POLITECNICO DI TORINO

Master's Degree in Data Science and Engineering

Master's Degree Thesis

Deep Learning-Based Estimation of Regional Snow Water Equivalent and Snow Depth Variations from InSAR Sentinel-1 Observations

Supervisors

Prof. Paolo GARZA

Luca BARCO

Lorenzo INNOCENTI

Candidate

Bianca BARTOLI

March 2026

Abstract

Accurate estimation of Snow Water Equivalent(SWE) and Snow Depth(HS) is essential for water resource management and climate monitoring in mountainous regions. The retrievement of these variables at regional scale is challenging due to the complex interaction between snow properties and radar signals. This thesis work investigates the potential of Interferometric Synthetic Aperture Radar (InSAR) data combined with temperature information and deep learning techniques for the estimation of regional SWE and HS variations.

The study is structured into two main parts. The first one focuses on dataset construction through a preprocessing pipeline, which integrates multiple data sources: InSAR data (coherence, displacement, elevation, intensity, local incidence angle, phase, unwrapped phase bands) from Sentinel-1 mission, temperature data from Sentinel-3 mission, temporal embeddings for seasonal dynamics and a validity mask to exclude invalid regions. IT-SNOW reanalysis dataset constitutes the snow-related ground-truth values for the study.

The second part presents a comparative analysis of machine learning and deep learning approaches. The tree-based ensemble model XGBoost is adopted as the baseline, while two deep architectures are explored: convolutional encoder-decoder model (U-Net with ResNet encoder) and transformer-based model(SegFormer with MiT encoder). All models are trained on a multi-band feature stack derived from the constructed dataset.

An ablation study is performed to quantify the contribution of individual InSAR features to the prediction task, showing how their relevance varies across models and that not all features improve results in every architecture. The results demonstrate how deep learning models exploit spatial dependencies in interferometric data, while the feature importance analysis highlights the varying predictive relevance of different radar inputs.

Table of Contents

List of Tables

List of Figures

Acronyms

AOI

Area Of Interest

CRS

Coordinate Reference System

DEM

Digital Elevation Model

DInSAR

Differential Interferometric Synthetic Aperture Radar

DL

Deep Learning

ESD

Enhanced Spectral Diversity

EUMETSAT

European Organisation for the Exploitation of Meteorological Satellites

HS

Snow Depth

InSAR

Interferometric Synthetic Aperture Radar

IW

Interferometric Wide

LOS

Line-Of-Sight

MAE

Mean Absolute Error

MiT

Mix Transformer

ML

Machine Learning

MLP

Multilayer Perceptron

MODIS

Moderate-resolution Imaging Spectroradiometer

PDD

Positive Degree Day

ResNet

Resisual Network

RMSE

Root Mean Squared Error

SAR

Synthetic Aperture Radar

SLC

Single Look Complex

SWE

Snow Water Equivalent

XGBoost

eXtreme Gradient Boosting

Chapter 1

Introduction

Snow plays a fundamental role in the global hydrological cycle and in the water reservoir for a large portion of the world's population. In particular, in mountainous regions, seasonal snowpack regulates water availability by accumulating in winter and melting in spring and summer seasons. This storage mechanism influences rivers' flow, groundwater recharge, hydropower generation and ecosystem stability. It also influences agricultural productivity, environmental and societal economy. Under ongoing climate change, analyzing snowfall patterns, accumulation and melting dynamics, and extreme events is increasingly fundamental for mitigating flood hazards, drought control and hydrological power reliability.

Problem Statement and Proposed Approach

Among the variables used to characterize snow conditions, Snow Depth (HS) and Snow Water Equivalent (SWE) are extremely important. HS and SWE are informative and complementary descriptors of seasonal snowpack dynamics, employed in hydrological and climate monitoring studies. HS is the vertical thickness of the snow layer, expressed in meters or centimeters, while SWE, typically expressed in millimeters, represents the amount of liquid water stored in a snow layer: it is the water obtained if a column of snow were completely melted. SWE is more informative than HS as it relates to the availability of water, whereas HS is easier to observe, but it is highly correlated to water density.

Remote sensing emerged as a promising tool for snow monitoring, in particular Synthetic Aperture Radar (SAR) and Interferometric SAR (InSAR). This type of observation provides sensitivity under all-weather and illumination conditions to snow structure and mass changes. However, these approaches are not always reliable when used alone due to snow conditions, low coherence and the geometric complexity of mountainous terrain, which can introduce additional uncertainties

in the measurements. At the same time, purely data-driven meteorological or statistical predictors do not exploit physical information, introducing a gap between physically interpretable models and flexible learning-based methods capable of handling nonlinear interactions and environmental variability. The main problem addressed in this thesis is the development of robust framework for regional HS and SWE estimation by proposing an hybrid model that combines multi-source remote sensing data with meteorological information in modern learning-based architectures. This hybrid approach is achieved in two main stages: first, a processing pipeline is developed to build a consistent multi-band dataset composed of InSAR and meteorological data, then a comparative learning study is implemented to evaluate both classical machine learning and deep learning architectures.

In the first stage, the dataset is composed by seven InSAR bands, including coherence, phase, phase, unwrapped, displacement, intensity, elevation, and local incidence angle, extracted by repeat-pass satellite acquisitions. InSAR bands are completed by temperature information and temporal embeddings designed to introduce seasonal variability and influence the learning framework to distinguish between wet and dry snow regions.

The second part implements a study to detect which model architecture is the most suitable for the task: starting with tree-based ensemble model XGBoost as the baseline, encoder-decoder networks as U-Net and Transformer-based architectures SegFormer demonstrate a higher capability of detect spatial dependencies and global contextual information. Models performances are assessed using the regression metrics MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) and an ablation study is conducted to quantify the contribution of individual InSAR-derived features in order to understand which one is more informative .

Our Contributions

This thesis contributes to the field of snow monitoring by proposing an hybrid framework for regional SWE and HS estimation through experimental rigor, the creation of a combined dataset, and methodological innovation.

The first contribution lies in the creation of the dataset. The developed pipeline contributes to advancing in a rigorous processing chain to produce InSAR products, which nowadays are difficult to process due to computational costs. In addition, the processing pipeline allows compatibility between multiple software tools that were not fully integrated, fragmented and with a complex workflow.

The proposed multi-source dataset aligns with recent studies highlighting the importance of integrating interferometric observables with environmental variables. By explicitly merging these information streams, the proposed framework provides a coherent foundation for modeling snow dynamics in complex alpine environments.

In particular, combining HS and SWE within a unified framework represents a novelty in the field of snow dynamics modeling via InSAR data.

Furthermore, a systematic evaluation of different architectural strategies is conducted to assess which modeling approach is most suitable for the task, which remains relatively underexplored in the literature. Pioneer studies focuses on physically based methods and, more recently, on traditional machine learning models. The application of spatial deep learning architectures to InSAR-based snow parameter estimation is still comparatively limited. In addition, an ablation study is performed to quantify the contribution of each input band, providing insights into the informational content of interferometric observables and their relative importance for snow parameter prediction.

Document Structure

Chapter 2 - Related Works

This chapter introduces state-of-the-art studies on the task and describes the used architectures (XGBoost, U-Net and SegFormer).

Chapter 3 - Dataset

It is divided into two parts: the first technical one that describes the pipeline to generate InSAR products and the second part that describes the creation of the data set. Details about data sources are analyzed.

Chapter 4 - Methodology

Methodology describes data preprocessing of the dataset, as harmonization techniques, bands concatenation and mask generation. Then, models exploited in the study are presented, together with training and evaluation strategies.

Chapter 5 - Results

In this chapter, the quantitative and qualitative results are illustrated and commented, and the ablation study reports the contribution of each band.

Chapter 6 - Conclusions

Finally, Conclusions summarizes the findings of the study and proposes next steps for future works to enrich the work and overcome the intrinsic limitations of the task.

Chapter 2

Related Works

Remote sensing techniques, including Synthetic Aperture Radar and Interferometric SAR, include a wide range of methodologies for snow monitoring and hydrological applications, spanning physical modeling approaches, empirical techniques, and data-driven methods based on machine learning and deep learning.

This chapter provides an overview of the state-of-the-art techniques for snow parameter estimation, as SWE and HS retrieval from radar observations. Special attention is given to studies that exploit SAR and InSAR observables for snow-related predictions. It also discusses the main technical components relevant to this thesis, including the adoption of modern learning architectures for spatial regression tasks.

2.1 Snow Water Equivalent and Snow Depth studies

In order to contextualize data and methodological choices of this thesis, state-of-the-art studies are analyzed and grouped into three main categories: physically-based retrieval methods, classical machine learning approaches and deep learning architectures. This division allows comparison between models, type of input data and motivates hybrid strategies employed in this study.

2.1.1 Physically-Based Retrieval Approaches

Pioneer works investigate HS and SWE retrieval using physical models grounded in electromagnetic scattering and interferometric signals. Repeat-pass InSAR has shown to provide acceptable sensitivity to changes in dry-snow scenarios, due to phase variations induced by refraction within snowpack. This highlights theoretical relations between interferometric phase and snow changes [1].

Several studies have investigated SWE and HS retrieval through physically-based approaches based on the relationship between slope-correlated InSAR phase variation and signal refraction[2][3]. The study conducted over Idaho using Sentinel-1 repeat-pass interferometry[4] explored SWE retrieval under operational satellite conditions. In this study, the authors analyze the sensitivity of the C-band signal to SWE variations and confirm limitations in this approach due to temporal decorrelation and the lack of additional information about coherence and snow conditions. Beyond interferometric phase-based retrieval, alternative physically-based methods have focused on volume scattering mechanisms in SAR amplitude data[5][6].

Despite these studies underline how physical-based approaches offer strong interpretability and explicit relationships with snow physics, they often rely on strict mathematical formulations and underlying physical assumptions. Their validation rely on in-situ measurements, which are challenging to collect in mountainous regions due to harsh seasonal and weather conditions. Moreover, these observations are sparse and unevenly distributed, limiting their representativeness at regional scales.

Physically based methods depend on restrictive assumptions of snow conditions and sufficient interferometric coherence. These conditions are not always satisfied in real-world scenarios, for example during melt periods.

These challenges highlight the need for hybrid methodologies, which combine physically meaningful explanatory variables with learning-based models designed to accommodate environmental heterogeneity.

2.1.2 Machine Learning Approaches

In recent years, Machine Learning (ML) models have been adopted as an alternative to purely physical models. Compared to physical retrieval, they estimate HS and SWE by learning empirical mappings, allowing multi-source predictors. In fact, ML can integrate heterogeneous inputs such as InSAR data and meteorological information. State-of-the-art models include ensemble tree models (Random Forest, CatBoost, AdaBoost, XGBoost) which remain competitive baselines because of their robustness on limited data.

Boeykens et al.[7] present a study in which Sentinel-1 observations are combined with meteorological data and physical models to estimate HS using XGBoost, while Cistý et al.[8] explore machine learning techniques to predict SWE, in particular XGBoost. Both studies achieve satisfying results highlighting the benefit of combining remote sensing with auxiliary physical drivers. Similarly, Alabi et al. [9] applies machine learning algorithms, as LASSO, Random Forest, Support Vector Machines, and CatBoost, to InSAR-derived products at high spatial resolution, focusing on analyzing how InSAR prediction power is influenced by land-cover

conditions.

These studies suggest that ML improves physically-based models by allowing nonlinear modeling capability, while they benefit in combining physically meaningful radar observables, temperature and meteorological data that capture regime transitions and improve generalization through feature selection to mitigate redundancy.

2.1.3 Deep Learning Approaches

To overcome intrinsic limitations of classical machine learning, deep learning techniques directly exploit spatial structure by allowing global context understanding. Convolutional encoder-decoder structures are frequently adopted for high-resolution snow mapping because they are able to combine hierarchical features extraction and pixel-wise prediction. Recent works demonstrate that snow depth estimation can be achieved with U-Net-based regression models, using multi-source spatial inputs, illustrating effectiveness of Convolutional Neural Networks(CNN) for spatially distributed prediction tasks[10].

Moreover, Transformer-based architectures emerged as a competitive alternative for dense prediction and broader global context due to their ability to model long-range dependencies via attention mechanisms. Attention mechanisms have been introduced within deep learning frameworks for HS and SWE predictions, demonstrating that learned weighting of spatial and temporal dependencies can improve predictive skill compared to standard baselines [11].

Even if large-scale monitoring of HS and SWE remains challenging because of the scarcity of in-situ measurements, remote sensing, specifically InSAR, and data-driven approaches are confirmed to be the most reliable, especially in mountainous terrain[12][13]. Radar observations provide physically meaningful information on snow structure and mass changes, while meteorological variables capture regime-dependent processes such as accumulation and melting. Consequently, recent research trends increasingly favor hybrid strategies that combine interferometric information with temperature and auxiliary drivers within learning-based frameworks.

2.2 Architectures

2.2.1 XGBoost

XGBoost(eXtreme Gradient Boosting) is a scalable and regularized implementation of gradient boosted decision trees, introduced by Chen and Guestrin in 2016 [14]. It builds upon the gradient boosting framework formalized by Friedman [15], which constructs an ensemble of weak learners in a sequential manner to approximate complex nonlinear functions, as illustrated in Figure 2.1.

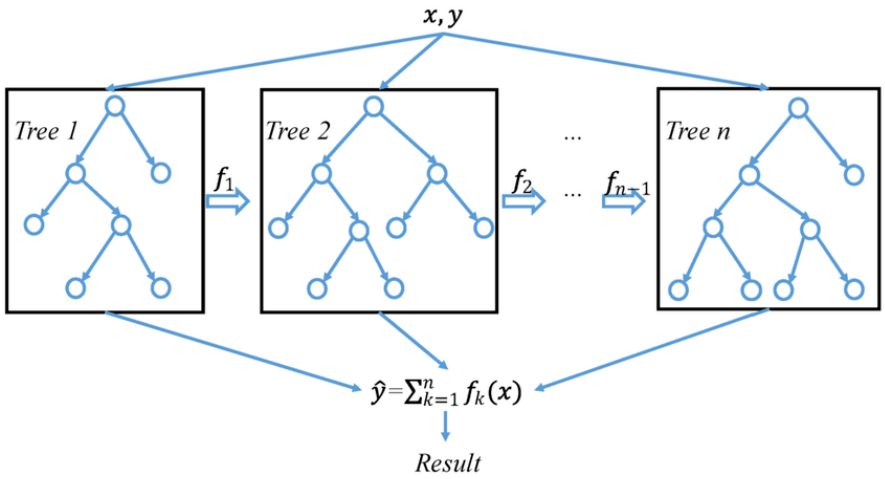

Figure 2.1: General architecture of XGBoost, [16]

Gradient boosting operates by iteratively adding decision trees to an ensemble. At each iteration, a new tree is trained to minimize the residual errors of the current model. The final model prediction is the weighted sum of multiple regression trees, and each new tree is fitted to the negative gradient of the loss function with respect to the previous prediction. This allows the ensemble to progressively correct its own mistakes and capture complex interactions among features.

XGBoost introduces methodological and computational improvements with respect to classical gradient boosting. XGBoost adds regularization terms on the tree complexity, which reduces the risk of overfitting. Optimization efficiency is achieved with the second-order Taylor approximation of the loss function, while it is also capable of optimizing handling of sparse inputs and missing values.

2.2.2 U-Net

U-Net is a convolutional neural network architecture introduced by Ronneberger et al. [17] for dense image prediction tasks. The architecture, as shown in Figure 2.2, follows a symmetric U-shaped structure composed of two main components: a contracting path (encoder) and an expanding path (decoder). The encoder reduces the spatial dimensions of the input image and increases the number of feature channels, through the repetition of convolutional operations followed by nonlinear activation functions.

The decoder mirrors the encoder by restoring spatial resolution by upsampling the feature maps using transposed convolutions, followed by concatenation with the corresponding high-resolution feature map from the encoder through skip connections. The connections between the two paths aim to capture high-level features and details. The concatenated feature map is refined with two additional convolutions, which reduce the number of feature channels. The output of the network is generated by a 1×1 convolution to the feature map at the last decoder layer, which assign each pixel to the corresponding class through a Softmax function for segmentation tasks. For regression tasks the final activation is removed, with a linear output suitable for continuous value prediction.

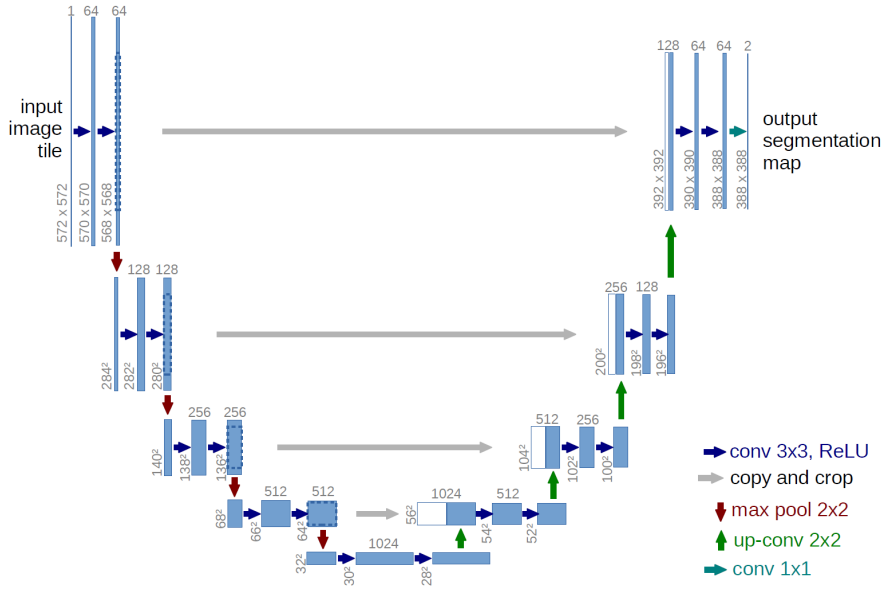

Figure 2.2: U-net architecture, consisting of a contracting encoder path, an expanding decoder path, and symmetric skip connections. [17]

The concatenation mechanism allows the decoder to combine high-level semantic features with precise spatial information from earlier layers. In this way, U-Net mitigates the loss of localization caused by repeated downsampling operations.

In modern implementations, the plain encoder is often replaced with a pretrained backbone while retaining the U-shape decoder and skip structure.

2.2.3 ResNet

Residual Networks (ResNet) [18] represent a solution to the vanishing gradients and performance degradation in standard convolutional networks as their depth increases.

ResNet addresses this issue through the introduction of residual learning. Instead of directly learning a mapping $\mathcal{H}(x)$ from input x to output, the network is structured to learn a residual function $\mathcal{F}(x)$ defined as:

$$\mathcal{F}(x) = \mathcal{H}(x) - x. \tag{2.1}$$

The original mapping can therefore be rewritten as:

$$\mathcal{H}(x) = \mathcal{F}(x) + x. \tag{2.2}$$

This reformulation is implemented through identity shortcut connections that skip one or more convolutional layers, as illustrated in Figure 2.3

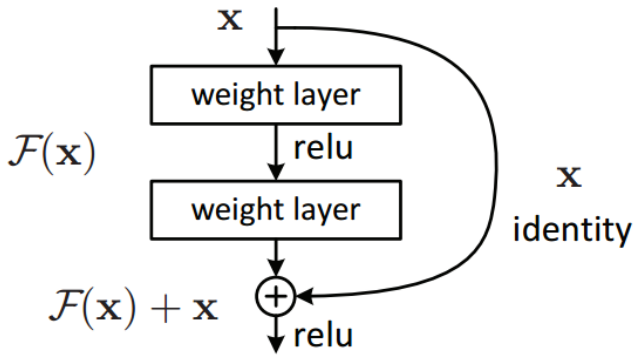

Figure 2.3: Residual Block of a ResNet, [18]

A residual block consists of two convolutional layers, each followed by batch normalization and a nonlinear activation function, typically a ReLU. The shortcut connection performs an element-wise addition between the input and the output of the residual branch. When the input and output dimensions differ, a linear projection is applied to the shortcut path to ensure dimensional consistency. Residual connections facilitate gradient flow during backpropagation and mitigate the vanishing gradient problem.

ResNet architectures has many variants depending on the dept and block design. Standard versions have 18, 34, 50, 101 and 152 layers. Shallower versions are

ResNet18 and ResNet34, which use basic blocks composed of two convolutions per residual unit. In contrast, deeper networks adopt a bottleneck block where each residual unit has three convolutions: the first layer reduces the feature dimensionality, the central one is designed for feature extraction, while the final layer restores the original dimensionality. In Figure 2.4 the design of ResNet18 is shown.

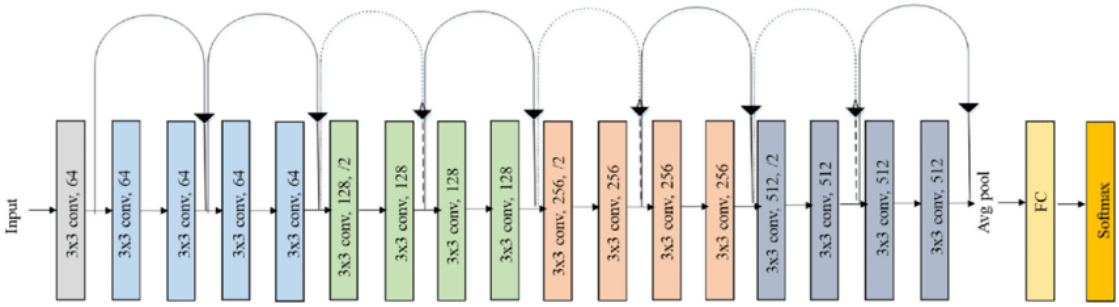

Figure 2.4: ResNet18 Architecture

2.2.4 SegFormer

SegFormer[19] is a Transformer-based architecture developed for segmentation task. It combines a hierarchical transformer encoder and a lightweight multilayer perceptron(MLP) decoder to achieve high efficiency while maintaining a high representational power. Transformer-based architectures rely on self-attention mechanisms to shape dependencies by estimating the relevance of each spatial feature representation with respect to all others.

Given an input feature matrix $X \in \mathbb{R}^{N \times d}$, where N represents the number of spatial feature and d the embedding dimension, self-attention computes:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V, \tag{2.3}$$

where W_Q, W_K, and W_V are learnable weight matrices that project the spatial feature matrix X into query, key, and value representations, allowing the computation of attention scores across all spatial embeddings. This enable global context understanding and captures long-range spatial dependencies.

As illustrated in Figure 2.5, the two key design elements are the hierarchical encoder and the lightweight decoder. The first operates ata single resolution and progressively reduces the spatial resolution while increasing channel dimensionality, allowing high-level global context understanding. Instead, the decoder aggregates multi-layer scale features through a linear layer and interpolation operations.

The core characteristic of SegFormer is that it does not rely on positional encoding interpolation, improving robustness and flexibility across different input scales. For this reason, it represents a semantic segmentation benchmark and has become a reference architecture in the field.

2.2.5 Mit

The Mix Transformer (MiT) is the hierarchical Transformer encoder introduced within the SegFormer framework [19], whose architecture is shown in Figure 2.5 in the encoder part.

It is inspired by Vision Transformer (ViT) [20] and Pyramid Vision Transformer (PVT) [21]. Unlike ViT, which processes fixed-size image patches at a single resolution, MiT adopts a hierarchical architecture. It begins with an overlapping patch embedding layer, implemented using convolutional projections. The convolution operates with stride and kernel size chosen to produce overlapping patches. This overlapping design improves local continuity modeling compared to non-overlapping patches.

The encoder is composed of multiple scale extraction, inspired by PVT architecture. Within each stage, Transformer blocks apply multi-head self-attention and

feed-forward networks to the embedded spatial features. To improve efficiency, MiT employs a spatial reduction mechanism in the attention computation. Instead of computing attention across all spatial locations at full resolution, keys and values are downsampled.

Several MiT variants differ in depth, embedding dimensions, and number of attention heads. Smaller variants provide computational efficiency, while larger ones increase representational capacity.

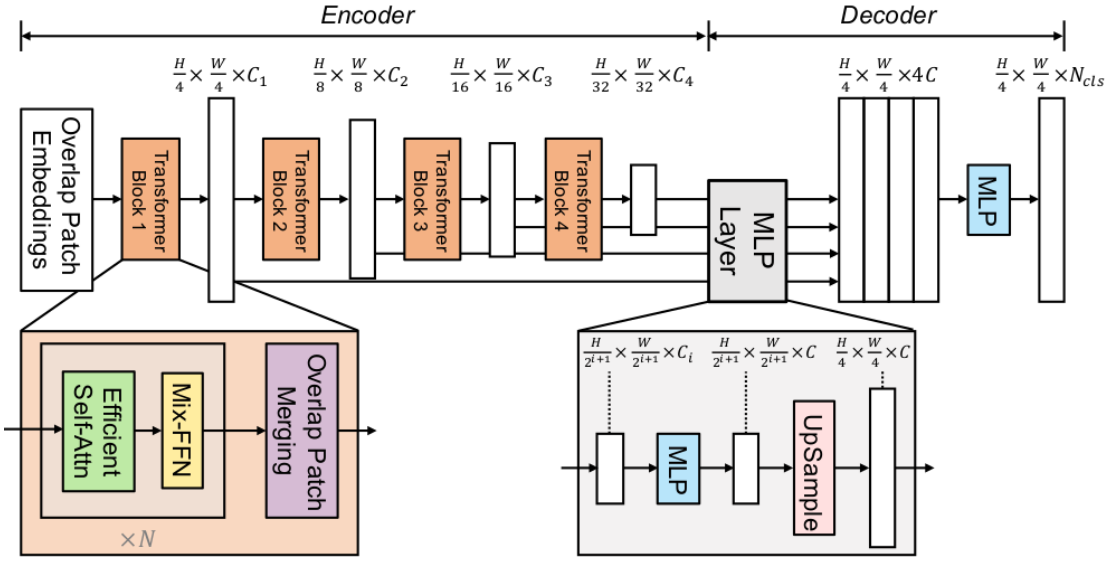

Figure 2.5: SegFormer Architecture

Chapter 3

Dataset

3.1 Data sources

3.1.1 Synthetic Aperture Radar and Interferometric SAR

Synthetic Aperture Radar

The basis concept behind the images used in this thesis work is Synthetic Aperture Radar[22], which is an active microwave imaging system designed to acquire high-resolution, light and weather independent images of the Earth's surface. The image acquisition is based on a pulsed radar installed on a moving platform, either airborne or spaceborne. The radar systematically transmits high-power electromagnetic pulses towards the ground and receives the backscattered signals.
The returning signal(s) recorded by the antenna is complex-valued, it contains both amplitude(A) and phase(π) information and the characteristics depend on the physical structure and dielectric properties of the imaged surface.

$$A = \sqrt{\Re(s)^2 + \Im(s)^2} \tag{3.1}$$

$$\phi = \frac{4\pi}{\lambda} R \tag{3.2}$$

where $\Re(s)$ denotes the real part of the signal, $\Im(s)$ denotes the imaginary part, λ is the radar wavelength, and R is the sensor-to-target distance traveled by the radar signal.

The amplitude is obtained by combining the real and imaginary components of the signal and represents the signal strength returned from the observed surface, which is influenced by its physical and electromagnetic properties. The phase is directly related to the distance traveled by the radar signal between the sensor and the target, and it is proportional to this range and depends on the radar wavelength,

accounting for the two-way propagation of the signal from the sensor to the ground and back, as illustrated in Figure 3.1. As a result, the phase provides very accurate information about changes in distance, even though it is ambiguous because it repeats periodically. This property makes phase measurements especially useful for interferometric analyses.

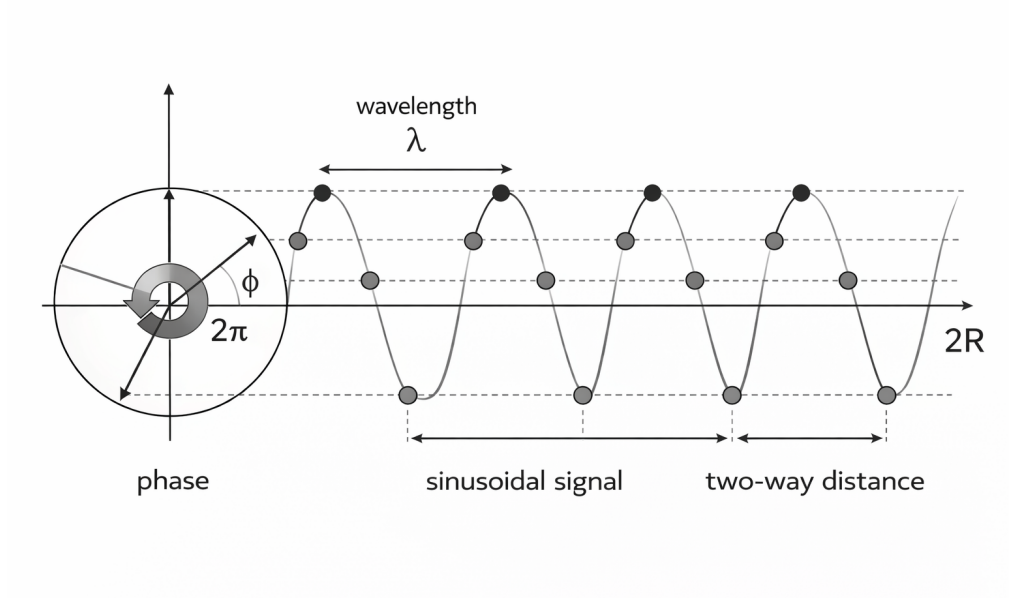

Figure 3.1: Relationship between phase, wavelength, and two-way distance in a sinusoidal signal, [23]

Depending on the frequency band, the radar signal with longer wavelenghts penetrates more in irregular surfaces as vegetation, ice, snow and dry soil. Since SAR operates in the microwave frequency domain, it is insensitive to clouds, precipitation and does not rely on solar illumination, making its observation consistent day-and-night and with all atmospheric conditions. The geometry of the acquisition, Figure 3.2, is defined by two orthogonal dimensions, along which the amplitude and phase of the backscattered signal are measured: the range and the azimuth. The range direction is aligned with the radar line of sight, while the azimuth direction corresponds to the track of the platform's movement. Since the radar platform moves, the target on the ground remains available for a limited interval of time, during which multiple echoes are collected from successive sensor positions.

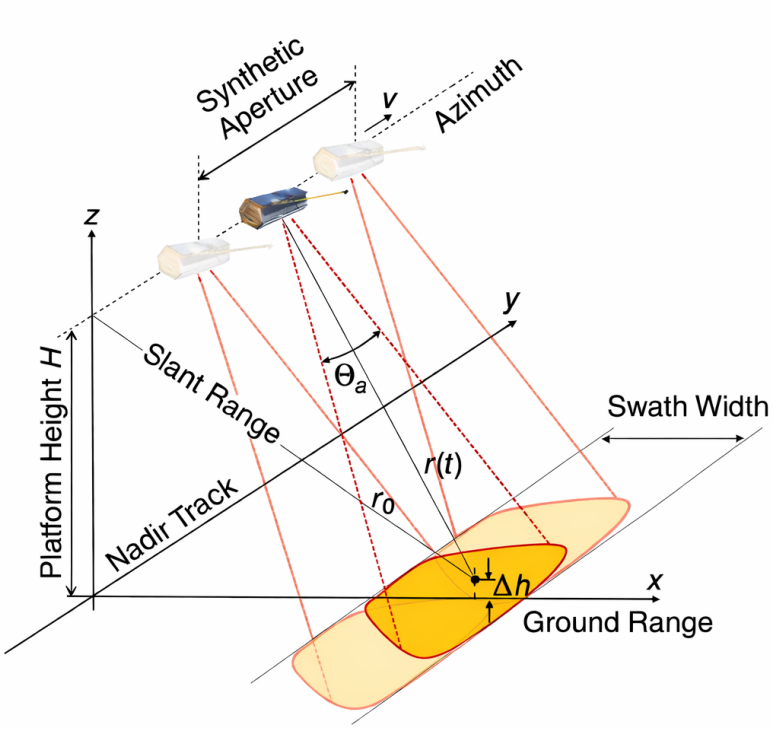

Figure 3.2: Illustration of geometry of a SAR acquisition, [22]

By coherently combining these echoes and exploiting their phase history, the SAR processor is able to synthesize a virtual antenna aperture that is significantly larger than the physical antenna, a process known as synthetic aperture formation. These raw SAR data are composed of a two-dimensional matrix sampled in fast time, associated with the propagation delay in range, and slow time, associated with the platform motion in azimuth. Subsequently, matched filtering operations in both dimensions are applied, commonly referred to as range compression and azimuth compression, which focus the energy of distributed echoes into localized image , as represented in Figure 3.3. The final SAR image is a calibrated reflectivity map of the target ground scene: bright areas correspond to strong backscattering, while dark areas indicate smooth or weakly scattering surfaces.

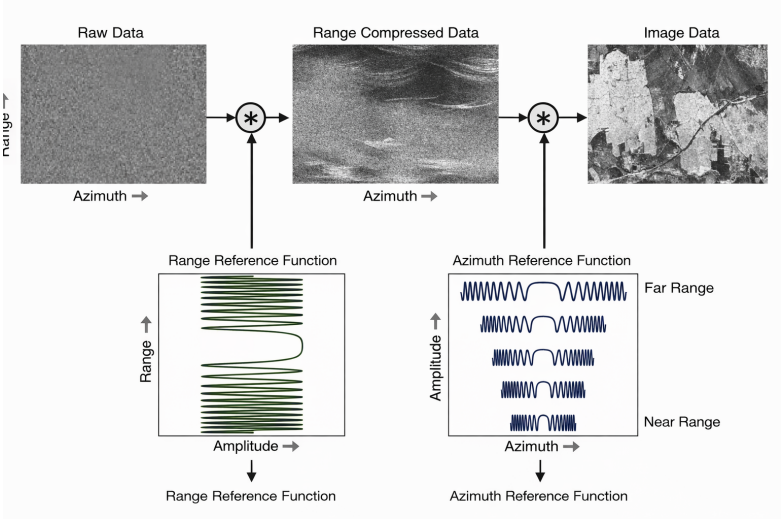

Figure 3.3: Summary of SAR processing steps where the range compressed data result from a convolution of the raw data with the range reference function, while the azimuth compression is performed through a convolution with the azimuth reference function, which changes from near to far range. [22]

Usually, SAR images are geometrically distorted due to the limitation of the radar ability to measure the projection of a three-dimesional scene in radar coordinates, represented by range and azimuth. This can cause a shadow effect on the areas that are hidden from the illumination. Moreover, SAR images are affected by speckle, a phenomenon caused by unresolved scatterers within each resolution cell. Within each resolution cell (e.g. a single pixel), many individual scatterers contribute to the total backscattered signal, such as oil particles, vegetation elements, rocks, snow or man-made structures. The exact position and properties of the scatterers inside a pixel are unknown, so their phase contributions can be considered random. When these signals are combined coherently, they produce variations in pixel intensity, creating the typical granular appearance known as speckle.

Speckle should not be considered noise, it contains information about sub-resolution structure and its effect can be mitigated through processing techniques, as multilooking and filtering.

Interferometric Synthetic Aperture Radar

Interferometric Synthetic Aperture Radar (InSAR) [23][24] extends the concept of SAR imaging by observing the same area from slightly different sensor positions, resulting in different viewing geometries. It can be done simultaneously using two sensors on the same moving platform or exploiting repeated orbits of the same satellite at different times [25], as shown in Figure 3.4.

Two coregistred complex-valued SAR images of the same target, either from two spatially separated apertures (single-pass interferometry) or from repeated passes of the same acquisition geometry (repeat-pass interferometry), are combined pixel-wise. The result of a complex conjugate multiplication is an interferogram, Figure 3.5: in particular, the interferometric phase is the phase difference between the two images and it is proportional to the difference in two-way propagation distance between the sensor positions and each ground resolution element.

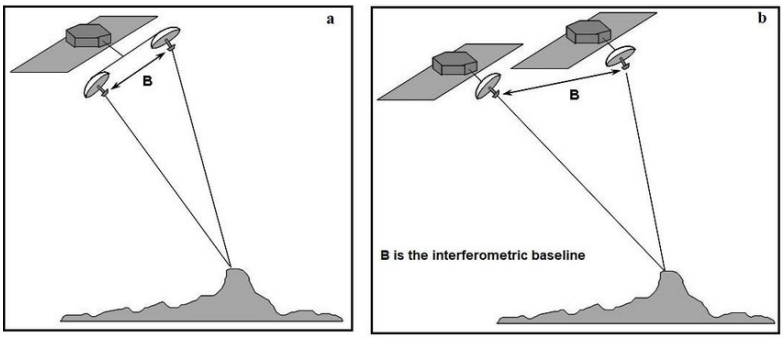

Figure 3.4: Single-pass interferometry vs. Repeat-pass interferometry

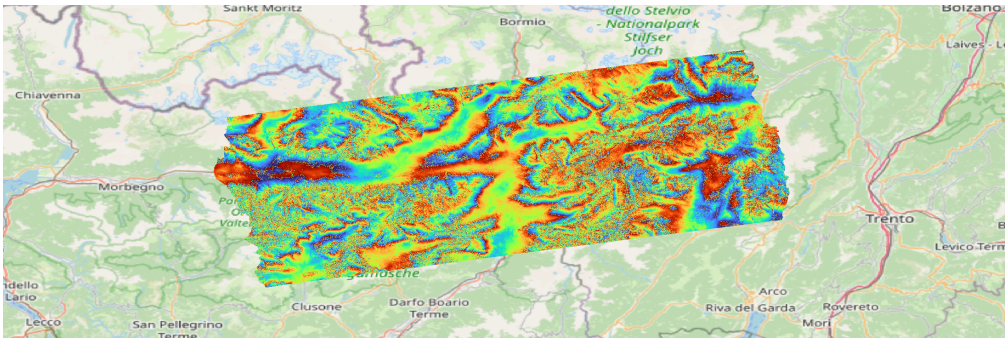

Figure 3.5: Example of Interferometric Phase, where color variations represent phase differences related to topography and surface displacement

Differential Interferometric Synthetic Aperture Radar

When the topographic contribution is removed from the interferometric phase, the technique is known as Differential Interferometric SAR (DInSAR). In DInSAR, an external Digital Elevation Model (DEM) is used to predict and subtract the topographic phase component, so that the resulting differential interferogram contains only phase variations related to the actual surface displacement and not geometric influence from the landscape. Thus, the residual interferometric phase

is proportional to the displacement of the ground projected along the radar LOS (Line Of Sight), allowing the measurement of surface motion with millimetric sensitivity. This approach is particularly suited for the analysis of slow or moderate deformation phenomena such as natural geohazards and human-induced activities.

Fundamental limitations are temporal and geometric decorrelation as InSAR interferometry requires a stable scattering behavior between acquisitions. Temporal decorrelation occurs when the physical properties or positions of the scatterers within a resolution cell change between two SAR acquisitions due to vegetation growth, soil moisture variations, snow dynamics, or human activity. For this reason, it is good practice to select SAR images that are not too far apart in time, in order to reduce temporal changes in the scene. Geometric decorrelation, on the other hand, is related to differences in the observation geometry between the two acquisitions, mainly due to a large spatial separation between the satellite positions.

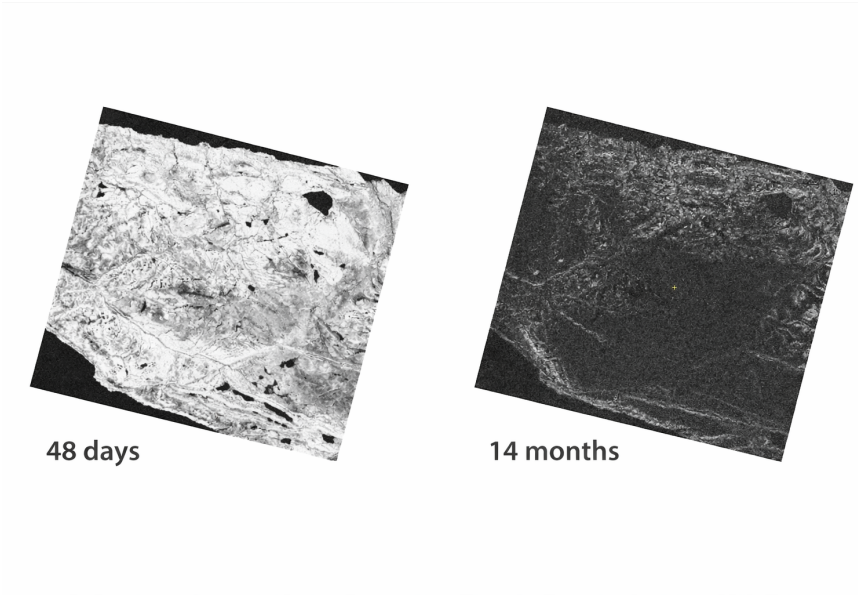

Figure 3.6: Example of temporal decorrelation due to temporal changes between SAR acquisitions, [26]

As visible in the Figure 3.6, brighter areas correspond to highly correlated images, indicating strong similarity between acquisitions, whereas darker areas as in the image on the right, show very low correlation.

Another limitation of the interferometric phase is that it is inherently ambiguous, as it is measured only within the interval $[0, 2\pi)$. As a result, the observed phase does not have a direct and unique interpretation in terms of topographic height or surface displacement. To represent the real elevation and displacement, it must be

unwrapped, a process that reconstructs a continuous phase field by resolving the 2π discontinuities, exemplified in Figure 3.7 .

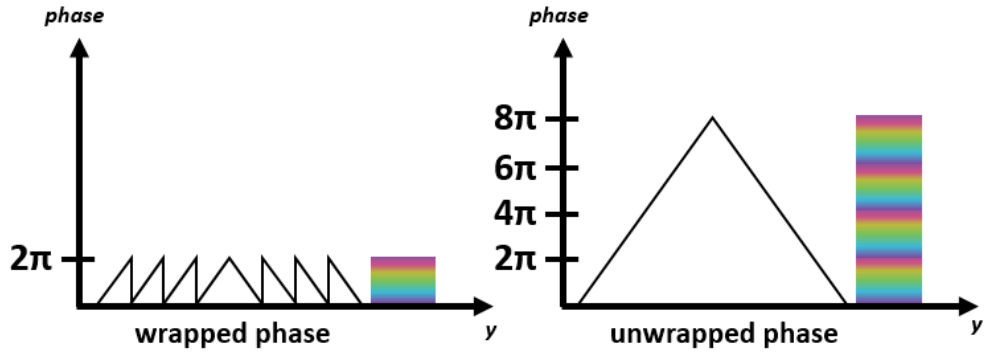

Figure 3.7: Illustration of phase unwrapping: the wrapped phase is confined within [0,2π), while the unwrapped phase reconstructs the continuous phase signal, [27]

Based on these concepts, SAR can be considered a fundamental framework for high-resolution Earth, allowing consistent data acquisition independently of illumination and atmospheric conditions. Interferometric SAR exploits the phase information of SAR signals to retrieve precise displacement and elevation measurements through the analysis of phase differences between multiple acquisitions. The ability of InSAR to measure topographic variations makes it a powerful tool for scientific research in geophysics, cryosphere studies, hydrology, and environmental monitoring. Despite a few limitations related to decorrelation, atmospheric effects, and phase ambiguity, the development of advanced processing techniques has extended the reliability of InSAR. In this thesis, these characteristics are exploited to predict snow depth and snow water equivalent, as SAR and InSAR observations are well suited to cryospheric applications due to their sensitivity to snow accumulation and redistribution processes; especially their capability to operate under persistent cloud cover and their potential to capture spatially continuous snow-related signals over complex alpine terrain.

3.1.2 InSAR Processing

Software and tools

The interferometric processing steps adopted in this work follow the TOPS Interferometry Tutorial [27] designed by the European Space Agency (ESA).

The processing chain, illustrated in Figure 3.8, is created to automate the entire workflow, from the download of images to the final interferometric product. All processing steps are implemented with SNAP[28], a software specifically developed for SAR data processing, accessed through its Python esa_snappy[29][30] library. The only step that requires an external tool is the phase unwrapping step, which is performed using snaphu[31], a widely used software for interferometric phase unwrapping.

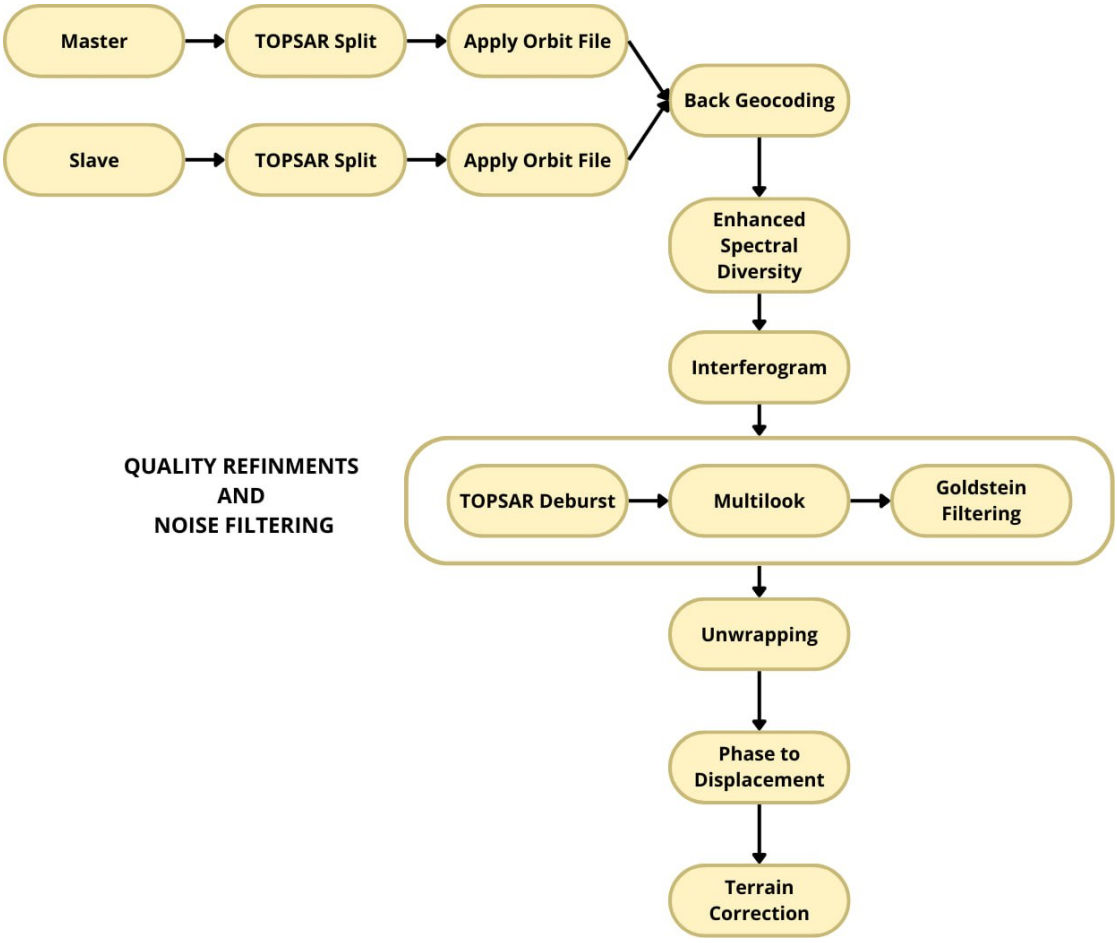

Figure 3.8: InSAR processing steps

Data Download

SAR products are selected and downloaded using the Copernicus Data Space Ecosystem via its OData API[32]. The data query is constructed to retrieve the desired type of acquisition belonging to one of the Copernicus Sentinel missions within a predefined temporal window, constrained to a specific orbit direction, relative orbit number, and slice number, allowing a precise area of interest.

TOPSAR Split

The processing begins with the TOPSAR (Terrain Observation with Progressive Scans SAR) Split operation. It extracts the relevant subswath and burst information from the SAR products. Interferometric Wide Swath (IW) mode is the standard acquisition mode over land. It provides a wide ground coverage (about 250 km swath width) while preserving a spatial resolution. With IW mode acquisition, as shown in Figure 3.9 the image is divided into three subswaths (IW1, IW2, IW3), each covering a different portion of the ground range. Each subswath is further subdivided into smaller units called bursts. A burst corresponds to a short acquisition segment during which the radar illuminates a specific ground area.

The TOPSAR Split operation reduces data volume by selecting only the subswath and bursts covering the area of interest.

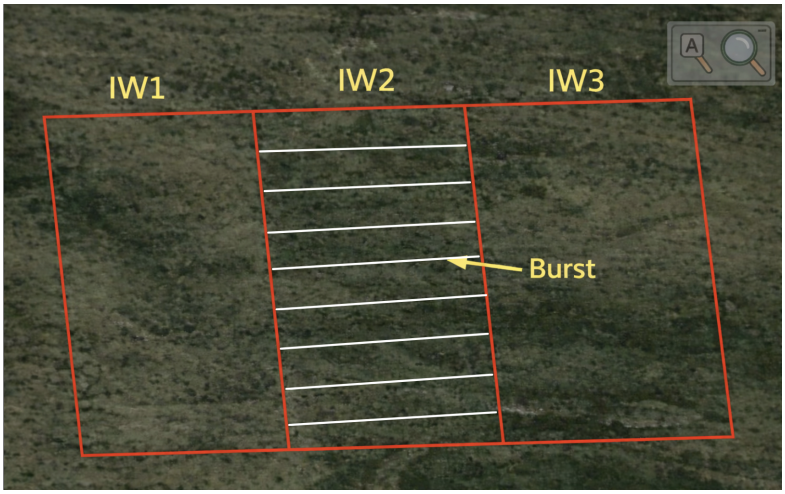

Figure 3.9: Example of Sentinel-1 Interferometric Wide (IW), swaths and bursts, illustrating the subdivision of each swath into multiple bursts.

At least two consecutive bursts must be considered in order to have a correct coregistration of the images because accurate geometric alignment requires sufficient spatial overlap and azimuth continuity.

Apply Orbit File

Then, precise orbit files are applied to both acquisitions to replace the restituted orbit information with refined orbital state vectors. Orbit files contain detailed information about the satellite trajectory during acquisition, including the position and velocity in space at each time step. Position and velocity are referred as the orbital state vectors, which describe where the satellite is located and how it is moving along its orbit. This allows the accuracy of the satellite position, which is critical for precise coregistration and interferometric phase computation.

Back Geocoding

From now onward, the two SAR images can be referred as master image(the earlier acquisition) and slave(the later acquisition). In this step, the master and the slave are coregistered using the Back Geocoding algorithm. Coregistration is a fundamental preprocessing step in InSAR, whose goal is to align the master and slave images at the pixel level, ensuring that each pixel in the master corresponds to the same ground location in the slave. Accurate coregistration is essential to avoid phase errors in the subsequent interferometric processing. The Back Geocoding algorithm performs this alignment by introducing an external Digital Elevation Model(DEM), a raster representation of the Earth's surface topography, providing elevation values for each ground location.

Enhanced Spectral Diversity

Coregistration accuracy is further refined using Enhanced Spectral Diversity (ESD), which corrects range and azimuth misregistration between bursts.

Interferogram

The interferogram is computed by cross multiplying the master image with the complex conjugate of the slave image. Thus, the amplitude of both images is multiplied while the phase is the phase difference between the two images. During this step, coherence is computed and stored as a separate raster.

The **coherence** band represents the similarity of the radar backscatter signal between two SAR images, the master and the slave, and represent the degree of correlation between two acquisitions. The similarity is represented by a number on a scale from 0.0 to 1.0, with 1.0 representing high coherence, indicating perfect similarity and a reliable interferometric phase. Instead, low coherence is typically associated with decorrelation due to vegetation, snow dynamics, surface changes or temporal effects. The value of the coherence is influenced by surface characteristics

and temporal changes between acquisitions. In particular, vegetation, irregular soil surfaces and snow cover introduce changes in the scattering properties of the scene.

The **elevation** band provides the topographic height information of the area. It is computed from an external Digital DEM, Copernicus 30m Global DEM in this work.

The **intensity** band represents the backscattered radar power, expressed in decibels (dB). It describes the reflectivity of the surface and depends on surface roughness, moisture content, and land cover.

The outputs product of this step are the interferometric phase, coherence, elevation and Intensity, shown in Figures 3.10, 3.11, 3.12, 3.13.

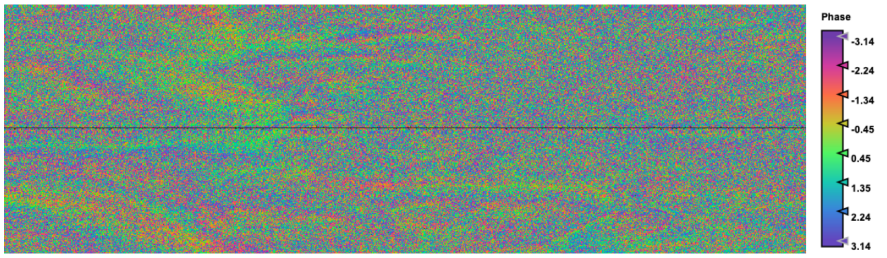

Figure 3.10: Interferometric phase

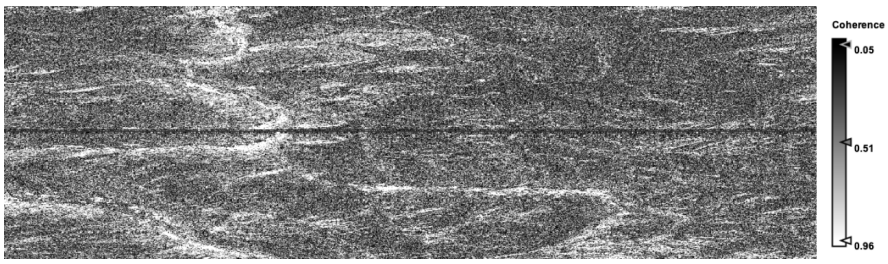

Figure 3.11: Coherence

Figure 3.12: Intensity

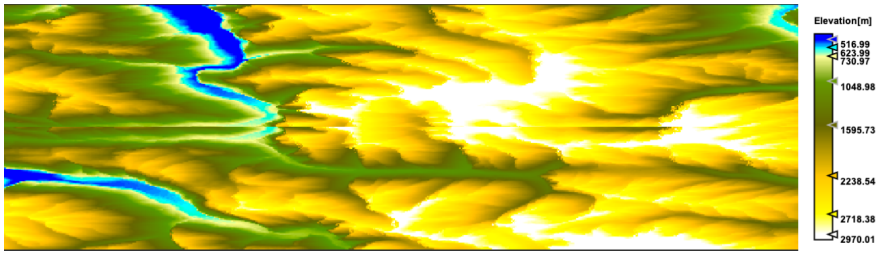

Figure 3.13: Elevation

TOPSAR Deburst

After interferogram formation, the TOPSAR Deburst operation is applied to remove burst boundaries and generate a continuous interferogram over the full subswath.

Multilook

Multilooking is applied in range and azimuth to reduce phase noise and improve the readability of the interferometric signal. This step reduces speckle and increases coherence, resulting in a more stable interferometric phase.

Goldstein Phase Filtering

Goldstein phase filtering is applied as an adaptive spectral filter, which suppresses high-frequency phase noise and preserves phase discontinuities associated with real topographic or deformation signals. Goldstein filter uses a Fast Fourier Transformation (FFT) to reduce the noise of the image, increasing the quality of the fringes of the interferogram. The resulting phase band contains the wrapped interferometric **phase** illustrated in Figure 3.14, defined within the interval $[0, 2\pi)$, and the coherence, Figure 3.15.

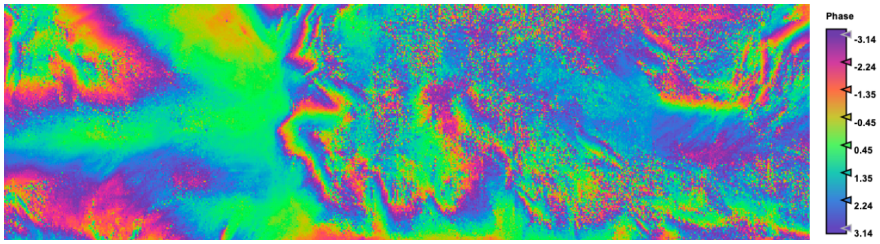

Figure 3.14: Interferometric phase after Goldstein filter and Multilook techniques

Figure 3.15: Coherence after Goldstein filter and Multilook techniques

As illustrated in Figure 3.14 compared to Figure 3.10, Deburst operation reduces

the separation between bursts, while multilooking and Goldstein filtering make the phase appear smoother, with a clear reduction of noise components.

Phase Unwrapping

Since the interferometric phase is ambiguous and only known within the scale of 2π, the phase unwrapping step solves this ambiguity by integrating the phase difference between neighboring pixels.

Unwrapping in SNAP follows three distinct steps:

1. Export of the wrapped phase from SNAP

2. Phase Unwrapping performed using snaphu software: it solves a global optimization problem that reconstructs a continuous phase field from the wrapped phase measurements

3. Import of the unwrapped phase back into SNAP

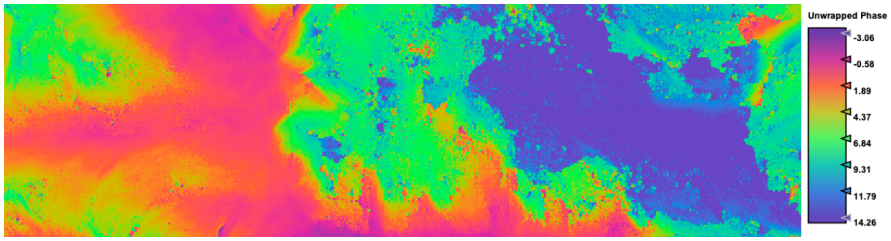

Figure 3.16: Unwrapped Phase

The **unwrapping** band corresponds to the unwrapped interferometric phase, obtained by resolving the 2π ambiguities present in the wrapped phase. It converts the ambiguous phase measurements into a continuous phase field, enabling the estimation of realistic physical quantities, illustrated in Figure 3.16.

Phase to Displacement

The unwrapped phase is now a continuous raster, but not a metric measure. It must be converted into a displacement by applying the standard **phase-to-displacement** operator, which yields displacement measurements expressed in metric units.

Terrain Correction

Finally, **Terrain Correction** is the last step applied to the image. It geocodes the interferometric product of the AOI from radar geometry into geographic coordinate system. In order to enforce spatial consistency across all all InSAR products, a fixed output pixel spacing is explicitly defined. The resampling to a common pixel size ensures that all bands and all processed images have the same spatial resolution and grid alignment. An additional band is computed in this step: Local Incidence Angle, angle between the radar LOS, the direction along which the radar signal travels from the sensor to the ground target and back, and the local surface normal.

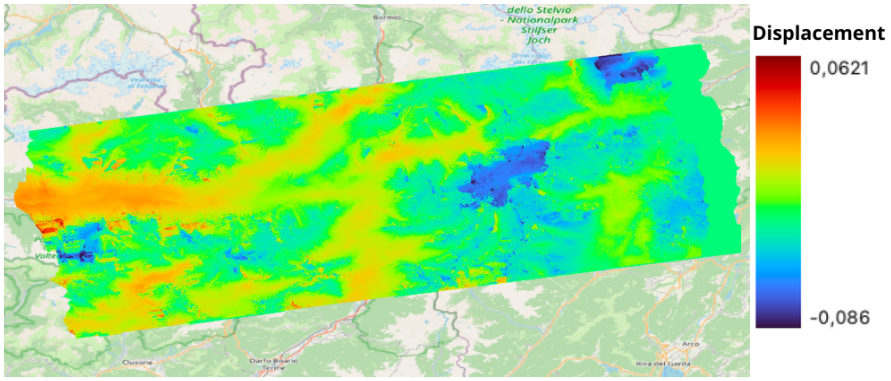

Figure 3.17: Displacement map after terrain correction of the same area, projected into real-world geographic coordinates

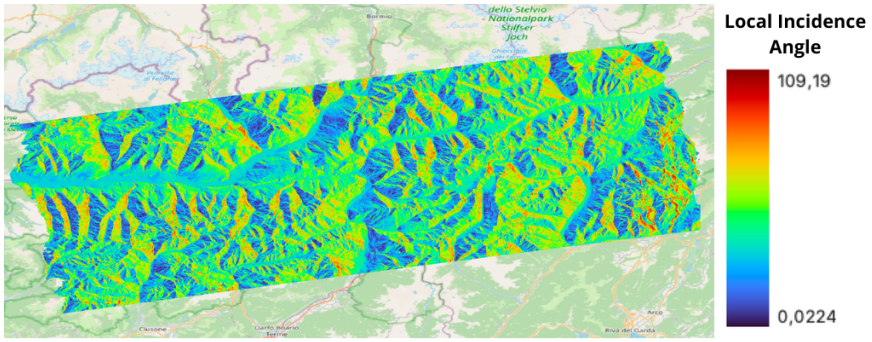

Figure 3.18: Local Incidence Angle map after terrain correction of the same area, projected into real-world geographic coordinates.

The **displacement**, Figure 3.17 band contains LOS surface displacement, which is derived from the unwrapped interferometric phase. The displacement, expressed in meters, represents the motion of the ground surface between two acquisitions: positive values correspond to a motion toward the satellite, while negative ones

represent a motion away form the satellites.

The **local incidence angle**, shown in Figure 3.18, describes the angle between the radar LOS and the local surface normal, taking into account terrain slope and aspect, influencing both backscatter intensity and coherence. It is particularly significant in mountainous or sloped terrain for correctly interpreting radar measurements, as in this specific AOI.

The final product of the InSAR processing is an Interferometric SAR image stored as a multi-band raster. Each band contains information derived from the combination of the master and slave SAR acquisitions and represents a specific physical or geometrical quantity relevant to interferometric analysis.

3.2 Dataset Creation

The dataset employed in this thesis work for the prediction of Snow Depth (HS) and Snow Water Equivalent (SWE) is created integrating three complementary data sources: Copernicus Sentinel-1 mission, Copernicus Sentinel-3 mission, both developed under the European Space Agency (ESA) Copernicus Programme, and ITSNOW[33], a dataset providing snow related measurements over Italy.

The Area Of Interst (AOI) is located in northern Italy, spanning between Lombardia and Trentino-Alto Adige regions, as illustrated in Figure 3.19. This area offers topographic heterogeneity, a wide range of elevations and a persistent seasonal snow cover. Particularly, the presence of the Adamello Glacier makes this area very suitable for snow studies.

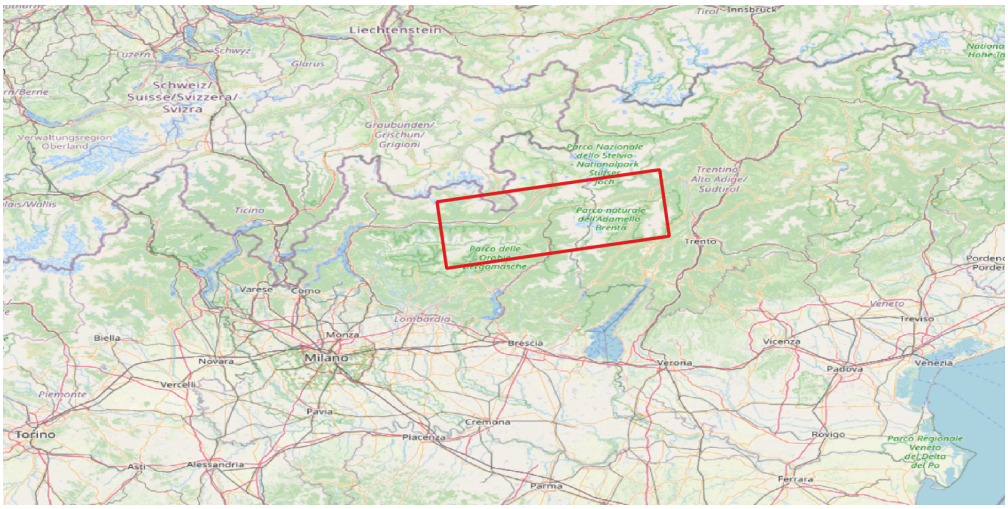

Figure 3.19: Area Of Interest located in the Central Italian Alps, highlighted in red

The temporal coverage of the dataset spans January 2021 to December 2022, which represents the period for training and validation of the model, including a total of 57 InSAR images. Instead the first quarter of 2023 is reserved for testing purposes, for a total of 7 images. The temporal separation is studied to guaranty realistic and unbiased evaluations and to allow generalization on new and unseen data, which may differ from the training one in snow and environmental conditions.

3.2.1 Copernicus Sentinel-1

Copernicus Sentinel-1[34] is the first mission of the ESA's program of satellite constellations for Earth observation. It is composed of four different satellites (Sentinel-1A, Sentinel-1B, Sentinel-1C and Sentinel-1D), which operate day and night to scan the whole Earth's surface. Sentinel-1B completed its disposal in 2024, due to data transmission issues started in 2022. Sentinel-1 is designed to use C-band synthetic aperture radar imaging, which allows them to capture images in all weather and light conditions. The C-band is a precise portion of the electromagnetic spectrum in the microwave range of frequencies: it ranges from 7.5 to 3.75 GHz.

Sentinel-1 satellites follow near-polar, sun-synchronous orbits and can acquire data in both ascending and descending passes. This means it can acquire data both northward or southward.

For this work, only Sentinel-1A mission's data are considered. Sentinel-1A provides a 12-day revisit time, which is a suitable trade-off between temporal resolution and simplicity in processing images. The analysis is based on Single Look Complex (SLC) products because it preserves amplitude and phase information of the radar signal. This characteristic makes SLC essential for interferometric applications, as the phase is required to compute interferograms.

Moreover, to spatial consistency in all the acquisitions, the orbit number, the slice number, the swath and the bursts indices are selected a priori with parameters shown in table 3.1. The orbit number represents the trajectory followed by the satellite around the Earth at the time of acquisition, while the slice number refers to temporal subdivision of a longer acquisition along the satellite ground track. By fixing these parameters, a precise rectangular area is obtained. Since this area is larger than the selected study area, the swath and the bursts are specified to isolate the AOI.

Orbit number	117
Slice number	11
Swath	IW1
Bursts	4 and 5

Table 3.1: Parameters for the area's selection

In this work, only data acquired in VV polarization are considered. The radar signal is transmitted and received with vertical polarization, indicating that the electric field oscillates in the vertical plane both during transmission and reception. Among the polarization modes available for Sentinel-1, VV is particularly suitable for interferometric applications due to its higher signal stability and stronger backscattering over natural surfaces. VV polarization generally provides higher coherence compared to cross-polarized channels, especially in mountainous and snow-covered areas. This is because of the reduced sensitivity to volume scattering effects caused by vegetation and complex surface structures. In northern Italy alpine regions, with steep slopes, rocky peek, and seasonal snow, VV polarization offers a more reliable phase signal.

In Table 3.2 the acquisition dates are reported. A total of 65 acquisitions is collected, resulting in 64 InSAR images, each computed between two consecutive acquisitions, as described in Data Processing paragraph. Each of the created interferometric product is composed of the 7 interferometric bands:

- Coherence
- Phase
- Elevation
- Intensity
- Unwrapping
- Local Incidence Angle
- Displacement

ID	Date	Sensor	Orbit	ID	Date	Sensor	Orbit
1	2021-01-09	S1A	Asc.	34	2022-03-05	S1A	Asc.
2	2021-01-21	S1A	Asc.	35	2022-03-17	S1A	Asc.
3	2021-02-02	S1A	Asc.	36	2022-03-29	S1A	Asc.
4	2021-02-14	S1A	Asc.	37	2022-04-10	S1A	Asc.
5	2021-03-10	S1A	Asc.	38	2022-04-22	S1A	Asc.
6	2021-03-22	S1A	Asc.	39	2022-05-04	S1A	Asc.
7	2021-04-03	S1A	Asc.	40	2022-05-16	S1A	Asc.
8	2021-04-15	S1A	Asc.	41	2022-05-28	S1A	Asc.
9	2021-04-27	S1A	Asc.	42	2022-06-21	S1A	Asc.
10	2021-05-09	S1A	Asc.	43	2022-07-03	S1A	Asc.
11	2021-05-21	S1A	Asc.	44	2022-07-15	S1A	Asc.
12	2021-06-02	S1A	Asc.	45	2022-07-27	S1A	Asc.
13	2021-06-14	S1A	Asc.	46	2022-08-08	S1A	Asc.
14	2021-06-26	S1A	Asc.	47	2022-08-20	S1A	Asc.
15	2021-07-08	S1A	Asc.	48	2022-09-01	S1A	Asc.
16	2021-07-20	S1A	Asc.	49	2022-09-13	S1A	Asc.
17	2021-08-01	S1A	Asc.	50	2022-09-25	S1A	Asc.
18	2021-08-13	S1A	Asc.	51	2022-10-07	S1A	Asc.
19	2021-09-06	S1A	Asc.	52	2022-10-19	S1A	Asc.
20	2021-09-18	S1A	Asc.	53	2022-10-31	S1A	Asc.
21	2021-09-30	S1A	Asc.	54	2022-11-12	S1A	Asc.
22	2021-10-12	S1A	Asc.	55	2022-11-24	S1A	Asc.
23	2021-10-24	S1A	Asc.	56	2022-12-06	S1A	Asc.
24	2021-11-05	S1A	Asc.	57	2022-12-18	S1A	Asc.
25	2021-11-17	S1A	Asc.	58	2022-12-30	S1A	Asc.
26	2021-11-29	S1A	Asc.	59	2023-01-11	S1A	Asc.
27	2021-12-11	S1A	Asc.	60	2023-01-23	S1A	Asc.
28	2021-12-23	S1A	Asc.	61	2023-02-04	S1A	Asc.
29	2022-01-04	S1A	Asc.	62	2023-02-16	S1A	Asc.
30	2022-01-16	S1A	Asc.	63	2023-02-28	S1A	Asc.
31	2022-01-28	S1A	Asc.	64	2023-03-12	S1A	Asc.
32	2022-02-09	S1A	Asc.	65	2023-03-24	S1A	Asc.
33	2022-02-21	S1A	Asc.				

Table 3.2: Sentinel-1 SAR acquisition dates used for InSAR processing.

Although the Sentinel-1A mission has a revisit time of 12 days, this time interval is not always respected (e.g. time interval between 2021-02-14 and 2021-03-10). Due to data availability on Copernicus API, some time intervals correspond to multiples of 12 days. However, this does not represent a limitation since interferograms

remain reliable. In the observations of longer time intervals, coherence values are comparable to those obtained with the nominal revisit time, suggesting usability and consistency of these interferometric products.

3.2.2 Copernicus Sentinel-3

Sentinel-3[35] is another Copernicus Program satellites mission for Earth's observations developed by ESA. It consists in 2 satellites Sentinel-3A and Sentinel3-B with a very short revisit time, allowing the earth to be imaged at least once a day.

The main objective of Sentinel-3 mission is to measure oceans, land, ice and atmosphere to provide ocean and weather forecasting, through temperature, sea-surface topography and ocean-surface colour. For this thesis work, Sentinel-3 SLSTR[36][37] Level 2 products are employed for the sea and land surface temperature, an information that enrich InSAR data.

Since it is a dual-view thermal infrared radiance signal, it has not the characteristics of SAR images. For these reasons, SLSTR data are interested in cloud cover and atmospheric effects.

Taking into account the study time interval, from January 2021 to March 2023, all available Sentinel-3 SLSTR products are considered. It is not possible to have one product per day due to atmospheric conditions, and also the majority of the products have missing values for the same reason. For these reasons, a filling technique is applied to the images: the Sentinel-3 images collection is treated as a multi-temporal stack, where each pixel is processed as an individual time series, performing a linear interpolation along the temporal axis for each pixel, illustrated in Figure 3.20. In this way, the temporal coherence in the temperature trend is maintained.

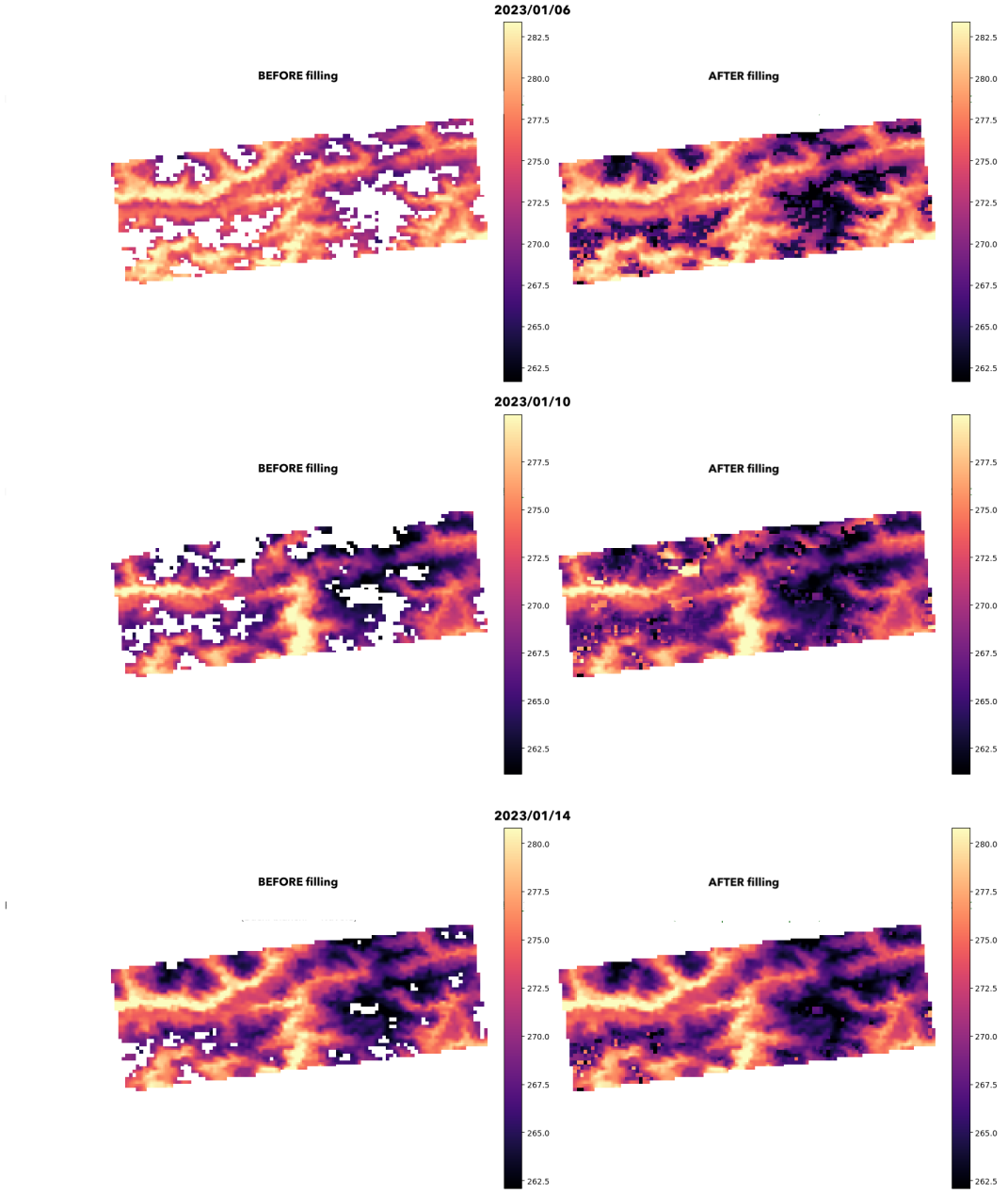

Figure 3.20: Example of before and after temporal filling strategy for Sentinel-3 acquisitions

Once all pixels have a valid value, the temperature data are processed to obtain two temperature bands, which give additional information for each time interval: Mean Temperature and Positive Degree Day(PDD).

Mean Temperature

The Mean Temperature represents the average thermal conditions of the surface over the considered time intervals. It provides essential information about the snowpack evolution.
Within each interval of time, computing the mean temperature suggests accumulation or melting trends of the snow. In particular, mean temperature indicates a melting tendency when temperatures are steadily above zero degree, while accumulation may occur when temperatures are lower than zero. In addition, it can also suggest the phase of precipitation, which influences the snowpack.

Positive Degree Day

The Positive Degree Day value represents a cumulative variable defined as the sum of daily temperatures above 0 °C over a specified period[38].

$$\mathrm{PDD} = \sum \max\left(0,\, T_{\mathrm{day}} - 273.15\right)$$

where T_{day} is the daily surface temperature expressed in Kelvin. This formulation converts temperature values to degrees Celsius and accumulates only positive contributions, excluding days with sub-freezing conditions.

PDD provides a physically meaningful indicator for the intensity and persistence of melting processes. Unlike mean temperature, which describes the instantaneous thermal state, PDD captures the cumulative impact of positive temperatures between two observation dates. It is strictly related with InSAR dynamics: high PDD values are linked to low coherence due to rapid structural changes in wet and meltring snow, while low values of PDD suggest higher coherence. Consequently, Positive Degree Day may help discriminate whether interferometric changes are valuable for the prediction of HS and SWE through the learning process.

3.2.3 ITSNOW

IT-SNOW[33] is a high resolution, multi-year reanalysis of spatial and temporal consistent estimates of snow conditions over the Italian territory. The dataset covers a period from September 2010 and is updated to August 2024. It is an openly available dataset, developed with objective of overcoming the limitations of existing snow observation sources, as sparsity of in-situ measurements, data gaps in satellite observations and uncertainties in sensing techniques.

IT-SNOW provides daily 500 m spatial resolution maps of four bulk snowpack variables: snow water equivalent (SWE), snow depth (HS), bulk snow density, and liquid water content. Since it is produced as a reanalysis output of the S3M Italy operational chain (Snow Multidata Mapping and Modeling system), used by the Italia Civil Protection for real-time snow monitoring, it is considered a valuable dataset in the cryospheric field. Indeed, IT-SNOW is not a purely model-driven product, it relies on the assimilation of heterogeneous observational data, allowing uncertainties reduction. Unlike purely satellite-based products, IT-SNOW explicitly reconstructs snow mass and depth, enabling quantitative analyses of water resources and snow dynamics, making it a reliable and appropriate data source.

In particular, IT-SNOW integrates three main sources of information:

1. Meteorological data including precipitation, air temperature, relative humidity, and solar radiation measured from thousands of weather stations across Italy

2. satellite-derived snow-covered area maps, on a daily basis, and combining observations from Sentinel-2, MODIS, and EUMETSAT H-SAF products. These measurements allow the system to capture the spatial extent of snow cover and to constrain snow presence or absence in areas with limited ground observations

3. Maps of snow depth derived from more than 350 in-situ snow depth sensors distributed across the country, which improve the realism of snow depth and SWE estimates

For these reasons, IT-SNOW is particularly well suited as a reference dataset for data-driven studies aimed at estimating snow depth and snow water equivalent. In this work, IT-SNOW is adopted as the target dataset for snow estimation; however, only the snow depth (HS) and snow water equivalent (SWE) variables are considered, as they represent the target values of the learning prediction process.

Moreover, the target quantities are processed in a differential form and they are not used their absolute values as provided in the reanalysis. Since the input data consist of interferometric SAR products computed over discrete time intervals,

the learning task is formulated to predict changes in snow conditions over time. Therefore, for each acquisition interval, the target HS and SWE values are computed as the difference between the IT-SNOW estimates at the beginning of the interval and at the end of it. The target is a variation of the snow conditions to ensure consistency between the temporal nature of the InSAR observations and the target variables. This allows the model to learn snow accumulation and melting dynamics rather than absolute snow states.

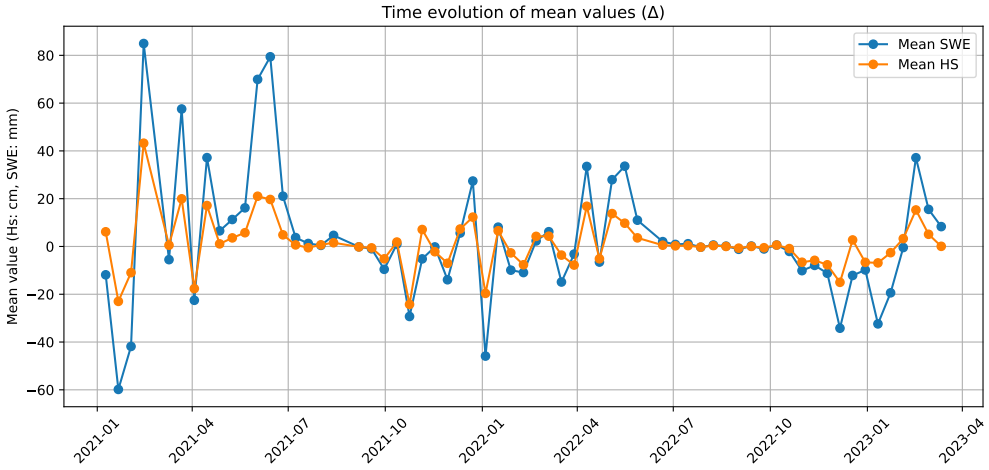

Figure 3.21: Temporal evolution of Snow Depth and Snow Water Equivalent evolution over the whole study period of time

Figures 3.21 and 3.22 provide an overview of the temporal dynamics and statistical variability of the IT-SNOW SWE and HS over the entire study period.

In particular, Figure 3.21 shows the temporal evolution of the mean spatial variation of SWE and HS between consecutive acquisitions. Both variables exhibit a clear seasonal pattern characterized by positive peaks during accumulation phases and negative peaks during melt periods. Early 2021 shows extreme values of SWE, meaning strong accumulation and intense melt processes. The close correspondence between SWE and HS confirms their physical correlation, although SWE generally exhibits higher variability due to its sensitivity to density variations and liquid water content.

An important characteristic emerges from the mean values: they remain consistently close to zero for most dates. This reflects the distribution of the dataset, where a large proportion of pixels correspond to snow-free areas, which assume zero values.

Figure 3.22 investigates the spatial distribution by reporting the minimum and maximum values for each date. The wide spread between extreme values highlights the strong spatial heterogeneity of snow conditions within the study area.

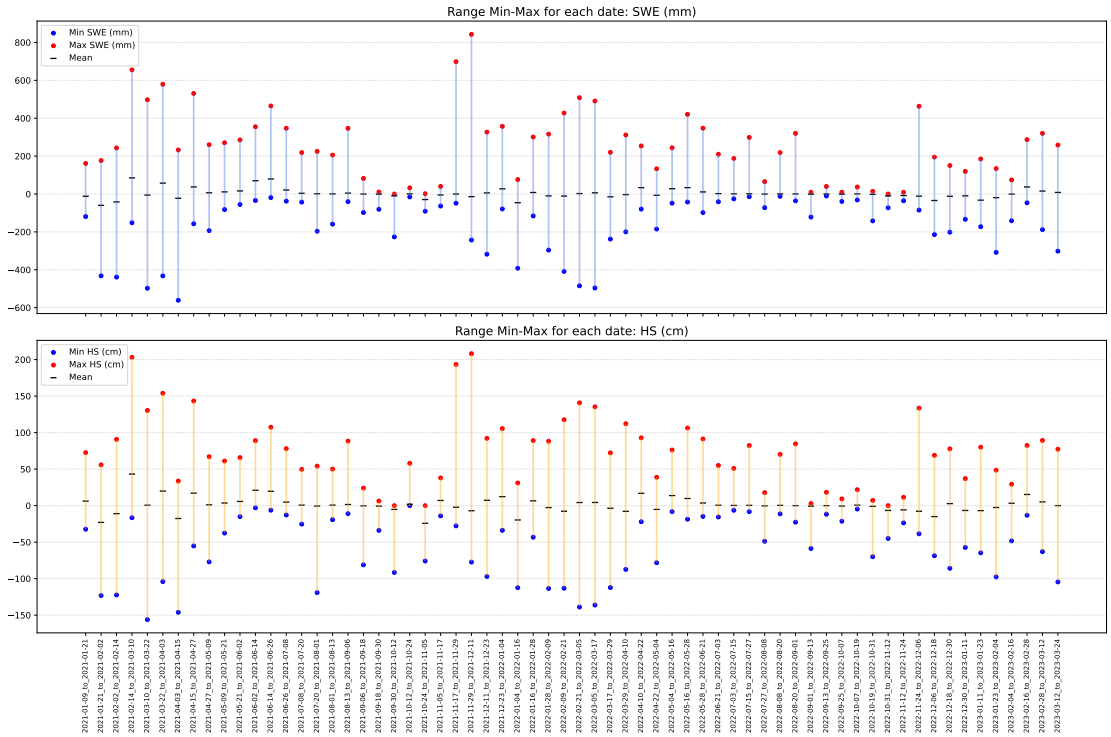

Figure 3.22: Minimum and Maximum values of Snow Depth and Snow Water Equivalent to inspect unusual behaviour.

Periods with large max–min ranges correspond to intense snow accumulation or rapid melting events, due to the influence of elevation, exposure, and temperature variability. During summer periods reduced variability and near zero values reflects the presence of snow-free conditions.

In Figure 3.23 all the input and output bands are illustrated.

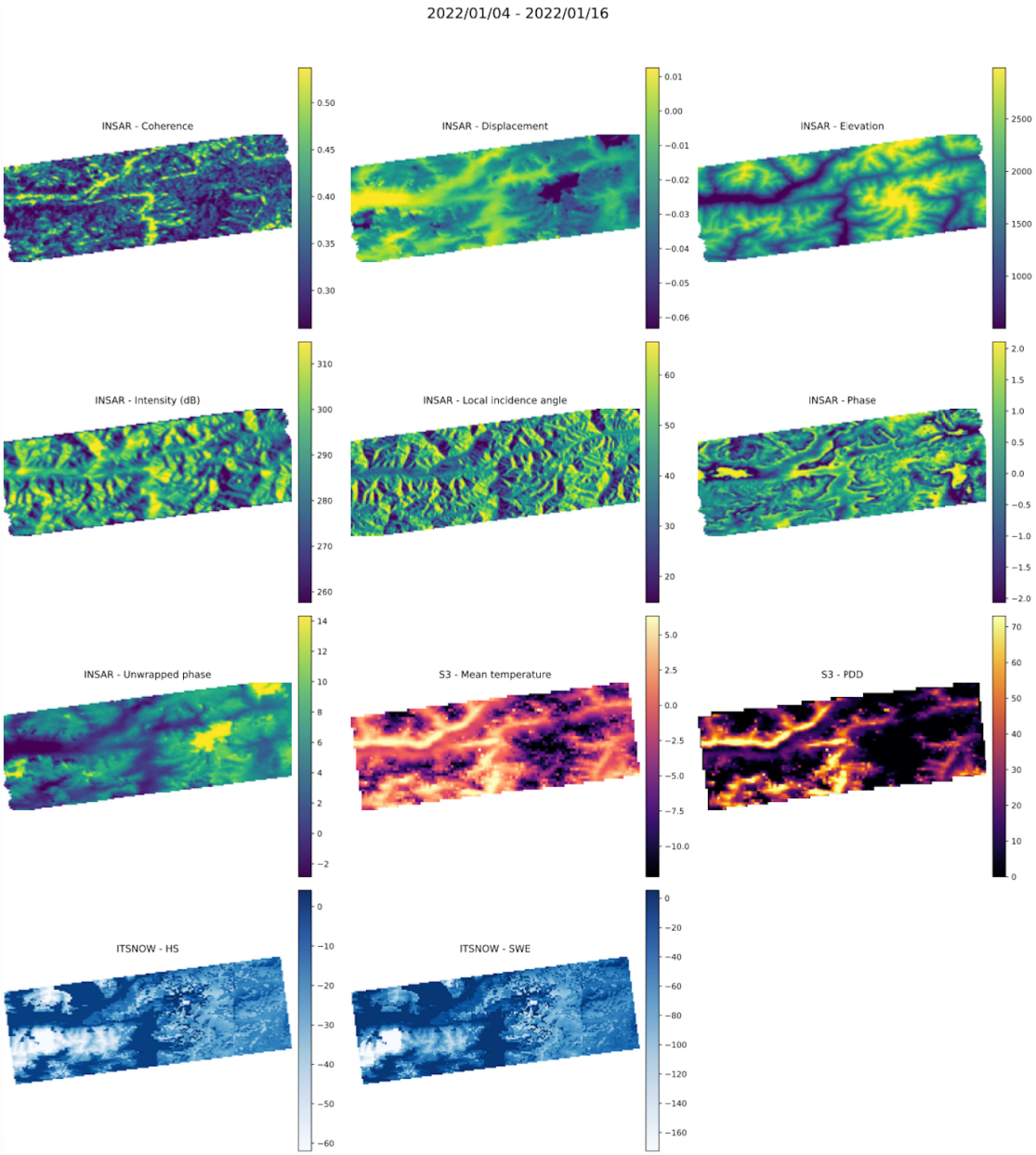

Figure 3.23: Example of all bands of 2022/01/04 - 2022/01/16

Chapter 4

Methodology

This chapter describes the methodology adopted in this study. A preprocessing pipeline is introduced to harmonize the different data sources: specific procedures are applied to ensure spatial alignment and dimensional consistency across all input bands. The preprocessing phase also includes the computation of temporal embedding features, designed to encode temporal context, and the concatenation of all input layers into a unified multi-band tensor suitable for model input.

In addition, a binary mask is generated to exclude invalid pixels from the following computations. These invalid pixels include border areas introduced by the terrain correction process and pixels located outside the Italian territory, which are not covered by the IT-SNOW dataset and filled with invalid target values.

Following data preparation, different predictive models are investigated for the estimation of SWE and HS. Specifically, three architectures are considered: XGBoost, U-Net, and SegFormer. These models are selected to represent different learning paradigms, allowing for a comparative evaluation of their performance on the proposed dataset.

Finally, the training strategy is described in detail. This includes the augmentations applied, the temporal split adopted for training, validation, and testing, and the formulation of the loss function and its computation over valid pixels only.

4.1 Data preprocessing

4.1.1 Data Harmonization

Since the final dataset integrates information from Sentinel-1 (InSAR products), Sentinel-3 and IT-SNOW reanalysis, which have different spatial resolutions and grid definitions, a spatial harmonization procedure is required to ensure geometric consistency across all inputs.

To guarantee pixel-wise comparability, all input layers are resampled to the spatial resolution, coordinate reference system (CRS), and affine transform of IT-SNOW, which is adopted as the reference grid. Thus, each pixel in the final dataset corresponds to the same geographic location and spatial extent across all data layers, preventing spatial mismatches that could introduce noise, artificial gradients, or bias into the learning process.

The InSAR products from Sentinel-1 and temperature data from Sentinel-3 are verified to match the IT-SNOW reference grid in terms of spatial dimensions, affine transformation matrix (geometric alignment) and coordinate reference system. Since Sentinel-3 has a coarse resolution compared to IT-SNOW, a bilinear interpolation strategy is used to ensure upscaling. Instead, Sentinel-1 data are downscaled to the approximately 500m IT-SNOW resolution.

The reason behind the choice of the selection of IT-SNOW as the harmonization reference is because it represents the target variable of the prediction task. Resampling all predictors to the target grid avoids interpolation of ground-truth values and preserves the physical consistency of SWE and HS.

4.1.2 Input Bands Concatenation

The input representation adopted in this work is designed to integrate spatially distributed physical information with explicit temporal context. Each training sample is constructed as a multi-channel tensor obtained by stacking interferometric features, auxiliary satellite temperature variables, and temporal embedding information. The final input consists of 13 bands:

- 7 Sentinel-1 InSAR bands
- 2 Sentinel-3 temperature bands
- 4 Temporal embedding bands

Temporal Embeddings

Temporal information is added using a cyclical embedding strategy to represent the time interval between which the interferometric product is produced. For each time interval, the start and end dates are extracted and converted into day-of-year, to which a sinusoidal encoding is applied using a yearly periodicity. The use of sine and cosine encoding ensures that temporal proximity is preserved in the feature space. In this way, the model receives global seasonal context while still learning spatial patterns from the interferometric and environmental features.

4.1.3 Mask Generation

A binary mask of the same size as the input bands is applied to them to distinguish between valid and invalid pixels. The mask is generated after the terrain Correction, the last step of the InSAR processing, and it reflects the actual spatial coverage of the images in geographic coordinates. As illustrated in Figure 4.1, white (1 value) pixels represent the valid region, while black pixels are set to 0 value.

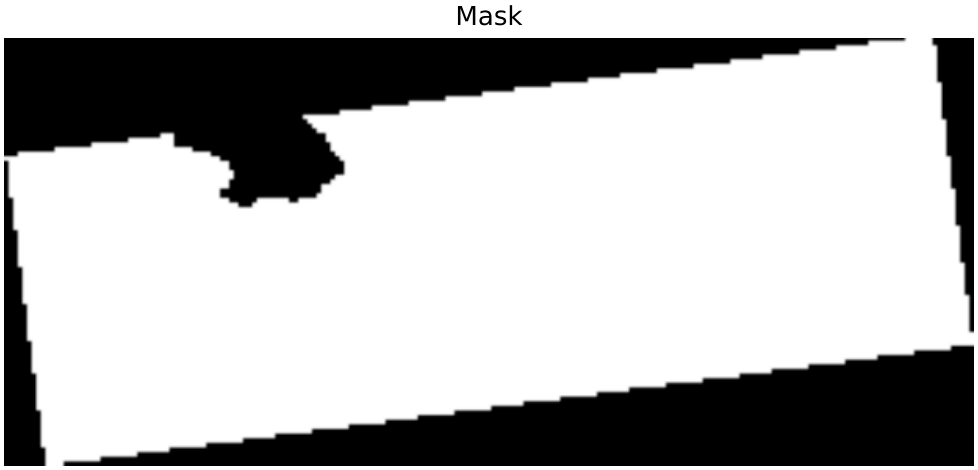

Figure 4.1: Binary mask of valid (white) and invalid(black) pixels

The invalid regions appear along the image borders, introduced by the geocoding procedure. Additionally, the upper-left portion of the mask corresponds to areas outside the Italian territory, specifically it is a part of Switzerland that falls within the radar acquisition trajectory. Since the analysis focuses on Italian territory covered by IT-SNOW reference data, these regions are deliberately excluded.

The mask is applied during model training and testing in order to consider only physically and geographically relevant pixels. This allows preventing the learning process from being influenced by borders values and areas outside the AOI.

4.2 Model Architectures

The design of the proposed models is driven by their suitability or a regression task aimed at predicting continuous snow-related variables in a unique framework. Multiple architectures are explored to evaluate which approach is more reliable and accurate for the task.

This section outlines the architectural configurations adopted in the experiments, while the theoretical foundations of the models are discussed in Chapter 2. Here the emphasis is on their practical implementation, structural properties, and the reasons why these architectures are suitable for this study.

Three established models are considered:

XGBoost

XGBoost is introduced as the baseline model due to its strong performance in tabular regression tasks. It operates on pixel-wise feature vectors derived from the stacked interferometric and environmental bands. Its strong regularization mechanisms, robustness to noise, and low computational cost make it a state-of-the-art reference point SWE and HS predictions.

U-Net with ResNet18 encoder

U-Net is selected as it is designed for dense prediction tasks. Its encoder–decoder structure allows the model to learn hierarchical spatial representations while preserving spatial details, making it suitable for snow metrics predictions as they are spatially correlated and exhibits a significant variability across different areas of the region under analysis.

ResNet18 is employed as the encoder for its lightweight backbone compared to deeper variants, which makes it suitable for datasets with limited size.

SegFormer with MiT b0 encoder

SegFormer leverages attention mechanisms to model global spatial dependencies more effectively than CNNs as U-Net.

The MiT b0 backbone is selected as the encoder due to its compact size and computational efficiency. Among the MiT variants, b0 is the lightest configuration, making it appropriate for scenarios with limited training samples.

4.3 Training Strategies

4.3.1 Padding and Data Augmentations

All the input bands are padded to a fixed spatial resolution of 256x256 pixels to allow encoder structures to perform successive downsampling operations. It is performed using reflective padding, which allows spatial continuity and does not introduce artifacts. The target mask was padded using constant zero values so that the artificially added pixels do not contribute to the learning process.

To enhance robustness and increase the number of available samples, spatial and geometric data augmentation is applied during training. The augmentation strategy was designed to preserve spatial consistency between inputs, targets, and validity masks. For this reason, all transformations are applied to input channels, target variables and the associated mask.

The adopted augmentations, implemented using Albumentation library[39], include a combination of random horizontal and vertical flips, random 90-degree rotations, and a combined shift–scale–rotate operation.

To increase the effective size of the training set, a virtual dataset expansion strategy was implemented. Each original sample is accessed multiple times during training, 20 times, with independent random augmentations applied at each iteration. This approach allows the model to observe diverse spatial variations of the same interferometric scene, improving generalization while maintaining computational efficiency and at the same time increase the number of training samples.

4.3.2 Loss Function

Huber loss

The base regression criterion used to compute the pixel-wise error is the Huber loss. Let y be the ground truth value and $\hat{y}$ the predicted value. The residual is defined as:

$$r = \hat{y} - y \tag{4.1}$$

The Huber loss with threshold parameter $\delta > 0$ is defined as:

$$\mathcal{L}_\delta(r) = \begin{cases} \frac{1}{2}r^2, & \text{if } |r| \leq \delta \\ \delta\left(|r| - \frac{1}{2}\delta\right), & \text{if } |r| > \delta \end{cases} \tag{4.2}$$

The Huber loss combines the advantages of mean squared error (MSE) and mean absolute error (MAE). For small prediction errors, The Huber Loss behaves as MSE, with precise fitting and smooth gradient updates, while for large errors, it behaves like MAE, reducing the influence of outliers. The parameter δ controls the transition point between the quadratic and linear regions.

This is relevant in this setting, where interferometric data may exhibit local artifacts caused by low-coherence areas or residual unwrapping errors, since its robustness to outliers and noisy targets. Moreover, the masked Huber formulation ensures that learning is driven by physically meaningful pixels and that the optimization process remains well balanced between the two predicted variables.

The loss is computed separately for Snow Depth and Snow Water Equivalent, and the final training objective is defined as the average of the two channel-wise losses as shown in 4.3. This allows balanced optimization between the two target variables.

$$\mathcal{L} = \frac{1}{2}\left(\mathcal{L}_{\text{HS}} + \mathcal{L}_{\text{SWE}}\right). \tag{4.3}$$

4.3.3 Splitting Strategy

The data employed in this the sis work have a temporal nature, so the splitting strategy follows time-consistent logic. The split is applied at temporal interval level and not pixel-level since it would have introduced temporal leakage and led to optimistic estimations.

The 64 available interferometric products are split as shown in Table 4.1. Among them, approximately 75% are allocated for the training, 14% for validation and approximately 11% are reserved for the final test set. The split is performed in chronological order: the model is trained on earlier temporal intervals and evaluated on temporally subsequent ones to have an adequate generalization throughout the year. This split strategy is applied for U-Net and Segformer, in particular, the validation set is utilized for implementing early stopping. The XGBoost model, on the contrary, is trained on the entire training and validation set, as it does not strictly require a validation set for its standard training procedure.

The test set, instead, consists of unseen future intervals of a different year with respect to training and validation sets. This ensures that performance is assessed on a coherent seasonal block rather than isolated samples, making the evaluation more robust and representative of real-world application scenarios.

This splitting strategy is used to avoid similar seasonal conditions across training and testing sets, which would have allowed the model to implicitly exploit temporal correlations rather than learning physically meaningful relationships between InSAR observables and snow variables. Moreover, consecutive interferograms are correlated because they share one common acquisition date and are computed over partially overlapping time intervals. By preserving chronological order, the adopted strategy reduces this dependency and provides a more rigorous evaluation of generalization capability.

InSAR data	**Split**	**InSAR data**	**Split**
2021-01-09 – 2021-01-21	Training	2022-02-21 – 2022-03-05	Training
2021-01-21 – 2021-02-02	Training	2022-03-05 – 2022-03-17	Training
2021-02-02 – 2021-02-14	Training	2022-03-17 – 2022-03-29	Training
2021-02-14 – 2021-03-10	Training	2022-03-29 – 2022-04-10	Validation
2021-03-10 – 2021-03-22	Training	2022-04-10 – 2022-04-22	Training
2021-03-22 – 2021-04-03	Validation	2022-04-22 – 2022-05-04	Training
2021-04-03 – 2021-04-15	Training	2022-05-04 – 2022-05-16	Training
2021-04-15 – 2021-04-27	Training	2022-05-16 – 2022-05-28	Training
2021-04-27 – 2021-05-09	Training	2022-05-28 – 2022-06-21	Training
2021-05-09 – 2021-05-21	Training	2022-06-21 – 2022-07-03	Validation
2021-05-21 – 2021-06-02	Training	2022-07-03 – 2022-07-15	Training
2021-06-02 – 2021-06-14	Validation	2022-07-15 – 2022-07-27	Training
2021-06-14 – 2021-06-26	Training	2022-07-27 – 2022-08-08	Training
2021-06-26 – 2021-07-08	Training	2022-08-08 – 2022-08-20	Training
2021-07-08 – 2021-07-20	Training	2022-08-20 – 2022-09-01	Training
2021-07-20 – 2021-08-01	Training	2022-09-01 – 2022-09-13	Validation
2021-08-01 – 2021-08-13	Training	2022-09-13 – 2022-09-25	Training
2021-08-13 – 2021-09-06	Validation	2022-09-25 – 2022-10-07	Training
2021-09-06 – 2021-09-18	Training	2022-10-07 – 2022-10-19	Training
2021-09-18 – 2021-09-30	Training	2022-10-19 – 2022-10-31	Training
2021-09-30 – 2021-10-12	Training	2022-10-31 – 2022-11-12	Training
2021-10-12 – 2021-10-24	Training	2022-11-12 – 2022-11-24	Validation
2021-10-24 – 2021-11-05	Training	2022-11-24 – 2022-12-06	Training
2021-11-05 – 2021-11-17	Validation	2022-12-06 – 2022-12-18	Training
2021-11-17 – 2021-11-29	Training	2022-12-18 – 2022-12-30	Training
2021-11-29 – 2021-12-11	Training	2022-12-30 – 2023-01-11	Testing
2021-12-11 – 2021-12-23	Training	2023-01-11 – 2023-01-23	Testing
2021-12-23 – 2022-01-04	Training	2023-01-23 – 2023-02-04	Testing
2022-01-04 – 2022-01-16	Training	2023-02-04 – 2023-02-16	Testing
2022-01-16 – 2022-01-28	Validation	2023-02-16 – 2023-02-28	Testing
2022-01-28 – 2022-02-09	Training	2023-02-28 – 2023-03-12	Testing
2022-02-09 – 2022-02-21	Training	2023-03-12 – 2023-03-24	Testing

Table 4.1: Splitting of input data in Training, Validation e Testing

4.3.4 Ablation Study

An ablation study is conducted to assess the importance of the interferometric features, performed by removing one input band at a time. The removed bands are those from Sentinel-1, so the interferometric data, in order to quantify the individual contribution of each band to the regression task. The model is always retrained under the same training configuration and performance variations are analyzed. A degradation in predictive accuracy indicates that the removed feature carries relevant information for the estimation of HS and SWE, while a negligible variation suggests limited marginal contribution. This ablation study is important in this work, where the input stack includes multiple interferometric data because it helps to interpret the physical relevance of each radar-derived variable, reduce redundancy among correlated bands, and improve model efficiency. Temperature variables and temporal embeddings are excluded from the study since they represent seasonal evolution and distinguishing between dry and wet snow regimes. By removing them, the physical context of the problem would be altered.

4.4 Evaluation Strategy

The evaluation of regression models requires quantitative metrics capable of measuring the discrepancy between predicted and ground-truth values. In this context, where the model outputs spatially distributed continuous variables, the performance assessment captures both the magnitude of the average error and the sensitivity to large deviations.

In this work, model performance is evaluated on the 7 measurements from 2023 data using two standard regression metrics: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Since RMSE and MAE provide complementary information about accuracy of predictions, they are largely adopted in geophysical and environmental modeling studies[40].

While MAE measures the average absolute deviation between predictions and ground truth, RMSE emphasizes larger errors due to its quadratic formulation. The joint use of these metrics allows for a balanced assessment of model robustness and sensitivity to outliers.

4.4.1 Mean Absolute Error

The Mean Absolute Error is the computation of the average absolute differences between the expected and actual values. Given a set of n predictions $\hat{y}_i$ and corresponding ground truth values y_i, MAE is defined as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \,. \tag{4.4}$$

MAE provides a direct measure of the average prediction error expressed in the same physical units as the target variable. It does not square the residuals, so MAE treats all deviations linearly and is less sensitive to extreme errors and outliers.

The lower the value of MAE, the more accurate the model is.

4.4.2 Root Mean Squared Error

The Root Mean Squared Error (RMSE) measures the square root of the average squared differences between predictions and reference values:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}. \tag{4.5}$$

The square operation before computing the average allows RMSE to assign a greater weight to large errors, making it sensitive to outliers and extreme deviations

than MAE. The square root operation restores the metric to the original unit of measurement, enabling direct comparison and interpretation.

RMSE is informative when large prediction errors are undesirable or when robustness to extreme deviations must be explicitly evaluated, as in this work. The comparison between MAE and RMSE provides insight into the error distribution: if RMSE is significantly larger than MAE, it indicates the presence of large residuals.

Chapter 5

Results

This chapter presents the experimental results obtained following the methodology described in Chapter 4. The evaluation metrics adopted for the regression task are the ones presented in Section 4.4. The study is structured progressively, starting from a strong machine learning baseline, the tree-based ensemble method XGBoost, U-Net and SegFormer improve performances over classical tabular learning methods, allowing spatial consistency and global understanding. The results are analyzed in terms of overall predictive accuracy, robustness to large residuals, and contribution of individual input bands, with the objective of understanding both quantitative performance and the informational value of interferometric observables.

5.1 Implementation Details

Two separate XGBoost models are trained for HS and SWE, with pseudo Huber loss as objective, a maximum of 500 estimators, a maximum of 6 of tree depth and a conservative learning rate of 0.1. These choices provide a sufficient nonlinear modeling capacity while limiting the risk of overfit. To enhance model robustness, row sampling and feature sampling are set to 0.8, allowing generalization improvement.

Instead, deep learning architectures perform joint multi-target regression, simultaneously predicting HS and SWE changes, with the same training configuration: spatially continuous patches of 256x256, AdamW as optimizer, Cosine Annealing scheduler, batch size of 16, learning rate at 10^{-5} and a weight decay of 10^{-2}, a maximum of 200 epochs training with early stopping patience of 70 and a pixel-wise masked Huber loss ($\delta = 5$).

The value of δ is chosen to balance precision and robustness: in the context of SWE and snow depth estimation, where localized large errors may occur due to heterogeneous terrain or complex snow conditions, this choice prevents instability while preserving sensitivity to meaningful deviations. The U-Net architecture was

implemented using a ResNet-18 encoder backbone. ResNet-18 was selected as a compromise between representational capacity and model complexity, ensuring sufficient feature extraction depth while avoiding excessive parameterization given the available dataset size. The SegFormer model was implemented using the MiT-B0 encoder, corresponding to the smallest variant of the Mix Transformer family. The B0 configuration was chosen to limit computational complexity and reduce overfitting risk, while still benefiting from hierarchical self-attention and multi-scale feature extraction. To further improve generalization in the Transformer-based architecture, a decoder dropout rate of 0.2 os applied in the SegFormer decoder to reduce overfitting during the multi-scale feature aggregation process. A drop-path rate of 0.1 is used within the encoder for regularization. Each experiment was repeated three times with different random initializations (seeds: 42, 123, 2024), and results are reported as mean and standard deviation. Training is performed on NVIDIA GeForce RTX 2080 Ti GPUs (11 GB VRAM). The available memory constrained batch size selection but allowed efficient processing of multi-channel spatial inputs.

5.2 InSAR bands Importance Study

To investigate the contribution of individual InSAR features, a feature importance analysis is first conducted using the XGBoost baseline model. Due to its tree-based structure, XGBoost provides a natural mechanism to estimate the relative importance of input variables based on their contribution to decision splits and error reduction. This preliminary analysis offers an initial indication of which interferometric observables carried the strongest predictive signal for snow depth and SWE estimation.

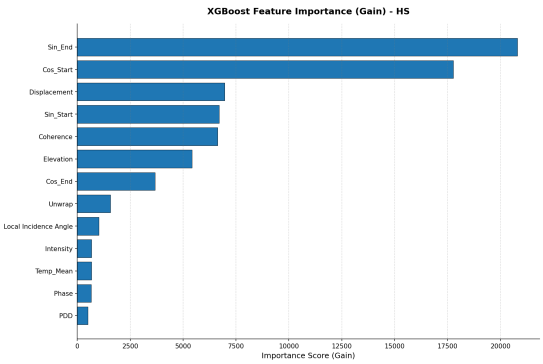

Figure 5.1: HS Feature Importance

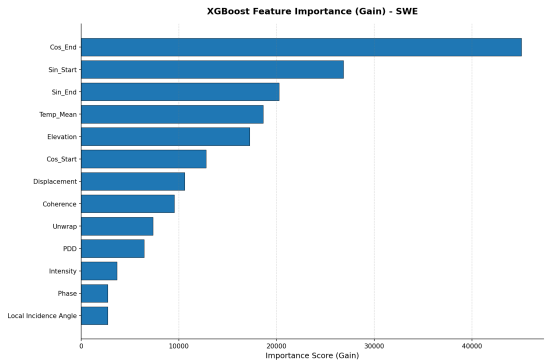

Figure 5.2: SWE Feature Importance

For HS, illustrated in Figure 5.1, temporal embeddings dominate the ranking, emerging as the most influential features. This indicates that seasonal information

plays a primary role in explaining snow depth variability. Among the InSAR observables, displacement and coherence appear as the most relevant interferometric predictors, suggesting that geometric and phase-related changes carry meaningful information for HS estimation. Elevation also contributes significantly, reflecting the strong dependence of snow accumulation on topographic gradients.

For SWE, shown in Figure 5.2, the dominance of temporal encodings is even more pronounced. Temperature-related variables gain relative importance compared to HS, confirming the sensitivity of SWE to thermodynamic conditions. Among radar-derived signals, displacement and coherence remain relevant.

Intensity, phase, and local incidence angle show limited contribution in both tasks, suggesting that their predictive signal may be redundant with other variables or less directly informative within a tree-based framework.

Building on these preliminary observations, an ablation study is performed to evaluate the importance of each InSAR band across learning paradigms. Excluding one feature at a time from the input stack and retraining the models under identical experimental conditions allows the identification of optimal feature combinations and which InSAR-derived inputs contributed most effectively to predictive accuracy.

Thermal variables and temporal embeddings are intentionally excluded from the ablation study because they encode seasonal variability and wet/dry snow regime transitions. Therefore, they are considered essential features. The objective of the study was to isolate the informational contribution of interferometric observables while maintaining consistent environmental context.

Figure 5.3 represents the graphic representation of Tables 5.1, 5.2, 5.3. It presents the average and standard deviation of the ablation study for HS and SWE prediction using MAE and RMSE. For each feature configuration, the bars represent the mean performance across runs, while the error bars indicate the associated standard deviation, providing a measure of variability and stability. The presence of relatively small standard deviations in several configurations suggests consistent model behavior, whereas larger deviations highlight greater sensitivity to feature selection. This representation allows a clear visual assessment of both central tendency and dispersion, offering insight into the robustness and reliability of the models under different input feature settings.

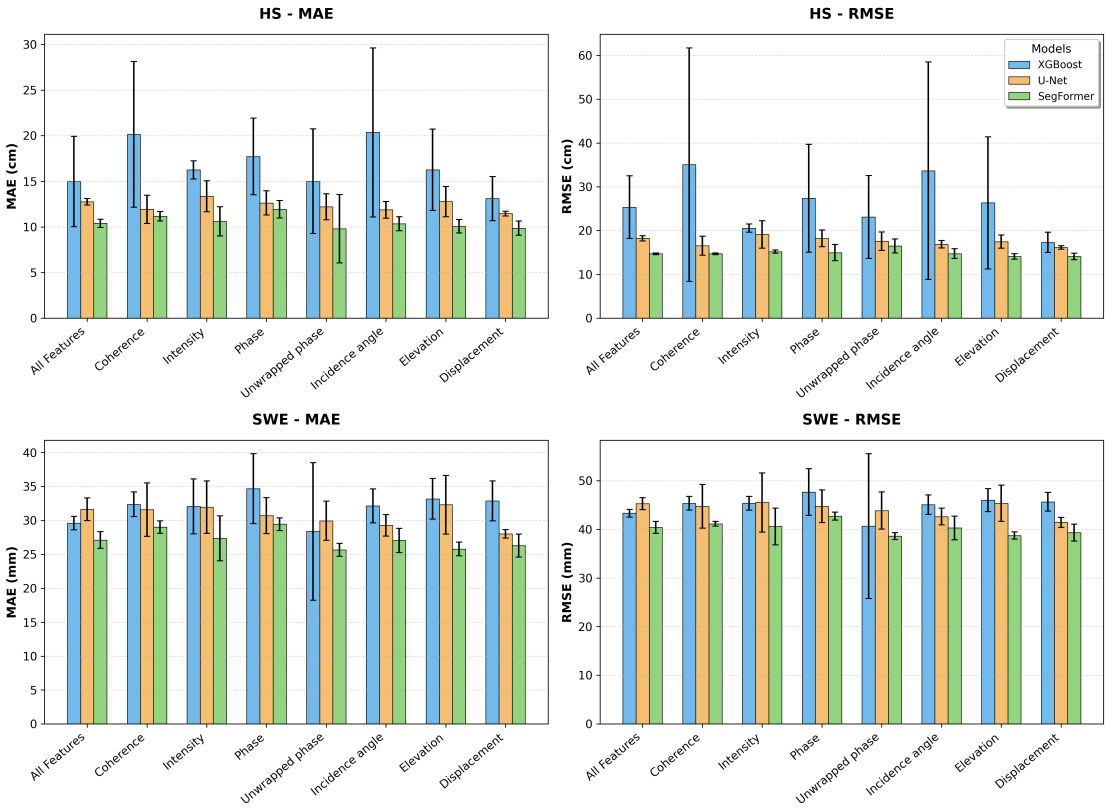

Figure 5.3: Average and standard deviation of the ablation study, the name of the bands represent the bands excluded from the running configuration

	HS		SWE	
	MAE	**RMSE**	**MAE**	**RMSE**
All Features	14.980 ± 4.952	25.346 ± 7.164	29.600 ± 0.983	43.297 ± 0.797
Excluded Feature				
Coherence	20.138 ± 7.988	35.058 ± 26.677	32.375 ± 1.820	45.337 ± 1.413
Intensity	16.251 ± 0.993	20.550 ± 0.921	32.065 ± 4.041	45.337 ± 1.413
Phase	17.720 ± 4.203	27.370 ± 12.313	34.693 ± 5.149	47.648 ± 4.792
Unwrapped phase	15.003 ± 5.739	23.099 ± 9.471	$\mathbf{28.379 \pm 10.135}$	$\mathbf{40.659 \pm 14.880}$
Incidence angle	20.354 ± 9.274	33.661 ± 24.835	32.135 ± 2.503	45.051 ± 2.011
Elevation	16.260 ± 4.468	26.325 ± 15.095	33.185 ± 2.993	45.993 ± 2.367
Displacement	$\mathbf{13.105 \pm 2.423}$	$\mathbf{17.302 \pm 2.291}$	32.869 ± 2.954	45.646 ± 1.903

Table 5.1: Ablation study results for XGBoost

Table 5.1 reports numerical results of the ablation study performed with XGBoost model. The exclusion of individual features produces heterogeneous effects. In the case of HS, removing Displacement leads to the lowest MAE (13.106 cm) and RMSE (17.302 cm) values, suggesting that this feature may introduce redundancy or

noise when combined with the full feature stack. In contrast, excluding Coherence or Incidence Angle results in a clear degradation of performance and a marked increase in standard deviation, indicating reduced stability stability of the model. For SWE, performance variations across feature removals are generally more contained, although excluding the Unwrapped Phase leads to a noticeable increase in variability, as reflected by the higher standard deviation. It is worth noting that Displacement and Unwrapped Phase are physically and mathematically related, as displacement is directly derived from the unwrapped interferometric phase. Their intrinsic correlation can introduce redundancy within the feature space, which may explain why HS benefits from excluding Displacement, while SWE from the removal of the Unwrapped Phase. These findings are consistent with the feature importance analysis where features identified as highly relevant as Coherence and Elevation produce stronger performance degradation when removed.

Table 5.2 presents the ablation for U-Net. In this case, U-Net exhibits a pattern, where most excluded features lead to improved or comparable performance relative to the all-features baseline for HS estimation. For both HS and SWE, removing Displacement leads to the lowest values of MAE and RMSE. HS without displacement reaches MAE of 11.464 and RMSE of 16.142, while the SWE has a MAE of 28.032 and a RMSE of 41.422. The removal of Incidence Angle and Unwrapped Phase also produces competitive results for SWE estimation, whereas excluding Elevation leads to a slight degradation, suggesting that topographic information contributes more consistently to SWE estimation.

	HS		**SWE**	
	MAE	**RMSE**	**MAE**	**RMSE**
All Features	12.754 ± 0.353	18.224 ± 0.580	31.635 ± 1.670	45.263 ± 1.251
Excluded Feature				
Coherence	11.919 ± 1.553	16.542 ± 2.176	31.589 ± 3.950	44.718 ± 4.481
Intensity	13.362 ± 1.700	19.097 ± 3.154	31.959 ± 3.872	45.507 ± 6.094
Phase	12.624 ± 1.323	18.209 ± 1.899	30.715 ± 2.647	44.715 ± 3.374
Unwrapped phase	12.207 ± 1.433	17.537 ± 2.110	29.949 ± 2.874	43.851 ± 3.817
Incidence angle	11.871 ± 0.922	16.877 ± 0.877	29.272 ± 1.591	42.619 ± 1.704
Elevation	12.770 ± 1.667	17.458 ± 1.512	32.310 ± 4.325	45.340 ± 3.715
Displacement	$\mathbf{11.464 \pm 0.262}$	$\mathbf{16.142 \pm 0.418}$	$\mathbf{28.032 \pm 0.625}$	$\mathbf{41.422 \pm 1.022}$

Table 5.2: Ablation study results for U-Net

Overall, U-Net benefits from a partially reduced feature space, where highly correlated inputs may introduce redundancy. The variability patterns further suggest that certain bands primarily contribute to model stability rather than direct predictive accuracy, as Intensity and Elevation.

	HS		SWE	
	MAE	**RMSE**	**MAE**	**RMSE**
All Features	10.391 ± 0.445	14.684 ± 0.210	27.113 ± 1.227	40.388 ± 1.254
Excluded Feature				
Coherence	11.167 ± 0.514	14.684 ± 0.210	29.012 ± 0.925	41.130 ± 0.508
Intensity	10.598 ± 1.607	15.188 ± 0.378	27.354 ± 3.309	40.578 ± 3.763
Phase	11.940 ± 0.963	14.955 ± 1.847	29.435 ± 0.925	42.695 ± 0.833
Unwrapped phase	9.799 ± 3.749	16.502 ± 1.594	$\mathbf{25.673 \pm 0.963}$	$\mathbf{38.594 \pm 0.712}$
Incidence angle	10.336 ± 0.771	14.737 ± 1.089	27.053 ± 1.784	40.269 ± 2.425
Elevation	10.070 ± 0.740	14.103 ± 0.627	25.790 ± 1.003	38.717 ± 0.742
Displacement	$\mathbf{9.856 \pm 0.783}$	$\mathbf{14.089 \pm 0.756}$	26.295 ± 1.685	39.316 ± 1.758

Table 5.3: Ablation study results for SegFormer

Table 5.3 reports the ablation results for SegFormer. With All Features, SegFormer achieves strong and stable performance, with HS MAE 10.391 ± 0.445 and RMSE 14.684 ± 0.210, and SWE MAE 27.113 ± 1.227 and RMSE 40.388 ± 1.254. The lowest HS MAE is obtained when excluding Unwrapped Phase (MAE 9.799 ± 3.749), but this configuration shows high instability and a worse RMSE, suggesting that the apparent improvement in MAE is not consistently stable across runs, making the best run the one excluding Displacement. For SWE, the exclusion of Unwrapped Phase (MAE 25.673 ± 0.963, RMSE 38.594 ± 0.712), closely followed by excluding Elevation(MAE 25.790 ± 1.003, RMSE 38.717 ± 0.742), represents the best configuration.

In general, the ablation highlights two recurring behaviors: removing Coherence or Phase consistently harms both HS and SWE, and excluding Unwrapped Phase or Displacement can improve the mean performance, with the SWE case showing the most consistent gains. These patterns suggest that Segformer benefits from selective feature engineering because certain features introduce noise rather than contributing useful information.

Overall, the ablation study focuses on analyzing the impact of excluding individual InSAR bands within each model separately, rather than comparing architectures, which is addressed in the next section. The results show that feature contributions are not uniform and that removing certain correlated inputs can improve performance, suggesting redundancy within the feature space. The analysis of standard deviations highlights differences in model stability across configurations. This section therefore emphasizes intra-model behavior and features effects.

5.3 Comparative Results

After examining the effect of band exclusion within each model, in this section a comparison between models is presented and analyzed. As illustrated in Table 5.4, the first comparison is based on all-features configuration with the focus on differences in predictive accuracy, stability and spatial reconstruction quality across the three learning paradigms.

	HS		**SWE**	
	MAE	**RMSE**	**MAE**	**RMSE**
XGBoost	14.980 ± 4.952	25.346 ± 7.164	29.600 ± 0.983	43.297 ± 0.797
U-Net	12.754 ± 0.353	18.224 ± 0.580	31.635 ± 1.670	45.263 ± 1.251
SegFormer	$\mathbf{10.391 \pm 0.445}$	$\mathbf{14.684 \pm 0.210}$	$\mathbf{27.113 \pm 1.227}$	$\mathbf{40.388 \pm 1.254}$

Table 5.4: Comparison between models when considering all bands features

From a quantitative perspective, SegFormer consistently achieves the lowest MAE and RMSE for both HS and SWE and the higher stability, followed by U-Net, while XGBoost shows higher errors and larger variability, in particular for HS estimation. U-Net achieves lower errors than XGBoost for HS, but is slightly worse for SWE (MAE 31.635 vs 29.600 and RMSE 45.263 vs 43.297). This behavior is explained by the tendency of CNNs as U-Net to models spatial structure and local variability through convolutional feature extraction, allowing to reconstruct more detailed patterns. Instead, XGBoost flattens the prediction around the mean leading to slightly higher MAE and RMSE values.

These differences between models' predictions can be explained by the architectural properties of the architectures. Transformer-based architectures, such as SegFormer, are designed to model global context through self-attention mechanisms, enabling long-range spatial dependencies to be captured effectively. In contrast, U-Net relies on hierarchical convolutional feature extraction, which models spatial structure locally and progressively expands the receptive field. XGBoost, being a tree-based ensemble model applied pixel-wise, lacks explicit spatial context modeling and therefore cannot directly exploit structured spatial dependencies.

Performance patterns can also be interpreted through a bias–variance perspective. XGBoost exhibits higher variance across configurations, indicating sensitivity to feature interactions and redundancy. U-Net demonstrates a more balanced trade-off, while SegFormer maintains both low bias, so reduced systematic error and low variance, stable standard deviations, suggesting improved generalization capacity. For this specific task, SegFormer through self-attention mechanism outperforms other architectures, making it the most robust and accurate model for HS and SWE predictions.

To support these considerations, Figure 5.4, 5.5, 5.6 shows the comparison between the ground truth and the predictions of HS and SWE.

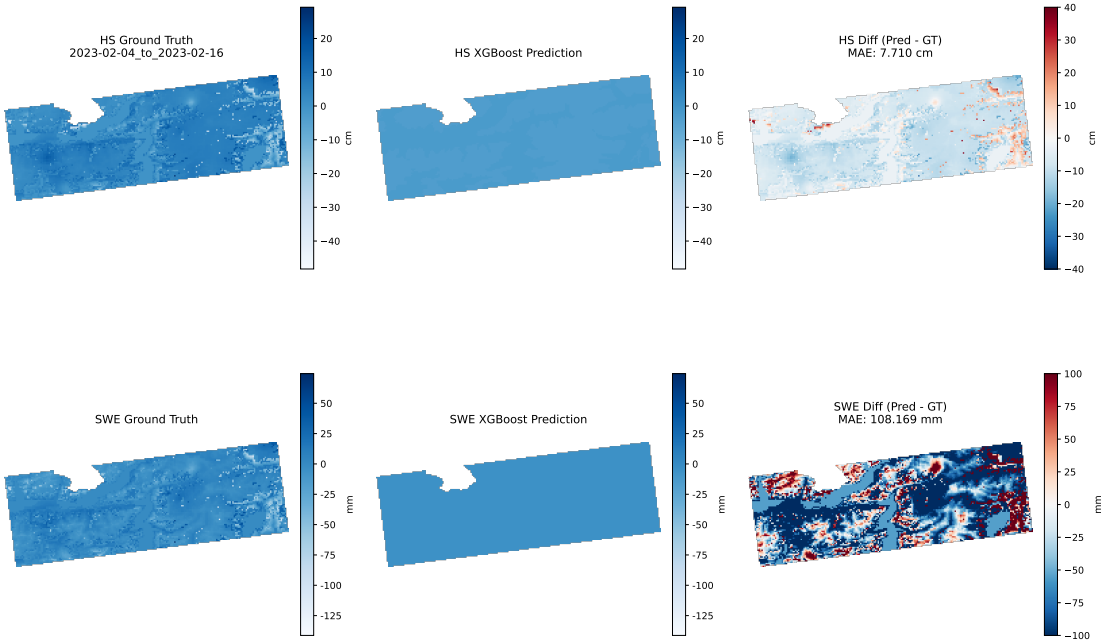

Figure 5.4: XGBoost qualitative results for pair 2023/02/04 and 2023/02/16. First row: HS GT, HS model prediction, difference GT vs. prediction. Second row: SWE GT, SWE model prediction, difference GT vs. prediction.

The qualitative maps produced by XGBoost in Figure 5.4 reveal a smoothing behavior in both HS and SWE predictions. While the overall spatial trend of snow distribution is captured but flattened towards the mean value, fine-scale spatial variability is strongly attenuated. The predicted fields appear nearly uniform compared to the ground truth, indicating limited responsiveness to local patterns and accumulation gradients. This effect is evident in the SWE maps, where the residual maps confirm this behavior: large and spatially structured errors are visible, especially in regions with higher accumulation.

Overall, XGBoost captures the global magnitude and general distribution of snow variables but fails to reconstruct detailed spatial structures. The results suggest a tendency toward conservative predictions, leading to reduced physical realism in complex terrain areas.

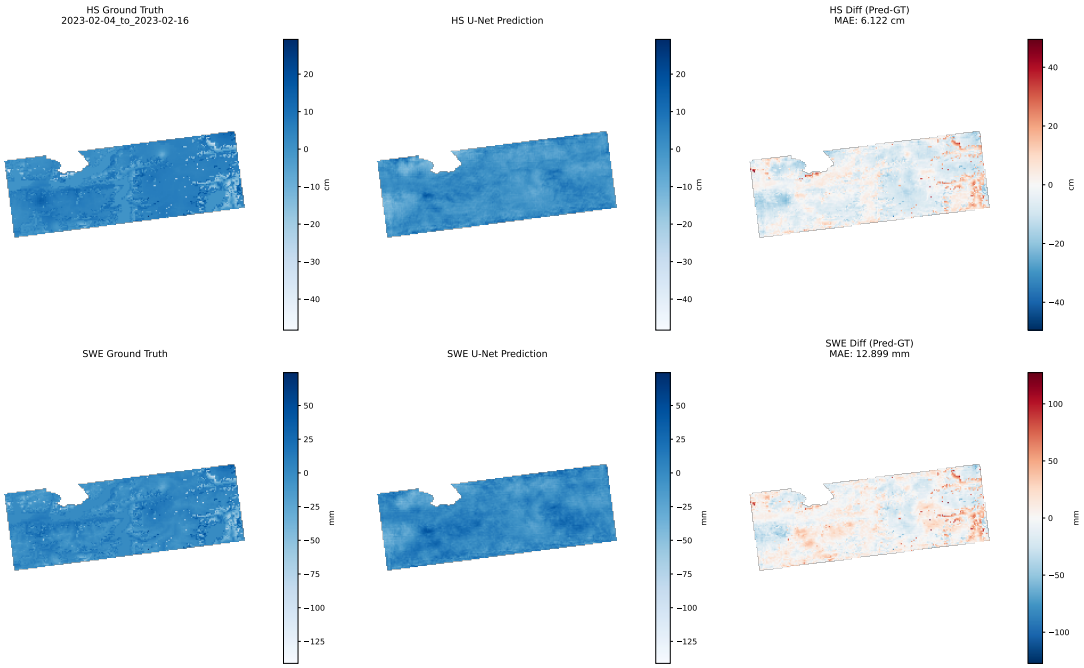

Figure 5.5: U-Net qualitative results for pair 2023/02/04 and 2023/02/16. First row: HS GT, HS model prediction, difference GT vs. prediction. Second row: SWE GT, SWE model prediction, difference GT vs. prediction.

U-Net's maps, Figure 5.5 show an improvement in spatial detail compared to XGBoost. Both HS and SWE predictions preserve terrain-driven gradients and local variability more effectively, resulting in reconstructions that visually resemble the ground truth more closely. However, residual maps reveal the presence of structured errors and localized artifacts. Even if the extreme deviations are less pronounced than in XGBoost, errors are still visible. U-Net achieves a good balance between spatial detail and stability. Reconstruction of fine-scale variability is more faithful than XGBoost, confirming that convolutional hierarchical feature extraction improves spatial realism.

Figure 5.6 shows the visual analysis for SegFormer architecture. Both HS and SWE predictions follow the physical structure of the ground truth in terms of large-scale gradients, terrain accumulation patterns and transitions across the scene. Unlike XGBoost, the predictions do not exhibit excessive flattening, and compared to U-Net, they appear less affected by localized artifacts or inconsistencies. The residual maps further confirm this behavior: errors are evenly distributed across the spatial domain with reduced clustering and fewer extreme deviations.

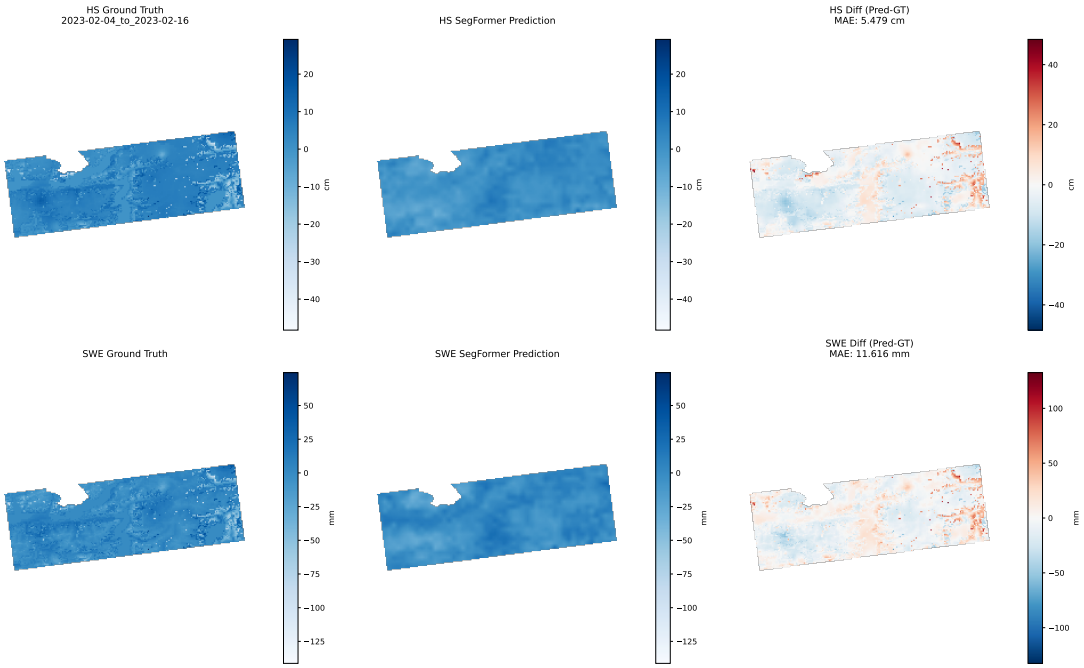

Figure 5.6: SegFormer qualitative results for pair 2023/02/04 and 2023/02/16. First row: HS GT, HS model prediction, difference GT vs. prediction. Second row: SWE GT, SWE model prediction, difference GT vs. prediction.

Overall, the comparative analysis between models highlights the advantages of deep learning approaches over traditional tree-based ensemble methods for spatial snow estimation tasks. Both U-Net and SegFormer benefit from spatial feature learning, improving the reconstruction of terrain-driven variability and patterns. SegFormer consistently demonstrates superior performance across quantitative metrics and spatial reconstructions due to transformer-based architecture explicitly modeling global context. This results in lower MAE and RMSE values, but also in a reduced discrepancy between them, indicating fewer extreme residuals and improved robustness.

The results reported in Table 5.5 further reinforce the conclusions drawn from the full-feature comparison.

	HS		**SWE**	
	MAE	**RMSE**	**MAE**	**RMSE**
XGBoost	Displacement Excluded		Unwrapped Phase Excluded	
	13.105 ± 2.423	17.302 ± 2.291	28.379 ± 10.135	40.659 ± 14.880
U-Net	Displacement Excluded		Displacement Excluded	
	11.464 ± 0.262	16.142 ± 0.418	28.032 ± 0.625	41.422 ± 1.022
SegFormer	Displacement Excluded		Unwrapped Phase Excluded	
	$\mathbf{9.856 \pm 0.783}$	$\mathbf{14.089 \pm 0.756}$	$\mathbf{25.673 \pm 0.963}$	$\mathbf{38.594 \pm 0.712}$

Table 5.5: Comparison between models when considering the best bands feature configuration

When each model is evaluated under its best-performing band configuration, SegFormer remains the most accurate architecture for both HS and SWE estimation. This confirms that the performance gap is not merely due to suboptimal band selection, but reflects intrinsic architectural advantages. Therefore, even under favorable conditions for all models, SegFormer confirms itself as the most robust and accurate solution for reconstructing snow-related variables in this experimental setting.

Chapter 6

Conclusions

While InSAR has been less explored for quantitative snow mass and depth estimation than for soil deformation or landslides, this study explicitly aims to investigate and formalize its potential in this context. Estimating Snow Depth and Snow Water Equivalent over mountainous terrain remains a scientifically demanding problem considering the variability of snow processes and the structural complexity of alpine environments. Factors influencing the radar signal over snow areas include temperature evolution, the intrinsic structure of the snow, surface and topographic geometry.

The effective use of learning-based methods depends on how input information is structured, harmonized and constrained by physical consistency to avoid ambiguity and improve robustness. Building upon this perspective, this thesis establishes the foundation for snow parameter estimations starting from InSAR observables embedded into a structured multi-source modeling framework.

Our Contributions

The main contribution of this work lies in the construction of a dataset which is physically consistent and multi-source, designed to combine interferometric observations and physical measures. This strategy ensures particular attention to preserve physical coherence of snow characteristics while maintaining compatibility within sources.

Another key contribution of this work is represented by the evaluation of different modeling strategies. The analysis begins with traditional machine learning approaches to establish a baseline and is then extended to deep learning architectures to assess their value and performance improvements.

In addition, features importance analysis performed through an ablation study quantifies the contribution of each interferometric bands. It provides insights of the

informational content of InSAR observables for HS and SWE predictions, allowing an informed feature selection for future developments.

Future Works

Despite the robustness of the proposed dataset and experimental framework, several directions can be pursued to further advance the field of snow estimation through remote sensing techniques. Some precautions can be inspected, both in relation to the input data, to the adopted reference dataset and learning frameworks experimented to strengthen the present research.

Even if the selected area still captures significant variability in elevation, snow regimes, and accumulation patterns, making it representative for methodological evaluation, as demonstrated by this study, expanding the dataset in both spatial and temporal dimensions would enhance statistical robustness and improve model generalization. Increasing the number of interferometric pairs and including additional geographic contexts would allow a broader validation of the proposed framework. The spatial extent of the study area,constrained by the high computational demands of InSAR processing, remains limited to a regional scale. Expanding the geographical coverage would enable the extension of the framework beyond a localized alpine context, allowing for large-scale and global investigations of snow dynamics.

An interesting direction for future developments concerns the extension of the framework to multi-frequency SAR data. In this study, Sentinel-1 C-band acquisitions were adopted due to their open availability and systematic temporal coverage. However, C-band SAR data implies a reduced sensitivity, limited penetration capability in dense vegetation and reduced sensitivity under wet snow conditions, which may constrain performances during melt periods or in vegetated areas. Future research may investigate the integration of additional SAR frequencies characterized by longer wavelengths and enhanced penetration capabilities. The combination of C-band and complementary radar frequencies within a unified multi-source framework would therefore represent a promising step toward more comprehensive and globally scalable snow monitoring approaches.

Future methodological developments could focus on more expressive deep learning architectures specifically designed for geophysical regression problems. In particular, deeper hierarchical Transformers, hybrid CNN–Transformer frameworks, or models explicitly designed for multi-scale environmental processes may enhance the representation of heterogeneous snow conditions and interactions between topography, radar observables and physical drivers.

A further important direction for future work concerns the validation strategy and the quality of reference data. In this study, the IT-SNOW dataset is adopted

as target variable, as it integrates in-situ measurements, physical drivers, and modeling components to provide spatially consistent snow estimates. While such hybrid products are highly valuable for regional-scale analyses, they inherently combine observational data with model-based assumptions, potentially propagating uncertainties. The ideal validation framework for this task would rely on dense and spatially distributed in-situ measurements of snow depth and snow water equivalent. However, accurate ground observations at high spatial resolution across complex mountainous terrain is logistically demanding, time-consuming, and unfeasible at regional scales. Future research could therefore focus on validation strategies that enhance physical consistency while remaining scalable. Increasing the physical reliability of validation data would also enable a clearer understanding of whether learning-based models capture true snowpack dynamics rather than characteristics inherited from the reference product itself.

Overall, the proposed framework demonstrates methodological robustness and scientific relevance. Given the rapid evolution of SAR technologies and data-driven modeling strategies, this field offers substantial opportunities for further investigation.

Bibliography

[1] Tore Guneriussen et al. «InSAR for Estimation of Changes in Snow Water Equivalent of Dry Snow». In: *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING* (Oct. 2001) (cit. on p. 4).

[2] Jayson Eppler, Bernhard Rabus, and Peter Morse2. «Snow water equivalent change mapping from slope-correlated synthetic aperture radar interferometry (InSAR) phase variations». In: *The Cryosphere* (Apr. 2022) (cit. on p. 5).

[3] Vasco Conde et al. «On the Estimation of Temporal Changes of Snow Water Equivalent by Spaceborne SAR Interferometry: A New Application for the Sentinel-1 Mission». In: (Mar. 2019) (cit. on p. 5).

[4] Shadi Oveisgharan, Robert Zinke, Zachary Hoppinen, and Hans Peter Marshall. «Snow water equivalent retrieval over Idaho – Part 1: Using Sentinel-1 repeat-pass interferometry». In: *The Cryosphere* (Feb. 2024) (cit. on p. 5).

[5] Hans Lievens et al. «Sentinel-1 snow depth retrieval at sub-kilometer resolution over the European Alps». In: *The Cryosphere* (Mar. 2021) (cit. on p. 5).

[6] Zachary Hoppinen et al. «Evaluating snow depth retrievals from Sentinel-1 volume scattering over NASA SnowEx sites». In: *The Cryosphere* (Nov. 2024) (cit. on p. 5).

[7] Lucas Boeykens et al. «Machine learning for snow depth estimation over the European Alps, using Sentinel-1 observations, meteorological forcing data and physically-based model simulations». In: *EGUsphere* (July 2025) (cit. on p. 5).

[8] Milan Cistý, Michal Danko, Silvia Kohnová, Barbora Považanová, and Andrej Trizna. «Machine Learning Enhanced by Feature Engineering for Estimating Snow Water Equivalent». In: *MDPI* (Aug. 2024) (cit. on p. 5).

[9] Ibrahim Olalekan Alabi, Hans-Peter Marshall, Jodi Mead, and Ernesto Trujillo. «Machine Learning Enhanced by Feature Engineering for Estimating Snow Water Equivalent». In: (Jan. 2025) (cit. on p. 5).

[10] Alejandro Betato, Hernán Díaz Rodríguez, Niamh French, Thomas James, and Beatriz Remeseiro. «MAPunet: High-resolution snow depth mapping through U-Net pixel-wise regression». In: *Remote Sensing Applications* (Feb. 2025) (cit. on p. 6).

[11] Krishu K Thapa, Bhupinderjeet Singh, Supriya Savalkar, Alan Fern, Kirti Rajagopalan, and Ananth Kalyanaraman. «Attention-based Models for Snow-Water Equivalent Prediction». In: (Nov. 2023) (cit. on p. 6).

[12] Samuel Schilling, Andreas Dietz, and Claudia Kuenzer. «Snow Water Equivalent Monitoring—A Review of Large-Scale Remote Sensing Applications». In: *MDPI* (Mar. 2024) (cit. on p. 6).

[13] Faye Hsu, Ziheng Sun, Gokul Prathin, Sanjana Achan, and Leah Zhang. «A Review of Machine Learning in Snow Water Equivalent Monitoring». In: (May 2024) (cit. on p. 6).

[14] Tianqi Chen and Carlos Guestrin. «XGBoost: A Scalable Tree Boosting System». In: (June 2016) (cit. on p. 7).

[15] Jerome Friedman. «Greedy function approximation: a gradient boosting machine». In: *The annals of statistics* (2001) (cit. on p. 7).

[16] Y. Wang, Z. Pan, J. Zheng, L. Qian, and M. Li. «A hybrid ensemble method for pulsar candidate classification». In: (2019) (cit. on p. 7).

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. «U-Net: Convolutional Networks for Biomedical Image Segmentation». In: (May 2015) (cit. on p. 8).

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. «Deep Residual Learning for Image Recognition». In: (Dec. 2015) (cit. on p. 9).

[19] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. «SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers». In: (Oct. 2021) (cit. on p. 11).

[20] Alexey Dosovitskiy et al. «AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE». In: (June 2021) (cit. on p. 11).

[21] Wenhai Wang et al. «Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions». In: (Aug. 2021) (cit. on p. 11).

[22] Alberto Moreira, Pau Prats-Iraola, Marwan Younis, Gerhard Krieger, Irena Hajnsek, and Konstantinos P. PapathanaSSiou. «A Tutorial on Synthetic Aperture Radar». In: *IEEE Geoscience and remote sensing magazine* (Mar. 2013) (cit. on pp. 13, 15, 16).

[23] Alessandro Ferretti, Andrea Monti-Guarnieri, Claudio Prati, and Fabio Rocca. «InSAR Principles: Guidelines for SAR Interferometry Processing and Interpretation». In: *ESA Publications* (Feb. 2007) (cit. on pp. 14, 16).

[24] Richard Bamler and Philipp Hartl. «Synthetic aperture radar interferometry». In: (Feb. 1998) (cit. on p. 16).

[25] Paul A. Rosen and Scott Hensley Et al. «Synthetic Aperture Radar Interferometry». In: (Mar. 2000) (cit. on p. 16).

[26] Kevin Jones. *Advanced SAR Training*. https://www.udemy.com/course/sar-course/. Accessed: 2024-02-11 (cit. on p. 18).

[27] Andreas Braun and Luis Veci. «TOPS Interferometry Tutorial». In: *ESA and Skywatch* (June 2021) (cit. on pp. 19, 20).

[28] ESA. *SNAP software*. https://step.esa.int/main/doc/. Accessed: 2024-02-06 (cit. on p. 20).

[29] Olaf Danne. *Configure Python to use the SNAP-Python (esa_snappy) interface (SNAP version 10+)*. https://senbox.atlassian.net/wiki/spaces/SNAP/pages/2499051521/Configure+Python+to+use+the+SNAP-Python+esa_snappy+interface+SNAP+version+10. Accessed: 2024-02-06 (cit. on p. 20).

[30] Olaf Danne et al. *esa$_s$nappy*. https://github.com/senbox-org/esa-snappy. Accessed: 2024-02-06 (cit. on p. 20).

[31] Stanford University. *snaphu*. https://web.stanford.edu/group/radar/softwareandlinks/sw/snaphu/snaphu_man1.html. Accessed: 2024-02-06 (cit. on p. 20).

[32] European Space Agency. *Copernisus API*. https://documentation.dataspace.copernicus.eu/APIs/OData.html. Accessed: 2024-02-06 (cit. on p. 21).

[33] Francesco Avanzi et al. «IT-SNOW: a snow reanalysis for Italy blending modeling, in situ data, and satellite observations (2010–2021)». In: *Copernicus pubblications* (Feb. 2023) (cit. on pp. 29, 37).

[34] European Space Agency. *Sentinel-1*. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1. Accessed: 2024-02-06 (cit. on p. 30).

[35] European Space Agency. *Sentinel-3*. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-3. Accessed: 2024-02-07 (cit. on p. 34).

[36] European Space Agency. *Copernicus S3 Documentation*. https://documentation.dataspace.copernicus.eu/Data/SentinelMissions/Sentinel3.html. Accessed: 2024-02-07 (cit. on p. 34).

[37] European Space Agency. *Sentinel-3 SLSTR Documentation*. https://sentiwiki.copernicus.eu/web/slstr-products. Accessed: 2024-02-07 (cit. on p. 34).

[38] Regine Hock. «Temperature index melt modelling in mountain areas». In: *Journal of hydrology* (2003) (cit. on p. 36).

[39] *Albumentation*. `https://albumentations.ai/docs/`. Accessed: 2024-02-13 (cit. on p. 45).

[40] Timothy O. Hodson. «Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not». In: *Geoscientific Model Developement* (July 2022) (cit. on p. 50).