https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

SnowGalileo: A Pre-trained Earth Observation Transformer for Daily, 100 m Fractional Snow Cover Mapping

Marlena Reil[1, 2,★], Julia Kaltenborn[1, 2,★], Donovan J. M. Allum[4], Francis Pelletier[1], Sebastian Roessler[3, 10], Samip Shrestha[3], Zhibang Lv[4], John Truckenbrodt[3], Gabriele Schwaizer[8], Thomas Nagler[8], John W. Pomeroy[4], Christopher B. Marsh[7], Benoit Montpetit[5], Tobias Jonas[6], Gabriel Tseng[9, 1, 2], David Rolnick[1, 2], Andreas J. Dietz[3], and Celia A. Baumhoer[3]

[1]Mila - Quebec AI Institute, Montreal, Canada
[2]School of Computer Science, McGill University, Montreal, Canada
[3]German Remote Sensing Data Center, Earth Observation Center, German Aerospace Center (DLR), Weßling, Germany
[4]Centre for Hydrology, University of Saskatchewan, Canmore, Alberta
[5]Environmental Research and Modelling Directorate, Environment and Climate Change Canada, Gatineau, Canada
[6]WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland
[7]Climate Processes Section, Climate Research Division, Environment and Climate Change Canada, Saskatoon, Canada
[8]ENVEO, Innsbruck, Austria
[9]Allen Institute for AI, Seattle, USA
[10]Bavarian Environment Agency
[★]These authors contributed equally to this work.

Correspondence: Marlena Reil (marlena.reil@mail.mcgill.ca) and Julia Kaltenborn (julia.kaltenborn@mila.quebec)

Abstract. Mountain snow is an important component of the cryosphere that directly affects downstream livelihoods. Accurate monitoring of mountain snow is crucial, yet remains challenging due to complex topography and frequent cloud cover. Although combining data from multiple Earth Observation (EO) satellites can improve spatial and temporal coverage, extracting Fractional Snow Cover (FSC) from sensors with different spatial and temporal resolutions remains difficult. AI-based Earth foundation models can process diverse sensor inputs and have been successfully applied to various remote sensing tasks; however, they often lack snow-specific design considerations. This paper presents SnowGalileo, a pre-trained transformer model designed to integrate EO data to map snow-covered areas. We evaluated SnowGalileo's potential for generating daily, gap-free FSC maps at 100 m resolution in mountainous regions. SnowGalileo combines one week of satellite time-series data from multiple sensors, including Sentinel-1, Sentinel-2, Landsat, Sentinel-3, MODIS, and VIIRS, with topographic, land cover, and meteorological information to predict FSC for a given day. The model was pre-trained using masked autoencoding and fine-tuned using labeled data from various mountain ranges across the Northern Hemisphere. In addition to clear-sky conditions, SnowGalileo is evaluated under a wider variety of conditions than was previously possible, including cloud cover and lack of high-resolution ($\approx$ 10–30 m) satellite imagery. For the Canadian Rockies and the Swiss Alps, respectively, SnowGalileo achieves RMSEs of 0.099 and 0.124 on clear days, 0.144 and 0.209 on cloudy days, 0.185 and 0.262 on days without high-resolution imagery, and 0.200 and 0.299 on cloudy days without high-resolution imagery. SnowGalileo also consistently outperforms random forests, support vector regressors, and multi-layer perceptrons. While the current product is a proof of concept that has undergone limited validation across geographic regions, operates on 1 km × 1 km tiles rather than full maps, and

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

has restricted capabilities in challenging conditions, this approach could ultimately enable the continuous, gap-free operational generation of FSC time series for mountain regions worldwide.

1 Introduction

Mountain snow is a critical component of the cryosphere, affecting communities and the climate system in various aspects. During the winter months, more than 45% of the global mountainous area is covered in snow (Gascoin et al., 2024). The seasonal snowpack in mountainous regions accumulates large quantities of frozen water. Combined with glacial melt, spring snowmelt is a primary source of freshwater, supplying approximately 2 billion people with water for domestic, agricultural, and industrial use (Immerzeel et al., 2020). Beyond its hydrological role, snow acts as a thermal insulator and habitat buffer, thereby determining the distribution and diversity of flora and fauna in mountain ecosystems (Jones, 2001; Huang et al., 2023; Niittynen and Luoto, 2018). For local and indigenous communities, snow is linked to cultural practices, values, and local knowledge (Riseth et al., 2011; Gorman-Murray, 2010; Niemi and Ahlstedt, 2012). Its role in the albedo feedback loop, its insulating properties, and its impact on the global water cycle make snow a highly sensitive indicator of global climate variability (Warren, 1982; Zhang, 2005; Barnett et al., 2005; Brown and Mote, 2009). As a result, changes in snow cover extent induce considerable impacts at both local and global scale (Johnson et al., 2025; UNESCO World Water Assessment Programme, 2025).

Over the past several decades, climate change has notably altered snow distribution, with impacts already being seen downstream (IPCC, 2022). The impacts of these changes on river discharge, surface water, soil moisture, vegetation and fire hazards are noticeable even beyond the snow covered season. Snow cover trends over the Northern Hemisphere for the 1981–2018 period are negative for all months of the year (Mudryk et al., 2020), with trends exceeding $-50 \cdot 10^3\ \mathrm{km}^2 \cdot \mathrm{yr}^{-1}$ indicating that mountain regions are facing a massive snow loss. Climate models from the Coupled Model Intercomparison Project Phase 6 show a linear relationship between spring snow extent and changes in global surface air temperature, demonstrating their dependency. To reflect the impact of a rapidly changing snow cover in hydrological models accurately, hydrologists need advanced snow monitoring products. Accurate snow monitoring, especially at high resolutions, is thus key to improving our decision-making in preparation for future water availability challenges.

The Global Climate Observing System (GCOS) named snow cover one of the most important climate variables to be monitored by satellite remote sensing (Bertoncini and Pomeroy, 2024; WMO, 2022). Snow cover is highly variable in space and time and often occurs in remote, hard-to-access terrain. Remote sensing is thus an ideal tool to overcome these challenges, which is why the snow community has been active in using current sensors to estimate snow cover parameters (Pomeroy et al., 2026; Nolin, 2010), even though snow is the only component of the water cycle without a dedicated Earth Observation mission (Derksen et al., 2019). These works have also shown the constraints of remote sensing for snow monitoring: Cloud cover hinders direct observation relying on multispectral sensors, SAR sensors are missing out on dry snowpacks, and the availability of high spatial-resolution remote sensing sensors is limited. Consequently, new methodological approaches are needed to address the current limits of remote-sensing-based snow monitoring.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

A key variable in snow monitoring is the fraction of ground covered in snow, called Fractional Snow Cover (FSC). FSC is important to hydrology because it indexes the redistribution of snow by wind, vegetation, and gravity (Pomeroy and Gray, 1995) and also the effect of spatially inhomogeneous ablation on the areal snowpack (Pomeroy et al., 1998). Knowing the FSC permits estimation of the snowmelt contributing area for runoff generation during the melt period - an essential state variable for hydrological models (DeBeer and Pomeroy, 2010). Estimating FSC in mountainous terrain is particularly challenging due to the effects of rugged topography and the implicated mixed pixel effects, persistent cloud cover, shadow effects, and limited ground truth for validation (Gascoin et al., 2024). Multiple algorithms and methods have been developed to derive FSC mainly from multispectral sensors such as Landsat (Missions and Tools), MODIS (Painter et al., 2009), or AVHRR (Zhao et al., 2024). One remote sensing-based dataset is the Snow CCI (Climate Change Initiative) by the European Space Agency (ESA). It offers daily global snow cover fraction products from MODIS for the 2000-2020 period (Nagler et al., 2022). Although these products are extremely valuable, they still present challenges with spatial and temporal coverage for downstream applications, including climate change monitoring, especially in mountainous areas (Sun et al., 2025). Further successful approaches are based on physically based models (e.g. MODSCAG) (Painter et al., 2009), topographic corrections (Sirguey et al., 2009), multi-sensor fusion (Berman et al., 2018), and robust uncertainty quantification (Salminen et al., 2018). However, for the application in hydrological models, high resolution (HR) is desired in both the spatial (minimum pixel resolution of 100 m) and temporal (daily) domains (WMO, 2022), a requirement that cannot be met by existing remote sensing missions due to the lack of available high-resolution observations. Thus, there is a need for novel, sophisticated methods that can accurately predict FSC at the required 100 m resolution.

Meanwhile, the AI community is developing new ways to combine heterogeneous, multi-sensor Earth observation data through so-called Earth Foundation Models (EFMs). EFMs use advanced deep learning techniques such as self-attention mechanisms and self-supervised pre-training to learn a general representation of the Earth from large-scale data. These can subsequently be adapted to specific downstream applications through fine-tuning. EFMs have been applied very successfully to improve a wide range of downstream tasks, but their potential for cryospheric applications, and specifically snow mapping, remains largely unexplored (see Sect. 2.5). This paper adapts and re-trains Galileo (Tseng et al., 2025), a well-established EFM, for snow tasks and evaluates it on FSC mapping.

This paper introduces SnowGalileo, a pre-trained snow-specific transformer model designed for daily, gapless FSC mapping at 100 m resolution. SnowGalileo integrates multi-sensor satellite data and auxiliary information from the previous week, including Sentinel-1, Sentinel-2, Sentinel-3, Landsat, MODIS, VIIRS, ERA5-derived meteorological variables, land cover, NDSI, NDVI, topography, and location information. By including observations across time, spatial context, and complementary sensors, SnowGalileo is designed to overcome challenging conditions such as cloud cover, forest cover, and limited HR[1] satellite availability.

SnowGalileo is evaluated as a method towards creating daily, gapless FSC maps at 100 m spatial resolution. To produce these maps, the model must perform well under cloudy conditions, with varying vegetation cover, and under different satellite availabilities, e.g. due to low revisit times of HR sensors. Performance is evaluated under all these conditions. Furthermore,

[1]Here, the term "high-resolution" (HR) refers to a spatial resolution between 10–30 m, i.e. the resolution of Landsat and Sentinel-1 and -2 sensors.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

SnowGalileo's capability is assessed across two different geographical regions and seasons, as Machine Learning (ML) models, in contrast to more traditional statistical approaches, are at higher risk of not generalizing well in such scenarios. SnowGalileo is also benchmarked against a range of simpler ML approaches used by prior work. SnowGalileo consistently outperforms every single model across all conditions and experiments, providing a promising new avenue for daily FSC mapping at higher resolutions.

The main contributions are as follows:

1. A methodology for daily, gapless, multi-sensor FSC mapping at 100 m resolution.
2. A systematic comparison of different methods for the FSC mapping task, conducted under a wide range of conditions in mountain regions such as the Canadian Rockies and the Swiss Alps. The comparison takes into account factors such as cloud cover, limited availability of HR imagery, vegetation, and seasons.
3. An open-sourced, pre-trained transformer model fine-tuned to the task of FSC mapping in mountain regions.

The paper starts with an introduction of relevant concepts (Sect. 2); followed by an explanation of the various data sources (Sect. 3) used to train and validate the models described in Sect. 4. The results of SnowGalileo and other models across a range of experiments are presented in Sect. 5 and discussed in Sect. 6.

2 Background

The studies presented here are all evaluated on FSC mapping, a task in which each pixel iss assigned a percentile representing the amount of snow present in that pixel. A range of methods exists to address this task, ranging from field surveys to remote sensing-based systems and novel ML approaches.

2.1 Earth Observation for FSC mapping

Before satellite remote sensing, snow cover was commonly observed using field surveys and low-altitude aerial photography. These methods provided useful local information, but they were costly, time-consuming, spatially limited, and often affected by oblique viewing geometry, making them unsuitable for repeated regional-scale monitoring (Barnes and Bowley, 1974; Rango and Itten, 1976).

Earth observation satellites transformed snow mapping by enabling repeated, large-area observations at arguably much lower cost and with greater consistency. Among them, Landsat-derived snow cover maps have been used for decades as inputs to, and validation data for, snow and hydrological models (Thompson, 1975; Rango and Martinec, 1979; Margulis et al., 2016; Lv and Pomeroy, 2019; Marsh et al., 2024). Its 30 m spatial resolution allows Landsat to capture patchy snow patterns more effectively than coarse-resolution sensors, making it a common higher-resolution reference for calibrating and validating FSC products derived from MODIS and VIIRS (Salomonson and Appel, 2004, 2006; Masson et al., 2018; Rittger et al., 2021a). However, its relatively infrequent revisit cycle and sensitivity to cloud cover and terrain effects limit its ability to monitor rapidly changing snow conditions, particularly in mountainous environments.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

MODIS substantially improved the temporal dimension of snow monitoring by providing near-daily global observations, making it highly valuable for tracking seasonal snow dynamics (e.g. Huang et al., 2017; Liu et al., 2021) and for a wide range of hydrological applications (e.g. Nagler et al., 2008; Largeron et al., 2020). It also provides operational FSC products developed using empirical relationships between the Normalized Difference Snow Index (NDSI) and subpixel snow fraction, calibrated with Landsat reference data (Salomonson and Appel, 2004, 2006; Hall et al., 2010). Despite this advantage, the coarse spatial resolution of MODIS introduces pronounced mixed-pixel effects in mountains, forests, and fragmented snow environments, reducing FSC accuracy under complex surface conditions (Rittger et al., 2021a; Stillinger et al., 2023).

In summary, optical snow mapping remains constrained by cloud cover, terrain shadow, changing illumination, and forest canopy interference, all of which contribute substantial uncertainty in mountainous and boreal regions (Dietz et al., 2012; Wang et al., 2018; Gascoin et al., 2024). Most importantly, these observation systems illustrate a persistent trade-off in snow monitoring: sensors with finer spatial resolution often lack sufficient temporal frequency, whereas sensors with frequent coverage often provide insufficient spatial detail and relatively lower accuracy.

2.2 Traditional FSC retrieval

Upon receiving satellite data, snow information must be extracted in order to be operationally useful. Early satellite snow mapping relied largely on visual interpretation, manual delineation, and simple multispectral thresholds. Such approaches helped establish the operational value of satellite snow mapping, but they were labor-intensive, subjective, and primarily suited to binary snow detection rather than FSC estimation (Wiesnet and McGinnis, 1974; Thompson, 1975; Rango and Itten, 1976; Rango et al., 1977). As interest shifted toward more realistic representations of heterogeneous snow conditions, FSC emerged as a more informative metric because it quantifies the proportion of snow within a pixel rather than classifying it simply as snow or snow-free.

Most FSC retrieval methods are based on the spectral properties of snow recorded on satellite data. The most widely used approach is the NDSI, which exploits snow's high reflectance in visible wavelengths and lower reflectance in shortwave infrared wavelengths. Initially developed for binary snow detection, NDSI was later extended to empirical FSC estimation and became especially important in MODIS-based operational products (Dozier, 1989; Hall et al., 1995; Hall and Riggs, 2007; Salomonson and Appel, 2004, 2006). A second major approach is spectral mixture analysis, which estimates snow fraction by unmixing the spectral contributions of snow and other surface components within a pixel. This approach is generally better suited to patchy and heterogeneous snow conditions and has been progressively refined from early hyperspectral studies to more advanced multispectral applications (Nolin et al., 1993; Painter et al., 2009). However, both NDSI-based and mixture-based methods remain sensitive to canopy cover, terrain shading, snow property variation, cloud contamination, and georegistration uncertainty, particularly in complex terrain and forested regions (Dietz et al., 2012; Rittger et al., 2013; Lv and Pomeroy, 2019; Bertoncini et al., 2022; Gascoin et al., 2024).

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

2.3 Machine Learning for FSC Retrieval

To address these limitations, recent studies have increasingly adopted machine learning approaches. Here, pixel-based approaches such as random forests, support vector regressors, and artificial neural networks (ANNs) are widely applied in the field (e.g. Kuter, 2021; Liu et al., 2020; Hou et al., 2020; Yang et al., 2023). These methods can integrate spectral, temporal, terrain, and ancillary information and are therefore more flexible than fixed-threshold or purely reflectance-based methods. Other studies have employed methods based on convolutional neural networks (CNNs) (e.g. John et al., 2022; Cannistra et al., 2021; Hou et al., 2026). Although machine learning methods often improve performance in complex environments, their success depends strongly on the quality and representativeness of the training data, and their transferability across regions and conditions remains an important concern.

2.4 Data Fusion for FSC Retrieval

Recent studies have adopted multi-sensor approaches to overcome the limitations of individual systems. Examples include Rittger et al. (2021b), who trained a two-stage random forest model on combined MODIS and Landsat data to map FSC daily and at 30 m spatial resolution. Berman et al. (2018) used dynamic time warping to reorder historic Landsat data informed by MODIS NDSI. Furthermore, Zhang et al. (2026b) combined Sentinel-1 and Sentinel-2 data to improve FSC retrieval under cloud cover, using a combination of a GAN and a temporal ConvNeXt. Others have downscaled MODIS data using Sentinel-2, e.g. by implementing a probabilistic method (Revuelto et al., 2021) or a random forest-based approach (Kollert et al., 2024; Richiardi et al., 2023).

These approaches improve both spatial detail and temporal continuity and represent an important step toward more operationally useful FSC products. Nevertheless, the integration of a large variety of sensors across spatial and temporal resolutions is largely unexplored and uncertainties remain related to cross-sensor differences, cloud contamination, and forest cover (Masson et al., 2018; Gascoin et al., 2024). Exceptions include most recent multi-sensor approaches such as Zhang et al. (2026a) and Wu et al. (2026), that used multi-stage and traditional data fusion methods to combine a larger variety of sensors, including MODIS, VIIRS, Sentinel-2, and Landsat data.

2.5 Earth Foundation Models

Earth Foundation Models (EFMs) are AI models pre-trained on vast Earth observation datasets that can be adapted to a wide variety of downstream tasks. In other words, EFMs use advanced deep learning techniques to integrate remote sensing data for mapping and monitoring tasks, improving the ability to combine different data sources, as well as the generalizability of previous methods.

At the design level, EFMs learn to encode images into a unified representation space that can be used to create outputs, such as classifications or regressions. EFMs are often based on vision transformer (ViT) models (Dosovitskiy et al., 2020). ViTs divide images into patches, which are linearly projected into “tokens“. Positional embeddings are added to the tokens to preserve information about the original image location of each token. During encoding, ViTs use self-attention mechanisms

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

(Vaswani et al., 2017) to model long-range dependencies in the input data. The resulting representation can be combined with a prediction head to produce outputs. Unlike CNNs, which process data using local operations over fixed image grids, ViTs operate on sequence data and can model long-range spatial and temporal context (Wang et al., 2024). This makes them particularly promising for multi-modal remote sensing mapping, as it enables the integration of heterogeneous input sources at different resolutions.

One disadvantage of ViT models is that they need to be trained on sufficiently large datasets (Dosovitskiy et al., 2020), but remote sensing labels are both rare and costly to obtain (Hauser et al., 2025). Therefore, the learning process of EFMs typically involves two stages. Firstly, during *pre-training*, large amounts of unlabeled remote sensing data are used with self-supervised learning techniques, in which labels are derived from the data itself (Wang et al., 2022). These data are often sampled from various locations around the world, improving transferability across regions and conditions. This also enables the same pre-trained model to be reused for multiple downstream tasks, thereby reducing the overall training requirements and making the model accessible to a range of downstream users. The model is then, in a second step, adjusted to the specific downstream task in a process called *fine-tuning*, where smaller amounts of labeled examples and supervised learning techniques are used to give domain-specific knowledge to the model. For a comprehensive overview of EFMs and their pre-training procedures for learning from unlabeled data, see Xiao et al. (2025).

EFMs have been successfully deployed for tasks such as land cover mapping, crop type monitoring, and flood mapping (Szwarcman et al., 2025). However, their deployment in the cryospheric domain remains sparse, with applications being limited to tasks such as glacier mapping and sea ice mapping or calving front extraction (e.g. Kaushik et al., 2025; Jiang et al., 2025; Kaushik et al., 2026), but not specific to snow. This is in part because appropriate training and evaluation datasets are unavailable (Kaushik et al., 2026), but also because snow-specific considerations are lacking in their design process. To account for the high spatial and temporal variability of snow, a highly flexible method is required that can integrate sensors with different spatial and temporal resolutions. However, EFMs that simultaneously provide multi-sensor, multi-resolution, and spatio-temporal anchoring are rare (Zhu et al., 2026). Previous work has largely focused on processing only high-resolution imagery with long revisit times, such as Sentinel-1 and Sentinel-2 (e.g. Cong et al., 2022; Fuller et al., 2023; Bastani et al., 2023; Jakubik et al., 2025). Although others can process multi-modal data, they do not incorporate time-based information (e.g. Noman et al., 2024; Xiong et al., 2024). Since snow cover mapping relies on a time series of satellite images to account for missing observations or frequent cloud cover (Dietz et al., 2015), an EFM that can incorporate time series at a daily scale is required.

The Galileo EFM model (Tseng et al., 2025) extends previous developments. It can simultaneously integrate images and time series (although not on a daily basis) and various remote sensing modalities, making it highly flexible for downstream adjustments. It should be noted that AnySat (Astruc et al., 2025) can provide similar flexibility to Galileo and should be regarded as comparable.

SnowGalileo builds on Galileo's architecture, while making snow-specific adaptations. As SnowGalileo has been fine-tuned and evaluated specifically for FSC mapping, it no longer fits the definition of an EFM, which should serve a variety of downstream tasks. Consequently, SnowGalileo is referred to as a “pre-trained transformer“ while still building on EFM principles.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

3 Data

Developing and evaluating a pre-trained transformer for snow requires multiple datasets. Pre-training the model requires a large amount of unlabeled, multi-modal data relevant to snow covered area detection. Fine-tuning requires input data similar to that used in the pre-training phase, but this time paired with labeled FSC scenes – the "targets" or "ground truth". During pre-training and fine-tuning, the input data consists of remote sensing observations and derived products that can identify snow at high spatial or temporal resolutions or provide additional information about topography, land cover and weather. Rather than being pre-trained on arbitrary Earth Observation (EO) data, the pre-training data used by SnowGalileo comes from locations that are specifically relevant to snow cover mapping. "Sampling" was conducted in mountainous regions with seasonal snow cover. Furthermore, some of the input data used for fine-tuning is obscured by the addition of generated clouds, since "ground truth" can only be accessed on cloud-free days. SnowGalileo receives samples from 1×1 km ground areas. In the following, to avoid confusion between Earth observation and machine learning terminology, the entire 1×1 km image that constitutes a data point is referred to as a "tile". The term "patch" is used to refer to a segment of an image processed by the model (i.e., the parts that are projected onto tokens; see Sect. 2.5 and Sect. 4.1.1).

All data sources and processing steps are described below.

3.1 Input Data: Earth Observation Data

3.1.1 Input Data Sources: Earth Observation and Reanalysis Products

A multitude of remote sensing sensors, derived EO products, and reanalysis data carry snow-relevant information, also relevant to FSC. The sensors used are Sentinel-1, Sentinel-2, Landsat, Sentinel-3, MODIS, and VIIRS. The derived products include topographical variables, land cover, NDSI, and the normalized difference vegetation index (NDVI), as well as meteorological variables from ERA5, a reanalysis product. A full overview is provided in Table 1.

Direct Earth Observation data at the Level-1 and Level-2 processing stage constitute the largest part of SnowGalileo's input data. This is because snowpacks act as reflectance and backscatter surfaces for most remote sensors, enabling snow information to be retrieved from them. MODIS and VIIRS sensors are crucial for snow monitoring due to their daily revisit time, which is needed to capture the quickly changing characteristics of snow. However, with a nominal spatial resolution of 500 m, they are more suitable for monitoring low-relief snowpacks (e.g. in the Canadian prairies) than for mountain snowpacks. Sentinel-2 and Landsat sensors, with spatial resolutions of 10–30 m but lower revisit times (5 days for Sentinel-2 and 8 days for Landsat at the equator), are therefore used to complement MODIS and VIIRS with high-resolution data, overcoming these challenges (Gascoin et al., 2019). Additionally, Sentinel-3 offers important advantages for snow monitoring with its high spectral resolution, 300 m spatial resolution, and daily global coverage. From its optical sensor (OLCI) and thermal radiometer (SLSTR), it is possible to estimate snow albedo, snow grain size, and FSC using a spectral unmixing technique (Tanhapour et al., 2026). C-band synthetic aperture radar (SAR) imagery from the Sentinel-1 sensor is also employed, even though it is not as commonly used for monitoring snow conditions due to its limited sensitivity to dry snow (Duguay and Bernier, 2012). However, recent studies show a link between snow depth and SAR backscatter for snowpacks exceeding 1 m thickness, which is

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

often the case in mountainous regions (Lemos and Riihelä, 2024). Furthermore, unlike other sensors, SAR data is independent of lighting conditions and cloud cover, the latter of which is especially frequent and persistent in mountain regions. Sentinel-1 thus provides complementary, weekly (6-day revisit time at the equator), 10 m, cloud-independent information on deep snowpacks. Combining all the above-mentioned sensors can provide snow-related information with high spatial and temporal resolution.

The input data is supplemented with several products derived from EO data to provide more direct access to snow variables. These include NDSI and NDVI, which compare how much light is reflected versus absorbed by snow or vegetation. The NDSI provides more direct access to snow information. For example, MODIS, VIIRS, Sentinel-2, and Landsat can all derive FSC from this index (Rittger et al., 2021a). The NDVI provides information about vegetation, which affects the reflectance of snowpacks. For similar reasons, landcover information is also included (ESA Worldcover (Zanaga et al., 2022)), since the spectral characteristics of snow vary by land cover type (Moody et al., 2007; Hall et al., 1995), and the reflectance of snow-free areas depends on surface type. Furthermore, topographic information from the Copernicus DEM GLO-30 Ecosystem (2022) is included, such as slope, elevation, and aspect. These variables are particularly relevant in mountainous regions, as past studies have shown strong dependencies between snow conditions and topography. In fact, it has been demonstrated that snow depth can be predicted from topographic features alone (Meloche et al., 2022). The contextual information about topography and vegetation, as well as the more direct data about snow, provides the necessary information to understand how snow characteristics are affected by their environmental context.

As snow is a meteorological phenomenon, meteorological data is also included. Precipitation, temperature, and wind are crucial variables that govern snow accumulation, melt, and redistribution processes (Marsh et al., 2024; Vionnet et al., 2025). While a high spatial resolution of these variables is important, the meteorological dataset also needs to be globally available. ERA5 C3S (2019) is the most suitable product for this purpose, as it is a daily, global, 11 km reanalysis product.

With revisit times ranging from daily to 8-day windows, and spatial resolutions ranging from 10 m to above 11 km, the assembled multi-modal dataset requires fusion on three different axes: 1) The different input types need to be harmonized (sensors and their channels, derived products, and reanalysis products); 2) they need to be spatially harmonized (multiple resolutions, different spatial availabilities); and 3) they need to be temporally harmonized (different temporal resolutions). To achieve this successfully, additional metadata is added to the processed data during modeling. This includes channel identifiers, the month of acquisition, the relative position within the eight-day time series, and the relative position within the image tile (see Sect. 4.1.1). Rather than being provided as direct input data, this information is encoded as embeddings and added to the tokenized representations, enabling the model to distinguish between different sensors, spectral channels, acquisition times, and spatial positions.

3.1.2 Input Data Processing

SnowGalileo's input data for training, evaluation, and inference is structured as an 8-day time series, containing 1×1 km tiles from each input source for each day that the source is available, at a given location. To obtain this data in an appropriate format

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Table 1. SnowGalileo's input data. The table lists all input data ingested by SnowGalileo, describing their data category, the related dataset or product, their spatial resolution in meters, as well as their temporal resolution and all bands or variables used. GRD stands for ground range detected, TOA stands for top-of-atmosphere, and SR stands for surface reflectance. The spatial and temporal resolutions listed here refer to the Google Earth Engine-derived data sources used, which are listed in Appendix A1.1. The temporal resolution (revisit time) of satellite sensors is reported at the equator. In terms of Landsat, the reported revisit time is the revisit time of the combined Landsat 8 and Landsat 9 observations.

Input data category	Dataset / Product	Spatial resolution [m]	Temporal resolution	Bands / Variables
Satellite sensor	Sentinel-1 GRD σ^0 backscatter	10	6 days	VV, VH, incidence angle
	Sentinel-2 TOA	10	5 days	B2, B3, B4, B8
		20		B11, B12
	Landsat 8 & 9 OLI TOA Collection 2	30	8 days (combined)	B2-B7
	Sentinel-3 OLCI TOA	300	daily	Oa17, Oa21
	MODIS Terra SR	500	daily	B1-B7
	VIIRS SR	500	daily	I1, I3
		1000	daily	M5, M7, M10, M11
Sensor-derived indeces	NDSI	500	daily	
	NDVI	500	daily	
Topography	Copernicus DEM	30	static	elevation, slope, aspect
Landcover	ESA Worldcover	10	static (state 2021)	
Meteorological variables	ERA5	11132	daily	skin temperature, temperature 2m, total precipitation sum, u component of wind, v component of wind,
Location	latitude, longitude		static	

280 for ML modeling, multiple processing steps are necessary, ranging from reprojections, handling missing data, handling special variables (precipitation), and normalizing the data. Further processing details are reported in Appendix A1.3.

All input data is acquired through Google Earth Engine. The individual data sources are first reprojected to a common coordinate system and resampled to a spatial resolution of 10 m. This enables consistent data export in the form of single GeoTIFF files per data point, despite the variety of native projections and spatial resolutions in the data sources. To reduce the

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

computational requirements for subsequent model processing, the lower-resolution products are then downsampled to a grid close to their native resolution using the mean: 5×5 for Sentinel-3; 2×2 for MODIS, VIIRS-500 m, NDSI, and NDVI; and 1×1 for VIIRS-1 km and ERA5. The grid size is retained to be divisible by the image height and width, which facilitates tokenization (see Appendix B1). All pre-training data are obtained in WGS84 following Tseng et al. (2025), whereas fine-tuning data are provided in UTM projection to match the coordinate system of the Landsat target images (see Sect. 3.2.2).

Data gaps occur in the time series for sensors providing samples less than once a day. Therefore, validity masks are created to enable the ML algorithms to handle the NaN values accordingly.

As the target data were exclusively acquired on precipitation-free days (see Sect. 3.3.2), the presence of precipitation during inference would be an out-of-distribution case and likely difficult to handle for the ML model. To partially mitigate this, all meteorological information, including precipitation, is masked out on the prediction day during fine-tuning. This way, the model does not rely on precipitation-free ERA5 data and can instead infer FSC from other sources.

Input normalization is an important pre-processing step in ML to handle data with different scales and facilitate model training. Following Tseng et al. (2025), all input data are normalized using 2σ-clipped min-max normalization, with per-channel statistics computed over the pre-training dataset. For the ESA land cover data, one-hot encoding is used instead to prepare the data for ML while preserving the categorical nature of the values. Similarly, NDSI and NDVI are not normalized as they already have a fixed bound of -1 to 1, and neither are the Cartesian coordinates, following Tseng et al. (2025). Normalization statistics for all other channels are reported in Appendix A1.4.

3.1.3 Input Data Adaptations: Simulating Challenging Conditions during Fine-Tuning

One of the main challenges of mapping snow cover in mountainous regions is the frequent presence of clouds, alongside limited high-resolution imagery and forest cover. While FSC mapping on clear days has been successfully addressed (e.g. Salomonson and Appel, 2004; Gascoin et al., 2020), the same task under cloudy conditions proves to be more difficult. However, target data is typically only obtained on cloud-free days, resulting in a lack of appropriate training and test images (Gascoin et al., 2024). Similarly, retrieving high-resolution FSC maps from (cloud-free) HR satellites has already been addressed, however, retrieving high-resolution FSC maps from low-resolution imagery is more challenging. Thus, to make the models robust and assess their performance under such challenging conditions, additional adapted input data sets are provided for training the models in different setups, as explained below. In terms of forest cover, the data sets remain unmodified, as labels for forested areas are available. See Sect. 3.2.2 for a description of how FSC labels were obtained in forested terrain, and see Sect. 6.2 for a discussion on the reliability of these labels.

Cloud Simulation

To make the models robust to clouds and assess their performance under cloudy conditions, clouds were simulated using an open-source generator tool (Czerkawski et al., 2023). In the training set used for fine-tuning, clouds were generated in the input data for the prediction day with a probability of 0.8. The probability of clouds in any other image in the time series was set to 0.5, since these images are more likely to already contain clouds. Various cloud types were randomly sampled (see Appendix

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

A1.2). In the cloud test sets, clouds were generated in the input data for the prediction day with a probability 1.0, while all other images in the time series remain cloud-free. For the test sets, a manual configuration was used to simulate higher occlusion in the images (also reported in Appendix A1.2). Clouds were generated on Sentinel-2, Sentinel-3, Landsat, MODIS, and VIIRS imagery, as these sensors are sensitive to clouds. Cloud data was generated at the highest resolution (10 m) and subsequently downsampled for sensors with lower resolutions. The NDSI and NDVI inputs were removed for any image on which clouds had been generated to avoid inconsistencies. The synthetic cloud data provides input images of cloudy days, enabling the methodology to be evaluated under such conditions.

HR Imagery Removal

To train the models to predict high-resolution FSC maps even when only low-resolution imagery is available on the prediction day, HR imagery were artificially removed from the last time step. Since FSC labels were obtained from Landsat imagery (see Sect. 3.2.1), the fine-tuning training and test data typically contain an HR image on day 8 of the time series, the day on which the FSC map is predicted. By removing Landsat and Sentinel-2 data on day 8, the models need to rely on low-resolution imagery and, if available, past HR imagery instead. This procedure ensures that the ML models receive data, which, in theory, enables them to perform HR-FSC mapping on a daily basis.

3.2 Target Data: Fractional Snow Cover Labels

3.2.1 Target Data Sources: Landsat Imagery

The fine-tuning dataset was derived from FSC products generated from Landsat 8 and Landsat 9 imagery. Source scenes were adopted from a prior validation study (Dietz and Roessler, 2025), which inherently constrained their temporal and spatial distribution; within this framework, scenes were preselected to minimize cloud contamination and avoid complete snow cover, thereby excluding peak winter conditions. The training dataset was sampled from 42 scenes acquired between 2020 and 2023 across spring, summer, and autumn, covering multiple mountain regions worldwide and spanning diverse landcover conditions. Model evaluation was performed on two geographically distinct test sets from the Canadian Rockies (2022, summer and autumn) and from the Swiss Alps (2020–2021, spring and autumn), which were sampled from 8 scenes per test region.

3.2.2 Target Data Processing

To generate FSC labels from these scenes, multiple thresholding approaches were dynamically combined. FSC retrieval from Landsat imagery followed a temperature- and shadow-thresholding approach developed by the ESA Satellite Snow Product Intercomparison and Evaluation Exercise (SnowPEx) (Ripper et al., 2015), complemented by an index-based thresholding scheme adapted from MODIS applications consistent with previous studies (Hu et al., 2019; Koehler et al., 2022b, a). This procedure also includes a dynamic NDSI threshold adjustment based on NDVI to account for lower NDSI values in forested areas (Poon and Valeo, 2006). Cloud-contaminated pixels were identified using the FMASK algorithm (Zhu et al., 2015). The resulting per-pixel classification included FSC values, while ancillary classes (“clouds”, “water/shadow”, and “no data”) were

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

consolidated into a single no-data category. All samples were exported in their original UTM projection derived from each Landsat tile. The final target dataset needed for fine-tuning is thus cloud-free.

3.3 Sampling

The pre-training data was sampled from the input data sources described above (Sect. 3.1). To fine-tune SnowGalileo on the FSC target data (Sect. 3.2), the input and target datasets were sampled together. The sampling processes – i.e. which data are selected from the available data sources and why – are described below.

3.3.1 Pre-Training Sampling

SnowGalileo was pre-trained on mountain regions worldwide, encompassing a broad range of snow conditions across elevations. Several filtering criteria were used to select the mountain ranges on which SnowGalileo was pre-trained. First, mountain regions were derived from the Global Mountain Biodiversity Assessment Inventory (Snethlage et al., 2022) and filtered to include only large and prominent mountain systems. Secondly, as this study focuses on snowpack data, GlobalSnowpack (Dietz and Roessler, 2025) was used to select only regions with an average snow cover duration of at least 50 days per year. Thirdly, to focus on major mountain systems, an additional minimum area threshold of 35,000 km^2 was applied. And fourthly, to focus on latitudes where the Canadian Rockies and Swiss Alps are located, the sampling points were clipped to latitudes between 42° and 60°. These filtering criteria yielded a set of 48 mountain ranges, also depicted in Fig. 1. The number of pre-training samples drawn from those regions is proportional to the total area of each mountain range. In terms of elevation, however, a balanced dataset is desired to ensure that the model gets equal access to different snow patterns (e.g. high-elevation snowpacks persist longer throughout the season, and might show different melt-refreeze cycles than low-elevation snowpacks). Thus, three elevation bins were created (0–1000 m, 1000–2000 m, and >2000 m), containing roughly the same number of location points. During sampling, random points are drawn from the filtered mountain ranges, representing them proportionally across different elevations.

All pre-training samples were acquired between late 2021 and early 2024, starting with the availability of Landsat 9. For each point, a random week within the snow onset period of the Northern Hemisphere was sampled (from 01. October to 15. December), high winter (from 16. December to 28. February), and the snowmelt period (from 01. March to 30. June). As a result, SnowGalileo was pretrained across three seasons, namely snow onset, high winter, and snowmelt.

In total, 152,559 pre-training samples were obtained via Google Earth Engine (Gorelick et al., 2017). Please refer to Appendix A3.1 and Appendix A1.1 for further details.

3.3.2 Fine-Tuning Sampling

SnowGalileo was fine-tuned (and all other ML baselines were trained directly) on a set of FSC targets from mountains in the Northern Hemisphere. Across the 58 Landsat scenes providing the target data (see Sect. 3.2.1), initially 500,000 random locations for each scene were iteratively sampled to create the training and test samples. A 1×1 km tile was extracted for each

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

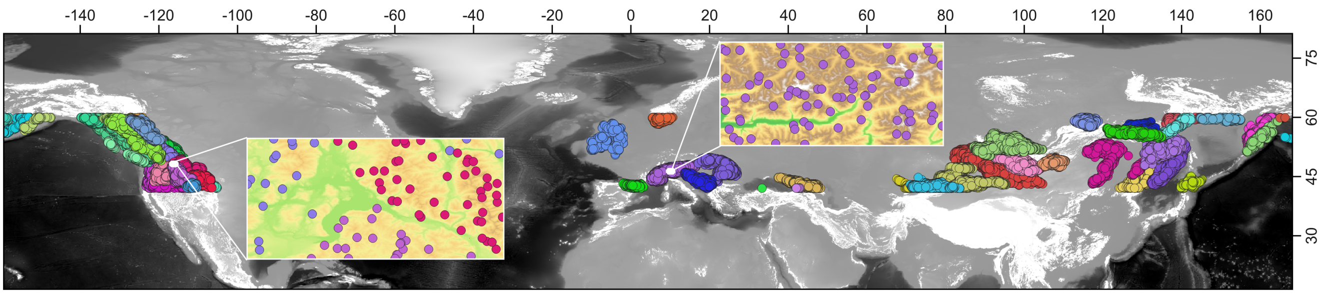

Figure 1. SnowGalileo's pre-training data was sampled from various mountain ranges of the Northern Hemisphere. The figure shows the spatial distribution of the pre-training samples over the 48 largest global mountain ranges. Each mountain range is shown with a different color. The two zoom panels show the random distribution of points sampled over the Canadian Rockies (left panel) and the Swiss Alps (right panel). Global mountain ranges outside the pre-training data are highlighted in white. All pre-training samples were retrieved between 2021 and 2024.

location, centered on the location's coordinates. Tiles were selected only if no cloud cover and glaciated area were present, to ensure the reliability of the resulting FSC targets. Glacierized areas were excluded using the Randolph Glacier Inventory (RGI) 7.0 inventory for the Rockies (Consortium, 2023) and Sentinel-2-based glacier boundaries for Switzerland (Paul et al., 2020). The training and test data were separated by ensuring that all test samples came from the Swiss Alps or the Rocky Mountains, and that these samples did not contain any data from the mountain ranges used during training. For this reason, test data for the Swiss Alps was additionally cropped to the country borders to avoid data leakage from the training to test data, as the training data from nearby countries could otherwise overlap with test scene boundaries. The training data is thus globally distributed, whereas the test data was sampled from the Swiss Alps and the Canadian Rockies, as depicted in Fig. 2.

In general, FSC values are not uniformly distributed: 0% and 100% FSC are far more frequent than patchy snow areas. As patchy snow areas are the more challenging prediction task, it needs to be ensured that the training data includes a sufficient amount of patchy snow images. Thus, tile-level FSC values were stratified into discrete bins (0%, 1–10%, 11–20%, . . . , 100%), with a maximum of 100 samples per bin to ensure a more uniform distribution across snow cover conditions. Sampling continued until all bins were filled or the maximum number of random points was reached. This way, the strongly bi-modal FSC target distribution became a more uniform one.

All fine-tuning samples were acquired between 2020 and 2023, spanning spring, summer, and autumn, to cover the onset and offset of the snow season, exhibiting the more challenging to predict, patchy snow cover. The test data in the Rocky Mountains was acquired for June, August, and September 2022, and the test data in the Swiss Alps was acquired for September to April (inclusively) 2020–2021.

In total, 29,436 fine-tuning samples, 793 test samples from the Canadian Rockies, and 894 test samples from the Swiss Alps were obtained and paired with the corresponding 8-day time-series EO data. This data was used to fine-tune SnowGalileo and train the baselines.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

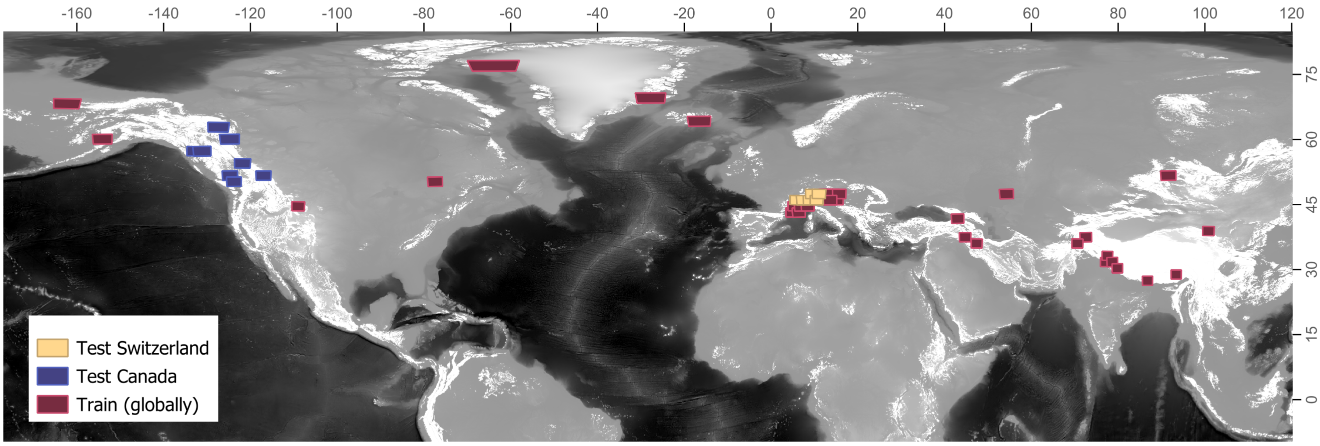

Figure 2. The fine-tuning data used to train all models was sampled from various mountain ranges in the Northern Hemisphere. All FSC labels (and their corresponding input data) for training were retrieved from multiple mountain regions in the Northern Hemisphere. Glaciated areas were excluded. The test data was limited to the Swiss Alps and the Canadian Rockies. All fine-tuning samples were retrieved between 2020 and 2023.

4 Methodology

To address FSC mapping in mountainous regions, the EFM Galileo is adapted for this task. The following section introduces the methods behind Galileo and its extensions of Galileo to SnowGalileo. SnowGalileo is pre-trained and fine-tuned, and subsequently evaluated across different test setups and against diverse ML baselines. The following section also defines data set splits, metrics, experiments, and software and hardware requirements.

4.1 Model

4.1.1 Galileo

Galileo (Tseng et al., 2025) is the EFM on which SnowGalileo's developments rely. At a high level, Galileo is a pre-trained transformer model that processes multi-modal data cubes to create feature representations. A single input data cube consists of remote sensing products acquired over a ground area of variable input sizes, stacked as a time series of monthly images. Galileo learns a representation of the data cube's features by being pre-trained on a large set of those data cubes. These representations can then be used for a wide variety of prediction tasks, by fine-tuning Galileo on such tasks.

Galileo's data processing for modeling comprises data tokenization and encoding. For data tokenization, input modalities are separated into one of four modality groups according to their spatiotemporal resolution. These modality groups are:

- **space-time-varying**: data with high spatial resolution that vary over time (e.g. Sentinel-1 and Sentinel-2),
- **space-varying**: data with high spatial resolution that remain constant over time (e.g. topography and land cover),

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

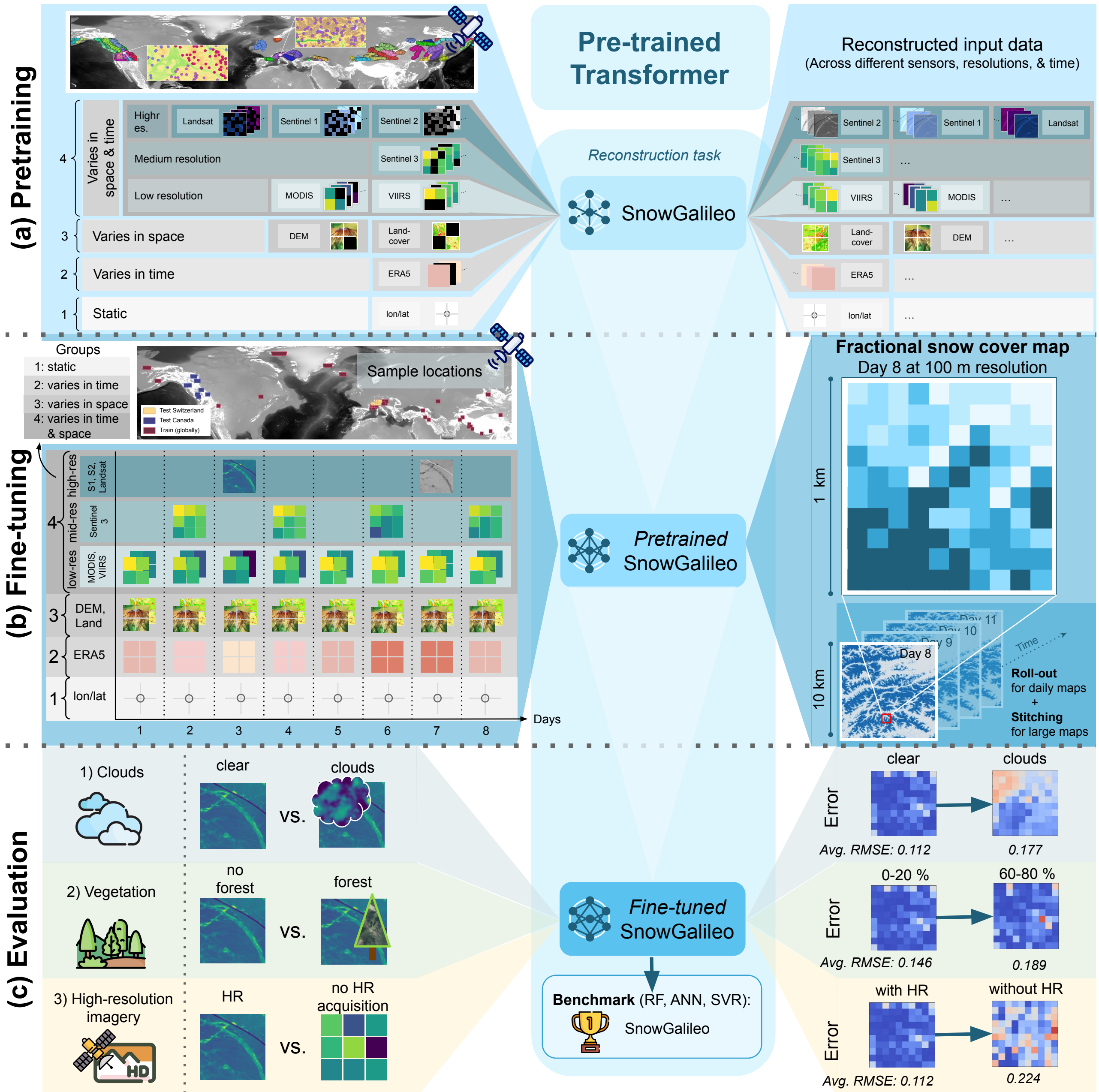

Figure 3. Visual abstract of SnowGalileo, demonstrating its pre-training, fine-tuning, and evaluation. (a) SnowGalileo is a transformer pre-trained on multi-modal time-series data from snow-covered mountain ranges across the Northern Hemisphere (NH). The model acquires snow-relevant knowledge by reconstructing removed data parts (black segments) via a self-supervised machine learning task (masked auto-encoding). **(b)** SnowGalileo is fine-tuned for fractional snow cover (FSC) mapping by being trained in a supervised manner on FSC labels across the NH and tested on the Canadian Rockies and Swiss Alps. From a time series of 8 days with various Earth observation inputs, SnowGalileo learns to predict a 100 m resolution FSC map for the 8th day. The outputs can be stitched to provide daily, gapless, 100 m resolution FSC maps. **(c)** The performance of SnowGalileo is compared against several other ML models under different setups,

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Figure 3. including cloud and forest cover, and limited availability of HR images. At the bottom right, RMSE performance under different challenging conditions is illustrated.

- **time-varying**: data with low spatial resolution (at least 1 km) that vary over time (e.g. ERA5),
- **static**: data that are constant (i.e. geographic coordinates).

As described in Sect. 2.5, standard ViT architectures tokenize images into smaller image segments. However, when making predictions from multi-modal and multi-temporal remote sensing data, the structure of the input data for the transformer model is more complex (i.e. containing multiple time steps and channels). To this end, Galileo subdivides the data of each modality group not only into spatial segments (if the data has a spatial resolution higher than the 1×1 km tile), but also into individual channel groups and time steps (if the data varies over time). A *channel group* is a set of spectrally similar channels from a specific sensor (e.g. Sentinel-2 B2, B3, and B4 are grouped into the Sentinel-2 RGB channel group). In the case of variables from derived products, it is a set of similar variables. After tokenization, contextual information about the token's location in the original image and position within the time series is added to each token using sinusoidal positional encodings (Vaswani et al., 2017). The model also receives learnable information about the channel group and sinusoidal positional encodings for the acquisition month. Consequently, each resulting token encodes a specific spatial region, time point, sensor, and spectral region. Finally, during encoding, Galileo applies self-attention (Vaswani et al., 2017) to embed the tokenized data into a global representation space.

In terms of pre-training, Galileo receives globally sampled input data, comprising Sentinel-2, Sentinel-1, and various auxiliary modalities. Galileo employs a dual learning strategy that enables the model to capture features at various scales and adapt to a variety of downstream tasks. As this work focuses on a uniform setting (snow in mountains), the pre-training process is simplified to masked autoencoding in this study (He et al., 2022), where parts of the input data are obscured and the model must reconstruct them based on the visible parts (see Appendix B). Further information on Galileo can be found in the original publication (Tseng et al., 2025).

4.1.2 SnowGalileo

SnowGalileo is an adaptation of the Galileo framework for FSC mapping. When considered for fine-tuning on an FSC mapping task, several limitations of Galileo become apparent, as it has not been designed for snow tasks. SnowGalileo (1) extends the range of input modalities and spatial resolutions that can be processed to include more sensors relevant to snow monitoring; (2) adapts the time series component to cover a week in daily intervals instead of a year in monthly intervals, to address the high temporal variability of snow; (3) adapts the pre-training dataset to focus on mountainous regions; and (4) is fine-tuned for snow cover mapping. We detail these proposed adjustments of Galileo towards snow tasks in the following.

Extending the range of input sources: SnowGalileo extends the range of Galileo's input modality groups to enable the processing of sensors at lower resolutions, such as MODIS, VIIRS, and Sentinel-3, which also carry relevant characteristics for the discrimination of snow/non-snow. The sensors are divided into a *space-time-varying medium resolution* modality

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

group to process the Sentinel-3 sensor with medium spatial resolution as well as a *space-time-varying low resolution* group to process data with a spatial resolution of 500 m but high temporal resolution (MODIS, VIIRS, NDSI, and NDVI). The space-time-varying group of Galileo is further referred to as *space-time-varying high-resolution* for clarity. A complete table of the modality grouping of the input data can be found in Appendix B1, along with a table of SnowGalileo's channel groups.

Handling daily timesteps: Due to the high variability of snow conditions over time, the model needs to consider data at high temporal granularity. However, Galileo operates on monthly timesteps over the course of an entire year. The input data considered in this study are daily time series acquired over 8 days. For space-time varying sensors with a revisit time lower than daily (i. .e. Sentinel-1, Sentinel-2, and Landsat), timesteps for which there is no valid data acquisition per channel are masked out. These will not be considered by the model further.

Pre-training specific to mountain snow: A snow-specific pre-training process is introduced by tailoring a mountain-specific dataset as described in Sect. 3.3.1. Masked autoencoding is used as the pre-training strategy (He et al., 2022). After pre-training, SnowGalileo has learned to map observational sources onto a global representation space that captures semantics about snow-covered mountains. This space can then be used to make predictions for specific downstream tasks, such as FSC mapping, through fine-tuning and subsequent additions. Further details are described in Appendix B3.

Fine-tuning: In the fine-tuning stage, SnowGalileo is adjusted to predict FSC at 100 m pixel resolution. This is achieved by grouping all embeddings per output pixel and mapping these to a single value between 0 (0% FSC) and 1 (100% FSC) using a simple linear regression and a sigmoid layer. The final model is fine-tuned using the training dataset introduced in Sect. 3.2. Further details are described in Appendix B2 and B4.

4.2 Evaluation

Here, evaluation refers to the evaluation of the models after fine-tuning, or, in the case of the baselines, after training. Standard ML evaluation procedures are followed by dividing the available data into separate training and testing sets. The training set is used for fine-tuning, and the test sets are used to quantify model performance. To provide context for this performance, it is compared with that of simpler machine learning models, or baseline models, which have been trained and tested on the same fine-tuning datasets and thus provide a performance benchmark. All benchmark results are reported as average (and standard deviation) across three model runs, since ML models are stochastic during training. To ensure that each model is getting a fair chance at competing, hyperparameter tuning is conducted, i.e. each model is run across a set of internal parameters, and the best one is chosen. Details of hyperparameter tuning are not discussed in this section, but are reported in Appendix D.

4.2.1 Data Splits

To assess how well the model generalizes to unseen regions, the fine-tuning data is split into geographically independent training and test sets. For testing purposes only, FSC input and labels from the Swiss Alps and the Canadian Rockies are withheld. These test sets provide an independent means of assessing the model's generalization, as they were not seen during training. The training set (the remaining FSC inputs and labels) is further split at random into an 80/20 training/validation set.

The validation set is used to monitor performance during training and observe model convergence. This yields 23,548 samples for training and 5,888 samples for validation.

As pre-training follows a self-supervised learning strategy, train/test splitting is not required during the pre-training stage.

4.2.2 Baseline Models

To benchmark SnowGalileo's performance on FSC mapping, it is compared with two sets of baseline methods. Firstly, the performance of the pre-trained SnowGalileo model is compared with a version of SnowGalileo that was not pre-trained, but was instead trained directly on the FSC mapping task, starting with randomly initialized weights. Secondly, it is compared with ML methods commonly used for FSC mapping in the literature (e.g. Kuter, 2021; Rittger et al., 2021b; Hou et al., 2020; Liu et al., 2020). These encompass random forests (RFs), support vector regressors (SVRs), and a multi-layer perceptron (MLP) regressor. RFs (Breiman, 2001) are ensembles of decision trees, where each tree processes different random parts of the data to make a prediction, and all predictions are aggregated to make a final choice. SVRs (Drucker et al., 1996; Smola and Schölkopf, 2004; Vapnik, 2013) are ML algorithms that look for a function to maximize the margin between predictions and labels. MLPs (Rumelhart et al., 1986; Kingma and Ba, 2014) are simple feedforward artificial neural networks comprising full connected layers and non-linear activation functions. None of the baseline models can be pre-trained, i.e. all results are reported only on the fine-tuning data.

To enable a fair comparison across the different models, the baseline models should use the same training (fine-tuning) data sets as SnowGalileo. To make the training data applicable to the RF, SVR, and MLP models, multiple processing steps are necessary. For the randomly initialized SnowGalileo, the input data pipeline remains unmodified.

Pixel processing: RFs, SVRs and MLPs all operate on a per-pixel basis. This means that they predict FSC independently for every pixel, with each pixel being represented by a number of features. To this end, all input data must first be represented on a common spatial grid. Therefore, pixel-wise samples are created from the input tiles by dividing them into 100 m grids spaced equally (i.e. the output grid of the FSC mapping method). Input data at a resolution higher than 100 m (e.g. Sentinel-2) is aggregated using average pooling to reach this grid, while input data at a lower resolution (e.g. MODIS) is repeated. For each resulting pixel, all time series data and channels belonging to that pixel are concatenated. Thus, the final input data consists of per-pixel samples, each represented by a concatenated feature vector containing the aggregated input sensors and time series information for the respective pixel.

Temporal information: To incorporate temporal information into the baseline models, a variable is added for each input feature indicating the number of days until the prediction day. For instance, a data sample acquired two days ago would have a value of two.

Missing data imputation: A forward-filling/aggregation imputation strategy is used to handle missing input data due to low revisit times. Details are described in Appendix C.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

4.2.3 Metrics

Evaluation metrics quantify the performance of a model. Root mean squared error (RMSE) is the main metric employed here, since it has been used extensively in related work (e.g. Rittger et al., 2021b; Gascoin et al., 2020; Stillinger et al., 2023). RMSE is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}, \tag{1}$$

where N is the number of samples, y_i the unit-less FSC label and $\hat{y}_i$ is the prediction of the ML model for the FSC target value.

The results are relatively consistent when using different regression metrics (R2, mean absolute error, and median absolute error). These are reported in detail in Appendix E.

4.3 Experiments

In addition to benchmarking SnowGalileo's performance using several ML models, it is analyzed how transferable SnowGalileo is and how well it performs under challenging conditions. Its transferability is tested across different geographical regions and different seasons, as explained below. It is further analyzed to what degree it can maintain performance under forest cover, cloud cover, and limited availability of high-resolution satellite imagery. Only if SnowGalileo can perform well under these challenging, yet realistic conditions, it can be used for skillful FSC mapping on a daily basis.

4.3.1 Geographical Regions

To understand whether SnowGalileo can be used across different (hold-out) regions, it is tested not only on one region but on two. This enables the comparison of performance across two mountain regions and the assessment of the extent to which regional performance differences need to be expected when deployed globally. As pointed out in the data-related sections, model performance is compared between the Canadian Rockies and the Swiss Alps.

4.3.2 Seasons

Similarly, SnowGalileo's performance across different seasons needs to be understood, as do its seasonal limitations. Thus, model performance is quantified according to the season, considering "snow onset" (1 October–15 December), "high winter" (16 December–28 February), "snowmelt" (1 March–30 June) and "summer" (1 September–30 December) seasons. This analysis is performed on the Swiss Alps and Canadian Rockies test data, however, the data obtained from these two regions is almost disjunct in terms of seasonality, which makes it difficult to disentangle seasonal and geographic effects.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

4.3.3 Forest Cover

As forest cover presents a major challenge for FSC mapping products, it is important to understand how SnowGalileo performs in areas with varying degrees of forest cover. To this end, the performance scores of tiles with different levels of forest cover is compared. The amount of forest cover in a given tile is determined by calculating its fractional forest cover (FFC). Here, FFC is defined as the ratio of pixels in a 1×1 km tile that are labeled as forest in the ESA Worldcover dataset (Zanaga et al., 2022). The shrubland class is excluded, as in Deschamps-Berger et al. (2025). However, this approach to evaluating the performance of FSC mapping in forested areas remains limited to visible forest cover; snow cover under the forest canopy cannot be inferred in this study (see Sect. 3.2).

4.3.4 Clouds and Limited HR Coverage

One of the main limitations of current FSC mapping is cloud cover. SnowGalileo aims to address this issue, and so its performance under cloud cover is extensively evaluated by adding diverse simulated clouds to the input sources. This allows the ML models to be trained and tested in cloudy conditions – an experimental setup that has largely been unexplored in previous literature. Further details on how the clouds are generated and added to the imagery can be found in Sect. 3.1.3.

Furthermore, since HR coverage is not available on a daily basis, it is important to assess model performance when no HR observations are available on the prediction day. To simulate this scenario, Sentinel-2 and Landsat imagery are excluded from the prediction-day inputs (see Sect. 3.1.3).

Accordingly, four input data setups are defined. “Clear-HR“ corresponds to the default input configuration under clear-sky conditions without HR imagery removal. “Cloudy-HR“ adds synthetic clouds to the prediction-day imagery. “Clear–no HR“ removes prediction-day HR imagery while retaining clear-sky conditions. Finally, “cloudy–no HR“ combines synthetic cloud cover with the removal of prediction-day HR imagery.

4.4 Implementation and Computational Requirements

SnowGalileo is based on the ViT-Tiny encoder and comprises approximately 5.9 million model parameters in total. Initial experiments showed that the larger ViT-Base encoder, which contains more model parameters, did not outperform the smaller SnowGalileo model while being more computationally expensive.

Hardware: Each pre-training run was executed on an A100 GPU with 92 GB RAM and took about 2 weeks to complete with checkpointing. Each fine-tuning run was executed on a GPU with a minimum of 48 GB GPU memory, 48GB RAM, and took about 4 days to complete with checkpointing. The inference execution time is about 0.06 seconds per 1×1 km sample.

Software: All code requires Python. Export of input data for pre-training, fine-tuning and inference requires an account with Google Earth Engine (Gorelick et al., 2017). SnowGalileo was implemented using PyTorch (Paszke et al., 2019), and all baseline models were implemented using scikit-learn (Pedregosa et al., 2011).

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

5 Results

First, the ability of SnowGalileo to produce daily, gapless FSC maps at 100 m resolution is presented. These results are compared to three different baselines, including a multi-layer perceptron, support vector machine, and random forest, to assess SnowGalileos' performance against established methods. Further, the capacity of SnowGalileo across different seasons and regions is investigated to evaluate its generalizability. Lastly, results from three experiments are presented which show how SnowGalileo outperforms other models under relevant conditions, namely cloud cover, vegetation cover, and limited high-resolution satellite coverage. SnowGalileo provides the methodology for producing large-scale FSC over mountain regions at a high level of spatial resolution and robustness across different conditions.

5.1 FSC Mapping with SnowGalileo

To demonstrate SnowGalileo's ability to predict FSC, its RMSE distribution is presented alongside a visual qualitative representation of the predicted FSC tiles within that distribution in Fig. 4. The figure shows the performance of SnowGalileo trained on the input data setup "cloudy-HR" (see Sect. 4.3.4). This training setup has been chosen because it yields the most robust model across all input setups (see Sect. 5.2.2). The "cloudy-HR" SnowGalileo is then tested across all possible test setups ("clear-HR", "cloudy-HR", "clear-no HR", and "cloudy-no HR"), and Fig. 4 shows the combined results of SnowGalileo tested on all test samples in each available setup. The results show that most of SnowGalileo's test samples fall within a low-RMSE range (mean of 0.112 with 0.003 std), comparable to the best-performing results reported in previous studies (Stillinger et al., 2023). Notably, prior work achieved similar RMSE ranges only for clear conditions, whereas SnowGalileo maintains this performance across all test conditions. In terms of performance distribution, there is a peak at 0.0 and a second peak around 0.15, close to the mean performance. The distribution rapidly decays to zero to the right of the second peak. It is suspected that the small spike at an RMSE of 1 may be due to tiles with full cloud coverage. Fig. 4 qualitatively shows how FSC maps look under the different reported RMSE scores. RMSE values around 0.1 visually match the target FSC maps very well, with quality degrading slowly above an RMSE of 0.15. In comparison, under more difficult input data setups than clear-HR, the performance of current state-of-the-art (SOTA) models degrades more rapidly than that of SnowGalileo (see Table 2 and Appendix E1.2): E.g., for the Rockies, the RF performance degrades by an average of 0.118 RMSE points compared to the clear-HR setup, whereas SnowGalileo's performance only degrades by 0.077 RMSE points, maintaining performance consistently <= 0.2.

Under the clear setup with HR coverage, SnowGalileo achieves an RMSE below 0.1 (see Table 2). SnowGalileo is a promising FSC mapper, showing improved FSC skill in mountain regions under various conditions, compared with SOTA models.

Fig. 5 demonstrates SnowGalileo's ability to produce daily, gapless, high-resolution FSC maps in mountain regions. SnowGalileo's inference module can be used to produce daily roll-outs, i.e. all available data of the past seven days (and the current day) is used to predict an FSC tile, effectively performing a sensor fusion task to produce FSC predictions. This means that SnowGalileo is not simply interpolating between two known maps, but instead relies on satellite and meteorological information from the last week to predict FSC on any given day. Fig. 5 A) shows that even on days without high-resolution imagery,

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

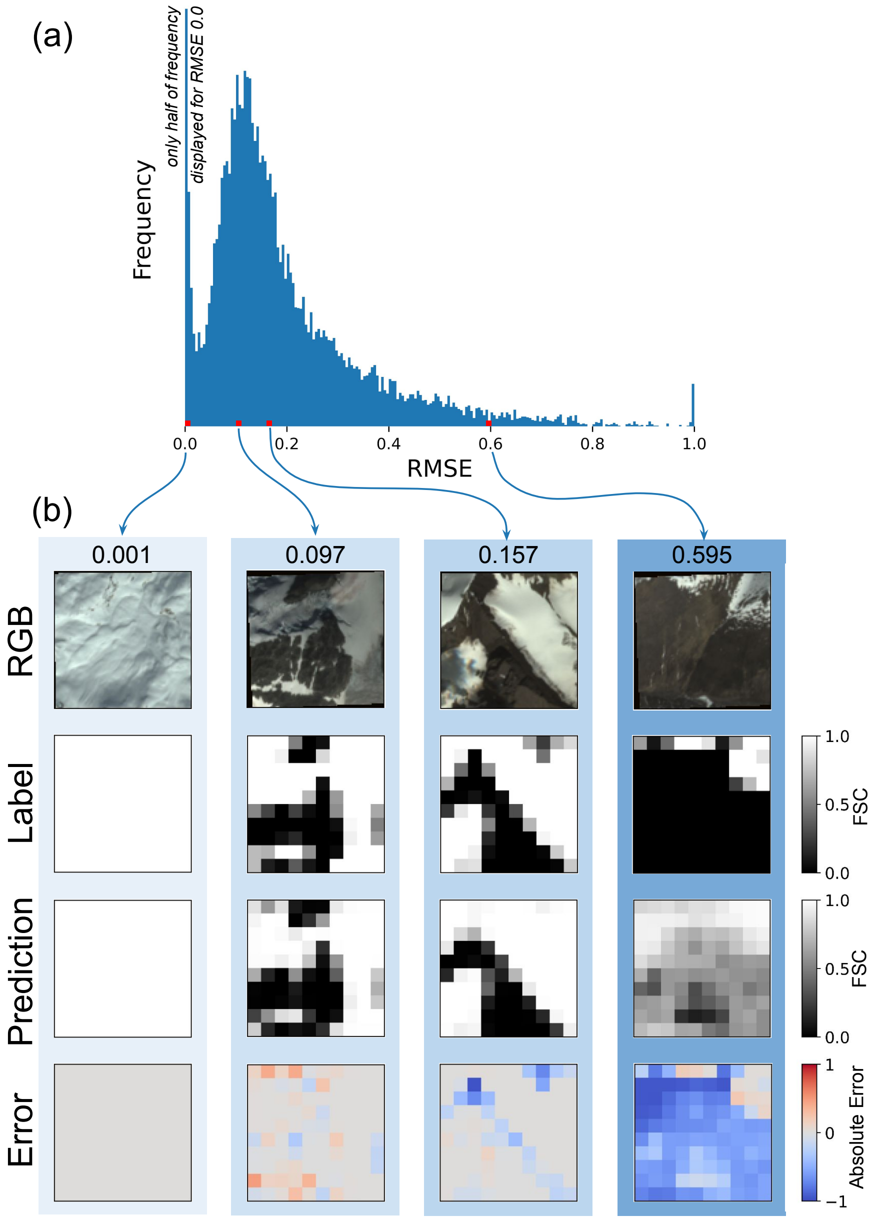

Figure 4. SnowGalileo's ability to predict Fractional Snow Cover (FSC). (a) The upper plot shows SnowGalileo's (cloudy-HR version) RMSE distribution over the test set, demonstrating its overall ability to predict FSC. Included are results for each test sample across all input data setups, as predicted by the model that has been trained on the cloudy-HR input setup across three seeds. **(b)** The lower plot shows FSC samples (1×1 km, at 100 m resolution) drawn from different parts of the RMSE distribution to visualize the RMSE capacities. The first row shows the RGB channels (Sentinel-2) of each sample, the second row shows the target FSC map annotated with the SnowPEx algorithm from Landsat 9 imagery, the third row shows the predictions of SnowGalileo during inference, and the fourth row shows the difference between the predictions and the targets (the error).

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Table 2. SnowGalileo consistently outperforms all baseline models in terms of RMSE. RMSE is reported as the mean (± sample standard deviation) of the overall RMSE per run, averaged across three seeds. All values show model performance on a test set with equivalent input data setups like the training set (i.e. the clear-HR results report test scores of the clear-HR model on clear-HR test data). All values are reported separately for the Canadian Rockies and the Swiss Alps. Clear-HR refers to clear sky conditions with high-resolution imagery available on the day of prediction. Cloudy-HR refers to the same condition but with clouds present on the day of prediction. Clear-no HR refers to clear conditions without HR imagery available, and cloudy-no HR refers to cloudy conditions without HR imagery available. The pre-trained SnowGalileo model consistently outperforms all models (best values are **embolded**), and the randomly initialized SnowGalileo comes in second across all conditions (second-best values are underlined).

Region	**Model**	**RMSE**			
		Clear-HR	Cloudy-HR	Clear-no HR	Cloudy-no HR
Canadian Rockies	RF	0.149 (std <0.001)	0.220 (std <0.001)	0.286 (std <0.001)	0.295 (std <0.001)
	MLP	0.164 (±0.002)	0.238 (±0.006)	0.288 (±0.004)	0.268 (±0.008)
	SVR	0.477 (±0.003)	0.481 (±0.001)	0.477 (±0.001)	0.480 (std <0.001)
	SnowGalileo random	0.106 (±0.007)	0.193 (±0.013)	0.255 (±0.018)	0.245 (±0.006)
	SnowGalileo pre-trained	**0.099 (±0.003)**	**0.144 (±0.022)**	**0.185 (±0.005)**	**0.200 (±0.004)**
Swiss Alps	RF	0.176 (std <0.001)	0.259 (std <0.001)	0.356 (std <0.001)	0.383 (±0.003)
	MLP	0.191 (std <0.001)	0.286 (±0.009)	0.335 (±0.015)	0.360 (±0.015)
	SVR	0.465 (±0.001)	0.467 (std <0.001)	0.465 (±0.001)	0.466 (std <0.001)
	SnowGalileo random	0.133 (±0.002)	0.229 (±0.002)	0.263 (±0.007)	0.302 (±0.018)
	SnowGalileo pre-trained	**0.124 (±0.003)**	**0.209 (±0.008)**	**0.262 (±0.005)**	**0.299 (±0.004)**

SnowGalileo can provide reliable FSC predictions, and can effectively use other information (such as past data or meteorological data) to produce trends in FSC, such as during the melting season. Fig. 5 B) shows that SnowGalileo's prediction tiles can be stitched together to provide FSC maps of larger regions, such as in the Mackenzie River Basin, Canada (62.17° N, 126.29° W). Overlaid on the satellite image is the FSC prediction, which fits very well with the observed snow cover. FSC is highest over the mountain peaks and lower at the valley bottom, and even the snow-covered lakes are detected by SnowGalileo. The stitching does not produce any edge artifacts in this example. Fig. 5 C) shows the contrast between SnowGalileo and the Global Snow Pack product (GSP) (Dietz and Roessler, 2025), a currently deployed operational snow cover product at 500 m resolution. Both show the satellite imagery and a half-transparent prediction for the same day, with GSP clearly displaying a much lower resolution and overpredicting the snow cover in the valley bottom. In summary, utilizing SnowGalileo for daily, gapless, high-resolution FSC mapping could provide a new and promising level of quality for FSC products.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

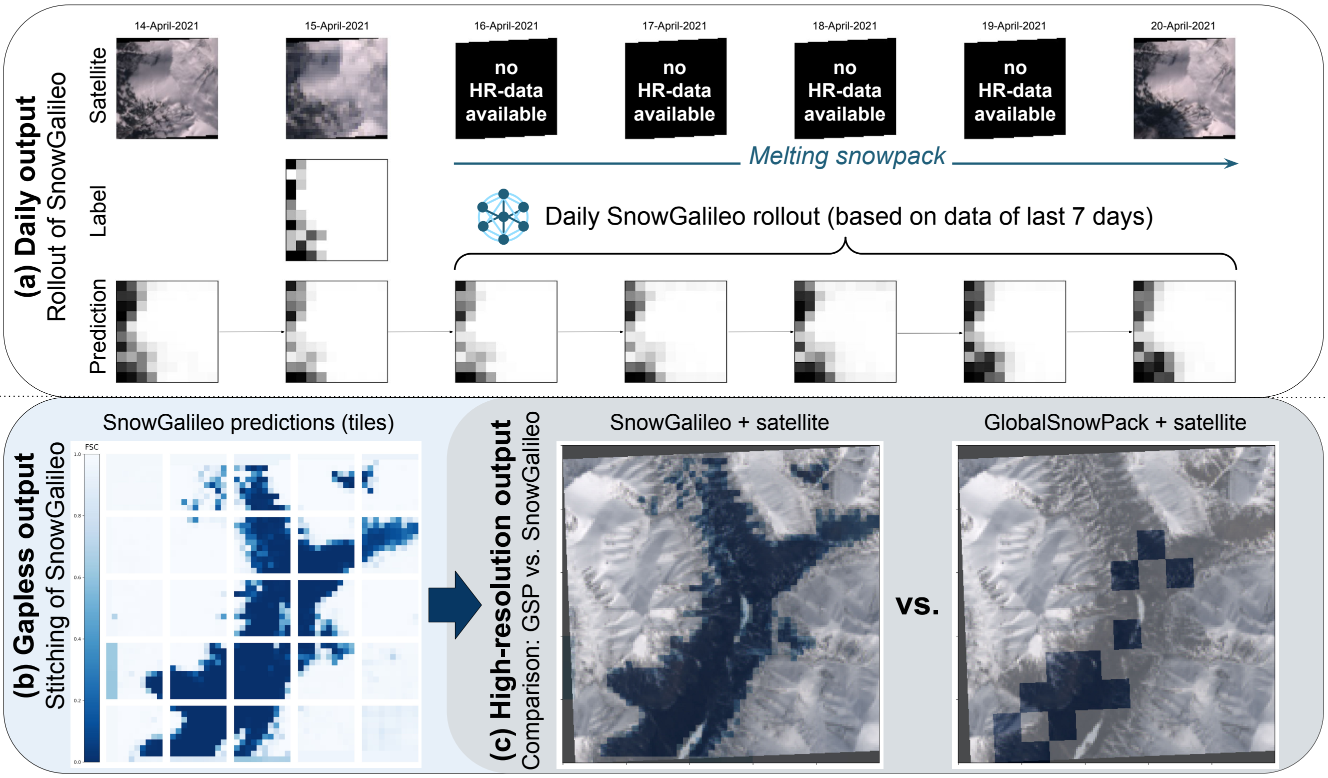

Figure 5. SnowGalileo can produce daily, gapless, and high-resolution FSC maps. (a) Shows daily roll-outs of SnowGalileo. The first row shows RGB images from Sentinel-2 and Landsat when available, the second row shows the ground-truth label when available, and the third row shows the prediction of SnowGalileo. Each column represents another day (14-Apr-2021 to 20-Apr-2021) within a melting week, i.e. a decline in FSC is expected (see satellite image), which can indeed be observed in SnowGalileo's predictions. **(b)** Shows that the single tiles of SnowGalileo's prediction can be reliably stitched to larger scenes. On the left, individual samples predicted by SnowGalileo are shown that are geographic neighbors (Mackenzie River Basin, Canada (62.17° N, 126.29° W)). In the middle, these samples are stitched together to create gapless FSC maps (the background shows the corresponding satellite image for visual comparison). **(c)** On the right, the snow cover map from the Global Snowpack Product (GSP) at 500 m resolution is shown, for visual comparison with the 100 m resolution output that SnowGalileo can produce. In summary, SnowGalileo can bridge HR gaps in time-series, stitch predictions, and provide FSC at a higher resolution than current SOTA.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

5.2 Model Intercomparison for FSC Mapping

5.2.1 Baseline Comparison

We compare SnowGalileo against an RF, an MLP, and an SVR to assess its performance relative to SOTA FSC mapping approaches. Table 2 shows the RMSE scores of all baseline models across different input data setups, averaged across three model runs. In this study, the RF, MLP, and SVR models of prior work are re-implemented, giving them access to the same information as SnowGalileo to ensure a fair comparison (see Sect. 4.2.2). Two different versions of SnowGalileo are also compared: Both share the same architecture and are trained on the same FSC mapping data, but one has been pre-trained (Sect. 4.1.2) on a global snow dataset, whereas the other has not received such pre-training and instead received randomly initialized weights prior to the FSC-specific training. Fig. 6 visualizes these RMSE scores for the Canadian Rockies alongside a qualitative sample for each model. Here, all models are trained on the same input data setup they are evaluated on. As shown in the figure (as well as Tab. 2), SnowGalileo consistently outperforms all baselines, with the random-initialized version coming in second across all experiments and the pre-trained version coming in first (RMSE for clear-HR in the Rockies: 0.099). The RF and the MLP perform comparatively well under clear conditions (RMSE of 0.149 and 0.164), whereas the SVR does not achieve skillful FSC prediction (RMSE of 0.477). A visual inspection reveals that all models except the SVR are able to capture the spatio-temporal pattern of FSC, and that only slight deviations of the FSC prediction from the label lead to higher RMSE values. In contrast, the SVR cannot capture spatio-temporal patterns and instead predicts an average FSC over all pixels. The strong performance of both SnowGalileo versions shows that its architecture provides new and promising FSC prediction capabilities. In sum, SnowGalileo consistently outperforms all other baselines across all input data setups, regions, and model runs.

5.2.2 Model Analysis

It has been established that an increased labeled training dataset size usually considerably improves model performance (Dosovitskiy et al., 2020; Sun et al., 2017), however, labels are often difficult to acquire, especially in cryospheric regions (clouds, polar night, inaccessible terrain), which motivates pre-training on larger unlabeled datasets (Wang et al., 2022) in the first place. Here, it is investigated whether pre-training is necessary for FSC mapping.

To understand the degree to which pre-training is helpful for SnowGalileo, different SnowGalileo versions are compared with each other. Here, SnowGalileo with “randomly initialized weights” means that SnowGalileo was not pre-trained and instead receives random weights before it is fine-tuned on the FSC mapping task like all other models. Table 2 shows that both SnowGalileo with randomly initialized weights and the pre-trained SnowGalileo model outperform all SOTA models; however, pre-training reliably results in the strongest performance gain, while the baseline SnowGalileo comes in second across all scenarios (see Table 2). This shows that both the architectural choices and the pre-training itself are key to SnowGalileo’s performance. The table also shows that pre-training yields stronger performance when less data is available and under more difficult input data setups. When the pre-trained model is compared with the randomly-initialized model, we can see a very strong error reduction of 19.5% on average across all input data setups in the Canadian Rockies, but only an error reduction of

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

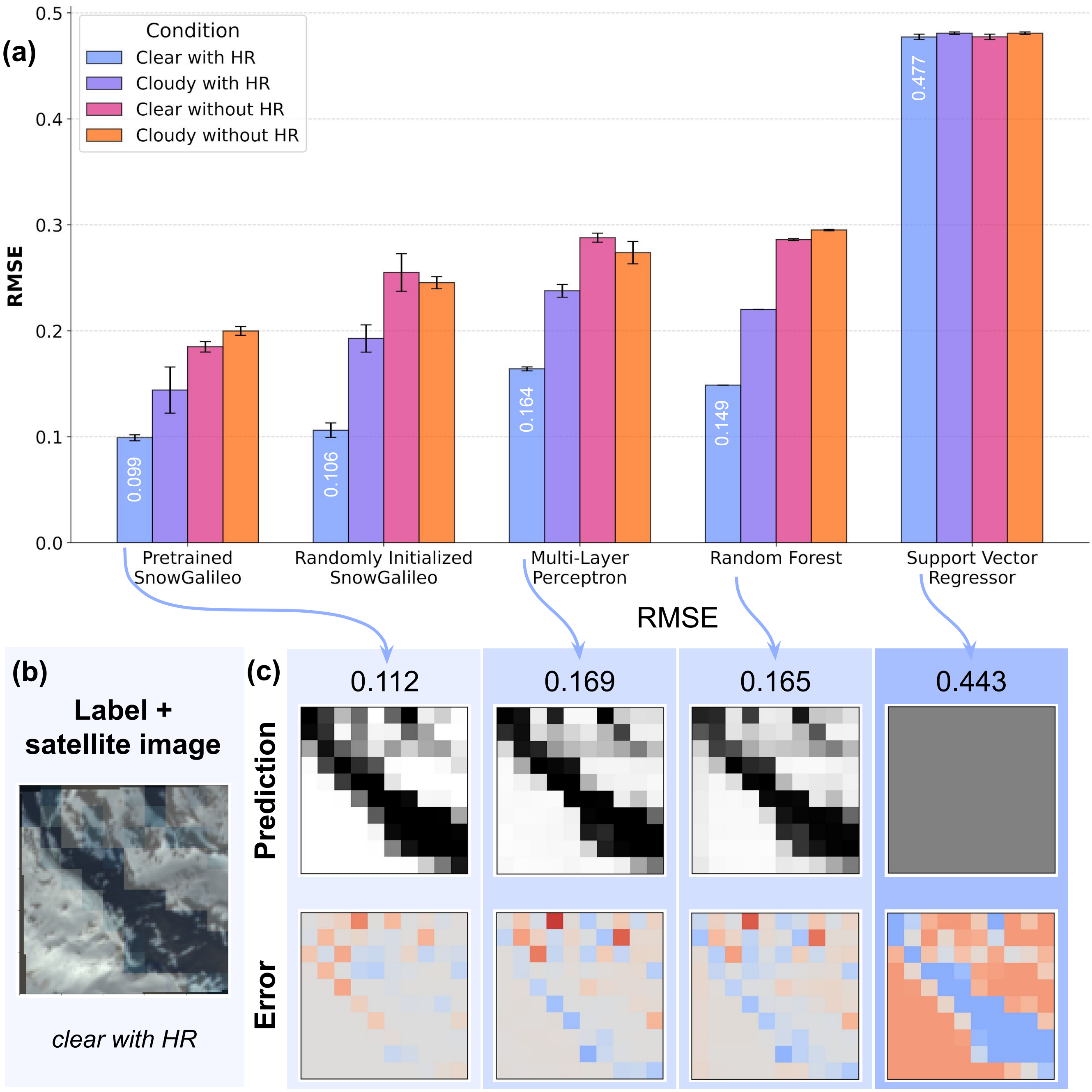

Figure 6. SnowGalileo outperforms SOTA across all input data setups in the Canadian Rockies. (a) shows the average RMSE score (and std) in the Rockies as a bar plot. Different colors indicate different input data setups. Here, each model was trained on the same input data setup it is evaluated on. **(b)** shows an example of a satellite image (gray), combined with the FSC label (blue). (c) depicts the predictions of each model (trained on clear-HR) for this satellite image. The RMSE scores in (c) refer to the RMSE scores reported for each specific sample depicted here.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Table 3. Pre-Training SnowGalileo helps more in data-scarce fine-tuning regimes. The table shows the RMSE of SnowGalileo trained and evaluated on clear-HR data, with one version being fine-tuned on 5,000 subsamples of the entire fine-tuning set and the other on 10,000 subsamples of the entire fine-tuning set. RMSE is reported as mean (± sample standard deviation) across three model runs. The difference in performance between the randomly initialized and the pre-trained versions of SnowGalileo is also reported. The performance difference is larger with fewer fine-tuning samples.

Model	Canadian Rockies		Swiss Alps	
	5,000	10,000	5,000	10,000
SnowGalileo random	0.177 (± 0.010)	0.139 (± 0.020)	0.175 (± 0.003)	0.151 (± 0.004)
SnowGalileo pre-trained	**0.112 (± 0.003)**	**0.104 (± 0.002)**	**0.137 (± 0.003)**	**0.131 (± 0.003)**
Difference of random & pre-trained	Δ 0.065	Δ 0.035	Δ 0.038	Δ 0.020

4.2% for the Swiss Alps (see Appendix E1.1). While the Swiss Alps and the Canadian Rockies are both present in the test set in roughly equal amounts, the fine-tuning training data contains more samples in geographic proximity to the Swiss Alps, while almost no training samples are geographically close to the Canadian Rockies test data (see Fig. 2). As pre-training is particularly helpful when less data is available, pre-training could be more powerful for the Canadian Rockies test case. Furthermore, it can be seen that pre-training has a greater impact on the Canadian Rockies under difficult conditions: Under the clear-HR coverage setup, pre-training only reduces the RMSE by 6.6%, but under the clear setup *without* HR coverage, pre-training results in a 27.5% improvement. These effects may be less pronounced when sufficiently similar data is available during fine-tuning. This could explain why such performance gains cannot be observed for the Swiss Alps. However, other, unrelated factors, such as the fact that the Rockies' data is mostly from the Northern Hemisphere summer season, could confound these results. Overall, pre-training reliably delivers the best results across multiple regions, input data setups, and model runs, indicating its superiority to all other models.

To further investigate the above hypothesis that SnowGalileo particularly improves performance in data-scarce fine-tuning regimes, its performance across varying fine-tuning dataset sizes is analyzed. Table 3 compares the performance of pre-trained and randomly initialized SnowGalileo models when fine-tuned using 5,000 and 10,000 random subsamples, respectively. An increase in sample size leads to a slight improvement in performance for both the pre-trained and the randomly initialized SnowGalileo versions. However, the randomly initialized version benefits more from additional samples. The performance difference between the randomly initialized and the pre-trained versions is much larger at smaller sample sizes and decreases at larger sample sizes. Notably, in this table, the pre-trained version on less data still outperforms the non-pre-trained model that has been provided with more sample data, showing that pre-training provides information that cannot be directly accessed through more fine-tuning samples. In summary, although pre-training provides only marginal improvements on the full dataset, it proves particularly helpful in data-scarce fine-tuning regimes.

Different SnowGalileo model versions further arise from training SnowGalileo on different fine-tuning conditions. Table 4 reports results under different test input data setups; here, each model is trained on the exact input data setup it is later tested

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Table 4. SnowGalileo trained on the cloudy-HR setup (SG Cloudy-HR) performs best across all input data setups on average. The table shows how models trained under different input data setups generalize onto other setups. Results are reported as RMSE on the Canadian Rockies (R) and the Swiss Alps (SW) test set for one pre-trained SnowGalileo model run. For example, the first row shows what happens if SnowGalileo has been trained only on the clear-HR setup, and is exposed to clouds and missing high-resolution imagery during testing. While the clear-HR SnowGalileo model achieves the best performance score under the clear-HR test setup (0.096/0.122, best values for each test input data setup are *italic*), it does not actually perform best across all input data setups. The last column shows the average RMSE performance of each trained model across all testing input data setups, with SnowGalileo trained on cloudy-HR data being the best model choice on average (best model in **bold**, second best is <u>underlined</u>).

Training Input Data Setup	Test Input Data Setup (R / SW)				Mean (R / SW)
	Clear-HR	Cloudy-HR	Clear-no-HR	Cloudy-no-HR	
SG Clear-HR	*0.096 / 0.122*	0.633 / 0.615	0.217 / 0.366	0.217 / 0.366	0.291 / 0.367
SG Cloudy-HR	0.108 / 0.147	*0.129 / 0.204*	0.221 / 0.350	0.201 / 0.283	**0.165 / 0.246**
SG Clear-no-HR	0.153 / 0.198	0.253 / 0.318	*0.179 / 0.267*	*0.190 / 0.275*	<u>0.194 / 0.265</u>
SG Cloudy-no-HR	0.150 / 0.204	0.222 / 0.350	0.274 / 0.315	0.202 / 0.295	0.212 / 0.291

on. For example, clear-HR results were obtained using models trained on the clear-HR setup, while cloudy-HR results were retrieved using models trained on the cloudy-HR setup. In practice, a model should be selected that performs well across all input data setups. To this end, SnowGalileo was trained on all setups, after which each version was “cross-validated” across all testing scenarios (see Table 4). Each model is expected to perform best on the same test input data setup on which it was trained (the diagonal of the table, in italic). However, under the cloudy-no HR setup, the clear-no HR SnowGalileo model outperforms the cloudy-no HR SnowGalileo model. This may be due to insufficient model runs (only one model run is reported here), or because the cloudy-no HR data does not provide enough information to learn an accurate representation, meaning the clear-no HR SnowGalileo model can best extrapolate best to the cloudy-no-HR case. Of all SnowGalileo versions, the one trained on the cloudy-HR setup performs best across all input data setups (mean RMSE of 0.165 for the Canadian Rockies and 0.246 for the Swiss Alps).

In summary, the pre-trained version of SnowGalileo, fine-tuned on 23,548 partially cloudy samples that include HR imagery, is the best-performing model across the testing scenarios.

5.3 Generalizability of SnowGalileo

5.3.1 Spatial Generalizability

For a global application of FSC mapping in mountainous regions, SnowGalileo needs to demonstrate spatial generalizability. For this purpose, SnowGalileo was designed to be spatially transferable by including pre-training data from locations across major mountain ranges in the Northern Hemisphere. All test results from the previous sections have been reported for the Swiss

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

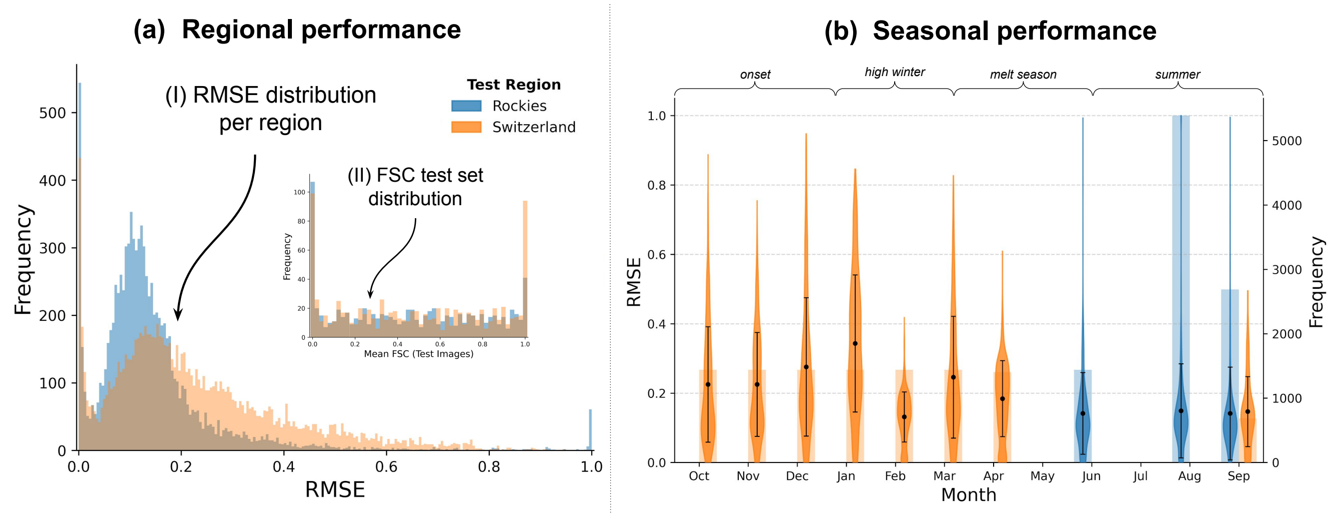

Figure 7. The performance of SnowGalileo (pre-trained and cloudy-HR) varies depending on the region and season.(a.I) shows the performance of SnowGalileo on the Canadian Rockies vs. the Swiss Alps test set. The two RMSE distributions overlap and show a peak at 0.0 RMSE, but they differ more than the distribution of the mean target FSC value per tile in **(a.II)**, indicating a real performance difference. SnowGalileo performs on average better in the Canadian Rockies, with its second peak exhibiting more lower RMSE values than in the Swiss Alps. **(b)** shows violin plots of RMSE values per month over one hydrological year for the Canadian Rockies (blue) and the Swiss Alps (orange). The bars in the background show the number of samples available, and the dots and lines indicate the mean and standard deviation for each month. When no violin plot is available, there is no test data available for that month. Performance appears to be higher during the summer months and the high winter season, and lower during snow onset and snow melt.

Alps and the Canadian Rockies; both mountain regions have been completely excluded from the fine-tuning dataset and are only present during testing (see also Fig. 2). This makes it possible to investigate how well SnowGalileo performs on mountain ranges it has seen during pre-training (when no labels are required) but not during fine-tuning (when expensive labeling is required). Fig. 7.A shows the RMSE distributions of SnowGalileo (cloudy-HR), as well as the FSC target distribution of the Canadian Rockies and the Swiss Alps. Assuming that the difficulty of FSC mapping depends on the structure of the snow cover (i.e. FSC tiles with a constant value of 0 or 1 are easier to predict than tiles with values in between), it is expected that similar FSC distributions result in similar RMSE / performance distribution between different regions. However, although the Canadian Rockies and the Swiss Alps test sets show a similar distribution of mean FSC value per tile, they clearly exhibit a performance difference (RMSE distributions show both skillfull FSC mapping, but they do not fully overlap). Thus, regional performance differences are likely driven by other data or mountain characteristics (see also Sect. 6.2) and need to be taken into account in global applications. Fig. 7.B suggests that the seasonal difference (summer season for Canadian Rockies vs. full season of Swiss Alps) could also be the main driver for the regional difference, since the seasonal and regional performance are correlated (see Sect. 5.3.2). Setting aside slight differences in regional performance scores, SnowGalileo clearly shows that it can skillfully perform FSC mapping on unseen mountain ranges, demonstrating its ability to generalize spatially.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

5.3.2 Temporal Generalizability

Similarly, for an all-year-round FSC application, it is necessary to investigate how well SnowGalileo performs in different seasons. To investigate differences in prediction accuracy over the course of the hydrological year, the mean RMSE value for each month is given in Fig. 7.B. Since the Canadian test data is only available in the summer months, and the Swiss Alps data is almost only available outside the summer months, it is not possible to differentiate between regional and summer/non-summer performance differences. This means that the low RMSE values in summer do not necessarily indicate better performance of SnowGalileo in summer, and further investigation is necessary (additionally, the winter and summer season is shifted between Canada and Switzerland, further complicating the direct seasonal performance comparison). However, data is available for both regions in September (summer month for both), showing very similar performance between the two regions this month. In any case, the Swiss Alps data make it possible to investigate the performance differences of SnowGalileo across snow onset, high winter, and snowmelt seasons. Here, the lowest RMSE values are during February, the month with the most reliable snowpacks. Less stable months, such as March and April, with their diurnal melt-freeze cycles and sudden temperature shifts, show higher RMSE scores, as expected. The snow onset window shows a similar RMSE score slightly above 0.2 across October, November, and December, implying that scattered FSC during the snow onset is more challenging. Notably, the RMSE score peaks in January, which is a high winter month when it would have been expected to perform particularly well. One factor that could explain this phenomenon is that the Swiss Alps are strongly impacted by large-scale atmospheric drivers such as the Arctic Oscillation (e.g. Polar Vortex instabilities leading to snow storms in the Alps) and North-Atlantic Oscillation (e.g. Azores High blocking Atlantic storm tracks, leaving the Alps dry and warm) in January, leading to volatile weather conditions (Scherrer and Appenzeller, 2006; Beniston, 1997). In addition, January snowpacks are low-density, i.e. they are more strongly affected by wind drift (Li and Pomeroy, 1997), which could lead to rapid changes in FSC that are neither captured in the satellite nor ERA5 data. In contrast, the performance drop at the onset and offset of the season is smaller than expected, indicating that SnowGalileo learns to handle snowfall and snowmelt to some degree. Overall, while performance varies throughout the year, the RMSE scores remain within an RMSE range of below 0.3 (except in January), demonstrating that SnowGalileo can indeed be used throughout the whole hydrological year.

5.4 Performance under Different Conditions

The following investigation compares the performance of SnowGalileo (pre-trained version, fine-tuned on cloudy-HR imagery) with SOTA under varying vegetation and cloud cover and satellite image resolution. Changes in HR imagery have the greatest impact on this particular version of SnowGalileo, followed by increasing forest cover, with additional cloud cover having the least impact.

5.4.1 Performance under Forest Cover

As mountain ranges are often covered in forest, and FSC mapping in forested areas is known to be a challenging task (Nolin, 2004; Metsämäki et al., 2005), SnowGalileo's predictive capabilities in forested areas are investigated. Fig. 8 (a) shows ridge

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

line plots of SnowGalileo in comparison with other models under different forest cover percentages (intensifying green with increasing forest cover). One can indeed observe a performance decline across all models except the SVR for increasing forest cover, as has been established in related literature (Hall et al., 1995; Salomonson and Appel, 2004; Gascoin et al., 2020). However, there is an exception at 80–100%, where RMSE suddenly improves, especially for SnowGalileo and the MLP. The reason could be that dense forest cover makes prediction easier since the FSC on the ground may be so obscured that it is clamped in both the labels and the raw signal. In other words, the labeled data might indicate that no snow is present, even though some snow is present, hidden away under the canopy. Given the labels used to train SnowGalileo, it can only predict and map viewable snow on top of vegetation.

While the results remain limited by the targets used here, SnowGalileo's mean performance across all forest cover types stays below 0.2 RMSE, clearly outperforming all the other models in terms of forest robustness.

5.4.2 Performance under Cloudy Conditions

Clouds are particularly present in mountain regions, and thus, an aspiring daily FSC product must perform exceptionally well under cloudy conditions. As made visible in Fig. 8 (b), SnowGalileo (fine-tuned on the cloudy-HR input data setup) outperforms all other models under clear conditions, local clouds, and cloud coverage. Generally, increased cloud cover is more challenging for each model, except for the SVR, which only predicts average pixel values and lacks overall prediction skill. The difference between clear and local conditions is consistent for all models. However, SnowGalileo's scores under cloudy conditions remain comparable to those of other models under clear conditions, demonstrating its capacity to maintain performance under cloudy conditions. A noticeable performance drop occurs between local clouds and increased cloud coverage for both MLP and RF, with an RMSE increase of 0.072 and 0.066, respectively. A smaller drop is observed for SnowGalileo, with an RMSE increase of 0.040. The performance drop for SnowGalileo between local and cloudy conditions is comparable to the difference in performance between clear and local cloud conditions (an RMSE increase of 0.026), demonstrating that SnowGalileo can handle clouds more robustly and reliably.

5.4.3 Performance under no-HR Conditions

A daily FSC method needs to be able to maintain performance under lower-resolution imagery (MODIS, VIIRS, Sentinel-3), as it is available on a daily scale – in contrast to HR imagery (Landsat, Sentinel 1-2). Fig. 8 (c) shows that the inaccessibility of HR imagery results in a stronger performance drop across all models (except for the underperformant SVR, as explained above). All models suffer a performance loss of >= 0.1 and exhibit a larger standard deviation than in the cloud experiments, however, SnowGalileo consistently outperforms all other models. Notably, even in the challenging no-HR case, SnowGalileo's scores are comparable to the scores of ML models with HR access. SnowGalileo's results on HR and no-HR imagery clearly show the novel and strong potential of SnowGalileo as a robust algorithm for daily FSC mapping.

In summary, SnowGalileo can perform daily, gapless, high-resolution FSC mapping under various challenging conditions. SnowGalileo consistently outperforms current SOTA, and demonstrates that pre-training – especially in data-scarce regimes

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

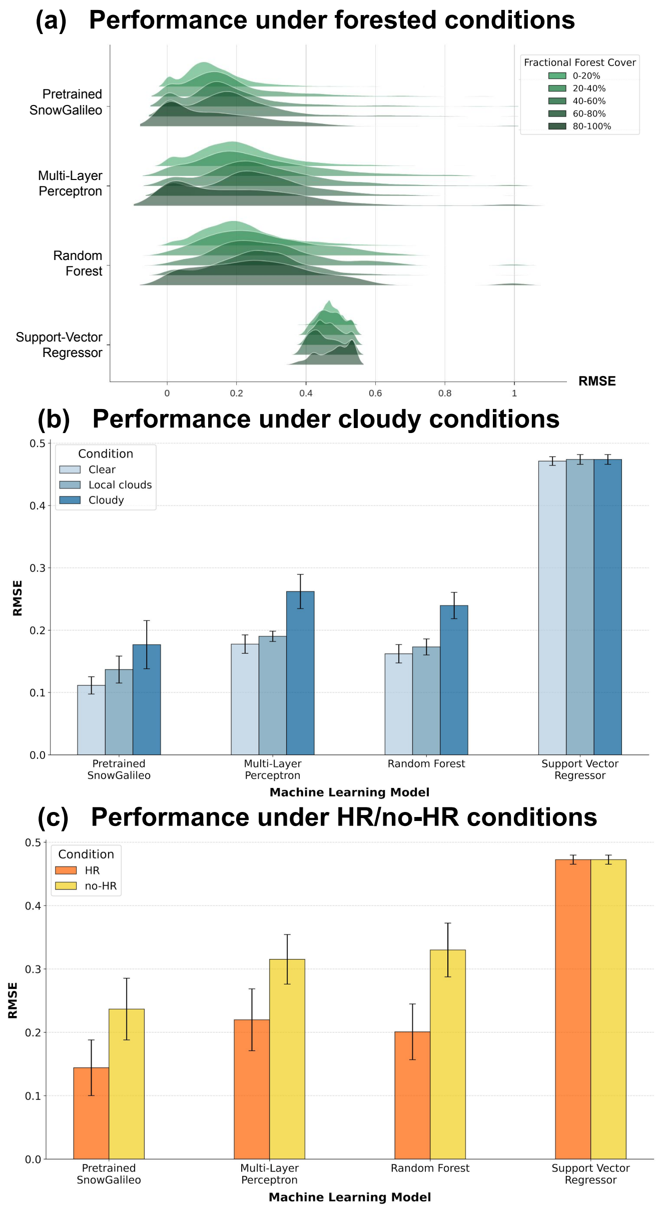

Figure 8. SnowGalileo is the best performing model under vegetation cover, clouds, and lack of HR imagery. (a) shows for each model (rows) the ridge line plots of the RMSE scores for each model for different forest cover fractions (shades of green indicate forest thickness). More forested area is more challenging for most models. SnowGalileo performs best across all forest cover percentages. **(b)** show for each model (columns) the mean RMSE scores and standard deviation across three model runs, for three different cloud categories (shades of blue indicate clear condition, local clouds, and clouds). More clouds are more challenging for all models. SnowGalileo outperforms all other models under all three cloud conditions. **(c)** shows for each model (columns) the mean RMSE scores and standard deviation across three runs for HR vs. no-HR imagery available on the prediction day. No-HR imagery is more challenging for all models. SnowGalileo outperforms all other models under both conditions. Overall, SnowGalileo shows robust performance across various challenging conditions. All results are reported with respect to the model trained under the same input data setup as was used for testing.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

– helps to achieve novel levels of performance. Regional and seasonal performance should be analyzed in more detail in future work, but are promising indicators of SnowGalileo's generalizability. Testing SnowGalileo in challenging yet realistic conditions in mountainous regions demonstrated its predictive ability and its robustness in the presence of forest cover, clouds and a lack of HR imagery.

6 Discussion

6.1 Towards Daily, Gapless, 100 m resolution FSC Mapping

The overall goal of the studies presented here is to develop a methodology for creating a daily, gapless, 100 m resolution FSC mapping product. As the results show, SnowGalileo is a promising approach towards achieving this goal. The following discusses SnowGalileo's readiness for FSC mapping (i.e. different deployment criteria, listed below) of SnowGalileo for FSC mapping and the steps necessary to achieve an operational product.

High-Resolution FSC mapping in the mountains is achieved through SnowGalileo at a spatial resolution of 100 m. In contrast, Global Snowpack (Dietz et al., 2015), a current SOTA operational snow cover product, achieves only 500 m resolution. As demonstrated in Fig. 5, SnowGalileo does not just provide "more pixels" with the same old de-facto resolution, but can capture more meaningful spatial information at a 100 m resolution, reliably matching up topographic and snow features. This success may be attributed to the sensor fusion methodology of SnowGalileo that can retrieve information from several sensors across an extended time-series. Thus, SnowGalileo's methodology successfully addressed the desired high-resolution criteria.

Daily FSC mapping is possible through SnowGalileo's setup. SnowGalileo can perform daily roll-outs (see Fig. 5), and can deal with cloudy, no-HR data (see Fig. 8), a key capability to achieve reliable performance on the daily level under varying data availability. In contrast, other ML SOTA approaches that operate at high-resolution cannot provide daily roll-outs as they have not been adapted for the challenging conditions of a daily method. While this issue is addressed here on the methodological level, software engineering is still needed to scale this solution efficiently, such that it can be run easily on large regions for extended periods of time.

Gapless FSC mapping is possible with SnowGalileo. SnowGalileo has shown a promising capacity to seamlessly stitch small prediction tiles into larger maps, as demonstrated in Fig. 5. However, more technological development is needed to achieve a reliable gapless algorithm. Some scenes still produce artifacts, e.g. when no high-resolution imagery has been available, and clouds last for extended periods of time. As such, single tiles can have a visible performance drop in comparison to the remaining predicted scene. Future work could investigate several simple interpolation approaches for these singular outliers (since the remaining scene performs well), as well as CNNs and UNets (on the temporal and spatial axis) to fill the remaining gaps. At the moment, SnowGalileo provides a gapless product in the sense that it can provide high-resolution imagery even under cloudy conditions on any given day, while technical work remains to be done before deploying SnowGalileo at a large scale.

Mountain regions provide a major challenge for traditional FSC methods, due to their topologically challenging terrain, and indeed SnowGalileo has demonstrated considerable skill in those regions. By testing SnowGalileo on the Canadian Rockies

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

and Swiss Alps (and completely holding out these regions from the fine-tuning training set), results are exclusively reported on mountaineous areas. Prior to deployment, SnowGalileo should further be evaluated in depth in all mountain regions. Additionally, future studies should aim to evaluate SnowGalileo's skill in other terrain, such as the Canadian prairies, glaciated areas, and on sea ice, as SnowGalileo could also contribute to FSC mapping outside mountainous regions.

Spatial and temporal generalizability is key for global, year-around deployment, and SnowGalileo shows promising results in this direction. However, as previously pointed out, spatial generalizability would need to be analyzed in more depth across all mountain ranges, and a more in-depth seasonal analysis across different regions would be necessary to fully comprehend the capacity as well as the limits of SnowGalileo on the global, multi-annual scale.

Robustness of FSC prediction to a diverse set of challenging conditions is also essential for deployment – SnowGalileo has demonstrated such robustness in comparison with other models. While some performance degradation cannot be avoided under challenging conditions, SnowGalileo has largely reduced it, keeping FSC predictions within a narrower window while maintaining higher performance compared to the other models tested in this paper. The studies presented here have subjected all models to particularly harsh test conditions to match the realistic conditions that are expected during deployment, thereby increasing confidence in SnowGalileo's robustness. Prediction robustness over forested terrain remains limited. The labels used to train SnowGalileo account for snow cover beneath forest canopies through an NDSI adjustment based on NDVI. However, snow on the forest floor that produces no spectral signal at the canopy top remains undetectable in optical remote sensing imagery. Resolving this limitation would require in-situ ground truth data, both for generating more reliable training labels and for validating model performance in these areas.

Extendability of a methodology for FSC mapping seems like a pragmatic concern; however, it is helpful when new sensors are launched that could improve performance. Extendability is, in principle, provided through SnowGalileo. Its sensor fusion approach and code setup allow developers to add new sensors (at different resolutions) or new meteorological and other relevant data products for FSC mapping. The future NISAR sensor, for example, is a promising candidate for snow products and could be integrated into subsequent versions of SnowGalileo.

Other needs of SnowGalileo before deployment encompass 1) fully re-training SnowGalileo on the full dataset, 2) extensively validating the FSC maps together with hydrologists and modelers, 3) building awareness on the role of HR imagery for prediction performance, and 4) understanding the limits of SnowGalileo for Near-Realtime-Prediction. Firstly, re-training SnowGalileo on the full dataset may be helpful and should be evaluated before deployment. At the moment, all results have been reported in the Swiss Alps and the Canadian Rockies which were held out during fine-tuning, and training samples were limited to 23,548. Re-training SnowGalileo with all data available, instead of subsets of it, will likely increase performance on the reported regions. Similarly, it could help to re-train SnowGalileo with all input data setups present at once, rather than training it on separate input data setups. Secondly, hydrological modelers will extensively validate the FSC maps provided through SnowGalileo, by comparing them with physically-based snow model outputs, high-resolution drone-based imagery, and LiDAR data, as well as in-situ meteorological and snow survey transect data for a specific mountain catchment. This is a key step to understanding if further methodological development is needed before moving towards engineering a deployable software. Thirdly, although the impact of low-resolution sensors on performance can be mitigated to a much greater extent

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

than with other models, addressing the lack of HR imagery remains more challenging than addressing forest cover or cloud coverage. Thus, missions should be advocated for that would enable high-resolution imagery to be captured on a more frequent basis. Fourth, SnowGalileo's potential for Near-Realtime (NRT) prediction is currently limited: All data ingested as input time-series must be available for prediction, i.e. variables subject to a processing time (such as ERA5-based variables) might not be suitable for NRT prediction. SnowGalileo can, in principle, be re-trained for NRT applications to ingest sources from more readily available sources (e.g. regional meteorological services).

In summary, SnowGalileo provides a methodological basis for daily, gapless, high-resolution FSC mapping, while still requiring additional validation through domain experts and software-engineering resources to move to a deployable and scalable product.

6.2 On SnowGalileo's Skill

In the following, several aspects are discussed of why SnowGalileo is able to consistently outperform all other models reported here. SnowGalileo's limitations are also discussed in Sect. 6.4.

Multi-resolution sensor fusion is a key aspect of why SnowGalileo can achieve high-resolution and quasi-gapless FSC predictions (see Fig. 5). In contrast to other sensor fusion models, SnowGalileo does not simply ingest interpolated/accumulated data from different sensors (i.e. resolutions have been matched), but instead learns a representation of how the imagery from different sensors relates to each other in space, in time, and in terms of resolutions. SnowGalileo is able to map data points to each other, i.e. it can encode the location of an image segment within different imagery of various resolutions. This, combined with the vision transformer's attention mechanisms cross-relating different information, gives SnowGalileo a methodological advantage over traditional ML methods, such as simple MLPs, RFs, SVR (Kuter, 2021), and also CNNs, UNets, and ResNets from more recent literature cannot provide (Hou et al., 2026).

SnowGalileo has successfully extended other early Earth Foundation Model approaches by including not only high-resolution and low-resolution sensors during pre-training, but also medium-resolution, as they were considered necessary for FSC mapping (AnySat (Astruc et al., 2025) has achieved this similarly). SnowGalileo has leveraged state-of-the-art machine learning, providing a novel methodological approach to sensor fusion for FSC mapping and clearly outperforms more traditional approaches for this task.

The time-series based approach of SnowGalileo likely drives its success in achieving daily FSC prediction (see Fig. 5). SnowGalileo has developed its underlying model approach, Galileo (Tseng et al., 2025), further by moving it from a monthly to a daily timestep. Both its pre-training and fine-tuning operate on ingesting 8 days of satellite and other input data. While all other models have been adapted to be able to ingest the same data for a fair comparison, these models do not have the same capacity to relate and temporally interpolate the different inputs. It is assumed that this is particularly helpful in challenging conditions, i.e. when the model cannot rely on high-quality data on the prediction day due to cloud cover or the absence high-resolution imagery. SnowGalileo is also assumed to benefit from pre-training on the temporal context here, i.e. it can extract such patterns in advance, providing a pre-trained architecture that directly incorporates temporal patterns. However, the exact role of the time-series-based approach in challenging conditions needs to be analyzed in more detail in future work.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Pre-training While the performance gain from pre-training might look small at first, it appears to improve generalizability and robustness. As discussed in the results section, pre-training is particularly helpful when fewer samples are available during fine-tuning. These samples stem from manual labels by domain experts. Providing such labeled data for all mountain regions across the globe is challenging and resource-intensive. The fact that a pre-trained version of SnowGalileo maintains performance under fewer labels is key to spatial generalizability at a more global scale. In terms of robustness, pre-training is likely contributing towards robustness to different challenges, as it has seen more samples from various conditions not present in the fine-tuning dataset alone. However, this remains to be investigated in more detail by future research. Pre-training has proven to be very helpful across various cryospheric tasks (Kaushik et al., 2026), as well as in remote sensing tasks more generally (Wang et al., 2022), and it has proven to be helpful for FSC mapping. The pre-trained SnowGalileo version might also prove useful for other snow remote-sensing tasks, as discussed in the next section (Sect. 6.3).

The underlying data fulfill a role in the success of SnowGalileo that needs to be emphasized strongly. As alluded to in the previous paragraph, the fine-tuning data set was processed by experts, and the quality of these samples drives the success of SnowGalileo. As for the pre-training data, SnowGalileo benefits greatly from access to open-source data, enabling it to ingest terabytes of satellite imagery, as well as meteorological and other data (see Table 1). Without the engineering efforts and community products for EO data, advanced ML approaches such as SnowGalileo could rarely realize their potential.

6.3 On SnowGalileo's Usefulness

SnowGalileo could be used for a wide array of remote-sensing-based snow parameter retrieval algorithms, if fine-tuned respectively. For any snow-related task that requires the same set of satellite sensors used here, the pre-trained SnowGalileo version can directly be fine-tuned and evaluated, hopefully with fewer samples than traditional ML approaches. If the new task requires additional sensor inputs, adjustments to the input data pipeline will need to be made. In general, SnowGalileo is a ML model particularly suitable for snow tasks, not only because it has been pre-trained on snow data, but because 1) it operates on daily time-series data to capture the high variability of snow; 2) it incorporates radar data that penetrates clouds, 3) as well as high-resolution imagery needed in complex mountain terrain; 4) it uses specific spectral ranges relevant to snow; and 5) it includes additional variables (topography, meteorology, land cover) that are relevant for snow characteristics. Evaluating SnowGalileo's performance on tasks other than FSC mapping would provide further insight into the hypothesis that it has acquired knowledge of mountain snow cover that can be applied to other tasks.

Another contribution of this study that goes beyond the FSC mapping task is the training and evaluation framework on no-HR data and cloudy conditions. Dropping the last HR image can easily be realized in diverse setups of remote-sensing-based snow tasks. More challenging is the handling of clouds: In this study, a recent tool by Czerkawski et al. (2023) is employed to add artificial clouds (as outlined in Sect. 3.1.3). This makes it possible to provide challenging and realistic training setups to models, while still having access to the ground truth. Usually, large-scale high-resolution ground truth data cannot be accessed under cloudy conditions. It is hoped that the evaluation scheme presented here contributes to the community's efforts to move towards more realistic training regimes and stricter evaluation settings.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Next to using SnowGalileo's pre-training weights and evaluation framework, SnowGalileo's outputs could prove useful to address certain snow tasks. For example, hydrological modelers are interested in using high-resolution FSC inputs to evaluate the performance of models on a large scale. Instead of using it as an input, numerical models can also condition on the FSC outputs and assimilate satellite-based observations this way. ML-based emulators or data-assimilators can similarly profit from SnowGalileo's outputs. However, useful outputs can only be achieved by scaling SnowGalileo to the global, annual level.

It is hoped to see SnowGalileo being used beyond FSC mapping, and that it can realize its potential further in future research.

6.4 On SnowGalileo's Limitations

The setup presented here limits training to the Northern Hemisphere and evaluation to the Swiss Alps and the Canadian Rockies, restricting the ability to draw conclusions beyond these areas. For future work, the pre-training set should be broadened beyond the Northern Hemisphere and the evaluation range should be extended to additional mountain ranges. Furthermore, while the pre-training dataset has been strictly confined to 42–$60°$, the fine-tuning data also includes samples outside these latitudinal boundaries — an artifact of different engineering pipelines. Since large-scale atmospheric patterns outside the 42–$60°$ range, such as the Polar vortex, can affect snow patterns in the regions of interest of this study, pre-training on a broader latitudinal range might further improve the results presented here. In sum, the pre-training and evaluation spatial window should be broadened and aligned more consistently in future work.

Another important restriction in the setup presented here is that the seasonal and regional subsets are correlated. In other words, it is currently not possible to separate the effects of seasonality and region because only test data from the summer months was available for the Canadian Rockies and only test data during snow melt, high-winter, and snow onset was available for the Swiss Alps. Thus, future work should investigate seasonal and regional performance differences in more detail for a broader evaluation setup.

In terms of data, the main limitations are the availability of ground truth data, the accuracy and coverage of artificial clouds, and the limitations of the sensors themselves. Ground truth data is rarely "true", and also in this study, it is not perfect. Labels are based on Landsat 8/9 data provided by the SnowPEx algorithm, following ESA's best-practice approach. This algorithm was used on clear Landsat imagery to prevent any inaccuracies due to clouds. Nevertheless, even the SnowPEx approach has limitations, especially in forested areas. It should therefore be highlighted that a deployment in forested areas should not precede a more in-depth training and evaluation phase on forested terrains with more reliable labels. To remove errors in the snow classifications reliably, ground truth measurements need to be acquired through in-situ field campaigns in close collaboration with snow experts. Furthermore, the "artificial clouds" presented here are not perfect replicas of real clouds, but rather approximations. While the approach proved extremely helpful, it introduces assumptions about what clouds look like, potentially weakening the internal representation of clouds in SnowGalileo. Furthermore, although the cloudy test input data setup produces clouds in every image, this does not mean that the whole image is covered by clouds. Future work could address this by evaluating how SnowGalileo performs under complete cloud coverage. The data also restricts SnowGalileo on a more profound level: Most of the remote-sensing data used by SnowGalileo is optical, providing information primarily about the snow surface and the upper part of the snowpack. These optical sensors do not provide the information needed to characterize

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

snow variables "within" the snowpack (such as snow depth, snow water equivalent, density, etc.). While limitations in the setup can be addressed in future work, the data limitations are much more of a given, and others, too, will need to work with them.

SnowGalileo, as a model itself, is limited on several levels. First of all, it requires larger amounts of resources (implementation, running time, compute, and more) — an RF can be trained more quickly and can be implemented more easily. For the same amount of data that SnowGalileo uses for fine-tuning, it takes the random forest about 1.5 days, compared to about 4 days for SnowGalileo. Some of this training time could be reduced by improving the efficiency of the data input pipeline. Conversely, inference time remains quick and can be run on a single CPU, i.e. during large-scale deployment, SnowGalileo could even be less resource-intensive than more traditional approaches. As previously discussed, artefacts can appear during stitching. This is another technical limitation of SnowGalileo that needs to be addressed. Calculating averages over prediction boundaries, CNN or UNet approaches, or penalizing FSC jumps in neighboring prediction tiles might help to mitigate this limitation. A more general limitation of SnowGalileo is its inability to provide physical explanations for its performance scores. Machine learning models are almost always black-box algorithms, with limited explainability. Further research could investigate the possibility of integrating physical constraints into SnowGalileo. While SnowGalileo shows promising results and might be useful for multiple snow remote-sensing tasks, it is not a silver bullet. It should never be used without consulting a snow expert, i.e. a careful analysis to which snow variables and sensors are key for a snow task in a particular region, cannot be replaced with a pre-trained snow transformer. In summary, despite its promising results, SnowGalileo needs to be regarded with the same careful skepticism as any other novel algorithm that has not yet been established.

7 Conclusion

SnowGalileo is the first transformer model for daily, gapless, 100 m FSC mapping that has been pre-trained on mountain regions in the Northern Hemisphere. It successfully employs a time-series-based, multi-resolution sensor fusion methodology, which has been evaluated on the FSC mapping task. SnowGalileo does not only consistently outperform current baselines for FSC mapping, it also demonstrates new capabilities for mapping in cloudy conditions and with limited availability of high-resolution satellite imagery. The next steps towards deployment are to rigorously validate SnowGalileo's output with domain experts and in-situ data, efficiently scale its model and data pipelines, and eliminate remaining artifacts.

It remains to be seen how SnowGalileo will contribute to the snow community over time. Specifically, it is hoped to see it being developed into an operational daily, gapless, 100 m resolution FSC product. Such a product would have a broad impact, ranging from providing new scientific insights and being used by snow-hydrological models and emulators to helping water managers predict changes in water availability more accurately. In times of climate change, it is crucial that water availability can be predicted as accurately as possible, so that policymakers can implement adaptive and protective strategies.

Code and data availability. All code is provided at https://github.com/RolnickLab/snowgalileo. The pre-training data, fine-tuning data, and model checkpoints of the SnowGalileo models are published at https://zenodo.org/records/20735656 (Reil et al., 2026). FSC labels were derived from Landsat data available at https://earthexplorer.usgs.gov/. Exact scene names can be derived from Table A5, Table A4, Table A6, and Table A7 in the Appendix. To map snow over glaciers we used RGI 7.0 for Canada available at https://nsidc.org/data/nsidc-0770/versions/7 (Consortium, 2023) and Sentinel-2 boundaries for the Alps available at https://doi.org/10.1594/PANGAEA.909133 (Paul et al., 2019).

Author contributions. Conceptualization: C.B., J.K., A.D., J.P., G.T., C.M., T.J. ; Data Curation: G.S., T.N.; Funding Acquisition: A.D.; Investigation: C.B.; Methodology: M.R., J.K., C.B., G.T., D.R.; Project Administration: A.D., C.B.; Resources: S.R., S.S., C.M., J.T.; Software: M.R., F.P., J.K.; Supervision: D.R., A.D., C.B., J.P., C.M., T.J.; Validation: D.A., C.M.; Visualization: M.R., J.K., C.B.; Writing - original draft: J.K., M.R., C.B., Z.L., A.D., B.M.; Writing - review and editing: All authors.

Competing interests. The authors declare no competing interests.

Acknowledgements. The authors are grateful to Mikolaj Czerkawski for their support with the SatelliteCloudGenerator.

The authors acknowledge the use of GitHub Copilot (free version), ChatGPT (free version), and DeeplWrite (free version) to optimize specific code functions, literature research, perform grammar checks, and rephrase specific sentences. The outputs were thoroughly validated, and the authors take full responsibility for the final content. LLMs were not used beyond that, i.e. the text and code body were drafted manually by the authors.

This work used data provided by the Copernicus Programme (Copernicus) via the Copernicus Data Space Ecosystem and the Copernicus Climate Change Service (C3S). The results contain modified Copernicus Climate Change Service information [2026]. Sentinel data was processed via the Sentinel Hub API. MODIS and VIIRS data were accessed courtesy of NASA's LP DAAC. Landsat imagery was obtained through the USGS EarthExplorer and the Copernicus Browser. The authors thank the contributing space agencies for making these Earth observation datasets freely available.

The authors gratefully acknowledge the computational and data resources provided through the joint high-performance data analytics (HPDA) project "terrabyte" of the German Aerospace Center (DLR) and the Leibniz Supercomputing Center (LRZ).

This research was enabled in part by compute resources provided by Mila (mila.quebec), including material support from NVIDIA Corporation.

MR was supported by a fellowship of the German Academic Exchange Service (DAAD).

DR is supported by Canada CIFAR AI Chairs.

This work was funded by the European Space Agency (ESA) under contract/grant number (AI4Science 4000143295/23/I-DT).

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

References

Astruc, G., Gonthier, N., Mallet, C., and Landrieu, L.: AnySat: One Earth Observation Model for Many Resolutions, Scales, and Modalities, in: Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 19 530–19 540, 2025.

Barnes, J. C. and Bowley, C. J.: Handbook of techniques for satellite snow mapping, Environmental Research & Technology, 1974.

Barnett, T. P., Adam, J. C., and Lettenmaier, D. P.: Potential impacts of a warming climate on water availability in snow-dominated regions, Nature, 438, 303–309, 2005.

Bastani, F., Wolters, P., Gupta, R., Ferdinando, J., and Kembhavi, A.: Satlaspretrain: A large-scale dataset for remote sensing image understanding, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 16 772–16 782, 2023.

Beniston, M.: Variations of Snow Depth and Duration in the Swiss Alps Over the Last 50 Years: Links to Changes in Large-Scale Climatic Forcings, p. 49–68, Springer Netherlands, Dordrecht, ISBN 978-94-015-8905-5, https://doi.org/10.1007/978-94-015-8905-5_3, 1997.

Berman, E. E., Bolton, D. K., Coops, N. C., Mityok, Z. K., Stenhouse, G. B., and Moore, R. D.: Daily estimates of Landsat fractional snow cover driven by MODIS and dynamic time-warping, Remote Sensing of Environment, 216, 635–646, 2018.

Bertoncini, A. and Pomeroy, J.: Assimilation of Satellite Albedo to Improve Simulations of Glacier Hydrology, Hydrological Processes, 38, e15 329, https://doi.org/https://doi.org/10.1002/hyp.15329, e15329 HYP-24-0068.R2, 2024.

Bertoncini, A., Aubry-Wake, C., and Pomeroy, J. W.: Large-area high spatial resolution albedo retrievals from remote sensing for use in assessing the impact of wildfire soot deposition on high mountain snow and ice melt, Remote Sensing of Environment, 278, 113 101, 2022.

Biewald, L.: Experiment Tracking with Weights and Biases, https://www.wandb.com/, software available from wandb.com, 2020.

Breiman, L.: Random forests, Machine learning, 45, 5–32, 2001.

Brown, R. D. and Mote, P. W.: The response of Northern Hemisphere snow cover to a changing climate, Journal of Climate, 22, 2124–2145, 2009.

C3S: ERA5-Land hourly data from 1950 to present, https://doi.org/10.24381/CDS.E2161BAC, 2019.

Cannistra, A. F., Shean, D. E., and Cristea, N. C.: High-resolution CubeSat imagery and machine learning for detailed snow-covered area, Remote Sensing of Environment, 258, 112 399, 2021.

Cong, Y., Khanna, S., Meng, C., Liu, P., Rozi, E., He, Y., Burke, M., Lobell, D., and Ermon, S.: Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery, Advances in Neural Information Processing Systems, 35, 197–211, 2022.

Consortium, R.: Randolph Glacier Inventory - A Dataset of Global Glacier Outlines, Version 7, https://doi.org/10.5067/F6JMOVY5NAVZ, 2023.

Czerkawski, M., Atkinson, R., Michie, C., and Tachtatzis, C.: Satellitecloudgenerator: controllable cloud and shadow synthesis for multi-spectral optical satellite images, Remote Sensing, 15, 4138, 2023.

DeBeer, C. and Pomeroy, J.: Simulation of the snowmelt runoff contributing area in a small alpine basin, Hydrology and Earth System Sciences, 14, 1205–1219, 2010.

Derksen, C., Lemmetyinen, J., King, J., Belair, S., Garnaud, C., Lapointe, M., Crevier, Y., Burbidge, G., and Siqueira, P.: A Dual-Frequency Ku-Band Radar Mission Concept for Seasonal Snow, in: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 5742–5744, https://doi.org/10.1109/IGARSS.2019.8898030, 2019.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Deschamps-Berger, C., López-Moreno, J. I., Gascoin, S., Mazzotti, G., and Boone, A.: Where Snow and Forest Meet: A Global Atlas, Geophysical Research Letters, 52, e2024GL113 684, https://doi.org/https://doi.org/10.1029/2024GL113684, e2024GL113684 2024GL113684, 2025.

Dietz, A. J. and Roessler, S.: Brief communication: Daily, gap-free snow cover information based on a combination of NPP VIIRS and MODIS data, The Cryosphere, 19, 3873–3877, https://doi.org/10.5194/tc-19-3873-2025, 2025.

Dietz, A. J., Kuenzer, C., Gessner, U., and Dech, S.: Remote sensing of snow–a review of available methods, International Journal of Remote Sensing, 33, 4094–4134, 2012.

Dietz, A. J., Kuenzer, C., and Dech, S.: Global SnowPack: a new set of snow cover parameters for studying status and dynamics of the planetary snow cover extent, Remote Sensing Letters, 6, 844–853, https://doi.org/10.1080/2150704X.2015.1084551, 2015.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale, arXiv preprint arXiv:2010.11929, 2020.

Dozier, J.: Spectral signature of alpine snow cover from the Landsat Thematic Mapper, Remote sensing of environment, 28, 9–22, 1989.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., and Vapnik, V.: Support vector regression machines, Advances in neural information processing systems, 9, 1996.

Duguay, Y. and Bernier, M.: The use of RADARSAT-2 and TerraSAR-X data for the evaluation of snow characteristics in subarctic regions, in: 2012 IEEE International Geoscience and Remote Sensing Symposium, pp. 3556–3559, IEEE, 2012.

Ecosystem, C. D. S.: Copernicus DEM, https://doi.org/10.5270/esa-c5d3d65, 2022.

Fuller, A., Millard, K., and Green, J.: Croma: Remote sensing representations with contrastive radar-optical masked autoencoders, Advances in Neural Information Processing Systems, 36, 5506–5538, 2023.

Gascoin, S., Grizonnet, M., Bouchet, M., Salgues, G., and Hagolle, O.: Theia Snow collection: High-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data, Earth System Science Data, 11, 493–514, 2019.

Gascoin, S., Barrou Dumont, Z., Deschamps-Berger, C., Marti, F., Salgues, G., López-Moreno, J. I., Revuelto, J., Michon, T., Schattan, P., and Hagolle, O.: Estimating Fractional Snow Cover in Open Terrain from Sentinel-2 Using the Normalized Difference Snow Index, Remote Sensing, 12, https://doi.org/10.3390/rs12182904, 2020.

Gascoin, S., Luojus, K., Nagler, T., Lievens, H., Masiokas, M., Jonas, T., Zheng, Z., and De Rosnay, P.: Remote sensing of mountain snow from space: status and recommendations, Frontiers in Earth Science, 12, 1381 323, 2024.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sensing of Environment, https://doi.org/10.1016/j.rse.2017.06.031, 2017.

Gorman-Murray, A.: An Australian feeling for snow: Towards understanding cultural and emotional dimensions of climate change, Cultural Studies Review, 16, 60–81, 2010.

Hall, D. K. and Riggs, G. A.: Accuracy assessment of the MODIS snow products, Hydrological Processes: An International Journal, 21, 1534–1547, 2007.

Hall, D. K., Riggs, G. A., and Salomonson, V. V.: Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data, Remote Sensing of Environment, 54, 127–140, https://doi.org/10.1016/0034-4257(95)00137-p, 1995.

Hall, D. K., Riggs, G. A., Foster, J. L., and Kumar, S. V.: Development and evaluation of a cloud-gap-filled MODIS daily snow-cover product, Remote sensing of environment, 114, 496–503, 2010.

Hauser, S., Augner, L., and Schmitt, A.: Perfect Labelling: A Review and Outlook of Label Optimization Techniques in Dynamic Earth Observation, Remote Sensing, 17, 1246, https://doi.org/10.3390/rs17071246, 2025.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., and Girshick, R.: Masked autoencoders are scalable vision learners, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 16 000–16 009, 2022.

Hou, J., Huang, C., Zhang, Y., and Guo, J.: On the value of available MODIS and Landsat8 OLI image pairs for MODIS fractional snow cover mapping based on an artificial neural network, IEEE Transactions on Geoscience and Remote Sensing, 58, 4319–4334, 2020.

Hou, J., Zhang, M., Hao, X., Guo, J., Dou, P., Zhang, Y., and Huang, C.: ChinaAI-FSC: a comprehensive AI-ready MODIS fractional snow cover dataset for China (2000–2022), Earth System Science Data, 18, 1995–2022, 2026.

Hu, Z., Dietz, A. J., and Kuenzer, C.: Deriving Regional Snow Line Dynamics during the Ablation Seasons 1984–2018 in European Mountains, Remote Sensing, 11, 933, https://doi.org/10.3390/rs11080933, 2019.

Huang, M., Wang, G., Bie, X., Jiang, Y., Huang, X., Li, J.-J., Shi, S., Zhang, T., and Peng, P.-H.: Seasonal snow cover patterns explain alpine treeline elevation better than temperature at regional scale, Forest Ecosystems, 10, 100 106, 2023.

Huang, X., Deng, J., Wang, W., Feng, Q., and Liang, T.: Impact of climate and elevation on snow cover using integrated remote sensing snow products in Tibetan Plateau, Remote Sensing of Environment, 190, 274–288, 2017.

Immerzeel, W. W., Lutz, A. F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B., Elmore, A., et al.: Importance and vulnerability of the world's water towers, Nature, 577, 364–369, 2020.

IPCC: Polar Regions, p. 203–320, Cambridge University Press, 2022.

Jakubik, J., Yang, F., Blumenstiel, B., Scheurer, E., Sedona, R., Maurogiovanni, S., Bosmans, J., Dionelis, N., Marsocci, V., Kopp, N., et al.: Terramind: Large-scale generative multimodality for earth observation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7383–7394, 2025.

Jiang, D., Li, S., Hajnsek, I., Siddique, M. A., Hong, W., and Wu, Y.: Glacial lake mapping using remote sensing Geo-Foundation Model, International Journal of Applied Earth Observation and Geoinformation, 136, 104 371, 2025.

John, A., Cannistra, A. F., Yang, K., Tan, A., Shean, D., Hille Ris Lambers, J., and Cristea, N.: High-resolution snow-covered area mapping in forested mountain ecosystems using PlanetScope imagery, Remote Sensing, 14, 3409, 2022.

Johnson, Z., DeBeer, C., Schuster-Wallace, C., Pomeroy, J., Thornton, J., Wangchuk, S., McPhee, J., Pradhananga, D., Black, K., and Wrona, F.: Changes in the Cryosphere and Impacts on Water, in: UNESCO World Water Assessment Programme (2025), pp. 25–41, https://doi.org/https://doi.org/10.54679/LHPJ5153, 2025.

Jones, H. G.: Snow ecology: an interdisciplinary examination of snow-covered ecosystems, Cambridge University Press, 2001.

Kaushik, S., Maurya, L., Tellman, E., Zhang, G., and Dharpure, J. K.: Debris covered glacier mapping using newly annotated multisource remote sensing data and geo-foundational model, Science of Remote Sensing, p. 100319, 2025.

Kaushik, S., Maurya, L., and Tellman, B.: Cryo-Bench: Benchmarking Foundation Models for Cryosphere Applications, arXiv preprint arXiv:2603.01576, 2026.

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

Koehler, J., Bauer, A., Dietz, A. J., and Kuenzer, C.: Towards Forecasting Future Snow Cover Dynamics in the European Alps—The Potential of Long Optical Remote-Sensing Time Series, Remote Sensing, 14, 4461, https://doi.org/10.3390/rs14184461, 2022a.

Koehler, J., Dietz, A. J., Zellner, P., Baumhoer, C. A., Dirscherl, M., Cattani, L., Vlahović, , Alasawedah, M. H., Mayer, K., Haslinger, K., Bertoldi, G., Jacob, A., and Kuenzer, C.: Drought in Northern Italy: Long Earth Observation Time Series Reveal Snow Line Elevation to Be Several Hundred Meters Above Long-Term Average in 2022, Remote Sensing, 14, https://doi.org/10.3390/rs14236091, 2022b.

Kollert, A., Mayr, A., Dullinger, S., Hülber, K., Moser, D., Lhermitte, S., Gascoin, S., and Rutzinger, M.: Downscaling MODIS NDSI to Sentinel-2 fractional snow cover by random forest regression, Remote Sensing Letters, 15, 363–372, 2024.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Kuter, S.: Completing the machine learning saga in fractional snow cover estimation from MODIS Terra reflectance data: Random forests versus support vector regression, Remote Sensing of Environment, 255, 112 294, 2021.

Largeron, C., Dumont, M., Morin, S., Boone, A., Lafaysse, M., Metref, S., Cosme, E., Jonas, T., Winstral, A., and Margulis, S. A.: Toward snow cover estimation in mountainous areas using modern data assimilation methods: A review, Frontiers in Earth Science, 8, 325, 2020.

Lemos, A. and Riihelä, A.: Snow depth derived from Sentinel-1 compared to in-situ observations in northern Finland, EGUsphere, 2024, 1–18, 2024.

Li, L. and Pomeroy, J. W.: Estimates of Threshold Wind Speeds for Snow Transport Using Meteorological Data, Journal of Applied Meteorology, 36, 205–213, https://doi.org/10.1175/1520-0450(1997)036<0205:eotwsf>2.0.co;2, 1997.

Liu, C., Huang, X., Li, X., and Liang, T.: MODIS fractional snow cover mapping using machine learning technology in a mountainous area, Remote Sensing, 12, 962, 2020.

Liu, Y., Fang, Y., and Margulis, S. A.: Spatiotemporal distribution of seasonal snow water equivalent in High Mountain Asia from an 18-year Landsat–MODIS era snow reanalysis dataset, The Cryosphere, 15, 5261–5280, 2021.

Lv, Z. and Pomeroy, J. W.: Detecting intercepted snow on mountain needleleaf forest canopies using satellite remote sensing, Remote Sensing of Environment, 231, 111 222, 2019.

Margulis, S. A., Cortés, G., Girotto, M., and Durand, M.: A Landsat-era Sierra Nevada snow reanalysis (1985–2015), Journal of Hydrometeorology, 17, 1203–1221, 2016.

Marsh, C. B., Lv, Z., Vionnet, V., Harder, P., Spiteri, R. J., and Pomeroy, J. W.: Snowdrift-permitting simulations of seasonal snowpack processes over large mountain extents, Water Resources Research, 60, e2023WR036 948, 2024.

Masson, T., Dumont, M., Mura, M. D., Sirguey, P., Gascoin, S., Dedieu, J.-P., and Chanussot, J.: An assessment of existing methodologies to retrieve snow cover fraction from MODIS data, Remote Sensing, 10, 619, 2018.

Meloche, J., Langlois, A., Rutter, N., McLennan, D., Royer, A., Billecocq, P., and Ponomarenko, S.: High-resolution snow depth prediction using random forest algorithm with topographic parameters: A case study in the Greiner watershed, Nunavut, Hydrological Processes, 36, e14 546, 2022.

Metsämäki, S. J., Anttila, S. T., Markus, H. J., and Vepsäläinen, J. M.: A feasible method for fractional snow cover mapping in boreal zone based on a reflectance model, Remote Sensing of Environment, 95, 77–95, https://doi.org/10.1016/j.rse.2004.11.013, 2005.

Missions, L. and Tools, W.: Landsat Collection 2 Level-3 Fractional Snow Covered Area Science Product.

Moody, E. G., King, M. D., Schaaf, C. B., Hall, D. K., and Platnick, S.: Northern Hemisphere five-year average (2000–2004) spectral albedos of surfaces in the presence of snow: Statistics computed from Terra MODIS land products, Remote Sensing of Environment, 111, 337–345, 2007.

Mudryk, L., Santolaria-Otín, M., Krinner, G., Ménégoz, M., Derksen, C., Brutel-Vuilmet, C., Brady, M., and Essery, R.: Historical Northern Hemisphere snow cover trends and projected changes in the CMIP6 multi-model ensemble, The Cryosphere, 14, 2495–2514, 2020.

Nagler, T., Rott, H., Malcher, P., and Müller, F.: Assimilation of meteorological and remote sensing data for snowmelt runoff forecasting, Remote sensing of environment, 112, 1408–1420, 2008.

Nagler, T., Schwaizer, G., Mölg, N., Keuris, L., Hetzenecker, M., and Metsämäki, S.: Daily global Snow Cover Fraction–viewable snow (SCFV) from MODIS (2000–2020), version 2.0., NERC EDS Centre for Environmental Data Analysis [data set], 2022.

Niemi, J. and Ahlstedt, J.: Finnish agriculture and rural industries 2012, MTT Taloustutkimus, 2012.

Niittynen, P. and Luoto, M.: The importance of snow in species distribution models of arctic vegetation, Ecography, 41, 1024–1037, 2018.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Nolin, A. W.: Towards retrieval of forest cover density over snow from the Multi-angle Imaging SpectroRadiometer (MISR), Hydrological Processes, 18, 3623–3636, https://doi.org/10.1002/hyp.5803, 2004.

Nolin, A. W.: Recent advances in remote sensing of seasonal snow, Journal of Glaciology, 56, 1141–1150, 2010.

Nolin, A. W., Dozier, J., and Mertes, L. A.: Mapping alpine snow using a spectral mixture modeling technique, Annals of Glaciology, 17, 121–124, 1993.

Noman, M., Naseer, M., Cholakkal, H., Anwer, R. M., Khan, S., and Khan, F. S.: Rethinking transformers pre-training for multi-spectral satellite imagery, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 27 811–27 819, 2024.

Painter, T. H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R. E., and Dozier, J.: Retrieval of subpixel snow covered area, grain size, and albedo from MODIS, Remote Sensing of Environment, 113, 868–879, 2009.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library, Advances in neural information processing systems, 32, 2019.

Paul, F., Rastner, P., Azzoni, R. S., Diolaiuti, G., Fugazza, D., Le Bris, R., Nemec, J., Rabatel, A., Ramusovic, M., Schwaizer, G., and Smiraglia, C.: Glacier inventory of the Alps from Sentinel-2, shape files, https://doi.org/10.1594/PANGAEA.909133, 2019.

Paul, F., Rastner, P., Azzoni, R. S., Diolaiuti, G., Fugazza, D., Le Bris, R., Nemec, J., Rabatel, A., Ramusovic, M., Schwaizer, G., and Smiraglia, C.: Glacier shrinkage in the Alps continues unabated as revealed by a new glacier inventory from Sentinel-2, Earth System Science Data, 12, 1805–1821, https://doi.org/10.5194/essd-12-1805-2020, 2020.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in Python, the Journal of machine Learning research, 12, 2825–2830, 2011.

Pomeroy, J. and Gray, D. M.: Snowcover: Accumulation, relocation, and management, Tech. rep., National Hydrology Research Institute, 1995.

Pomeroy, J., Gray, D., Shook, K., Toth, B., Essery, R., Pietroniro, A., and Hedstrom, N.: An evaluation of snow accumulation and ablation processes for land surface modelling, Hydrological Processes, 12, 2339–2367, 1998.

Pomeroy, J. W., Johnson, Z., and others members of the Ad Hoc Committee for the Decade of Action on Cryospheric Sciences: Synthesis of the Decade of Action for the Cryospheric Sciences – 2025-2034, https://unesdoc.unesco.org/ark:/48223/pf0000397486, 2026.

Poon, S. K. and Valeo, C.: Investigation of the MODIS snow mapping algorithm during snowmelt in the northern boreal forest of Canada, Canadian Journal of Remote Sensing, 32, 254–267, 2006.

Rango, A. and Itten, K.: Satellite potentials in snowcover monitoring and runoff prediction, Hydrology Research, 7, 209–230, 1976.

Rango, A. and Martinec, J.: Application of a snowmelt-runoff model using Landsat data, Hydrology Research, 10, 225–238, 1979.

Rango, A., Salomonson, V., and Foster, J.: Seasonal streamflow estimation in the Himalayan region employing meteorological satellite snow cover observations, Water Resources Research, 13, 109–112, 1977.

Reil, M., Kaltenborn, J., Pelletier, F., Rolnick, D., Dietz, A., and Baumhoer, C.: AI4SNOW SnowGalileo Datasets and Model Checkpoints, https://doi.org/10.5281/zenodo.20735656, 2026.

Revuelto, J., Alonso-Gonzalez, E., Gascoin, S., Rodriguez-Lopez, G., and Lopez-Moreno, J.: Spatial downscaling of MODIS snow cover observations using Sentinel-2 snow products, Remote Sens.-Basel, 13, 4513, 2021.

Richiardi, C., Siniscalco, C., and Adamo, M.: Comparison of three different random forest approaches to retrieve daily high-resolution snow cover maps from MODIS and Sentinel-2 in a Mountain Area, Gran Paradiso National Park (NW Alps), Remote Sensing, 15, 343, 2023.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Ripper, E., Schwaizer, G., Nagler, T., Metsämäki, S., Törmä, M., Fernandes, R., Crawford, C. J., Painter, T., and Rittger, K.: Guidelines for the Generation of Snow Extent Products from High Resolution Optical Sensors – FINAL Deliverable D8, https://snowpex.enveo.at/doc/D08_Guidelines_for_the_generation_of_snow_extent_products_from_HR_optical_sensors_FINAL_v2.1.pdf, 2015.

Riseth, J. Å., Tømmervik, H., Helander-Renvall, E., Labba, N., Johansson, C., Malnes, E., Bjerke, J. W., Jonsson, C., Pohjola, V., Sarri, L.-E., et al.: Sámi traditional ecological knowledge as a guide to science: snow, ice and reindeer pasture facing climate change, Polar Record, 47, 202–217, 2011.

Rittger, K., Painter, T. H., and Dozier, J.: Assessment of methods for mapping snow cover from MODIS, Advances in Water Resources, 51, 367–380, 2013.

Rittger, K., Bormann, K. J., Bair, E. H., Dozier, J., and Painter, T. H.: Evaluation of VIIRS and MODIS snow cover fraction in high-mountain Asia using Landsat 8 OLI, Frontiers in Remote Sensing, 2, 647 154, 2021a.

Rittger, K., Krock, M., Kleiber, W., Bair, E. H., Brodzik, M. J., Stephenson, T. R., Rajagopalan, B., Bormann, K. J., and Painter, T. H.: Multi-sensor fusion using random forests for daily fractional snow cover at 30 m, Remote Sensing of Environment, 264, 112 608, 2021b.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning representations by back-propagating errors, nature, 323, 533–536, 1986.

Salminen, M., Pulliainen, J., Metsämäki, S., Ikonen, J., Heinilä, K., and Luojus, K.: Determination of uncertainty characteristics for the satellite data-based estimation of fractional snow cover, Remote Sensing of Environment, 212, 103–113, 2018.

Salomonson, V. V. and Appel, I.: Estimating fractional snow cover from MODIS using the normalized difference snow index, Remote sensing of environment, 89, 351–360, 2004.

Salomonson, V. V. and Appel, I.: Development of the Aqua MODIS NDSI fractional snow cover algorithm and validation results, IEEE Transactions on geoscience and remote sensing, 44, 1747–1756, 2006.

Scherrer, S. C. and Appenzeller, C.: Swiss Alpine snow pack variability: major patterns and links to local climate and large-scale flow, Climate Research, 32, 187–199, 2006.

Sirguey, P., Mathieu, R., and Arnaud, Y.: Subpixel monitoring of the seasonal snow cover with MODIS at 250 m spatial resolution in the Southern Alps of New Zealand: Methodology and accuracy assessment, Remote Sensing of Environment, 113, 160–181, 2009.

Smola, A. J. and Schölkopf, B.: A tutorial on support vector regression, Statistics and computing, 14, 199–222, 2004.

Snethlage, M. A., Geschke, J., Ranipeta, A., Jetz, W., Yoccoz, N. G., Körner, C., Spehn, E. M., Fischer, M., and Urbach, D.: A hierarchical inventory of the world's mountains for global comparative mountain science, Scientific Data, 9, 149, https://doi.org/10.1038/s41597-022-01256-y, 2022.

Stillinger, T., Rittger, K., Raleigh, M. S., Michell, A., Davis, R. E., and Bair, E. H.: Landsat, MODIS, and VIIRS snow cover mapping algorithm performance as validated by airborne lidar datasets, The Cryosphere, 17, 567–590, 2023.

Sun, C., Shrivastava, A., Singh, S., and Gupta, A.: Revisiting Unreasonable Effectiveness of Data in Deep Learning Era, in: 2017 IEEE International Conference on Computer Vision (ICCV), p. 843–852, IEEE, Venice, ISBN 978-1-5386-1032-9, https://doi.org/10.1109/ICCV.2017.97, 2017.

Sun, H., Fang, Y., Margulis, S. A., Mortimer, C., Mudryk, L., and Derksen, C.: Evaluation of the Snow Climate Change Initiative (Snow CCI) snow-covered area product within a mountain snow water equivalent reanalysis, The Cryosphere, 19, 2017–2036, 2025.

Szwarcman, D., Roy, S., Fraccaro, P., Gíslason, E., Blumenstiel, B., Ghosal, R., Oliveira, P. H. d., Almeida, J. L. d. S., Sedona, R., Kang, Y., Chakraborty, S., Wang, S., Gomes, C., Kumar, A., Truong, M., Godwin, D., Lee, H., Hsu, C.-Y., Asanjan, A. A., Mujeci, B., Shidham, D., Keenan, T., Arevalo, P., Li, W., Alemohammad, H., Olofsson, P., Hain, C., Kennedy, R., Zadrozny, B., Bell, D., Cavallaro, G., Watson, C.,

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Maskey, M., Ramachandran, R., and Moreno, J. B.: Prithvi-EO-2.0: A Versatile Multi-Temporal Foundation Model for Earth Observation Applications, https://doi.org/10.48550/arXiv.2412.02732, arXiv:2412.02732 [cs], 2025.

Tanhapour, M., Parajka, J., Schwaizer, G., Vreugdenhil, M., Kohnová, S., Hlavčová, K., Vỳleta, R., and Szolgay, J.: Potential of Sentinel-3 snow cover fraction data for improving hydrological simulations at the regional scale, Scientific Reports, 2026.

Thompson, A.: Utilization of Landsat monitoring capabilities for snowcover depletion analysis, Tech. rep., 1975.

Tseng, G., Fuller, A., Reil, M., Herzog, H., Beukema, P., Bastani, F., Green, J. R., Shelhamer, E., Kerner, H., and Rolnick, D.: Galileo: Learning global and local features in pretrained remote sensing models, arXiv e-prints, pp. arXiv–2502, 2025.

UNESCO World Water Assessment Programme: The United Nations World Water Development Report 2025, United Nations Educational, Scientific and Cultural Organization (UNESCO), France, ISBN 978-92-3-100743-9, https://doi.org/https://doi.org/10.54679/LHPJ5153, 2025.

Vapnik, V.: The nature of statistical learning theory, Springer science & business media, 2013.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I.: Attention is all you need, Advances in neural information processing systems, 30, 2017.

Vionnet, V., Leroux, N. R., Fortin, V., Abrahamowicz, M., Woolley, G., Mazzotti, G., Gaillard, M., Lafaysse, M., Royer, A., Domine, F., et al.: Enhancing simulations of snowpack properties in land surface models with the Soil, Vegetation and Snow scheme v2. 0 (SVS2), Geoscientific Model Development, 18, 9119–9147, 2025.

Wang, R., Ma, L., He, G., Johnson, B. A., Yan, Z., Chang, M., and Liang, Y.: Transformers for remote sensing: A systematic review and analysis, Sensors, 24, 3495, 2024.

Wang, X., Wang, J., Che, T., Huang, X., Hao, X., and Li, H.: Snow cover mapping for complex mountainous forested environments based on a multi-index technique, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11, 1433–1441, 2018.

Wang, Y., Albrecht, C. M., Braham, N. A. A., Mou, L., and Zhu, X. X.: Self-supervised learning in remote sensing: A review, IEEE Geoscience and Remote Sensing Magazine, 10, 213–247, 2022.

Warren, S. G.: Optical properties of snow, Reviews of Geophysics, 20, 67–89, 1982.

Wiesnet, D. and McGinnis, D.: Snow-extent mapping and lake ice studies using ERTS-1 MSS together with NOAA-2 VHRR, in: NASA. Goddard Space Flight Center 3d ERTS-1 Symp., Vol. 1, Sect. B, PAPER-W2, 1974.

WMO: The 2022 GCOS ECVs Requirements, GCOS 245, Global Climate Observing System (GCOS), Geneva, https://library.wmo.int/records/item/58111-the-2022-gcos-ecvs-requirements-gcos-245, 2022.

Wu, J., Zhang, X., Xiao, P., Jia, Y., Tang, B., and Liu, Y.: Multi-sensor spatiotemporal fusion for 30-m daily gapless snow cover mapping, Water Resources Research, 62, e2025WR041 155, 2026.

Xiao, A., Xuan, W., Wang, J., Huang, J., Tao, D., Lu, S., and Yokoya, N.: Foundation models for remote sensing and earth observation: A survey, IEEE Geoscience and Remote Sensing Magazine, 2025.

Xiong, Z., Wang, Y., Zhang, F., Stewart, A. J., Hanna, J., Borth, D., Papoutsis, I., Saux, B. L., Camps-Valls, G., and Zhu, X. X.: Neural plasticity-inspired multimodal foundation model for earth observation, arXiv preprint arXiv:2403.15356, 2024.

Yang, K., John, A., Shean, D., Lundquist, J. D., Sun, Z., Yao, F., Todoran, S., and Cristea, N.: High-resolution mapping of snow cover in montane meadows and forests using Planet imagery and machine learning, Frontiers in Water, 5, 1128 758, 2023.

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., et al.: ESA WorldCover 10 m 2021 v200, 2022.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Table A1. The Google Earth Engine (GEE) collection names for the input products used. Note that "-" should be replaced by an underscore.

Dataset / Product	GEE collection
Landsat 8	LANDSAT/LC08/C02/T1-TOA
Landsat 9	LANDSAT/LC09/C02/T1-TOA
Sentinel-1	COPERNICUS/S1-GRD
Sentinel-2	COPERNICUS/S2-HARMONIZED
Sentinel-3	COPERNICUS/S3/OLCI
MODIS	MODIS/061/MOD09GA
VIIRS	NASA/VIIRS/002/VNP09GA
ERA5	ECMWF/ERA5-LAND/DAILY-AGGR
DEM	COPERNICUS/DEM/GLO30
ESA WorldCover	ESA/WorldCover/v200

Zhang, C., Jiang, L., Pan, J., Yang, J., Wang, J., and Jin, Z.: Daily seamless 30-m fractional snow cover mapping via an adaptive Time-Series approach, International Journal of Applied Earth Observation and Geoinformation, 146, 105 068, 2026a.

Zhang, T.: Influence of the seasonal snow cover on the ground thermal regime: An overview, Reviews of Geophysics, 43, 2005.

Zhang, Y., Lu, W., Weng, L., Lin, H., and Xia, M.: CMF-SnowNet: Cross-Modal Fusion with Attention-Guided GAN and Temporal ConvNeXt for Long-Term Snow Cover Prediction, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2026b.

Zhao, Q., Hao, X., Che, T., Shao, D., Ji, W., Luo, S., Huang, G., Feng, T., Dong, L., Sun, X., et al.: Estimating AVHRR snow cover fraction by coupling physical constraints into a deep learning framework, ISPRS Journal of Photogrammetry and Remote Sensing, 218, 120–135, 2024.

Zhu, X. X., Xiong, Z., Wang, Y., Stewart, A. J., Heidler, K., Wang, Y., Yuan, Z., Dujardin, T., Xu, Q., and Shi, Y.: On the foundations of Earth foundation models, Communications Earth & Environment, 2026.

Zhu, Z., Wang, S., and Woodcock, C. E.: Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images, Remote Sens. Environ., 159, 269–277, 2015.

Appendix A: Data

A1 Input Data

A1.1 Google Earth Engine Usage

Table A1 reports the Google Earth Engine collections used. For satellite imagery, top-of-atmosphere products are prioritized, while surface reflectance products are used otherwise, depending on the availability on Google Earth Engine.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

A1.2 Cloud Generation

During training, the wide, thick global, thick local, and foggy cloud configurations provided at https://github.com/strath-ai/SatelliteCloudGenerator/blob/main/src/configs.py were randomly sampled. During testing, the following configuration to simulate greater cloud coverage was used:

```
FULL_CONFIG={'min_lvl': [0.5,0.9],
             'max_lvl': 1.0,
             'const_scale':True,
             'decay_factor':1.0,
             'clear_threshold':0.0,
             'locality_degree':1,
             'cloud_color':[True,False],
             'channel_offset':2,
             'blur_scaling':2
            }
```

Clouds were not generated on Sentinel-1, ERA5, topography, and land cover inputs since these are not sensitive to clouds.

Because the appearance of a cloud depends on the wavelength of the image channel in question, magnitudes specific to each channel were generated and applied during cloud simulation (Czerkawski et al., 2023). These magnitudes were derived using clear and cloudy images of each sensor, which were sampled in snow-covered mountain areas to approximate realistic conditions for the model. Relationships were calculated using quantile values as described in Czerkawski et al. (2023). Further details can be found in SnowGalileo code.

Google Earth Engine provides some sensor products in reflectance and others in digital number format. While the native format of the images was used for training purposes, all images were temporarily scaled to the reflectance range for cloud simulation. The scaling factor was determined for each sensor using the product's scaling specifications in Google Earth Engine.

At this stage, cloud shadows were not simulated due to the effects of topography. However, this is relevant for realistic cloud simulations in mountainous terrain and should be addressed by future work.

A1.3 Export and Pre-processing Details

The following section details the export of the input data and the subsequent processing steps involved.

For the pre-training data, Tseng et al. (2025) was followed and input tiles of 1×1 km centered around the respective pre-training coordinate exported in WGS84 projection. For each pre-training point, three input time series were randomly selected, with one instance drawn from each of the following windows: 1 October–15 December, 16 December–28 February, and 1 March–30 June. Each input time series contains 8 days. Sampling starts after the onset of Landsat 9 and spans 16 Dec

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

2021–28 Feb 2024. Although the sampling strategy is tailored to the Northern Hemisphere, it can be adapted for the Southern Hemisphere by adjusting the seasonal boundaries. For the fine-tuning data, input tiles and the preceding time series were extracted in UTM projection, based on the region boundaries of the Landsat target data. To address sampling uncertainties, all pre-training images were cropped to a fixed grid of 100×100 pixels after export, and all fine-tuning data were cropped to the bounds of the Landsat target tiles.

For export purposes, all input data is derived as stacked GeoTIFFs from Google Earth Engine and resampled to a resolution of 10 m. In order to reduce computation during model processing, MODIS and 500 m-VIIRS channels are subsequently downsampled to a fixed 2×2 grid and Sentinel-3 data to a fixed 5×5 grid. These grid sizes are close to the native spatial resolution of the products and divide the source height and width evenly. All input products with a spatial resolution equal to or greater than the 1×1 km tile size are downsampled to a fixed 1×1 grid. This applies to ERA5 and the 1 km-VIIRS channels.

Data validity masks are created to identify missing input data due to infrequent coverage. Additionally, values beyond a valid range are masked that are either detected as outliers when investigating data distributions or are documented as a no-data value in the product specifications. More details are provided with the SnowGalileo code.

NDSI is computed from MODIS B4 and B6, and NDVI from MODIS B2 and B1. Geographic coordinates are transformed from latitude and longitude to a Cartesian representation.

A1.4 Input Normalization Statistics

A2 Target Data

A2.1 Landsat Fractional Snow Cover Scenes

A3 Sampling

A3.1 Pre-Training Sampling Details

Mountain regions were derived from the Global Mountain Biodiversity Assessment (GMBA) Inventory (Snethlage et al., 2022) and filtered to include only “mountain ranges with well-recognized names”, defined by non-empty Level_04 attributes and empty Level_05 attributes. To clarify, the GMBA first uses a DEM: Mountain ranges must meet elevation and ruggedness criteria to qualify as mountain ranges in the GMBA. Only afterward does the GMBA assign a “well-recognized” name to each mountain range that meets the DEM criteria. Thus, this filtering step indeed ensures the selection of large and prominent mountain systems. After this filtering step the initial set of 8,327 polygons was reduced to 875. These polygons were subsequently intersected with the GlobalSnowPack mean snow cover duration dataset (Dietz and Roessler, 2025), retaining only regions with an average snow cover duration of at least 50 days per year, which further reduced the dataset to 190 polygons. To focus on major mountain systems, an additional minimum area threshold of 35,000 km^2 was applied, resulting in a final set of 87 polygons. Each polygon was then rasterized according to its corresponding MODIS tile at the native spatial resolution of 461 $\times$ 461 m. Almost equal elevation bins were created by grouping pixels into three elevations (0–1000 m, 1000–2000 m, and

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Table A2. Per-channel normalization values used to train the SnowGalileo and baseline models. The statistics were computed based on the entire pre-training dataset. Input products that are one-hot encoded (ESA Worldcover), have a fixed bound of -1 to 1 (NDSI and NDVI), and location values are excluded from normalization and are therefore not listed in the table.

Product	Channel	mean	standard deviation
Sentinel-1	VV	-10.577118182546462	5.549875127263841
	VH	-17.413747730102024	5.52247086353095
	incidence angle	37.87672146761802	6.246272206074418
Sentinel-2	B2	4353.812924442672	2659.3241333545625
	B3	3882.4714859394135	2535.7876350987553
	B4	4143.164481541021	2898.523478436185
	B8	4553.686576667571	2815.647414817182
	B11	1984.6073453947486	1410.2812943148137
	B12	1691.285757395636	1215.6321478201032
Landsat 8/9	B2	0.40983663406386944	0.26012964358343504
	B3	0.36108484889534986	0.25280879395539274
	B4	0.3769700502033976	0.28397523263932617
	B5	0.44525105642564017	0.28897991416969265
	B6	0.18602104210359371	0.13838916959794115
	B7	0.15829798053673816	0.12082495412160361
Sentinel-3	Oa17	87.26304896500532	60.57194781681194
	Oa21	34.9697076889377	26.933913691634007
MODIS	B1	4643.811665095236	2890.8877149766913
	B2	5026.624925509647	2493.183929693802
	B3	4610.635207942517	2908.8608451539467
	B4	4655.987361096781	2887.4632988582116
	B5	3961.8422347875894	1880.7297166295346
	B6	2449.3430132621206	1439.2873447260633
	B7	1529.3952947142036	1056.4458841938367
VIIRS	I1	0.4489128557701946	0.2972102527320826
	I3	0.23627941508602207	0.14612073456521718
	M5	0.45816973215305734	0.29438185279294077
	M7	0.49394365576982674	0.24593392161424962
	M10	0.23513348776524087	0.13710222779644296
	M11	0.23887112761717894	0.15406434796497845

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Table A3. Normalization statistics continued.

Product	Channel	mean	standard deviation
DEM	elevation	1310.152802906603	785.4857516714201
	slope	17.89735732318929	11.366248795647541
	aspect	177.6182724630746	104.74749922704228
ERA5	skin temperature	266.840673083913	12.981804159101426
	temperature 2m	268.4170796454643	11.94776532518963
	total precipitation sum	0.002648917660436159	0.005732518625909269
	u component of wind	0.38598109602431546	1.331282418647898
	v component of wind	0.15003408501065887	1.2581386859148131

>2000 m), yielding 11,488,417, 10,283,177, and 11,136,673 pixels, respectively. Within each elevation class, random sampling points were generated at the center of each pixel, with the number of points weighted proportionally to the total area of each mountain range. To focus on latitudes where the Rocky Mountains and Alps are located, the sampling points were clipped to latitudes between 42 and 60 degrees, resulting in 51,213 points. Pre-training samples were acquired between late 2021 and early 2024, starting with the availability of Landsat 9. For each point, a random week was sampled within the snow onset period of the Northern Hemisphere (1 October – 15 December), high winter (16 December – 28 February) and the snowmelt period (1 March – 30 June). In total, 152,559 pre-training samples could be exported successfully with Google Earth Engine (Gorelick et al., 2017).

Appendix B: SnowGalileo Details

Here, SnowGalileo is described in more detail.

B1 Input Grouping, Tokenization, and Encoding

Tokenization is the process of dividing input sources into a series of tokens. In classic ViT approaches, this is achieved by dividing an image of size $H \times W$ into smaller patches with a patch size of P, where the patch size is defined as the number of pixels on each side of a patch. The patch size usually defines a trade-off between granularity and computational feasability, with smaller patch sizes achieving higher accuracy but requiring more compute (Dosovitskiy et al., 2020).

As SnowGalileo's remote sensing input sources cover a ground area of 1×1 km but differ in their spatial resolution, the magnitude of image height and image width vary per product. For example, for data from the space-time-varying high-resolution modality group, it is $H = 100$, $W = 100$ and $P = 10$ is used. In this case, each token of this group corresponds to the 100 m target resolution of one SnowGalileo output. For the space-time-varying low resolution group, it is $H = 2$ and $W = 2$ and $P = 1$ is used. Similarly, $P = 1$ is used for the space-time-varying medium resolution group, for which $H = 5$ and $W = 5$.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Table A4. Landsat training scenes (distributed across the Northern Hemisphere).

#	Training Scenes (globally)
1	LC08_L2SP_151034_20220626_20220706_02_T1_snowclass_QA.tif
2	LC08_L2SP_217015_20211011_20211019_02_T1_snowclass_QA.tif
3	LC08_L2SP_190027_20200909_20200919_02_T1_snowclass_QA.tif
4	LC09_L2SP_018025_20220918_20220920_02_T1_snowclass_QA.tif
5	LC08_L2SP_190027_20210421_20210430_02_T1_snowclass_QA.tif
6	LC09_L2SP_029005_20230326_20230328_02_T1_snowclass_QA.tif
7	LC08_L2SP_190028_20210421_20210430_02_T1_snowclass_QA.tif
8	LC09_L2SP_037029_20220923_20220925_02_T1_snowclass_QA.tif
9	LC08_L2SP_191027_20210223_20210303_02_T1_snowclass_QA.tif
10	LC09_L2SP_071018_20220601_20220603_02_T1_snowclass_QA.tif
11	LC08_L2SP_191028_20201018_20201105_02_T1_snowclass_QA.tif
12	LC09_L2SP_081012_20220927_20221003_02_T1_snowclass_QA.tif
13	LC08_L2SP_191028_20201119_20210315_02_T1_snowclass_QA.tif
14	LC09_L2SP_133033_20220604_20220606_02_T1_snowclass_QA.tif
15	LC08_L2SP_192027_20201009_20201016_02_T1_snowclass_QA.tif
16	LC09_L2SP_136040_20230119_20230313_02_T1_snowclass_QA.tif
17	LC08_L2SP_192027_20201025_20201106_02_T1_snowclass_QA.tif
18	LC09_L2SP_140041_20230216_20230310_02_T1_snowclass_QA.tif
19	LC08_L2SP_194029_20201124_20210316_02_T1_snowclass_QA.tif
20	LC09_L2SP_142024_20230318_20230320_02_T1_snowclass_QA.tif
21	LC08_L2SP_194029_20210316_20210328_02_T1_snowclass_QA.tif
22	LC09_L2SP_145039_20230102_20230315_02_T1_snowclass_QA.tif
23	LC08_L2SP_195029_20200928_20201006_02_T1_snowclass_QA.tif
24	LC09_L2SP_146038_20220802_20220804_02_T1_snowclass_QA.tif
25	LC08_L2SP_195029_20210424_20210501_02_T1_snowclass_QA.tif
26	LC09_L2SP_147037_20220606_20220608_02_T1_snowclass_QA.tif
27	LC08_L2SP_195030_20200912_20200919_02_T1_snowclass_QA.tif
28	LC09_L2SP_147038_20220606_20220608_02_T1_snowclass_QA.tif
29	LC08_L2SP_195030_20200928_20201006_02_T1_snowclass_QA.tif
30	LC09_L2SP_152035_20230119_20230313_02_T1_snowclass_QA.tif

As in Tseng et al. (2025), SnowGalileo creates tokens not only for each image patch, but also each timestep (if the input source varies over time), and each channel group. One channel group comprises spectrally similar channels of one modality (see Table B2). The number of resulting tokens per input source is listed in Table B1.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Table A5. Landsat training scenes continued (distributed across the Northern Hemisphere).

#	Training Scenes (globally)
31	LC08_L2SP_195030_20210323_20210402_02_T1_snowclass_QA.tif
32	LC09_L2SP_165027_20230303_20230307_02_T1_snowclass_QA.tif
33	LC08_L2SP_195030_20210408_20210416_02_T1_snowclass_QA.tif
34	LC09_L2SP_167035_20230301_20230308_02_T1_snowclass_QA.tif
35	LC08_L2SP_196029_20200903_20200918_02_T1_snowclass_QA.tif
36	LC09_L2SP_169034_20230227_20230308_02_T1_snowclass_QA.tif
37	LC08_L2SP_196029_20210415_20210424_02_T1_snowclass_QA.tif
38	LC09_L2SP_171031_20221223_20230317_02_T1_snowclass_QA.tif
39	LC08_L2SP_196030_20201005_20201016_02_T1_snowclass_QA.tif
40	LC09_L2SP_228011_20220629_20220701_02_T1_snowclass_QA.tif
41	LC08_L2SP_196030_20210226_20210304_02_T1_snowclass_QA.tif
42	LC08_L2SP_196030_20210314_20210328_02_T1_snowclass_QA.tif

Table A6. Landsat test scenes (Canadian Rockies).

#	Test Scenes Rockies
1	LC09_L2SP_044024_20220807_20220810_02_T1_snowclass_QA.tif
2	LC09_L2SP_052018_20220815_20220817_02_T1_snowclass_QA.tif
3	LC09_L2SP_048022_20220819_20220822_02_T1_snowclass_QA.tif
4	LC09_L2SP_055016_20220601_20220603_02_T1_snowclass_QA.tif
5	LC09_L2SP_048025_20220920_20220922_02_T1_snowclass_QA.tif
6	LC09_L2SP_055020_20220820_20220823_02_T1_snowclass_QA.tif
7	LC09_L2SP_049024_20220810_20220812_02_T1_snowclass_QA.tif
8	LC09_L2SP_056020_20220912_20220915_02_T1_snowclass_QA.tif

Each token contains $P \times P \times C$ pixels, where C is the number of channels, and P and C depend on the token's channel and modality group. Through linear projection, each token is embedded with a dimension of 192, determined by the ViT-Tiny model that is used for encoding (Dosovitskiy et al., 2020). Additional encodings that communicate information such as the original image location remain unmodified from Tseng et al. (2025). The token embeddings and encodings are fed as a 1D-sequence into the ViT encoder.

B2 Prediction

The aim of SnowGalileo is to produce FSC values at a spatial resolution of 100 m. To achieve this, SnowGalileo's encoder is used to compute the representations of a 1×1 km ground area. Then, a prediction head is attached to project these representations onto 10×10 images, with each pixel representing an area of 100×100 m. To this end, the representations produced by the

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Table A7. Landsat test scenes (Swiss Alps).

#	Test Scenes Switzerland
1	`LC08_L2SP_193027_20210325_20210402_02_T1_snowclass_QA.tif`
2	`LC08_L2SP_195028_20201030_20201106_02_T1_snowclass_QA.tif`
3	`LC08_L2SP_193028_20201117_20210315_02_T1_snowclass_QA.tif`
4	`LC08_L2SP_195028_20210118_20210306_02_T1_snowclass_QA.tif`
5	`LC08_L2SP_194027_20210228_20210311_02_T1_snowclass_QA.tif`
6	`LC08_L2SP_196028_20200903_20200918_02_T1_snowclass_QA.tif`
7	`LC08_L2SP_194028_20201226_20210310_02_T1_snowclass_QA.tif`
8	`LC08_L2SP_196028_20210415_20210424_02_T1_snowclass_QA.tif`

Table B1. The modality groups used by SnowGalileo. Each modality group contains input sources with a similar spatiotemporal resolution. For instance, data with high spatial resolution and temporal variability belongs to the space-time-varying high-resolution group, whereas data that varies over time but has a spatial resolution equal to or lower than the input tile size belongs to the time-varying group. The number of tokens corresponds to the potential number of tokens resulting from the tokenization stage for one input tile (number of patches per height $\times$ number of patches per width $\times$ number of timesteps $\times$ number of channel groups). In practice, however, the number of actual tokens processed depends on the product's revisit time. For example, in the case of Landsat, $10 \times 10 \times 8 \times 6$ tokens will be produced in the tokenization stage, but SnowGalileo will only process $10 \times 10 \times 1 \times 6$ tokens in most cases, since the other timesteps provide no valid data and will be masked out.

Group name	Input products	Spatial resolution	Number of tokens
space-time-varying high-resolution	Sentinel-1	10 m	$10 \times 10 \times 8 \times 1$
	Sentinel-2	10 – 20 m	$10 \times 10 \times 8 \times 3$
	Landsat	30 m	$10 \times 10 \times 8 \times 3$
space-time-varying medium resolution	Sentinel-3	300 m	$5 \times 5 \times 8 \times 1$
space-time-varying low resolution	MODIS	500 m	$2 \times 2 \times 8 \times 3$
	VIIRS-500 m	500 m	$2 \times 2 \times 8 \times 2$
	NDSI, NDVI	500 m	$2 \times 2 \times 8 \times 2$
space-varying	Copernicus DEM	10 m	$10 \times 10 \times 1 \times 1$
	ESA Worldcover	10 m	$10 \times 10 \times 1 \times 1$
time-varying	VIIRS-1 km	1 km	$1 \times 1 \times 8 \times 3$
	ERA5	11132 m	$1 \times 1 \times 8 \times 1$
static	geo-coordinates in cartesian form (x, y, z)		$1 \times 1 \times 1 \times 1$

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Table B2. The channel groups used by SnowGalileo. In the case of sensors, a channel group contains spectrally similar bands, while in the case of derived remote sensing products, it contains the products themselves.

Channel group	Bands
Sentinel-1	VV, VH, angle
Sentinel-2 RGB	B2, B3, B4
Sentinel-2 NIR	B8
Sentinel-2 SWIR	B11, B12
Landsat RGB	B2, B3, B4
Landsat NIR	B5
Landsat SWIR	B6, B7
Sentinel-3 NIR	Oa17, Oa21
MODIS RGB	B1, B2, B3
MODIS NIR	B4
MODIS SWIR	B5, B6, B7
VIIRS-500 m RGB	I1
VIIRS-500 m VNIR	I3
NDSI	NDSI
NDVI	NDVI
VIIRS-1 km RGB	M5, M7
VIIRS-1 km VNIR	M10
VIIRS-1 km SWIR	M11
ERA5	all ERA5 variables used
DEM	elevation, slope, aspect
Landcover	one-hot encoded ESA Worldcover classes
Location	x, y, z

encoder are grouped according to the position of each output pixel, repeating tokens that represent a lower spatial resolution. This ensures that each output pixel is informed by all input sources (if they are not masked due to missing coverage of the respective sensor). For each spatial position, the grouped representation is then projected onto the output using a simple linear layer. A sigmoid layer is added to ensure the values are clamped between 0 and 1, representing FSC values between 0 % and 100 %. Since the tokens were cross-related through self-attention during the encoding process, each output position can be informed by neighboring positions.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

B3 Pre-training Process

In a self-supervised learning process, the model is trained using unlabeled data, with labels being created from the data itself. SnowGalileo relies on the *masked autoencoding* approach (He et al., 2022), which involves masking parts of an image and using the encoder to embed the visible parts. A decoder is then used to reconstruct the original image based on these embeddings. This differs from the approach taken in Tseng et al. (2025), where a dual learning strategy is employed for greater flexibility. The simpler masked autoencoding approach was chosen because the setting is consistent (snow in mountains).

Masked autoencoding requires identifying which tokens should be used as inputs and which as reconstruction targets. Galileo deploys two masking strategies: *unstructured masking*, which uniformly and randomly masks out tokens, and *structured masking*, which consistently masks out specific groups of data (e.g. a consistent timestep across all pixels within an image, or a consistent pixel across all timesteps within a time series). For SnowGalileo, only unstructured masking was used because it excels in segmentation tasks (Tseng et al., 2025) and because it was easier to implement with the multi-resolution adjustments. During pre-training encoding, 0.9 % of all tokens were masked out and during pre-training decoding, 0.5 % of all tokens were reconstructed. These are the ratios used by Galileo. However, unlike Galileo, missing data was also masked out, e.g. due to low revisit times, as described in Sect. 3.1.2. This means that the effective ratio of masked tokens was higher. Missing data was neither seen by the encoder, nor by the decoder.

Despite Galileo, which uses different patch sizes for training, a fixed patch size setting is used for SnowGalileo since the output resolution remains constant (100 m).

Bfloat16 precision was used following Tseng et al. (2025) and the AdamW optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ and gradient clipping. Images were randomly augmented using vertical and horizontal flipping and 90-degree rotations. A batch size of 128 was used with gradient accumulation over 4 steps, resulting in an effective batch size of 512. Pre-training lasted for 100 epochs using a masked reconstruction loss based on the Smooth L1 loss. The maximal learning rate was set to 0.003 with 30 warmup epochs and an exponential cosine decay cooldown following Tseng et al. (2025).

B4 Fine-tuning Process

For fine-tuning, the decoder used during pre-training was discarded and the prediction head introduced in B2 was attached instead.

We used float32 precision and the AdamW optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.9$ and gradient clipping. We randomly augmented images using vertical and horizontal flipping and 90-degree rotations, and used a batch size of 8. We fine-tuned for 100 epochs using the mean squared error loss. The maximal learning rate was set to 0.006 with 10 warmup epochs and an exponential cosine decay cooldown following Tseng et al. (2025). The hyperparameter selection procedure and all final hyperparameters are listed in D.

After fine-tuning, the model is ready to be used for inference, that is, predicting FSC values for a given 1×1 km tile.

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Appendix C: Baseline Details

The following describes the forward-filling imputation strategy of the baseline models in detail. The aim here is to provide the baseline models with the same multi-sensor data as SnowGalileo. Since some of the sensors have a revisit time less frequent than daily, the methods need to be able to handle missing information. However, some of the baseline models that are used (MLP and SVR) cannot handle missing values natively. Therefore, a forward filling strategy is used that fills missing values based on the last valid value in the time series on a per-pixel and per-channel basis. If there is no prior data available (e.g. if the missing data point occurs at the beginning of the time series), for each pixel, the median value over the entire time series is calculated for the given channel. If no data is available for the channel, the median value across the spatial image is used. If a specific channel group contains no data at all, the per-channel mean of the pre-training data set is used.

Furthermore, each data point is enriched with a variable containing information about when the data was acquired, to enable the baseline models to integrate temporal information. This value is set to 0 when the data was acquired on the same day as the current timestep, to 1 when it was acquired the day before, and so on. When filling with a median or normalized value, this variable is set to a -1 placeholder.

Appendix D: Hyperparameter Tuning

In machine learning models, hyperparameters are the parameters that can be selected manually before training. As many of these parameters often depend on each other, hyperparameter tuning is used to select the optimal hyperparameter configuration. This involves testing different hyperparameter configurations in subsequent model runs. The following section specifies the tuning and final hyperparameter setups of the models.

D1 SnowGalileo

D1.1 Pre-training

Due to the high computational requirements of pre-training SnowGalileo, extensive hyperparameter tuning for this training stage is infeasible. Therefore, the same hyperparameter setup was used for pre-training SnowGalileo as for the original Galileo model. These hyperparameters are:

```
{"training": {
"max_lr": 0.003,
"num_epochs": 100,
"batch_size": 128,
"effective_batch_size": 512,
"warmup_epochs": 30,
"final_lr": 1e-06,
```

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026


```
    "weight_decay": 0.01,
    "grad_clip": true,
    "betas": [0.9, 0.999],
    "augmentation": {"flip+rotate": true},
    "encode_ratio": 0.1,
    "decode_ratio": 0.5,
    "loss_type": "MAE",
    "normalization": "std"
    },
    "model": {
        "encoder": {
            "embedding_size": 192,
            "depth": 12,
            "num_heads": 3,
            "mlp_ratio": 4,
            "max_sequence_length": 24,
            "freeze_projections": false,
            "drop_path": 0.1
        },
        "decoder": {
            "depth": 4,
            "num_heads": 3,
            "mlp_ratio": 4,
            "max_sequence_length": 24,
            "learnable_channel_embeddings": true,
            "embedding_size": 192,
            "use_fast_attn": true
            }
        }
    }
```

Here, `max_lr` specifies the maximum learning rate that is achieved after `warmup_epochs` number of epochs. Afterwards, the learning rate follows exponential cosine decay cooldown until it reaches the final learning rate, `final_lr`. The effective batch size is the number of training samples that are passed before gradient updates are computed, and is higher than the actual batch size due to gradient accumulation. `betas` are the beta values for the AdamW optimizer. `std` normalization identifies

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

the 2σ-clipped min-max normalization procedure described in Sect. 3.1.2. `encode_ratio` and `decode_ratio` describe the number of tokens masked during pre-training (see Sect. B3).

D1.2 Fine-tuning

For fine-tuning, hyperparameters were tuned using a subset of 3,000 samples from the fine-tuning dataset. Weights & Biases (Biewald, 2020) was used to sweep hyperparameters with 100 sweep configuration trials. Each training run lasted for 25 epochs, during which the following hyperparameters were optimized:

```
snowgalileo_sweep_configuration = {
    "method": "random",
    "metric": {"goal": "maximize", "name": "model.regression.r2"},
    "parameters": {
        "learning_rate": {"values": [1e-5, 3e-5, 6e-5, 1e-4, 3e-4,
        6e-4, 1e-3, 3e-3, 6e-3]},
        "lr_schedule": {"values": [True, False]},
        "warmup_fraction": {"values": [0.0, 0.1, 0.2, 0.3, 0.5]},
        "batch_size": {"values": [8, 4, 2]},
        "optimizer": {"values": ["Adam", "SGD"]},
        "weight_decay": {"values": [0, 1e-5, 1e-3]},
        "augmentation": {"values": [True, False]},
        "sigmoid_slope": {"values": [0.01, 0.1, 0.5, 1.0]},
        "adam_beta_2": {"values": [0.9, 0.95, 0.98, 0.99, 0.995, 0.999]},
    },
}
```

Here, `lr_schedule` defines whether an exponential cosine decay cooldown with a maximum learning rate of `learning_rate` should be used, or whether the learning rate should remain constant. `warmup_fraction` refers to the ratio of warmup epochs to the total number of epochs, which is fixed to 100. `sigmoid_slope` scales the input to the sigmoid function, which constrains the final FSC output values to the range `[0,1]`.

This results in the following final hyperparameter set for SnowGalileo.

```
"hyperparameters_snowgalileo": {
    "augmentation": true,
    "batch_size": 8,
    "weight_decay": 0.00001,
    "sigmoid_slope": 0.5,
```

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026


```
        "lr_schedule": true,
        "learning_rate": 0.006,
        "optimizer": "Adam",
        "adam_beta_2": 0.9,
        "warmup_fraction": 0.1,
    },
```

D2 Baselines

As with fine-tuning, Weights & Biases (Biewald, 2020) was used to tune hyperparameters of the baseline models on a subset of 3,000 samples. Hyperparameters were swept with 100 sweep configuration trials.

For random forest, the following hyperparameters were tuned:

```
rf_sweep_configuration = {
    "method": "random",
    "metric": {"goal": "maximize", "name": "r2"},
    "parameters": {
        "n_estimators": {"values": [50, 100, 200, 300, 400, 500]},
        "normalization": {"values": [None, "std"]},
        "max_features": {"values": ["feature_dependent", "sqrt"]},
        "min_samples_leaf": {"values": [1, 2, 5]},
        "max_depth": {"values": [None, 10, 20, 30]},
        "min_samples_split": {"values": [2, 5, 10]},
    },
}
```

Here, `n_estimators` refers to the number of trees in the forest, `max_features` refers to the maximum number of features to be considered at each split, `min_samples_leaf` to the minimum number of samples required for a leaf node, `max_depth` to the maximum depth of the tree, and `min_samples_split` to the minimum number of samples needed for a node split. For `max_features`, the value `feature_dependent` was tested, referring to $\lceil p/3 \rceil$, where p refers to the number of features, following (Kuter, 2021). `sqrt` was also tested, which is defined as the square root of the number of features.

This results in the following final hyperparameter set for the random forest.

```
"hyperparameters_rf": {
    "n_estimators": 400,
    "min_samples_leaf": 2,
```

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026


```
        "min_samples_split": 2,
        "max_depth": 30,
        "max_features": "feature_dependent",
        "normalization": "std",
    },
```

For the support vector regressor, the following hyperparameters were tuned, following the analysis of (Kuter, 2021):

```
    svr_sweep_configuration = {
        "method": "random",
        "metric": {"goal": "maximize", "name": "r2"},
        "parameters": {
            "kernel": {"values": ["linear", "poly", "rbf"]},
            "C_exponent": {"values": [-15, -10, -5, 0, 5, 10, 15]},
            "degree": {"values": [2, 3]},
            "gamma_exponent": {"values": [-5, 0, 5]},
            "normalization": {"values": [None, "std"]},
            "epsilon": {"values": [0.01, 0.1, 0.5, 1.0]},
            "max_iter": {"values": [500, 1000, 5000, 10000]},
            "bagging": {"values": [True, False]},
        },
    }
```

Here, `kernel` defines the kernel function and `C_exponent` specifies the degree of the regularization parameter C used (with base 2). `degree` specifies the degree of the polynomial, if the polynomial kernel function is used and `gamma_exponent` defines the degree of the RBF kernel (with base 2), if the RBF kernel is used. Different values for epsilon were also tested, and whether creating an ensemble using bagging helps.

This results in the following final hyperparameter set for the support vector regressor:

```
    "hyperparameters_svr": {
        "kernel": "rbf",
        "epsilon": 0.1,
        "bagging": true,
        "C_exponent": 0,
        "degree": 3,
        "gamma_exponent": -5,
        "max_iter": 5000
```

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026


```
    },
```

For the MLP, the architecture was set by subsequently adding layers with a different number of neurons. Three hidden layers with 256, 128 and 64 neurons respectively produced optimal results. For remaining hyperparameters, the following were tuned:

```
mlp_sweep_configuration = {
    "method": "random",
    "metric": {"goal": "maximize", "name": "r2"},
    "parameters": {
        "learning_rate_init": {"values": [0.0001, 0.001, 0.01, 0.1]},
        "activation": {"values": ["logistic", "tanh", "relu"]},
        "normalization": {"values": [None, "std"]},
        "alpha": {"values": [1e-5, 1e-4, 1e-3]},
        "batch_size": {"values": [64, 128, 256, "auto"]},
    },
}
```

Here, `activation` is the activation function for the hidden layers and `alpha` refers to the strength of the L2 regularization term. As optimizer, the Adam default was used.

This results in the following final hyperparameter set for the MLP regressor.

```
"hyperparameters_mlp": {
    "learning_rate_init": 0.001,
    "activation": "logistic",
    "max_iter": 1000,
    "hidden_layer_sizes": [256, 128, 64],
    "alpha": 0.00001,
    "batch_size": 128
},
```

Appendix E: Additional Results

E1 Summary Statistics

E1.1 Error Reduction

Percentage reduction in RMSE is computed as

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Table E1. Percentage error reduction in RMSE of the pre-trained SnowGalileo compared to the randomly initialized version. All values show model performance on a test set with equivalent input data setups like the training set.

Region	**Percentage Error Reduction**				**Mean**
	Clear-HR	Cloudy-HR	Clear-no HR	Cloudy-no HR	
Canadian Rockies	6.60%	25.39%	27.45%	18.37%	19.45%
Swiss Alps	6.77%	8.73%	0.38%	0.99%	4.22%

$$\text{Percentage Error Reduction} = \frac{\mathrm{RMSE}_{\mathrm{random}} - \mathrm{RMSE}_{\mathrm{pre-trained}}}{\mathrm{RMSE}_{\mathrm{random}}} \times 100\%, \tag{E1}$$

where $\mathrm{RMSE}_{\mathrm{random}}$ is the RMSE of the randomly initialized SnowGalileo and $\mathrm{RMSE}_{\mathrm{pre-trained}}$ of the pre-trained version, computed for each input data setup separately.

E1.2 Performance Degradation for Challenging Input Data Setups

The performance degradation in RMSE for challenging input data setups is computed as

$$\text{Performance Degradation} = \frac{\mathrm{RMSE}_{\mathrm{cloudy-HR}} + \mathrm{RMSE}_{\mathrm{clear-no-HR}} + \mathrm{RMSE}_{\mathrm{cloudy-no-HR}}}{3} - \mathrm{RMSE}_{\mathrm{clear-HR}}, \tag{E2}$$

where $\mathrm{RMSE}_{\mathrm{random}}$ is the RMSE of the randomly initialized SnowGalileo and $\mathrm{RMSE}_{\mathrm{pre-trained}}$ of the pre-trained version, computed for each input data setup separately.

E2 Additional Metrics

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Table E2. Average performance degradation reported as RMSE gain for challenging input data setups compared to the clear-HR setup. Clear-HR refers to clear sky conditions with high-resolution imagery available on the day of prediction. Challenging input data setups comprise cloudy-HR, clear-no HR, and cloudy-no HR. Cloudy-HR refers to the same condition but with clouds present on the day of prediction. Clear-no HR refers to clear conditions without HR imagery available, and cloudy-no HR refers to cloudy conditions without HR imagery available. All values show model performance on a test set with equivalent input data setups like the training set.

Region	**Model**	**Performance degradation (RMSE gain)**
Canadian Rockies	RF	0.118
	MLP	0.101
	SVR	0.002
	SnowGalileo random	0.125
	SnowGalileo pre-trained	0.077
Swiss Alps	RF	0.157
	MLP	0.136
	SVR	0.001
	SnowGalileo random	0.132
	SnowGalileo pre-trained	0.133

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Table E3. SnowGalileo and baseline model results reported with respect to R2 value (mean ± sample standard deviation). All values show model performance on a test set with equivalent input data setups like the training set. Clear-HR refers to clear sky conditions with high-resolution imagery available on the day of prediction. Cloudy-HR refers to the same condition but with clouds present on the day of prediction. Clear-no HR refers to clear conditions without HR imagery available, and cloudy-no HR refers to cloudy conditions without HR imagery available. The pre-trained SnowGalileo model consistently outperforms all other models with respect to R2 (higher values are better), with the randomly initialized SnowGalileo model coming in second.

Region	**Model**	**R2**			
		Clear-HR	Cloudy-HR	Clear-no HR	Cloudy-no HR
Canadian Rockies	RF	0.902 (std <0.001)	0.784 (std <0.001)	0.635 (±0.002)	0.611 (±0.002)
	MLP	0.880 (±0.003)	0.748 (±0.013)	0.631 (±0.011)	0.679 (±0.019)
	SVR	-0.015 (±0.011)	-0.030 (±0.005)	-0.016 (±0.011)	-0.027 (±0.001)
	SnowGalileo random	0.950 (±0.007)	0.834 (±0.022)	0.709 (±0.040)	0.732 (±0.012)
	SnowGalileo pre-trained	**0.960 (±0.002)**	**0.906 (±0.029)**	**0.848 (±0.008)**	**0.822 (±0.007)**
Swiss Alps	RF	0.857 (std <0.001)	0.689 (std <0.001)	0.411 (±0.002)	0.320 (±0.010)
	MLP	0.831 (std <0.001)	0.619 (±0.025)	0.479 (±0.046)	0.396 (±0.049)
	SVR	-0.004 (±0.005)	-0.011 (±0.003)	-0.004 (±0.005)	-0.009 (std <0.001)
	SnowGalileo random	0.918 (±0.003)	0.756 (±0.005)	0.678 (±0.018)	0.575 (±0.052)
	SnowGalileo pre-trained	**0.929 (±0.003)**	**0.796 (±0.016)**	**0.680 (±0.012)**	**0.584 (±0.012)**

https://doi.org/10.5194/egusphere-2026-3833
Preprint. Discussion started: 8 September 2026

Table E4. SnowGalileo and baseline model results reported with respect to mean absolute error (mean ± sample standard deviation). All values show model performance on a test set with equivalent input data setups like the training set. Clear-HR refers to clear sky conditions with high-resolution imagery available on the day of prediction. Cloudy-HR refers to the same condition but with clouds present on the day of prediction. Clear-no HR refers to clear conditions without HR imagery available, and cloudy-no HR refers to cloudy conditions without HR imagery available. The pre-trained SnowGalileo model consistently outperforms all other models with respect to mean absolute error (lower values are better), with the randomly initialized SnowGalileo model coming in second.

Region	**Model**	**Mean Absolute Error**			
		Clear-HR	Cloudy-HR	Clear-no HR	Cloudy-no HR
Canadian Rockies	RF	0.077 (std <0.001)	0.164 (std <0.001)	0.185 (std <0.001)	0.189 (std <0.001)
	MLP	0.073 (±0.001)	0.120 (±0.002)	0.166 (±0.008)	0.149 (±0.004)
	SVR	0.467 (±0.002)	0.469 (std <0.001)	0.467 (±0.002)	0.469 (std <0.001)
	SnowGalileo random	0.039 (±0.003)	0.086 (±0.006)	0.118 (±0.010)	0.110 (±0.003)
	SnowGalileo pre-trained	**0.035 (±0.001)**	**0.057 (±0.008)**	**0.074 (±0.003)**	**0.080 (±0.003)**
Swiss Alps	RF	0.103 (std <0.001)	0.191 (std <0.001)	0.295 (±0.001)	0.325 (±0.002)
	MLP	0.096 (std <0.001)	0.159 (±0.008)	0.227 (±0.004)	0.240 (±0.014)
	SVR	0.450 (std <0.001)	0.451 (std <0.001)	0.450 (std <0.001)	0.451 (std <0.001)
	SnowGalileo random	0.059 (std <0.001)	0.124 (±0.001)	0.144 (±0.003)	0.175 (±0.010)
	SnowGalileo pre-trained	**0.053 (±0.001)**	**0.102 (±0.003)**	**0.133 (±0.003)**	**0.155 (±0.003)**

Table E5. SnowGalileo and baseline model results reported with respect to median absolute error (mean ± sample standard deviation). All values show model performance on a test set with equivalent input data setups like the training set. Clear-HR refers to clear sky conditions with high-resolution imagery available on the day of prediction. Cloudy-HR refers to the same condition but with clouds present on the day of prediction. Clear-no HR refers to clear conditions without HR imagery available, and cloudy-no HR refers to cloudy conditions without HR imagery available. The pre-trained SnowGalileo model consistently outperforms all other models with respect to median absolute error (lower values are better), with the randomly initialized SnowGalileo model coming in second.

Region	**Model**	**Median Absolute Error**			
		Clear-HR	Cloudy-HR	Clear-no HR	Cloudy-no HR
Canadian Rockies	RF	0.020 (std <0.001)	0.121 (std <0.001)	0.097 (std <0.001)	0.093 (std <0.001)
	MLP	0.008 (std <0.001)	0.025 (±0.004)	0.053 (±0.022)	0.042 (±0.010)
	SVR	0.485 (±0.015)	0.464 (±0.007)	0.485 (±0.015)	0.468 (±0.002)
	<u>SnowGalileo random</u>	<u>0.001 (std <0.001)</u>	<u>0.009 (±0.001)</u>	<u>0.009 (std <0.001)</u>	<u>0.007 (std <0.001)</u>
	SnowGalileo pre-trained	**<0.001 (std <0.001)**	**0.003 (std <0.001)**	**0.004 (std <0.001)**	**0.004 (±0.001)**
Swiss Alps	RF	0.039 (std <0.001)	0.138 (std <0.001)	0.263 (±0.002)	0.297 (±0.001)
	MLP	0.014 (±0.002)	0.036 (±0.003)	0.129 (±0.017)	0.123 (±0.015)
	SVR	0.485 (±0.015)	0.464 (±0.007)	0.485 (±0.015)	0.468 (±0.002)
	<u>SnowGalileo random</u>	<u>0.007 (std <0.001)</u>	<u>0.032 (±0.001)</u>	<u>0.035 (±0.002)</u>	<u>0.052 (±0.009)</u>
	SnowGalileo pre-trained	**0.004 (std <0.001)**	**0.015 (±0.002)**	**0.019 (std <0.001)**	**0.023 (±0.001)**