 2020; Rußwurm et al., 2020) or GRU (Rußwurm and Körner, 2018; Pelletier et al., 2019a; Garnot et al., 2019), which follow similar principles to mitigate gradient vanishing and enable learning over long sequences. GRU is more efficient in terms of storage and computational cost, while previous work has shown that both variants tend to achieve similar performance (Rußwurm and Körner, 2018), on par with more recent self-attention approaches (Rußwurm and Körner, 2020). Convolutional variants of LSTM and GRU cells – ConvLSTM (Shi et al., 2015) and ConvGRU (Ballas et al., 2016) – have been developed to more efficiently extract spatio-temporal information when analyzing image time-series. In this work, we employ ConvGRU as a core component of our neural architecture.

2.3. Uncertainty estimation in deep learning

For many applications a desirable capability of a machine learning system is to deliver uncertainty margins for the predicted values (Gustafsson et al., 2020). But modern deep learning systems are known to often produce overconfident estimates (Guo et al., 2017; Lakshminarayanan et al., 2017). For this reason, several extensions have been proposed to enable well-calibrated uncertainty estimation (Niculescu-Mizil and Caruana, 2005; Kendall and Gal, 2017; Kuleshov et al., 2018). Recent work (Kendall and Gal, 2017) discerns between two forms of uncertainty: *aleatoric* uncertainty inherent in the observations, e.g., stemming from sensor noise or ambiguous spectra; and *epistemic* uncertainty, expressing the fact that the model parameters may not apply to unfamiliar input patterns.

Aleatoric uncertainty estimation can, under reasonable assumptions, be learned from the data themselves, by formulating the learning task in a probabilistic manner (Kendall and Gal, 2017; Becker et al., 2021). For regression, this amounts to representing the network output as a probability distribution rather than a point estimate, and training the network to simultaneously output the mean and the (log-)variance of that distribution, individually for each sample. The epistemic uncertainty, on the other hand, reflects a lack of knowledge caused by gaps in the training data. It is typically estimated with some form of model ensemble (Lakshminarayanan et al., 2017), building on the assumption that the stochastic nature of the optimization scheme itself and of its initialization will lead to distinct sets of weights for different ensemble members, whose disagreement will then reflect the knowledge gaps. While it has been shown that as few as five ensemble members already provide good estimates of epistemic uncertainty (Ovadia et al., 2019), recent work has tried to find ways to reduce the computational overhead for training and testing and entire ensemble (Turkoglu et al., 2022; Kushibar et al., 2022). We find that, for snow depth estimation

with the proposed data sources, aleatoric uncertainty is the dominating component. Therefore we refrain from the use of ensemble methods, so as to save computational effort.

The described framework for probabilistic deep learning has already been applied in the context of Earth observation, e.g., for estimating canopy height (Alagialoglou et al., 2022; Lang et al., 2022b,a) and other forest structure variables (Becker et al., 2021), and in the context of extrapolating time series to obtain forecasts (Rußwurm et al., 2020). To the best of our knowledge, they have not yet been explored in the context of snow depth estimation.

3. Methodology

The proposed method combines the spatiotemporal analysis capability of convolutional gated recurrent unit (ConvGRU) networks (Ballas et al., 2016) with a probabilistic formulation which enables the network to model heteroscedastic aleatoric uncertainty (i.e., varying variance for each prediction) (Kendall and Gal, 2017). The presence of aleatoric uncertainty in the considered problem is intuitive: it would be naive to assume that the considered data sources (SAR images, optical images, and elevation maps) contain all the information needed to perfectly estimate snow depth, even using a hypothetical “perfect” algorithm. The ability of the network to predict heteroscedastic uncertainties allows us to more reliably interpret the produced estimates. Such uncertainties can also help in the usage of such maps for downstream analyses or products.

This section describes how we produce snow depth estimates and uncertainty maps using the available data sources, which will be described in more detail in Section 4. The first challenge is to handle the input of differently and irregularly sampled inputs in time to produce regular outputs. We then provide a detailed description of the neural network’s architecture and the probabilistic modeling of uncertainties which is used.

3.1. Temporal stacking

The proposed method aims to produce spatially dense snow depth estimates every 7 days. We do not aim to go beyond the temporal resolution of the available dynamic data sources, and we find that weekly snow depth estimates suffice for most applications. Note that the method can be adjusted to produce snow depth estimates at any frequency if desired, up to producing daily results. Since data from satellite sources are acquired at different dates and at different rates, we propose a simple strategy for temporally stacking the data in order to include the maximum amount of information in the inputs to our RNN. The main idea is the following: for each pixel of each spectral band, the latest available valid data is considered.

More formally, we define an update process for each M by N color channel $\mathbf{x}_c \in \mathbb{R}^{M \times N}$ that is computed every day given the new observation \mathbf{h}_c and the associated binary validity mask \mathbf{m}_c which denotes the region of the image for which new data are available with 1, and contains 0 everywhere else. In the case of optical satellite images, \mathbf{m}_c also takes into account cloud masks. For a given day d , the temporally stacked image can then be obtained by computing the following equation:

$$\mathbf{x}_{c,d} = \mathbf{m}_{c,d} \odot \mathbf{h}_{c,d} + (1 - \mathbf{m}_{c,d}) \odot \mathbf{x}_{c,d-1}, \quad (1)$$

where \odot denotes element-wise product. This way we obtain a daily composite for \mathbf{x}_c , from which samples can be drawn at any rate (weekly in our case) while guaranteeing that the latest information for each location is included. A schematic illustrating this process can be seen in Fig. 1

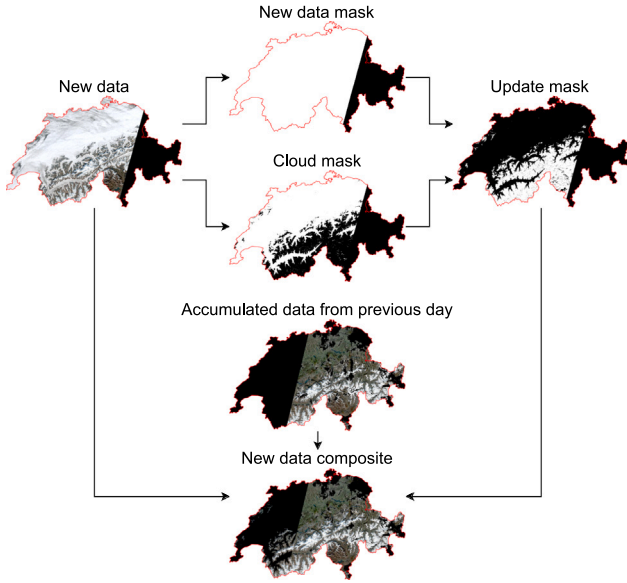


Fig. 1. Update procedure for temporal stacking of optical images. For SAR images the procedure is identical, except there is no need to consider a cloud mask. This procedure ensures that the network always has access to the latest valid information for each location.

3.2. Convolutional and recurrent network architecture

To best estimate snow depth, it is important that the neural network be capable of identifying both spatial and temporal patterns. Recurrent neural networks are neural architectures designed with the explicit purpose of handling ordered sequences of inputs, usually in the time dimension (Goodfellow et al., 2016). The proposed neural network architecture for snow depth estimation is based on ConvGRU cells (Ballas et al., 2016) for the reasons discussed in Section 2.2. Such cells are an extension of the GRU recurrent cells (Cho et al., 2014) using convolutions to promote spatial pattern recognition capacity.

The network receives as inputs the data \mathbf{x}_t^l sequentially, where $t \in \mathbb{N}$ denotes the time-step index. The ConvGRU update operations are given by:

$$\mathbf{z}_t^l = \sigma(\mathbf{W}_z^l * \mathbf{x}_t^l + \mathbf{U}_z^l * \mathbf{h}_{t-1}^l + \mathbf{b}_z^l) \quad (2)$$

$$\mathbf{r}_t^l = \sigma(\mathbf{W}_r^l * \mathbf{x}_t^l + \mathbf{U}_r^l * \mathbf{h}_{t-1}^l + \mathbf{b}_r^l) \quad (3)$$

$$\tilde{\mathbf{h}}_t^l = \tanh(\mathbf{W}^l * \mathbf{x}_t^l + \mathbf{U} * (\mathbf{r}_t^l \odot \mathbf{h}_{t-1}^l) + \mathbf{b}_h^l) \quad (4)$$

$$\mathbf{h}_t^l = (1 - \mathbf{z}_t^l) \odot \mathbf{h}_{t-1}^l + \mathbf{z}_t^l \odot \tilde{\mathbf{h}}_t^l \quad (5)$$

where l denotes the layer index, $*$ denotes the convolution operation, \odot denotes element-wise multiplication, σ is the sigmoid activation function, and \tanh is the hyperbolic tangent activation function. The learnable convolution weights and biases are denoted by \mathbf{W} , \mathbf{U} , and \mathbf{b} . As such, \mathbf{z} is the update gate tensor, \mathbf{r} is the reset gate tensor, and \mathbf{h} is the output/state of the ConvGRU cell. It is clear then how spatial patterns are recognized through the use of spatial convolutions, and temporal patterns are recognized by the information propagation in \mathbf{h} across time-steps. In the beginning of the time series, the previous state input is composed of zeros, i.e., $\mathbf{h}_{-1}^l = \mathbf{0}$. We refer the reader to Cho et al. (2014) and Ballas et al. (2016) for more information on gated recurrent units.

In practice, the chosen architecture uses 5 layers of ConvGRU cells with hidden depth of 128 channels. No encoder-decoder architecture is employed to maximize the spatial detail of the outputs, in accordance with previous work (Rodriguez and Wegner, 2018; Lang et al., 2019; Becker et al., 2021). A diagram of the proposed network can be seen in Fig. 2, along with a detailed schematic of a ConvGRU cell. For every

time-step, the dynamic inputs for that date are concatenated to the static inputs along the color channel, resulting in an “image” with several color channels. In the proposed system, this image contains 12 channels from optical data, 4 channels from SAR data, and 6 channels of static data derived from elevation maps, for a total of 22 channels. These data sources will be explained in more detail in the following section. The output layer consists of a 1×1 convolution to adjust the number of feature channels and the appropriate activation, which will be discussed in more detail in the next section.

3.3. Probabilistic modeling of aleatoric uncertainties

Given that we are dealing with a regression problem of strictly non-negative numbers, the most intuitive approach is for the output layer to yield a single feature map, followed by a ReLU activation to enforce strict non-negativity of the outputs. Supervision can then be done using a regression loss such as mean squared error (MSE) or mean absolute error (MAE). In Section 5 we describe experiments conducted using the MSE loss function, which can be defined as

$$\text{MSE}(\mathbf{y}, \hat{\mathbf{y}}) \triangleq \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}, \quad (6)$$

where $\hat{\mathbf{y}}$ is an estimate of \mathbf{y} , and $i \in [1, \dots, N]$ indexes all elements $\hat{\mathbf{y}}$ and \mathbf{y} . This definition can be trivially extended to a multidimensional case.

Recent work (Becker et al., 2021; Lang et al., 2022a; Alagialoglou et al., 2022; Lang et al., 2022b) has explored the usage of probabilistic formulations for allowing the network to predict an output distribution rather than a single point estimate (Kendall and Gal, 2017). The Gaussian model which is most often used requires the network to output two parameters which uniquely define a normal distribution: mean $\hat{\mathbf{y}}$ and log-variance s . The log-variance is used for numerical stability reasons, but one can calculate the distribution’s standard deviation, which is more frequently used to parametrize Gaussian distributions, simply by computing $\sigma = \sqrt{e^s}$. In this formulation, $\hat{\mathbf{y}}$ can be interpreted as our best estimate, and s can be interpreted as an estimate of aleatoric uncertainty.

Note that in this case no complete ground truth exists, strictly speaking, since there is no reference data for what the value of s should be. s represents the network’s heteroscedastic (spatially and temporally varying) estimation of aleatoric uncertainty in the predictions $\hat{\mathbf{y}}$, and is therefore not a property of the data. A network that uses such probabilistic modeling can be trained using the Gaussian negative log-likelihood (GNLL) loss function:

$$\text{GNLL}(\mathbf{y}, \hat{\mathbf{y}}, s) = \frac{1}{N} \sum_{i=1}^N \frac{1}{2} \left[\frac{(y_i - \hat{y}_i)^2}{e^{s_i}} + s_i + \ln(2\pi) \right], \quad (7)$$

where the constant term $\ln(2\pi)$ is usually discarded since it has no influence in the optimization process or the position or shape of a function’s minima.

In this approach, the network’s output layer is a 1×1 convolutional layer with 2 output channels associated with $\hat{\mathbf{y}}$ and s . The outputs $\hat{\mathbf{y}}$ go through a ReLU activation layer, since we are attempting to regress non-negative snow depth values.

3.4. Training and inference details

Given the two available sets of snow depth maps described in Sections 4.2.1 and 4.2.2, which differ in scale, availability, and quality, we opted to conduct the training of the proposed neural network in two stages. First, the network is trained using the large-scale, low-fidelity maps described in Section 4.2.1 as reference data. The aim here is to learn as much as possible from these data, especially temporal snow dynamics. This training stage is later referred to as pre-training or PT. Then, we use the high-fidelity maps described in Section 4.2.2 as reference data. The aim here is to learn the high-frequency details from

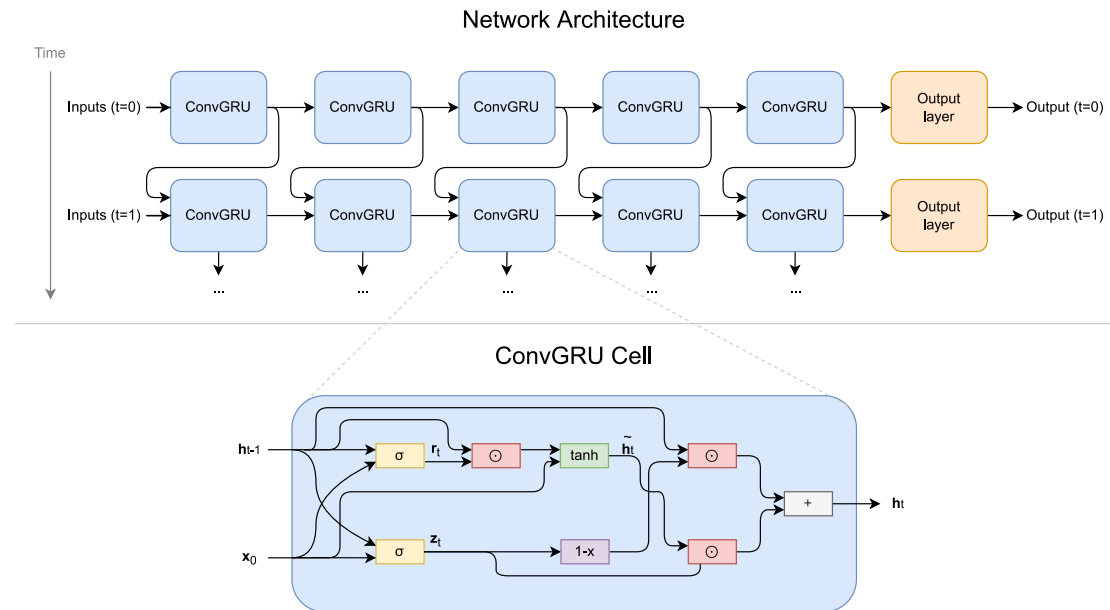


Fig. 2. Network architecture and ConvGRU cell details. Convolutions are used to better identify spatial patterns, while the recurrent cells allow for the accumulation of information over time.

these data, to calibrate the uncertainty estimation, as well as to correct any possible biases present in the first set of reference data. This stage will later be referred to as fine-tuning, or FT.

The network is trained using time-series of inputs ranging from the first of November until the last day of April using a time-step of 7 days. For the PT stage, only data from the 2020/21 winter season are available. At this stage we split the data into train and validation by longitude: the western 80% are used for training, and the eastern 20% are used for validation. The rationale behind this split is to avoid any overlap between data used for training in this stage and data used for cross-validation data in the FT stage. During the FT stage, data from three winter seasons are available: 2018/19, 2019/20, and 2020/21. All high-fidelity snow depth maps used at this stage are located in the eastern 20% of the Swiss territory. We conduct this stage in a cross-validation manner, i.e., we train using data from two years (e.g., 2018/19 and 2019/20) and evaluate it using the remaining data (e.g., 2020/21). Additionally, the results are evaluated against automated ground measurements over the entire Swiss territory.

4. Data

In this work we combine data from several different sources for both input data and reference data. The region of interest is the entire area of Switzerland and Liechtenstein (referred to as Switzerland from this point on for simplicity). In this region, accurately estimating snow depth with high spatial resolution would have a strong positive impact in energy management and winter sport activities. The total considered area amounts to approximately 41'500 km², although the high resolution snow depth maps do not cover the entire area. The total size of the dataset we use, after all preprocessing steps, is approximately 8 TB.

The compiled data spans three winter seasons: 2018/19, 2019/20, and 2020/21. All the available data from this period relative to the area of interest has been downloaded and preprocessed. The preprocessing details for each of the data sources are described below. All data have been resampled using an aligned grid with a footprint of 10 m GSD. The operations described in the following subsections are done using ESA's SNAP toolbox (SNAP, 2022), GDAL (GDAL/OGR contributors, 2022), or custom code written in Python.

4.1. Input data

We use input data from three sources. The two dynamic data sources (i.e., changing in time) are SAR images from the Sentinel-1 satellites (Sentinel-1, 2022) and multispectral optical images from the Sentinel-2 satellites (Sentinel-2, 2022). These have been chosen due to their high resolution, high revisit rate, and large-scale coverage. These data are also freely accessible, which would enable the application of the method proposed in this paper to many different regions. Furthermore, a static data source (i.e., not changing in time) is a digital elevation map (DEM), from which we derive and precompute certain features.

4.1.1. Sentinel-1

The Sentinel-1 GRD data (Sentinel-1, 2022) are composed of C-band SAR images in two polarizations: VV and VH. The obtained images are robust to variations in solar illumination, cloud cover, and other meteorological events, which makes them a reliable, timely tool for observing the Earth at regular intervals. For the considered time and region, the average revisit rate was approximately 3 days.

The downloaded Sentinel-1 data was preprocessed using ESA's SNAP toolbox (SNAP, 2022). The preprocessing pipeline followed the one proposed in Truckenbrodt et al. (2019) for maximum accuracy. Notably, terrain flattening and terrain correction are of high importance for maximizing the coregistration of SAR images with each other and with other sources of data, as well as for minimizing the distortions caused by differences in angle of incidence and reflection.

Furthermore, there is a significant difference in images obtained during ascending or descending orbits, as can be seen in Fig. 3. Due to these differences, we chose to separate the acquisitions into two groups depending on the orbit type. This then resulted in four total image channels to serve as inputs to the network: two polarization channels for each of the two orbit types.

4.1.2. Sentinel-2

The multispectral optical images obtained by the Sentinel-2 satellites (Sentinel-2, 2022) also serve as inputs to the network. The Sentinel-2 L2 A product contains 12 spectral bands in the visible and infrared range representing bottom-of-atmosphere reflectances. The revisit rate for these satellites is approximately 5 days, although optical

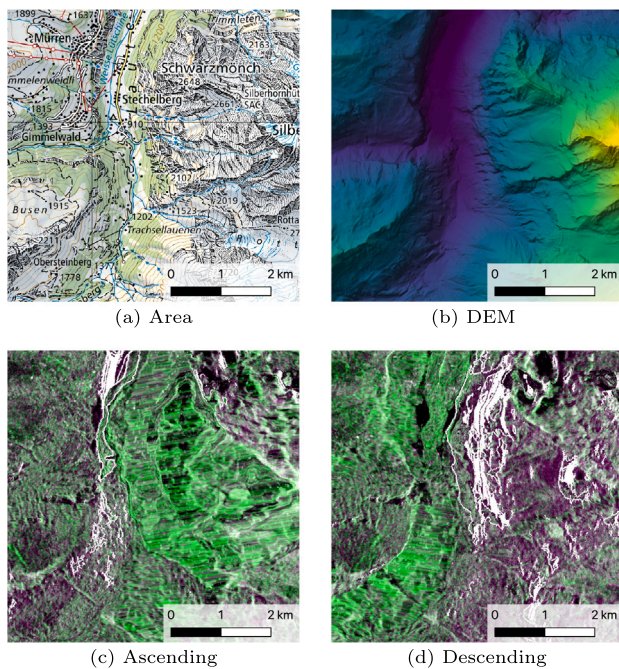


Fig. 3. Sentinel-1 SAR images obtained during (c) ascending and (d) descending orbits display very different profiles depending on the orientation of the imaged surface. Therefore, images acquired during different orbits are considered separately. In (c) and (d), green represents the VH polarization and magenta represents the VV polarization. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

imagery is affected by clouds and other types of occlusion, therefore the effective revisit rate is lower.

For this data source the preprocessing is kept to a bare minimum. Images are simply mosaiced and resampled in the appropriate grid at a resolution of 10 m GSD. This results in 12 color channels that serve as inputs to the neural network. Additionally, cloud masks were obtained using the s2cloudless algorithm (Skakun et al., 2022).

4.1.3. Digital elevation map

The last input data source is a digital elevation map (DEM). We use swissALTI3D (Swisstopo, 2018), a high quality DEM of the Swiss territory provided by SwissTopo, which we resample in the same 10 m grid as all other data sources. Furthermore, we use this DEM to compute topographic features of interest using GDAL. The chosen features are: slope, terrain ruggedness index (TRI) (Riley et al., 1999), topographic position index (TPI) (Weiss, 2001; Wilson et al., 2007), and aspect. Previous work has shown that topographic features of this kind strongly correlate with snow depth, especially TPI (Revelto et al., 2020). The aspect channel is further split into cosine and sine of the computed aspect angle to avoid step discontinuities in north-facing areas. Although these features can in theory be learned by the network if necessary given enough data, we opt to pre-compute them to facilitate the learning process and because the training data are limited. The set of DEM and derived features provide six input channels which are used by the neural network.

In Fig. 4 we see part of the elevation map and the associated features for the same region that have been calculated from it. The features capture characteristics which affect how snow accumulates and melts. Notably, TPI captures fine-scale terrain details that correlate with locations where snow accumulates, e.g., gullies.

4.2. Reference data

In this work, in accordance with previous work (Bühler et al., 2015, 2016, 2017; Eberhard et al., 2021; Jacobs et al., 2021), the term “snow

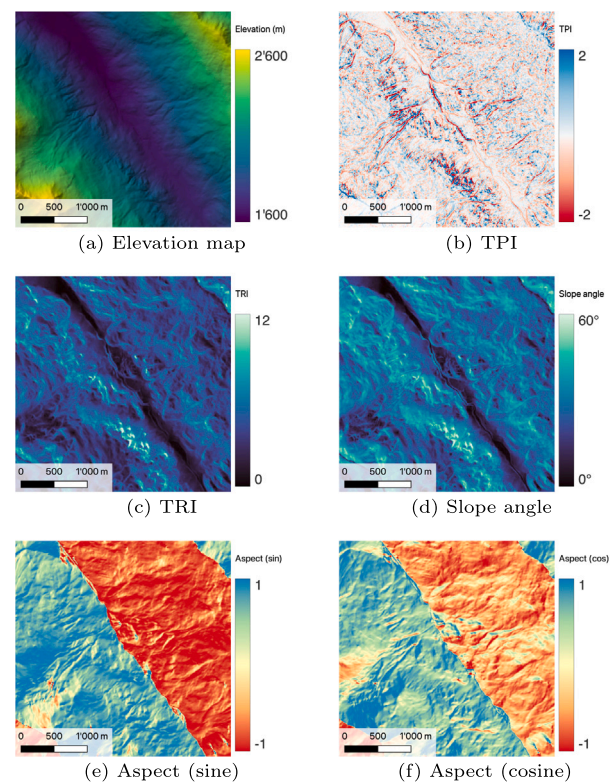


Fig. 4. DEM and extracted features in an area of interest near Davos. The extracted features (b–f) capture fine terrain characteristics that are not obvious in the original elevation map (a). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

depth” refers to the difference between measurement at a given date and reference data measured in the summer.

Two sets of snow depth maps are available to supervise the training of the proposed neural networks. The first set of maps is available daily for the winter of 2020/21 and covers the entire Swiss territory, but with limited accuracy and spatial detail. The second set of maps has been generated using high-fidelity photogrammetric methods (Bühler et al., 2022), but are limited to only few regions and dates. These two sources of data are therefore complementary. Examples can be seen in Fig. 5.

4.2.1. Large scale snow depth maps

We use snow depth maps that have been produced daily for the 2020/21 winter season, and which span the entire Swiss territory. These maps have been produced using TCAM (Wulf et al., 2020). The first step in that method is a first-order estimation of the correlation between the snow depth measurements from measurement stations, and whose residuals are used as inputs to a generalized additive model to account for spatial variations. This is then refined using learned residuals which are estimated using topographic features. We refer the reader to Wulf et al. (2020) for a more detailed description of this approach.

The snow depth estimates produced by this algorithm are currently the best available at country-scale, which is important for the use cases considered in this paper. Nevertheless, the produced maps lack the spatial detail that are seen when mapping the snow with photogrammetric methods. In particular, these maps are important for planning and safety analysis of winter sports. Furthermore, the reliance on measurement stations may limit the extension of such algorithm to regions that are not as densely monitored as Switzerland.

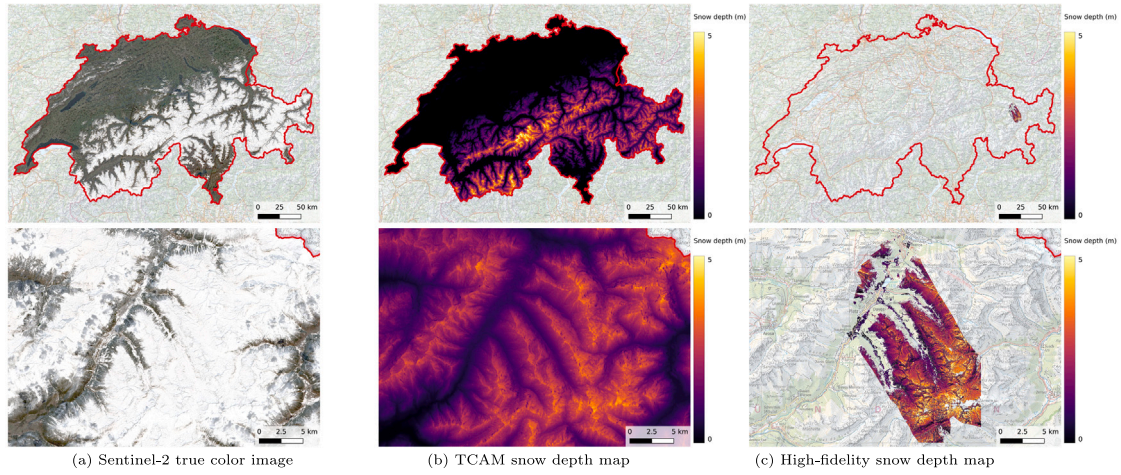


Fig. 5. Color image and snow depth maps available for 16/04/2021. The snow maps shown in (b) and (c) are complementary in terms of (c) spatial detail and high fidelity, and (b) spatiotemporal coverage. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

Locations and dates of available high-resolution snow depth maps obtained using photogrammetry.

Location	Date
Dischma valley	16/03/2019
	06/04/2020
	16/04/2021
Dorfberg	11/12/2020
	25/02/2021
Gaudergrat	12/12/2018
	12/03/2019
	07/02/2020
	17/02/2020
Latschülfurgga	18/12/2020
	24/02/2021
	26/03/2021
Schürlialp	18/02/2019
	22/04/2020

4.2.2. High fidelity snow depth maps

A second set of local snow depth maps is available. These have been obtained using photogrammetric techniques (Bühler et al., 2016, 2017; Eberhard et al., 2021; Bühler et al., 2022) and provide the best available snow depth mapping accuracy at very high spatial resolution (better than 0.5 m GSD). Due to the cost and planning required for the acquisition of the very high resolution images from aerial sources or uncrewed aerial vehicles (UAVs) needed for such methods, these maps are only available for sparse dates and regions of interest in the considered winter seasons.

Despite the low temporal resolution and limited area coverage of these maps, their strengths are complementary to the weaknesses of the maps described in the previous section. Such maps contain high spatial frequency details and are highly accurate, especially in regions where detailed snow distribution patterns are of special interest. The high accuracy of these maps is useful not only for supervising the training of the proposed machine learning method, but also to validate it by using a cross-validation scheme.

The dates for which measurements are available can be found in Table 1. The sparsity and scale of the described data is clear, but our results show that using such data is of paramount importance to boost the performance of our snow depth estimation algorithm. Finally, the approximate area covered by each of these datasets can be found in Fig. 6.

5. Experiments

5.1. Experimental setup

All the reported experiments were conducted using PyTorch on a Nvidia GeForce 3090 GPU card. The optimization during training was done using stochastic gradient descent (SGD) with momentum of 0.9 and a base learning rate of 10^{-4} . Note that the concept of an epoch is not properly defined in the described dataset, so here we define an epoch to be a set of 800 randomly sampled time-series of patches for training and 200 for validation. During the PT stage, the network was trained for 100 epochs at the base learning rate, followed by 10 epochs with a linearly decreasing learning rate. During the FT stage, the network was trained for 50 epochs at the base learning rate, followed by 50 epochs with a linearly decreasing learning rate. Batches consisted of two time-series of 128×128 patches (covering an area of approximately 1.6 km²) composed of data starting in the beginning of November and ending on the last available day in April.

5.2. Evaluation metrics

The accuracies of the snow depth estimates were evaluated using four main metrics: mean absolute error (MAE), root mean squared error (RMSE), Pearson's correlation coefficient (ρ), and mean error (ME). MAE and RMSE measure the general discrepancy between \mathbf{y} and $\hat{\mathbf{y}}$, with MAE being more sensitive to smaller errors and RMSE being more sensitive to larger errors. ρ measures the correlation between \mathbf{y} and $\hat{\mathbf{y}}$ while ignoring errors in overall shift and scale. ME measures whether there are biases in the errors. These metrics can be defined as:

$$\text{MAE}(\mathbf{y}, \hat{\mathbf{y}}) \triangleq \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (8)$$

$$\text{RMSE}(\mathbf{y}, \hat{\mathbf{y}}) \triangleq \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (9)$$

$$\text{ME}(\mathbf{y}, \hat{\mathbf{y}}) \triangleq \frac{1}{N} \sum_{i=1}^N y_i - \hat{y}_i \quad (10)$$

$$\rho(\mathbf{y}, \hat{\mathbf{y}}) \triangleq \frac{\sum_{i=1}^N (y_i - \mu(\mathbf{y})) \cdot (\hat{y}_i - \mu(\hat{\mathbf{y}}))}{N \cdot \sigma(\mathbf{y}) \cdot \sigma(\hat{\mathbf{y}})} \quad (11)$$

where $\mu(\cdot)$ denotes the mean operator and $\sigma(\cdot)$ denotes the standard deviation operator.

The calibration accuracy of the estimated uncertainties has been quantified in two ways. The estimated variance is, by definition, the estimated value of the squared error of a given snow depth estimate.

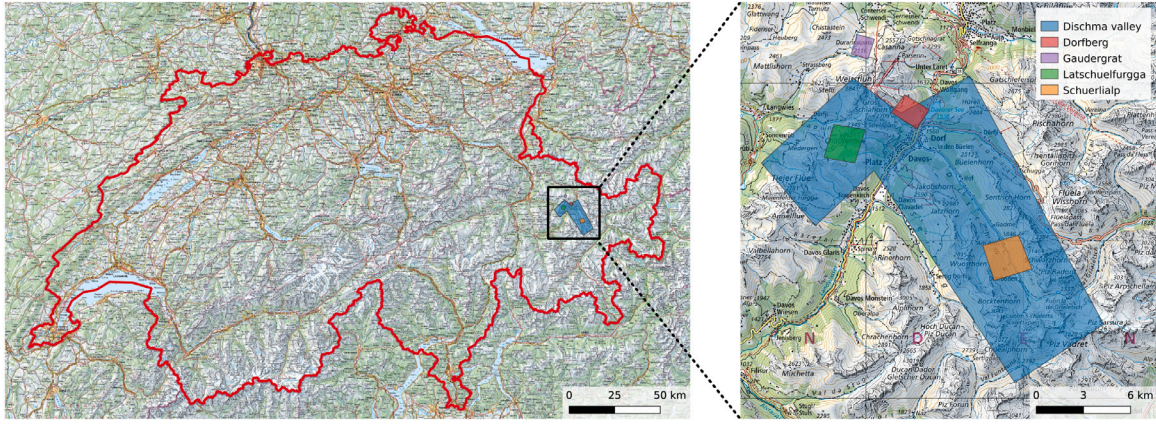


Fig. 6. Approximate locations for each of the regions for which high-fidelity snow depth maps are available. All the considered areas are near Davos, in eastern Switzerland. Despite the spatial overlap, no two maps have been produced for the same date. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

But given that this is a stochastic process, samples cannot be individually evaluated. Thus, samples can be grouped into M groups $G_i, i \in [1, M]$ according to estimated variance, and then each group's mean squared error (MSE) can be compared with the groups mean estimated variance. We split the samples into 100 quantiles for this computation. The expected calibration error (ECE) (Alagialoglou et al., 2022) can then be computed using

$$\text{ECE}(\mathbf{y}, \hat{\mathbf{y}}, \sigma^2) \triangleq \sum_{i=1}^M \frac{|G_i| \cdot |\text{MSE}(\mathbf{y}_{G_i}, \hat{\mathbf{y}}_{G_i}) - \mu(\sigma_{G_i}^2)|}{N}, \quad (12)$$

where $|G_i|$ denotes the number of elements in group G_i . MSE was defined in Eq. (6).

Another way to validate the estimated distributions is to evaluate the coverage of the reference data for different confidence margins. Intuitively, if we compute for each estimate a lower and an upper bound which give us an X% margin of confidence, we should expect X% of the reference values to fall within this margin. Since the probabilistic estimates \hat{y}_i and $\sigma_i^2 = e^{v_i}$ can be interpreted as a Gaussian probability distribution function $\mathcal{N}(\hat{y}_i, \sigma_i^2)$, the lower bound $l(c)$ and upper bound $u(c)$ defining a margin with expected coverage $c \in (0, 1)$ can be computed using the inverse cumulative distribution function $\text{icdf}(x)$:

$$l(c) = \text{icdf}(0.5 - \frac{c}{2}) \quad (13)$$

$$u(c) = \text{icdf}(0.5 + \frac{c}{2}) \quad (14)$$

We can then graph the actual coverage versus the expected coverage for several values in the range $(0, 1)$. Kuleshov et al. also evaluate the calibration of regression systems in a similar way (Kuleshov et al., 2018), although we center the intervals around the median, i.e., $\text{icdf}(0.5)$, since that is more relatable to real scenarios where a best estimate and a confidence margin around that value are used. We can then quantify the calibration error as the area between the obtained curve and the ideal (identity) curve, which approaches 0 for perfect calibration and 0.5 in the worst case scenario. This measure is henceforth named Area Between Curves (ABC).

5.3. Snow depth estimation results

After training the proposed neural network, we produced country-scale estimates for both qualitative and quantitative evaluations. One such output is presented in Fig. 7, along with the associated uncertainty estimates. We have also included the temporal-stacked true color composite Sentinel-2 image and the DEM for context. While the finer details cannot be seen in this country-scale image, it clearly shows several correlations that are to be expected. First, we can observe how regions in higher elevations often accumulate more snow during the winter.

Second, we see that the predicted snow cover follows very closely the snow that is visible in the optical images. Furthermore, we see that regions with higher snow depth estimates also tend to be attached to higher uncertainty values.

Fig. 8 contains zoomed in results of selected regions to highlight the difference in spatial detail contained in the presented results and those obtained by using TCAM (Wulf et al., 2020). This method is the only one found in the relevant literature whose results are comparable in scale and resolution, as was discussed in Section 2.1. Note that all snow depth estimates have been computed at 10 m GSD, thus differences in the amount of fine details in each result do not come from differences in image resolution. While both results generally agree in the distribution of snow at a macro scale, our method produces snow depth estimates with much finer spatial detail. These high-frequency details closely depend on the local topography, and their presence in the snow depth estimates is important for applications such as planning winter activities in nature. These details are likely to help in predicting avalanche risks more accurately.

Figs. 9 and 10 show a comparison of snow depth estimates to reference data around Davos on 16/04/2021, which is the date for which the most reference data are available. Snow depth estimates produced by our method and by TCAM are presented, as well as images containing the residual errors after comparing to the reference data. Fig. 10, which contains zoomed in crops of the results, is especially illustrative of the improvements that come from using our method. We can see that the high-frequency details that are present in our snow depth estimates correspond to those in the reference data. This leads to a general improvement in the image of residual differences for our snow depth estimates. We can also observe in these images the strong correlation between the TPI and the spatial detail in our snow depth estimates.

Quantitative evaluation of the proposed method is presented in Table 2. Results are computed using data from all available years (left half) or only data from the 2020/21 winter season (right half). This is done because snow depth estimates using TCAM are only available for this last winter season, and this way a quantitative comparison between these two methods is possible. We also present results for ablated versions of the proposed method to evaluate the contribution of each data source and training procedure detail. The first groups of ablations concern data sources: each input data source is ablated one at a time. The second group of ablations concerns training and architecture details: the absence of the PT and FT stages are studied, as well as replacing the probabilistic loss function with the more standard MSE loss function. We also explore the value of temporal information, by severing the temporal connections in the recurrent cells. Note that the network trained using the MSE loss is unable to produce uncertainty estimates.

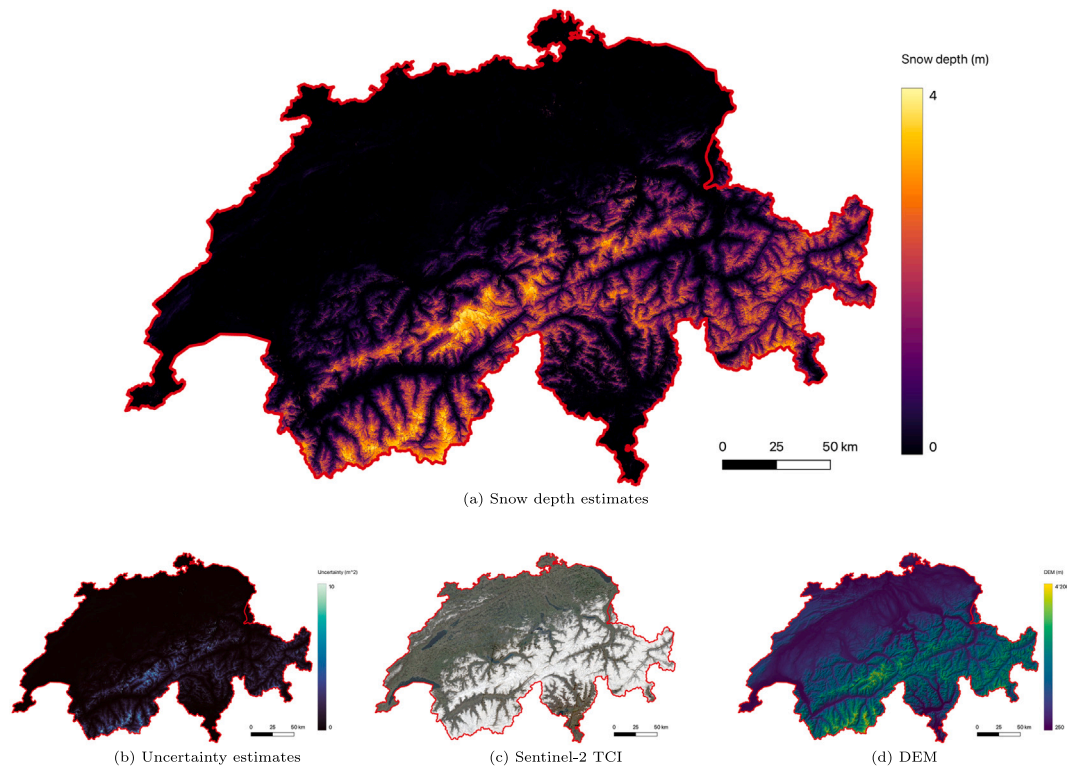


Fig. 7. Snow depth map, uncertainty map, optical image composite, and DEM for 16/04/2021. The proposed method allows us to produce such estimates at 10 m GSD on a weekly basis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

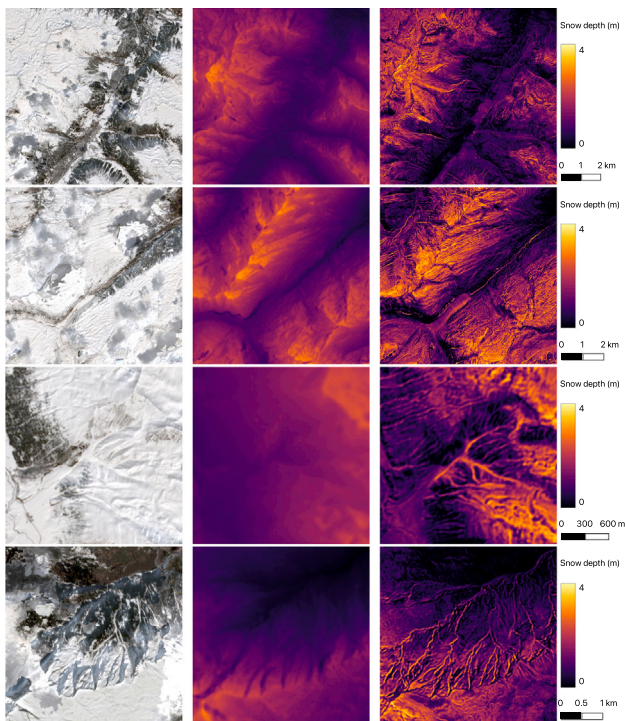


Fig. 8. Comparison between snow depth estimates obtained by using TCAM (center) and the presented method (right), along with a true color composite image for reference (left). There is a clear improvement in the amount of spatial detail in our results, especially in gullies which accumulate snow over the winter. All snow depth estimates refer to 16/04/2021 and are computed at 10 m GSD. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We have also repeated the training procedure three times using different seeds for generating random numbers in order to evaluate the consistency of the results across different runs. This was done for the proposed method, which also allows us to compute metrics for the “PT only” ablation. Results are reported in Table 3. We see that the combination of a relatively small network and an optimizer with slowly decaying learning rate leads to very consistent results. The computed standard deviation is minimal for MAE, RMSE, and ρ , with slightly larger values for ME.

Fig. 11 contains 2D histograms that show the distribution of estimated snow depth versus reference data for the entire dataset. We see that after the PT stage the neural network’s estimates closely match the distribution of the snow estimates produced by TCAM, which are used as reference data at this stage. In (c), we see that the distribution after the fine-tuning stage more adequately follows the identity diagonal, which is our target. Even if there remains a fair amount of uncertainty in the results, the final results are much less biased, which is important especially when using or analyzing aggregate estimates.

5.4. Uncertainty calibration evaluation

To evaluate the calibration of the estimated predictive uncertainties, we first plot the MSE versus the estimated variance after binning the predictions by quantiles using 100 bins. This curve is shown in Fig. 12, alongside the histogram of estimated predictive uncertainties. The plot shows that the calibration curve closely follows the identity diagonal, which is our target, for values up to approximately 1.5 m^2 . The histogram included in the figure shows that most values fall within this interval, which means the uncertainty is well calibrated for most predicted values. We see that the predictive uncertainty is slightly underestimated in the range of values up to approximately 1.2 m^2 . This gap can likely be explained by the fact that we are not estimating epistemic uncertainty, only aleatoric uncertainty (Kendall and Gal, 2017). Estimating epistemic uncertainties, for instance using ensembles

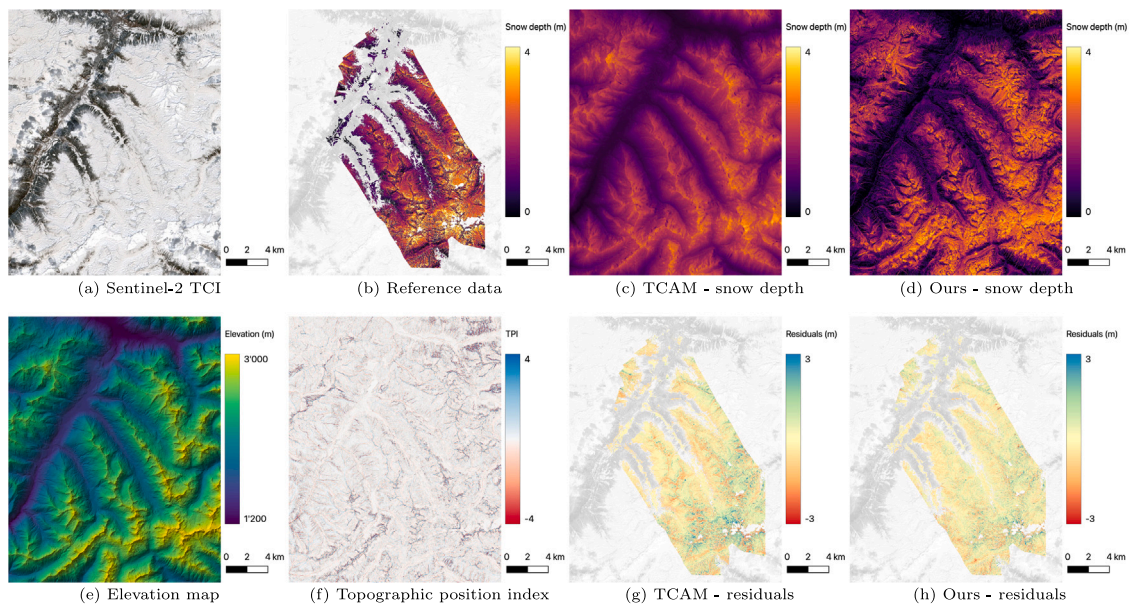


Fig. 9. Comparison between our method and the considered baseline, TCAM, for 16/04/2021. We see that at a macro scale, both algorithms are in agreement. The residual images show that our estimates are more accurate despite having no access to measurement station data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

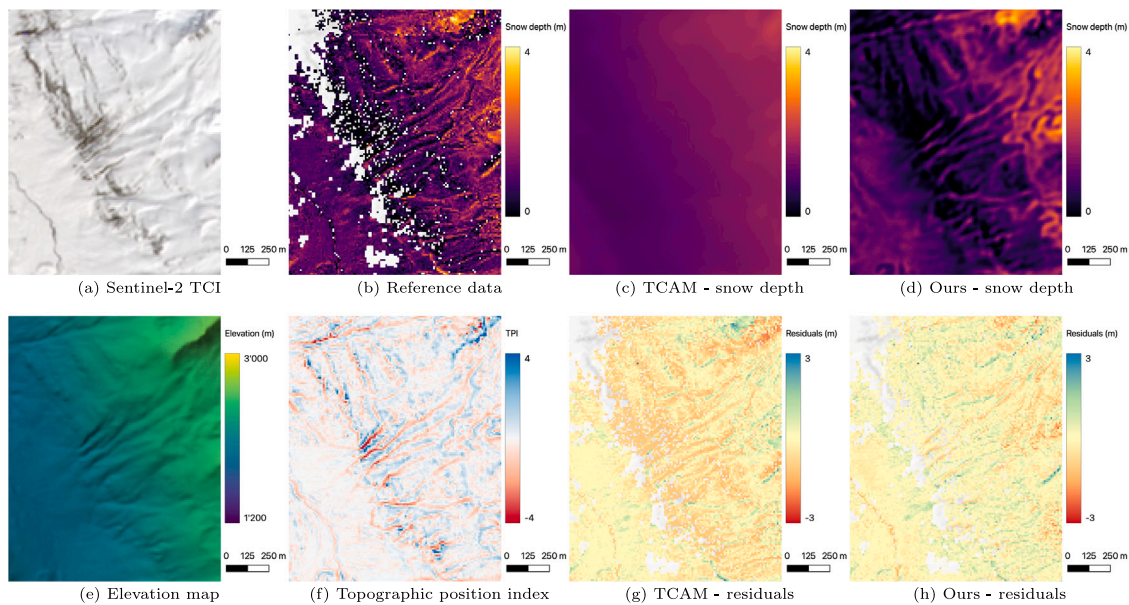


Fig. 10. The improvements with respect to TCAM become even clearer at a fine scale. Our snow depth estimates contain the fine-scale details that are missing from the baseline method, which reduces the errors observed in the residual images. The displayed images refer to 16/04/2021. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2

Quantitative evaluation of snow depth estimates. Units are in meters (MAE, RMSE, ME) or unitless (ρ). The proposed method outperforms TCAM in all metrics. Ablation results for used data sources (first group) and training procedure (second group) are also included. Numbers in parentheses highlight the difference to the proposed method's results.

Method	All dates				Winter 2020/2021			
	MAE (↓)	RMSE (↓)	ρ (↑)	ME (0)	MAE (↓)	RMSE (↓)	ρ (↑)	ME (0)
TCAM	–	–	–	–	0.69 (+0.11)	0.94 (+0.14)	0.51 (−0.17)	0.07
Proposed method	0.68	0.91	0.59	0.08	0.58	0.80	0.68	0.06
No S2	0.69 (+0.01)	0.93 (+0.02)	0.59 (0.00)	0.22	0.64 (+0.06)	0.88 (+0.08)	0.63 (−0.05)	0.26
No S1	0.72 (+0.04)	0.95 (+0.04)	0.53 (−0.06)	0.06	0.61 (+0.03)	0.83 (+0.03)	0.65 (−0.03)	−0.03
No DEM	0.70 (+0.02)	0.94 (+0.03)	0.55 (−0.04)	0.01	0.64 (+0.06)	0.87 (+0.07)	0.61 (−0.07)	−0.08
PT only	0.74 (+0.06)	1.00 (+0.09)	0.50 (−0.09)	0.19	0.67 (+0.09)	0.92 (+0.12)	0.56 (−0.12)	0.15
FT only	0.67 (−0.01)	0.91 (0.00)	0.59 (0.00)	0.11	0.58 (0.00)	0.79 (−0.01)	0.69 (+0.01)	0.06
MSE	0.66 (−0.02)	0.89 (−0.02)	0.62 (+0.03)	0.11	0.57 (−0.01)	0.79 (−0.01)	0.69 (+0.01)	0.07
No recurrence	2.23 (+1.55)	2.50 (+1.59)	0.19 (−0.40)	−1.85	2.58 (+2.00)	2.78 (+1.98)	0.30 (−0.38)	−2.53

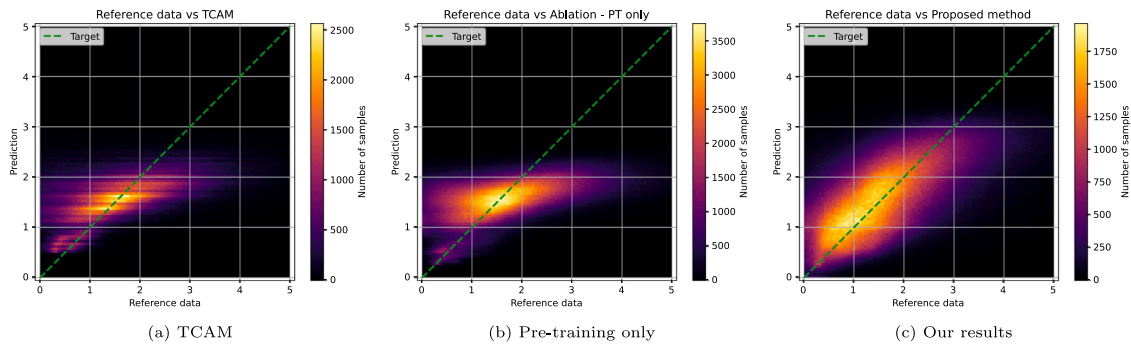


Fig. 11. 2D histograms of reference snow depth values versus predictions by TCAM and our method, including the “PT only” ablation. We see that our proposed method follows much more closely the ideal line depicted in dashed green. After the pre-training stage the results resemble those of TCAM, and further improve after fine-tuning. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3

Mean and standard deviation for each metric over multiple runs. We observe extremely consistent results in terms of MAE, RMSE, and ρ with slightly stronger variations in bias (ME).

Method	All dates				Winter 2020/2021			
	MAE (↓)	RMSE (↓)	ρ (↑)	ME (0)	MAE (↓)	RMSE (↓)	ρ (↑)	ME (0)
Proposed method	0.68 ± 0.01	0.91 ± 0.01	0.60 ± 0.01	0.09 ± 0.02	0.59 ± 0.01	0.80 ± 0.01	0.68 ± 0.00	0.08 ± 0.05
PT only	0.74 ± 0.00	0.99 ± 0.01	0.50 ± 0.00	0.18 ± 0.05	0.67 ± 0.00	0.92 ± 0.01	0.56 ± 0.00	0.14 ± 0.03

Table 4

Calibration metrics and mean variance for each experiment. We see that most ablations achieve worse uncertainty calibrations than the base method. The exception is the experiment which uses no elevation information, but at the cost of actual accuracy. Notably, we see that skipping the fine-tuning stage drastically decreases the uncertainty calibration by making the model over-confident.

Model	ECE (↓)	ABC (↓)	σ^2
Proposed method	0.190	0.011	0.80
No S2	0.241 (+0.051)	0.020 (+0.009)	0.78
No S1	0.389 (+0.199)	0.047 (+0.036)	0.79
No DEM	0.163 (−0.027)	0.002 (−0.009)	0.86
PT only	0.917 (+0.727)	0.279 (+0.268)	0.09
FT only	0.256 (+0.066)	0.024 (+0.013)	0.75
No recurrence	2291 (+2291)	0.273 (+0.262)	2296

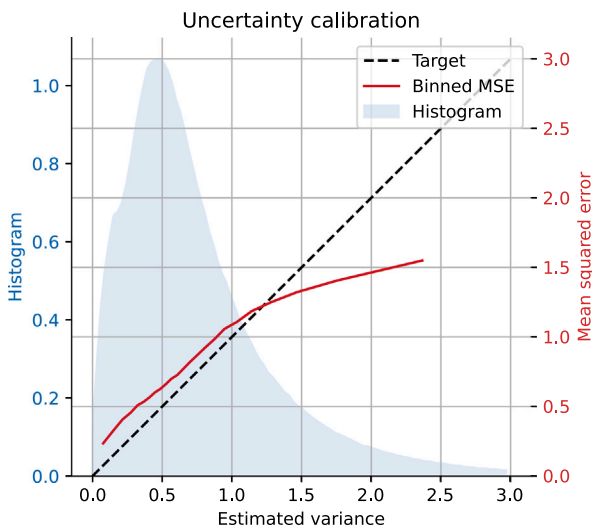


Fig. 12. Binned mean squared error versus estimated variance and histogram of predicted variances. We observe good calibration of the estimated predictive uncertainties in the range of variances that includes most samples.

(Kendall and Gal, 2017; Lakshminarayanan et al., 2017) or pseudo-ensembles (Turkoglu et al., 2022; Kushibar et al., 2022), could bring marginal improvements to the calibration of uncertainties.

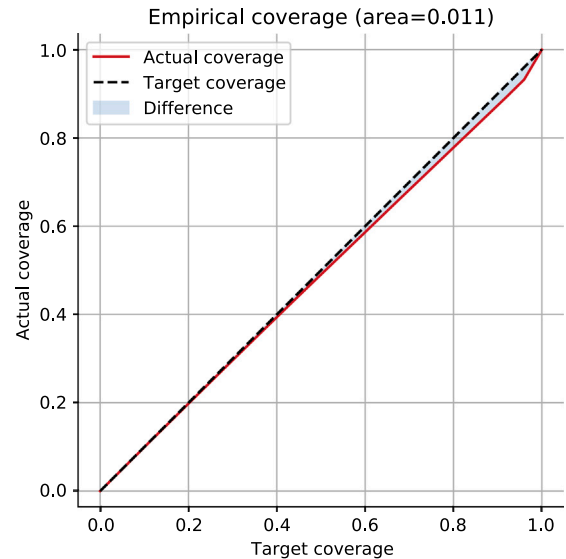


Fig. 13. Observed coverage for various target confidence margins. We see that the observed empirical coverage obtained using windows calculated using the estimated uncertainty follows closely the target, which means our uncertainties are very well calibrated.

In Fig. 13 we show the coverage plot, described in Section 5.2. Using this curve we can evaluate how accurate confidence margins used in practical applications would be. We see that the measured coverage curve follows very closely the ideal coverage, which means that estimated confidence margins computed for each location would very closely match the target coverage. For instance, computing the confidence margins for a target coverage of 50% achieves real coverage of 49.06%. The discrepancy between the empirical coverage and the ideal coverage curves is small, which is a sign of good uncertainty calibration.

Table 4 contains the quantitative metrics that evaluate the calibration of estimated predictive uncertainties. These numbers show that almost every single ablation experiment that was conducted led to worse calibration of predictive uncertainties. The only exception is

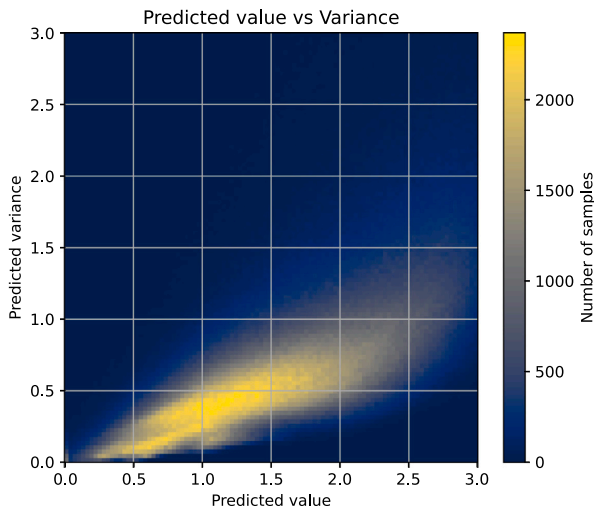


Fig. 14. We see a correlation between predicted values and uncertainties. This is intuitive: areas with higher snow depths are attributed proportionally higher uncertainties. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5

Accuracy and uncertainty calibration w.r.t. ground measurements from the IMIS network of ground stations. Deviations tend to increase for later months of the winter season, likely due simply to the larger amounts of snow. Deviations are also lower in the eastern region that was used for fine-tuning the network. *N* stands for the number of data points in each partition.

Data	N	MAE	RMSE	ρ	ME	ECE	ABC
All	3402	0.59	0.81	0.61	0.39	0.57	0.21
11/2020	512	0.13	0.18	0.45	−0.03	0.05	0.15
12/2020	528	0.24	0.31	0.56	0.00	0.06	0.06
01/2021	659	0.54	0.69	0.41	0.30	0.42	0.26
02/2021	524	0.80	0.95	0.32	0.60	0.82	0.35
03/2021	524	0.86	1.07	0.28	0.68	1.04	0.32
04/2021	655	0.90	1.10	0.32	0.74	1.11	0.34
Eastern 20%	698	0.48	0.59	0.71	0.30	0.27	0.18
Western 80%	2704	0.62	0.86	0.60	0.42	0.66	0.21

the experiment which did not use elevation data, but this also led to lower real accuracy, as seen in Table 2. The experiment which obtained the worst results in these uncertainty calibration metrics was the one which skipped the fine-tuning stage. We observed that in this case the estimates were more strongly biased towards underestimating snow depth, as can be seen in Fig. 11, while also being overly confident.

Finally, in Fig. 14 we show that there is a strong correlation between the predicted snow depth value and the estimated uncertainty. This is intuitive since we are estimating strictly positive values: it means that the relative uncertainty remains somewhat constant for values in different ranges.

5.5. Validation with ground measurements

To evaluate the estimated snow depth maps against truly independent observations, we used measurements from the Inter-cantonal Measurement Information System (IMIS) ground stations. We produced the full time-series of weekly snow depth maps from 11/2020 to 04/2021, spanning the entire Swiss territory. We have also obtained all available measurements in the area of interest from the IMIS network, in total between 128 and 132 measurements per day. The evaluation was only conducted for our proposed method. TCAM uses these measurements as inputs, so the comparison would be meaningless. Furthermore, too few IMIS data points (namely, five) fall in the area of the high-fidelity snow maps to compute reliable statistics.

The comparison to IMIS, reported in Table 5, is split into three sections. First, we report an evaluation using all data points. Then, a month-wise evaluation was conducted to see how reliable the produced snow depth maps are throughout the winter season. Finally, separate evaluations were conducted for the eastern and western splits of regions used for training and validation, as described in previous sections.

6. Discussion

Our results show that snow depth estimation using exclusively satellite data sources and elevation maps is not only feasible, but it can be more accurate than what was previously possible at a large scale. As discussed in Section 1, we aim to estimate snow depth for large areas (e.g., entire countries or mountain ranges) and at a high resolution (10 m GSD). The only other method which achieves this combination is TCAM. Our results show that we obtain better performance without relying on a measurement station network. By instead using data from satellite sources, our method is in principle scalable without a dense infrastructure of measurement stations. The results presented in Table 2 show that our method outperforms this baseline by a significant margin in terms of MAE, RMSE, and correlation. Even the results' bias, measured by the mean error, see a slight improvement even though we do not have access to ground measurements.

We also show that the level of spatial detail present in our results is unparalleled by those in the baseline method. These details come largely from terrain features extracted from the elevation map, which provide complementary data to those collected by satellite sources. Figs. 8 and 10 highlight those differences, and show how this positively impacts the quality of the results by reducing the residual errors. Fig. 10 also shows that there is a strong correlation between the TPI, which is extracted from the DEM, and the spatial detail in the produced snow estimates and in the reference data.

Our method achieved an RMSE of 0.91 m in the evaluation using the entire dataset, and 0.80 m when evaluated in the winter season of 2020/2021. This is comparable to the results obtained by Eberhard et al. using photogrammetric methods based on very high resolution satellite images, which achieved an RMSE of 0.44–0.92 m. Furthermore, our method uses open access data, which significantly reduces the cost, and is capable of scaling to a much wider area.

The proposed system was also capable of producing well-calibrated estimates of predictive uncertainties by harnessing a probabilistic framework for training the neural network, even though predicting well-calibrated uncertainties is often challenging (Niculescu-Mizil and Caruana, 2005; Guo et al., 2017; Becker et al., 2021; Lang et al., 2022a). The calibration can be verified visually in Figs. 12 and 13, and quantitatively in Table 4. The analysis presented in Fig. 14 also shows that there is a correlation between estimated snow depth values and predictive uncertainties, which is expected in a regression problem focused on strictly non-negative numbers.

In our ablation studies, most experiments led to a decline in performance relative to the proposed method. The first set of ablations in Table 2 shows that every source of data contributes to the final result in a different way:

- Optical images allow us to accurately estimate snow cover, given good atmospheric conditions;
- SAR images are especially useful for estimating snowmelt (Lievens et al., 2019) and are more robust to atmospheric effects, which means that data from SAR sources are always up to date;
- Elevation maps contain implicit information about the microclimate, as well as fine-grained terrain features which affect snow accumulation patterns.

We also showed that the pre-training stage is not enough to obtain the most accurate snow depth estimates possible. While at this stage we are already able to roughly match the performance of TCAM, the

fine-tuning stage is essential to obtain better results and calibrated uncertainty estimates.

Somewhat surprisingly, our results show that skipping the pre-training stage improves the accuracy of the produced snow depth estimates, albeit by a very small margin. It may be the case that skipping the pre-training stage increases the accuracy in the (more restricted) region used for fine-tuning at the cost of generalization performance to other regions which are only seen during the pre-training stage, but in this study we did not have access to data for validating that hypothesis. Replacing the probabilistic framework by a simple MSE loss function also leads to a small improvement in the results. But we note that these slight improvements come at the cost of uncertainty calibration: skipping the pre-training stage strongly decreases the accuracy of uncertainty calibration estimates, as is shown in Table 4, and using the MSE loss allows for no estimation of uncertainties at all. Furthermore, the observed bias (ME) is worse in these cases compared to our base case when computed using all the available data. Nevertheless, if this is deemed a worthy trade-off by a user, these variations on the proposed method may be preferred for a given application.

We have also studied the performance gains obtained by harnessing satellite time series, by severing the recurrent connections (vertical connections in Fig. 2). As expected, temporal information is essential to accurately estimate snow depth from the considered data sources. While the setting is similar to monocular depth estimation, a well-studied problem in computer vision, the visual cues present in the satellite images are not strongly affected by snow depth, but mainly by snow presence or absence. This forces the network to focus on other cues, i.e., temporal dynamics and seasonal priors.

Finally, the comparison to measurements from the IMIS ground stations in Table 5 provides us with a realistic evaluation against a reliable and completely independent data source. We see that the values for MAE, RMSE and ρ do not differ much from those of the high fidelity maps, but we do see a deterioration in ME, ECE, and ABC. We also see generally lower errors for earlier months in the season, despite the fact that the network uses less data to generate these estimates. This is likely due simply to the lower snow depth in these months. As for the east/west split, we do observe lower errors in the region used for fine-tuning the model. While these numbers are not directly comparable, they do suggest that having a wider variety in the data used for fine-tuning will likely improve the model's capability to generalize.

6.1. Limitations

While our results show that the proposed snow estimation method is a significant step forward when compared to previous work, there are also limitations that should be acknowledged.

At the moment, it is unlikely that the proposed method generalizes well to other regions outside the European Alps. Several characteristics specific to this mountain range are implicitly learned from the data. Given enough local data, the proposed method can be extended to other mountain ranges and snow-covered areas, but that has not been achieved yet. It also remains to be tested whether a single network would be capable of estimating snow depth in different climates or if different regions need to be treated separately by training individual networks.

We have also noted artifacts in the snow depth and uncertainty estimates in regions containing objects that are not present or are under-represented in the high-fidelity dataset used during the fine-tuning stage. Notably, artifacts were present in some urban areas, likely due to the unusually strong SAR signal, and in deciduous forests, which are not present in the fine-tuning dataset.

Another limitation comes from the fact that variations in snow depth do not have a direct effect on the information contained in the optical and SAR images. In other words, the data we have used do not contain details on all the processes related to snow depth and snow cover. Thus, there is an upper bound for the achievable accuracy for

any system that considers only the data sources discussed in this paper, even if superior methods are developed in the future. Furthermore, we are dependent on an accurate cloud detection algorithm for processing optical satellite images, and these are often inaccurate in snow-covered regions. Also, not having access to optical images at times where the snowline quickly changes, e.g., due to cloud cover in the spring or fall, may have a negative impact in the results.

Regarding uncertainty modeling, we note that the Gaussian model may not always be appropriate. For instance, where the network estimates an absence of snow (depth 0), 50% of the probability mass is attributed to negative snow depth values, which are physically meaningless. More sophisticated uncertainty modeling with strictly non-negative distributions is an interesting direction for the future.

Finally, the proposed method is not currently able to harness data from measurement stations. Such data, although sparse, are the most consistent source of information regarding snow height development. This information would be especially valuable to improve the system's response time to events that lead to a rapid change in snow depth and to produce more frequent snow depth estimates.

6.2. Future work

Despite the presented work being a considerable step forward in scalable, high-resolution snow depth estimation, several lines of work can already be seen for further improving the quality of snow depth estimation systems.

The most obvious way to improve the results is to find a way to include station measurements as inputs for the network. This could improve the estimates' accuracy and reduce the biases in the results in regions with dense networks of measurement stations. Station measurements are also valuable sources of information when optical images are missing for long periods due to cloud cover and during events that lead to fast-changing snow pack. Nevertheless, using point-measurements from stations is not directly compatible to the gridded data that have been used in this work, so approaches need to be developed to meaningfully fuse those data sources.

Another possible way to improve the results would be to use meteorological data, such as precipitation, sunshine duration, air temperature, wind speed and direction, and land surface temperature data. While such data are not available in such a fine grid as the 10 m GSD one used in this work, these variables could still provide additional information which helps to predict the temporal evolution of the snow pack.

The applicability of the produced estimates also needs to be explored further for different applications, such as for estimating the risks associated with outdoor activities and avalanches (Pérez-Guillén et al., 2021; Bühler et al., 2022). If coupled with the estimation of snow water equivalent, our snow depth estimates could also be used for hydrological studies of snowmelt and for forecasting energy availability from hydropower during the warmer months of the year (Jonas et al., 2009; Margulis et al., 2019).

Finally, it would be interesting to apply the proposed method to estimate snow depth in other regions by retraining the network using local data. As long as high-fidelity reference data are available, the proposed methodology is in principle applicable to any region within the range of the Sentinel-1 and Sentinel-2 satellites.

7. Conclusion

We have presented a method for estimating snow depth in very large regions (e.g., entire countries or mountain ranges) using time-series of optical and SAR satellite images and an elevation map, and without relying on ground measurement stations. Spatiotemporal information is extracted from these data sources using a recurrent convolutional neural network based on ConvGRU cells. We achieved unprecedented accuracy for this combination of large-scale and high spatial resolution

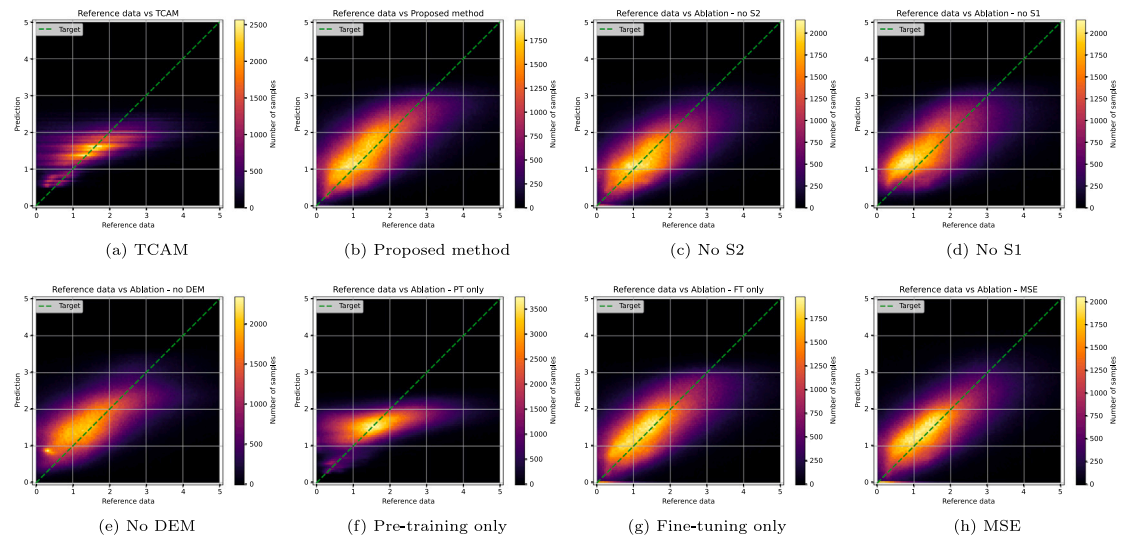


Fig. 15. 2D histograms for reference data versus predicted snow depth values for each ablation experiment and baseline algorithm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

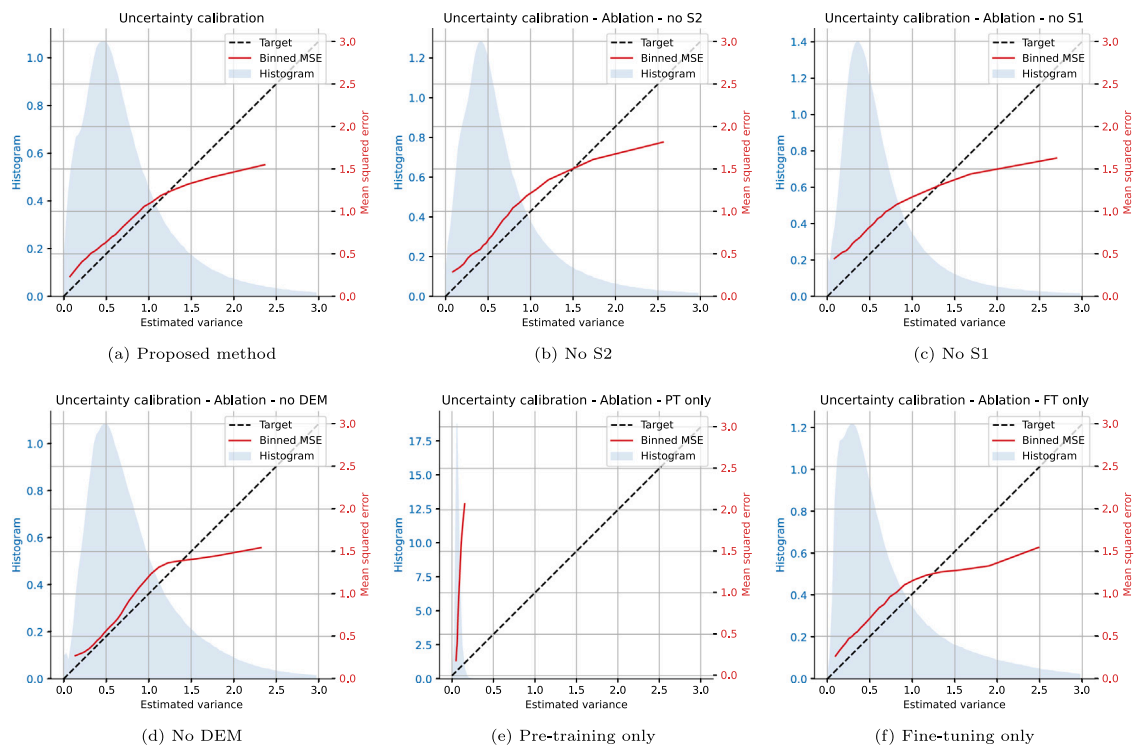


Fig. 16. Calibration curves for all ablation experiments.

snow depth estimation. Furthermore, we were able to produce well-calibrated uncertainty maps associated with the snow depth estimates that can provide useful information for users of such products as well as for downstream applications which make use of such estimates. We showed that all chosen data sources contribute towards more accurate results, and that the two-stage training scheme which combines large-scale reference data with high-fidelity reference data is essential for obtaining the best uncertainty calibration.

We hope that the presented work will not only provide better, more accurate snow depth estimates at a large scale, but also will inspire further research on this topic. A thorough monitoring and understanding

of the cryosphere and how it is changing with the climate is essential for managing water resources and to support a sustainable future.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: This work is funded by the Innosuisse innovation project 42258.1 “IP-ICT”, which aims at improved snow depth modeling in alpine regions. ExoLabs GmbH aims to use the methods developed in this work and to exploit these findings commercially. However, there are no competing financial interests between the authors.

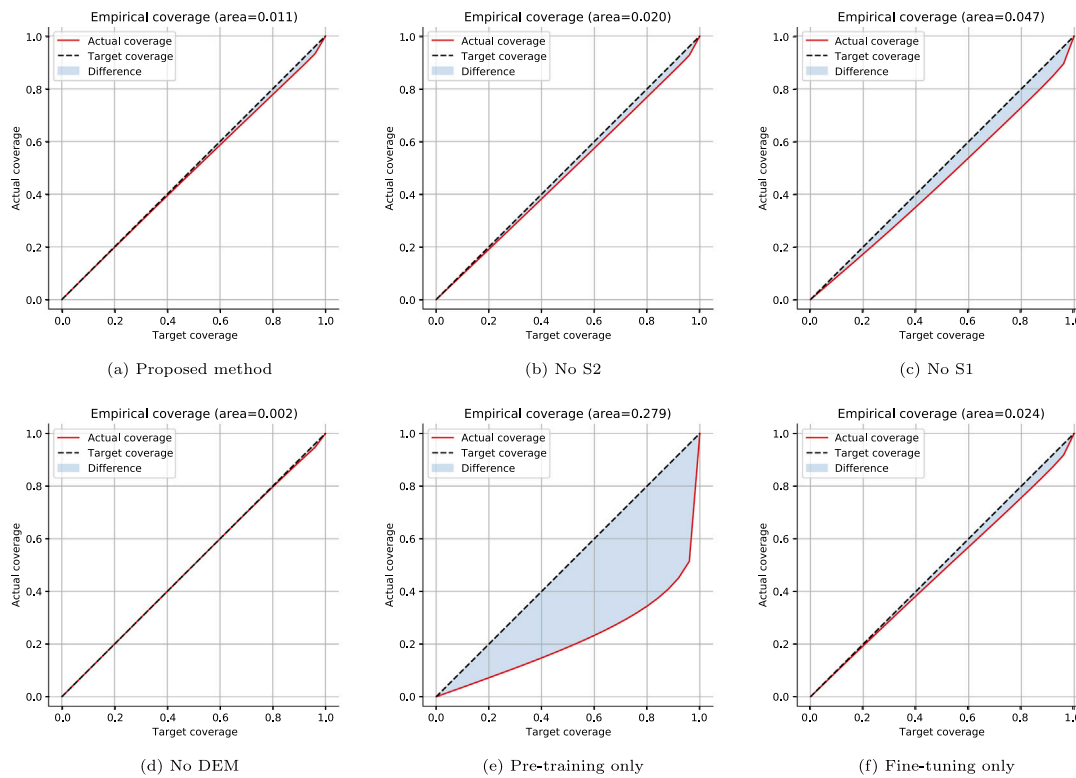


Fig. 17. Coverage curves for all ablation experiments.

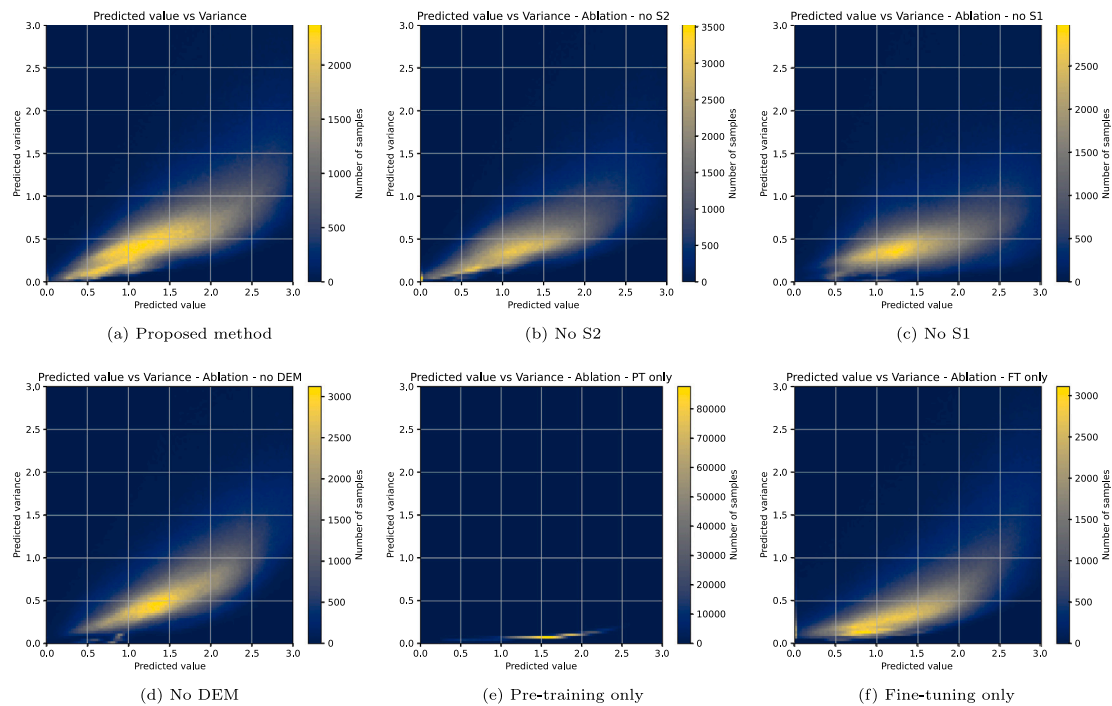


Fig. 18. Estimated snow depth versus uncertainty for all ablation experiments. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Acknowledgments

This work was supported by Innosuisse, Switzerland through the DeepSnow Innovation Project (42258.1 IP-ICT).

Appendix

In this section we include the plots for all the experiments described in Section 5. These include 2D histograms of snow depth estimates versus reference data (Fig. 15), calibration curves (Fig. 16), empirical

coverage curves (Fig. 17), and 2D histograms for estimated snow depth values and uncertainties (Fig. 18).

References

- Adams, M.S., Bühler, Y., Fromm, R., 2018. Multitemporal accuracy and precision assessment of unmanned aerial system photogrammetry for slope-scale snow depth maps in alpine terrain. *Pure Appl. Geophys.* 175 (9), 3303–3324.
- Alagialoglou, L., Manakos, I., Heurich, M., Červenka, J., Delopoulos, A., 2022. A learnable model with calibrated uncertainty quantification for estimating canopy height from spaceborne sequential imagery. *IEEE Trans. Geosci. Remote Sens.* 60, 1–13.
- Ballas, N., Yao, L., Pal, C., Courville, A.C., 2016. Delving deeper into convolutional networks for learning video representations. In: *International Conference on Learning Representations*.
- Becker, A., Russo, S., Puliti, S., Lang, N., Schindler, K., Wegner, J.D., 2021. Country-wide retrieval of forest structure from optical and SAR satellite imagery with Bayesian deep learning. *arXiv preprint arXiv:2111.13154*.
- Bühler, Y., Adams, M.S., Bösch, R., Stoffel, A., 2016. Mapping snow depth in alpine terrain with unmanned aerial systems (UASs): potential and limitations. *Cryosphere* 10 (3), 1075–1088.
- Bühler, Y., Adams, M.S., Stoffel, A., Boesch, R., 2017. Photogrammetric reconstruction of homogenous snow surfaces in alpine terrain applying near-infrared UAS imagery. *Int. J. Remote Sens.* 38 (8–10), 3135–3158.
- Bühler, Y., Bebi, P., Christen, M., Margreth, S., Stoffel, L., Stoffel, A., Marty, C., Schmucki, G., Caviezel, A., Kühne, R., Wohlwend, S., Bartelt, P., 2022. Automated avalanche hazard indication mapping on a statewide scale. *Nat. Hazards Earth Syst. Sci.* 22 (6), 1825–1843.
- Bühler, Y., Marty, M., Egli, L., Veitinger, J., Jonas, T., Thee, P., Ginzler, C., 2015. Snow depth mapping in high-alpine catchments using digital photogrammetry. *Cryosphere* 9 (1), 229–243.
- Bühler, L.J., Marty, M., Eberhard, L.A., Stoffel, A., Hafner, E.D., Bühler, Y., 2022. Spatially continuous snow depth mapping by airplane photogrammetry for annual peak of winter from 2017 to 2021. *Cryosphere Discuss.* 2022, 1–37. <http://dx.doi.org/10.5194/tc-2022-65>, URL <https://tc.copernicus.org/preprints/tc-2022-65/>.
- Cho, K., van Merriënboer, B., Bahdanau, D., Bengio, Y., 2014. On the properties of neural machine translation: Encoder-decoder approaches. In: *Workshop on Syntax, Semantics and Structure in Statistical Translation*. Association for Computational Linguistics, pp. 103–111.
- Eberhard, L.A., Sirguey, P., Miller, A., Marty, M., Schindler, K., Stoffel, A., Bühler, Y., 2021. Intercomparison of photogrammetric platforms for spatially continuous snow depth mapping. *Cryosphere* 15 (1), 69–94.
- Eker, R., Bühler, Y., Schlögl, S., Stoffel, A., Aydın, A., 2019. Monitoring of snow cover ablation using very high spatial resolution remote sensing datasets. *Remote Sens.* 11 (6), 699.
- Garnot, V.S.F., Landrieu, L., 2020. Lightweight temporal self-attention for classifying satellite images time series. In: *International Workshop on Advanced Analytics and Learning on Temporal Data*. Springer, pp. 171–181.
- Garnot, V.S.F., Landrieu, L., 2021. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 4872–4881.
- Garnot, V.S.F., Landrieu, L., Giordano, S., Chehata, N., 2019. Time-space tradeoff in deep learning models for crop classification on satellite multi-spectral image time series. In: *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 6247–6250.
- Garnot, V.S.F., Landrieu, L., Giordano, S., Chehata, N., 2020. Satellite image time series classification with pixel-set encoders and temporal self-attention. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 12325–12334.
- GDAL/OGR contributors, 2022. GDAL/OGR Geospatial Data Abstraction Software Library. Open Source Geospatial Foundation, <http://dx.doi.org/10.5281/zenodo.5884351>, URL <https://gdal.org>.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep Learning*. MIT Press, <http://www.deeplearningbook.org>.
- Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q., 2017. On calibration of modern neural networks. In: *International Conference on Machine Learning*. pp. 1321–1330.
- Gustafsson, F.K., Danelljan, M., Schon, T.B., 2020. Evaluating scalable bayesian deep learning methods for robust computer vision. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. pp. 318–319.
- Hugonnet, R., McNabb, R., Berthier, E., Menounos, B., Nuth, C., Girod, L., Farinotti, D., Huss, M., Dussailant, I., Brun, F., Kääb, A., 2021. Accelerated global glacier mass loss in the early twenty-first century. *Nature* 592 (7856), 726–731.
2022. Information about snow depth – SLF. <https://www.slf.ch/en/avalanche-bulletin-and-snow-situation/snow-maps/information-about-snow-depth.html>, Accessed: 2022-06-30.
- Jacobs, J.M., Hunsaker, A.G., Sullivan, F.B., Palace, M., Burakowski, E.A., Herrick, C., Cho, E., 2021. Snow depth mapping with unpiloted aerial system lidar observations: a case study in Durham, New Hampshire, United States. *Cryosphere* 15 (3), 1485–1500.
- Jonas, T., Marty, C., Magnusson, J., 2009. Estimating the snow water equivalent from snow depth measurements in the Swiss Alps. *J. Hydrol.* 378 (1–2), 161–167.
- Kelly, R.E., Chang, A.T., Tsang, L., Foster, J.L., 2003. A prototype AMSR-E global snow area and snow depth algorithm. *IEEE Trans. Geosci. Remote Sens.* 41 (2), 230–242.
- Kendall, A., Gal, Y., 2017. What uncertainties do we need in Bayesian deep learning for computer vision? *Adv. Neural Inf. Process. Syst.* 30, 1–12.
- Kuleshov, V., Fenner, N., Ermon, S., 2018. Accurate uncertainties for deep learning using calibrated regression. In: *International Conference on Machine Learning*. pp. 2796–2804.
- Kushibar, K., Campello, V.M., Moras, L.G., Linardos, A., Radeva, P., Lekadir, K., 2022. Layer ensembles: A single-pass uncertainty estimation in deep learning for segmentation. *arXiv preprint arXiv:2203.08878*.
- Lakshminarayanan, B., Pritzel, A., Blundell, C., 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. *Adv. Neural Inf. Process. Syst.* 30.
- Lang, N., Jetz, W., Schindler, K., Wegner, J.D., 2022a. A high-resolution canopy height model of the Earth. *arXiv preprint arXiv:2204.08322*.
- Lang, N., Kalischek, N., Armston, J., Schindler, K., Dubayah, R., Wegner, J.D., 2022b. Global canopy height regression and uncertainty estimation from GEDI LIDAR waveforms with deep ensembles. *Remote Sens. Environ.* 268, 112760.
- Lang, N., Schindler, K., Wegner, J.D., 2019. Country-wide high-resolution vegetation height mapping with sentinel-2. *Remote Sens. Environ.* 233, 111347.
- Lettenmaier, D.P., Alsdorf, D., Dozier, J., Huffman, G.J., Pan, M., Wood, E.F., 2015. Inroads of remote sensing into hydrologic science during the WRR era. *Water Resour. Res.* 51 (9), 7309–7342.
- Lievens, H., Demuzere, M., Marshall, H.-P., Reichle, R.H., Brucker, L., Brangers, I., de Rosnay, P., Dumont, M., Giotto, M., Immerzeel, W.W., et al., 2019. Snow depth variability in the Northern Hemisphere mountains observed from space. *Nature Commun.* 10 (1), 1–12.
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* 152, 166–177.
- Margulis, S.A., Fang, Y., Li, D., Lettenmaier, D.P., Andreadis, K., 2019. The utility of infrequent snow depth images for deriving continuous space-time estimates of seasonal snow water equivalent. *Geophys. Res. Lett.* 46 (10), 5331–5340.
- 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., Terzaghi, S., Valt, M., Beozzo, W., Cianfarra, P., Gouttevin, I., Marcolini, G., Notarnicola, C., Pettita, 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. *Cryosphere* 15 (3), 1343–1382.
- Metzger, N., Turkoglu, M.O., D'Aronco, S., Wegner, J.D., Schindler, K., 2021. Crop classification under varying cloud cover with neural ordinary differential equations. *IEEE Trans. Geosci. Remote Sens.* 60, 1–12.
- Muelchi, R., Rössler, O., Schwanbeck, J., Weingartner, R., Martius, O., 2021. River runoff in Switzerland in a changing climate—runoff regime changes and their time of emergence. *Hydrol. Earth Syst. Sci.* 25 (6), 3071–3086.
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choula, M., Harrigan, S., Hersbach, H., et al., 2021. ERA5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* 13 (9), 4349–4383.
- Niculescu-Mizil, A., Caruana, R., 2005. Predicting good probabilities with supervised learning. In: *International Conference on Machine Learning*. pp. 625–632.
- Olefs, M., Koch, R., Schöner, W., Marke, T., 2020. Changes in snow depth, snow cover duration, and potential snowmaking conditions in Austria, 1961–2020—A model based approach. *Atmosphere* 11 (12), <http://dx.doi.org/10.3390/atmos11121330>, URL <https://www.mdpi.com/2073-4433/11/12/1330>.
- Ovadia, Y., Fertig, E., Ren, J., Nado, Z., Sculley, D., Nowozin, S., Dillon, J., Lakshminarayanan, B., Snoek, J., 2019. Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. *Adv. Neural Inf. Process. Syst.* 32.
- Painter, T.H., Berisford, D.F., Boardman, J.W., Bormann, K.J., Deems, J.S., Gehrke, F., Hedrick, A., Joyce, M., Laidlaw, R., Marks, D., et al., 2016. The airborne snow observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sens. Environ.* 184, 139–152.
- Pelletier, C., Valero, S., Inglada, J., Champion, N., Dedieu, G., 2016. Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas. *Remote Sens. Environ.* 187, 156–168.
- Pelletier, C., Webb, G.I., Petitjean, F., 2019a. Deep learning for the classification of sentinel-2 image time series. In: *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 461–464.
- Pelletier, C., Webb, G.I., Petitjean, F., 2019b. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sens.* 11 (5), 523.
- Pérez-Guillén, C., Techel, F., Hendrick, M., Volpi, M., van Herwijnen, A., Olevski, T., Obozinski, G., Pérez-Cruz, F., Schweizer, J., 2021. Data-driven automated predictions of the avalanche danger level for dry-snow conditions in Switzerland. *Nat. Hazards Earth Syst. Sci. Discuss.* 1–38.

- Persello, C., Wegner, J.D., Hansch, R., Tuia, D., Ghamisi, P., Koeva, M., Camps-Valls, G., 2022. Deep learning and earth observation to support the sustainable development goals: Current approaches, open challenges, and future opportunities. *IEEE Geosci. Remote Sens. Mag.*
- Revuelto, J., Billecocq, P., Tuzet, F., Cluzet, B., Lamare, M., Larue, F., Dumont, M., 2020. Random forests as a tool to understand the snow depth distribution and its evolution in mountain areas. *Hydrol. Process.* 34 (26), 5384–5401.
- Riley, S., Degloria, S., Elliot, S., 1999. A terrain ruggedness index that quantifies topographic heterogeneity. *Intermount. J. Sci.* 5, 23–27.
- Rodriguez, A.C., Wegner, J.D., 2018. Counting the uncountable: Deep semantic density estimation from space. In: *German Conference on Pattern Recognition*. Springer, pp. 351–362.
- Rumpf, S.B., Gravey, M., Brönnimann, O., Luoto, M., Cianfrani, C., Mariethoz, G., Guisan, A., 2022. From white to green: Snow cover loss and increased vegetation productivity in the European Alps. *Science* 376 (6597), 1119–1122.
- Rußwurm, M., Ali, M., Zhu, X.X., Gal, Y., Körner, M., 2020. Model and data uncertainty for satellite time series forecasting with deep recurrent models. In: *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 7025–7028.
- Russwurm, M., Körner, M., 2017. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. In: *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. pp. 11–19.
- Rußwurm, M., Körner, M., 2018. Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS Int. J. Geo-Inf.* 7 (4), 129.
- Rußwurm, M., Körner, M., 2020. Self-attention for raw optical satellite time series classification. *ISPRS J. Photogramm. Remote Sens.* 169, 421–435.
- Rußwurm, M., Lefevre, S., Courty, N., Emonet, R., Körner, M., Tavenard, R., 2019. End-to-end learning for early classification of time series. *arXiv preprint arXiv:1901.10681*.
- Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., et al., 2010. The NCEP climate forecast system reanalysis. *Bull. Am. Meteorol. Soc.* 91 (8), 1015–1058.
- Schweizer, J., Bruce Jamieson, J., Schneebeli, M., 2003. Snow avalanche formation. *Rev. Geophys.* 41 (4).
2022. Sentinel-1 SAR technical guide. <https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-1-sar>, Accessed: 2022-06-30.
2022. Sentinel-2 MSI technical guide. <https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi>, Accessed: 2022-06-30.
- Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W.-c., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* 28.
- Shukla, P., Skea, J., Slade, R., Khourdajie, A.A., van Diemen, R., McCollum, D., Pathak, M., Some, S., Vyas, P., Fradera, R., Belkacemi, M., Hasija, A., Lisboa, G., Luz, S., Malley, J. (Eds.), 2022. *IPCC - Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., Frantz, D., Gascon, F., Gómez-Chova, L., Hagolle, O., López-Puigdollers, D., Louis, J., Lubej, M., Mateo-García, G., Osman, J., Peressutti, D., Pflug, B., Puc, J., Richter, R., Roger, J.-C., Scaramuzza, P., Vermote, E., Vesel, N., Zupanc, A., Žust, L., 2022. Cloud mask intercomparison exercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. *Remote Sens. Environ.* 274, 112990.
2022. SNAP - earth online. <https://earth.esa.int/eogateway/tools/snap>, Accessed: 2022-06-30.
- Stahl, K., Weiler, M., Van Tiel, M., Kohn, I., Haensler, A., Freudiger, D., Seibert, J., Moretti, G., Gerlinger, K., 2022. Climate Change Impact on Rain, Snow and Glacier Melt Components of Streamflow for the River Rhine: Synthesis of a Model Experiment and Relevance for Water Use. Tech. rep., Copernicus Meetings.
- Swisstopo, 2018. *swissALTI3D - das hoch aufgelöste terrainmodell der schweiz*.
- Thackeray, C.W., Derksen, C., Fletcher, C.G., Hall, A., 2019. Snow and climate: Feedbacks, drivers, and indices of change. *Curr. Clim. Change Rep.* 5 (4), 322–333.
2022. The 17 goals | sustainable development. <https://sdgs.un.org/goals>, Accessed: 2022-06-30.
- Truckenbrodt, J., Freemantle, T., Williams, C., Jones, T., Small, D., Dubois, C., Thiel, C., Rossi, C., Syriou, A., Giuliani, G., 2019. Towards Sentinel-1 SAR analysis-ready data: A best practices assessment on preparing backscatter data for the cube. *Data* 4 (3), 93.
- Turkoglu, M.O., Becker, A., Gündüz, H.A., Rezaei, M., Bischl, B., Daudt, R.C., D'Aronco, S., Wegner, J.D., Schindler, K., 2022. Film-ensemble: Probabilistic deep learning via feature-wise linear modulation. *arXiv preprint arXiv:2206.00050*.
- Turkoglu, M.O., D'Aronco, S., Perich, G., Liebisch, F., Streit, C., Schindler, K., Wegner, J.D., 2021a. Crop mapping from image time series: Deep learning with multi-scale label hierarchies. *Remote Sens. Environ.* 264, 112603.
- Turkoglu, M.O., D'Aronco, S., Wegner, J., Schindler, K., 2021b. Gating revisited: Deep multi-layer RNNs that can be trained. *IEEE Trans. Pattern Anal. Mach. Intell.*
- Weiss, A., 2001. Topographic position and landforms analysis. In: *ESRI User Conference*, Vol. 200. p. 1.
- Wilson, M.F., O'Connell, B., Brown, C., Guinan, J.C., Grehan, A.J., 2007. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Mar. Geod.* 30 (1–2), 3–35.
- Winkler, M., Schellander, H., Gruber, S., 2021. Snow water equivalents exclusively from snow depths and their temporal changes: the Δ snow model. *Hydrol. Earth Syst. Sci.* 25 (3), 1165–1187. <http://dx.doi.org/10.5194/hess-25-1165-2021>, URL <https://hess.copernicus.org/articles/25/1165/2021/>.
- Wipf, S., Stoeckli, V., Bebi, P., 2009. Winter climate change in alpine tundra: plant responses to changes in snow depth and snowmelt timing. *Clim. Change* 94 (1), 105–121.
- Wulf, H., Sassik, B., Milani, G., Leiterer, R., 2020. High-resolution snow depth monitoring for entire mountain ranges. *Swiss Conf. Data Sci.* 1–4.
- Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* 5 (4), 8–36.