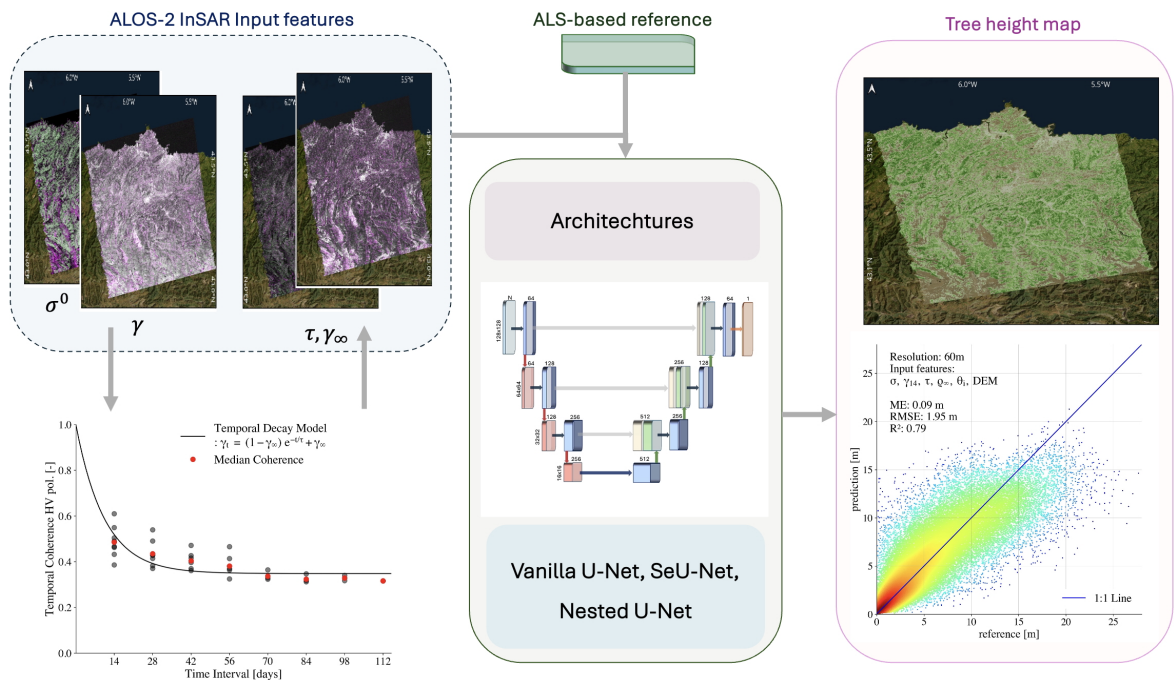


Graphical Abstract

High-Resolution Forest Mapping from L-Band Interferometric SAR Time Series using Deep Learning over Northern Spain

Chiara Telli, Oleg Antropov, Anne Lönnqvist, Marco Lavalle



Highlights

High-Resolution Forest Mapping from L-Band Interferometric SAR Time Series using Deep Learning over Northern Spain

Chiara Telli, Oleg Antropov, Anne Lönnqvist, Marco Lavallo

- Demonstrated the ability of joint multi-temporal interferometric and backscatter observations from PALSAR-2 to capture structural variations in forest canopies.
- Developed a hybrid physics–deep learning framework that combines InSAR coherence model–based inversion, polarimetric and auxiliary observables with UNet-family convolutional neural networks for forest height retrieval.
- Achieved RMSEs between ~ 2.8 m at 20 m resolution and 2.0 m at 60 m strongly outperform traditional machine learning or physics-based inversion approaches.
- Attention mechanisms and nested feature fusion yield more spatially coherent forest height maps consistent with airborne laser scanning measurements.
- Moderate spatial aggregation (40–60 m) provided an effective noise suppression enabling scalable monitoring of forest structure from satellite L-band InSAR imagery.

High-Resolution Forest Mapping from L-Band Interferometric SAR Time Series using Deep Learning over Northern Spain

Chiara Telli^{a,b,*}, Oleg Antropov^{c,*}, Anne Lönnqvist^c, Marco Lavallo^b

^a*Sapienza University, Dept. of Information Eng., Electronics and Telecommunications, 00184 Roma, Italy*

^b*Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91011, USA*

^c*VTT Technical Research Centre of Finland, Tekniikantie 1, 02150 Espoo, Finland*

Abstract

In this study, we examine the potential of high-resolution forest mapping using L-band interferometric time series datasets and deep learning modeling. Our SAR data are represented by a time series of nine ALOS-2 PALSAR-2 dual-pol SAR images acquired at near-zero spatial baseline over a study site in Asturias, Northern Spain. Reference data are collected using airborne laser scanning. We examine the performance of several candidate deep learning models from UNet-family with various combinations of input polarimetric and interferometric features. In addition to basic Vanilla UNet, attention reinforced UNet model with squeeze-excitation blocks (SeU-Net) and advanced UNet model with nested structure and skip pathways are used. Studied features include dual pol interferometric observables additionally incorporating model-based derived measures. Results show that adding model-based inverted InSAR features or InSAR coherence layers improves retrieval accuracy compared to using backscatter intensity only. Use of attention mechanisms and nested connection fusion provides better predictions than using Vanilla UNet or traditional machine learning methods. Forest height retrieval accuracies range between 3.1-3.8 m ($R^2 = 0.45-0.55$) at 20 m resolution when only intensity data are used, and improve to less than 2.8 m when both intensity and interferometric coherence features are included. At 40 m and 60 m resolution, retrieval performance further improves, primarily due to higher SNR in both the intensity and interferometric layers. When using intensity at 60 m resolution, best achieved RMSE is 2.2 m, while when using all suitable input features the achieved error is 1.95 m. We recommend this hybrid approach for L-band SAR retrievals also suitable for NISAR and future ROSE-L missions.

Keywords: ALOS-2 PALSAR-2, synthetic aperture radar, L-band, polarimetry, interferometry, deep learning, forest mapping, regression modeling

1. Introduction

Accurate methods for assessing forest carbon using Earth Observation (EO) sensors are increasingly needed to support climate change monitoring, carbon stock accounting, and to address the operational needs of diverse forest management stakeholders (Latterini et al., 2025). Forests are essential in maintaining healthy ecosystem interactions and biodiversity on Earth, and they play a critical role in tackling climate change due to their capacity to rapidly restore carbon stocks. This ecological and climatic importance has placed forests at the center of international environmental policies, which call for accurate, timely, and reliable methods for large-scale forest monitoring (Berger et al., 2025; Moersberger et al., 2024).

At present, forest carbon stock and its changes are typically estimated through manual inventories of structural forest variables, a process that remains highly labor-intensive and limited in spatial and temporal coverage (Neumann et al., 2016; GFOI, 2014; Penman et al., 2003).

EO sensors offer opportunities to overcome these limitations by providing consistent, wall-to-wall coverage with regular temporal sampling. The integration of satellite EO data with ground measurements has therefore become an established practice in forest mensuration and monitoring, enabling the estimation of forest variables across multiple spatial scales and mapping units (McRoberts and Tomppo, 2007), and improving national forest inventories (Hunka et al., 2025). Such estimates can be obtained either by inverting physics-based models, which aim to describe the interaction between transmitted signals and forest structure, or by applying data-driven approaches, which directly learn the underlying relationships from large sets of reference samples (GFOI, 2014).

Within this context, multiparametric satellite optical and synthetic aperture radar (SAR) imagery represent critical sources of information for large-scale forest mapping. Optical sensors provide valuable insights into vegetation phenology and health, but their applicability is often constrained by cloud cover and atmospheric conditions. In contrast, SAR observations, being independent of daylight and cloud cover, enable consistent monitoring across regions and seasons. The interaction between the radar signal and the observed scatterers is strongly influenced by the scatterers' geometric and dielectric proper-

*Corresponding authors

Email addresses: chiara.telli@uniroma1.it (Chiara Telli), oleg.antropov@vtt.fi (Oleg Antropov)

ties, making SAR particularly sensitive to key vegetation characteristics, such as forest height, growing stock volume, and above-ground biomass (Ulaby and Long, 2014; Le Toan et al., 2004)

To date, extensive studies demonstrated significant correlations between both SAR backscatter coefficient and SAR interferometric measurements, and forest variables (Schmullius et al., 2015). This sensitivity is influenced by several factors, including mission configuration, radar frequency, polarization, and the structural characteristics of the forest. Due to the enhanced penetration capability to vegetation structure compared with shorter wavelengths, L-band time series have proven particularly effective in estimating forest variables and are considered presently most suitable for SAR based forest variable retrieval (GFOI, 2014). Currently reported retrieval accuracies, however, indicate that further improvements are needed to achieve more accurate and reliable estimates. The integration of deep learning techniques into EO-based forestry, together with the advent of new remote sensing missions offering advanced imaging capabilities, such as NISAR, holds strong potential for methodological advances when improved modeling approaches are combined with novel imaging systems. Research in this field remains active, with continued efforts to refine physically based retrieval models using L-band time series and to develop deep-learning frameworks for data-driven forest estimation parameters. The following sections summarize recent progress in both approaches, highlighting key trends and remaining challenges.

1.1. L-band SAR time series in forest mapping

SAR intensity based approaches infer forest structural variables from SAR backscatter time series by leveraging the sensitivity of low-frequency radar to forest structure using physics-based models and, more recently, machine learning techniques. Early research on L-band time series for forest mapping focused primarily on estimating growing stock volume (GSV) in boreal forests using L-band SAR imagery, particularly from the JERS (Kurvonen et al., 1999; Pulliainen et al., 1999; Rauste, 2005; Santoro et al., 2006). Later studies utilized ALOS PAL-SAR and ALOS-2 PALSAR-2 sensors to map not only boreal forests (Cartus et al., 2012; Antropov et al., 2013, 2017; Santoro et al., 2024a) but also temperate and tropical forests (Englhart et al., 2012; Berninger et al., 2018; Rodríguez-Veiga et al., 2019), and even on global scales (Santoro et al., 2024b).

While models based approaches based on Water Cloud Model (Santoro et al., 2024a) provide a simplified representation of these processes, forward modeling of electromagnetic wave propagation within forested environments remains inherently complex due to multiple scattering interactions and the challenge of isolating contributions from distinct scattering components. In forested areas, L-band backscatter generally arises from three main components: (i) volume scattering from the canopy, (ii) surface scattering from the forest floor, particularly in canopy gaps, and (iii) double-bounce scattering caused by interactions between the ground and tree trunks. The sensitivity of radar backscatter intensity to forest height is further constrained by the well-known problem of signal saturation, which reduces the

dynamic range of tree height estimates (Dobson et al., 1992; Mermoz et al., 2014; Mitchard et al., 2009).

Potential improvement in estimating forest variables are possible by leveraging the interferometric potential of SAR acquisitions. Early and well-established approaches rely on the sensitivity of interferometric measurements to forest height, which arises from the angular separation between two SAR acquisitions (Treuhart and Siqueira, 2000). A widely used physical model is the Random Volume over Ground (RVoG), which represents the forest scattering profile as the sum of a Dirac-like ground component and a vertical distribution of randomly oriented scatterers (Cloude and Papathanassiou, 1998). Inversion of the RVoG model typically requires a single-baseline, fully polarimetric (quad-pol) interferometric SAR acquisition, allowing the retrieval of forest height. However, most current and planned interferometric SAR missions, both at higher frequencies (e.g., C-band Sentinel-1) and at lower frequencies (e.g., L-band ALOS-2, NISAR, and ROSE-L), are designed to operate in dual-polarization modes with near-zero spatial baselines. In these observation scenarios, sensitivity to vegetation structure due to the interferometric angular separation is reduced, and the influence of temporal vegetation dynamics on the interferometric phase and coherence becomes dominant. This has motivated recent efforts to retrieve forest parameters from time series of near-zero-baseline, dual-polarimetric interferometric observations at C- and L-band, where temporal effects play a major role (Lei and Siqueira, 2014; Seppi et al., 2022; Bhogapurapu et al., 2024; Cartus et al., 2022; Lavalley et al., 2023).

For zero- or near-zero interferometric baseline configurations, interferometric measurements are mainly affected by temporal decorrelation, i.e., the decrease in the correlation coefficient between two radar complex images acquired at different times albeit with the same observation geometry. Lavalley et al. (2012) extended the RVoG model by introducing a vertical motion profile of canopy scatterers to represent wind-induced temporal decorrelation, as well as dielectric variations that occur between acquisitions, linking these effects to forest structural properties. This model, known as the Random-Motion-over-Ground model (RMoG), has been used to describe temporal decorrelation mechanisms and to retrieve tree height from both Sentinel-1 C-band (Lavalley et al., 2023) and ALOS-2 L-band backscatter and interferometric coherence time series (Telli et al., 2025).

A simplified approach to characterize temporal decorrelation is to model temporal coherence as an exponential decay (Sica et al., 2019). Although this behavior may not always hold at the pixel level, temporal coherence, particularly when considered in aggregated form, is generally expected to decrease over time due to the combined influence of multiple decorrelation sources. In this formulation, two parameters are defined: the *temporal decay rate*, which describes how rapidly coherence decreases with increasing temporal baseline, and the *long-term coherence*, which represents the residual coherence observed at very large time intervals. Together, these parameters absorb the effects of various decorrelation sources, including vegetation dynamics, soil moisture variability, and system noise.

Given a time series of interferometric coherence maps, these two parameters can be estimated in a pixel-wise basis through curve fitting, thereby providing spatially explicit insights into the sensitivity of temporal coherence to forest structure. Compared with more complex models, this simplified approach requires fewer assumptions and less ancillary information, while still offering a practical way to exploit temporal decorrelation for forest monitoring. However, since the different sources of decorrelation are not explicitly separated, directly deriving forest variables from these parameters is not straightforward. This limitation motivates the deep learning approach proposed in this study, which aims to investigate whether hidden, non-linear relationships between temporal coherence dynamics and forest structure can be effectively captured to enable tree height estimation.

1.2. Deep learning modeling and SAR data in forest mapping

Deep learning (DL) methodologies are widely adopted in the remote sensing community for image classification and semantic segmentation tasks (Persello et al., 2022; Zhu et al., 2021, 2017). To date, several fully convolutional and recurrent neural network architectures were investigated for regression tasks in the context of EO-based forest inventory mapping (Astola et al., 2021; Illarionova et al., 2022; Ge et al., 2022a,b; Bolyn et al., 2022). These models enable the integration of spatial and/or temporal context in addition to spectral or radiometric (polarimetric) observables within the modeling process. Thus, they typically achieve higher accuracy in forest classification, biophysical variable estimation (e.g., biomass, height, and volume), and forest change mapping (Kuzu et al., 2024) compared to traditional machine learning, statistical, or model-based approaches. Training these DL models generally requires a representative extensive dataset consisting of fully segmented reference labels, such as image patches derived from airborne laser scanning (ALS) digital surface models or ALS-based forest maps. Nevertheless, these models can be adapted to new sites through transfer learning strategies (Astola et al., 2021; Ge et al., 2023).

To date, deep learning approaches in the context of forest mapping were studied using primarily Sentinel-1 images (Ge et al., 2022b), sometimes augmented by optical image datasets (Ge et al., 2022a).

In addition, several studies explored the potential of deep learning for forest height mapping using TanDEM-X data. Although these studies reported excellent results with accuracies reaching 10-20% RMSE (Mazza et al., 2019; Carcereri et al., 2023), the availability of TanDEM-X data remains limited, restricting its applicability for large-scale or operational forest monitoring.

In contrast, L-band time series have been investigated less frequently due to data availability constraints. In this study, we address this gap by utilizing a time series of PALSAR-2 interferometric acquisitions for forest mapping, leveraging advanced U-Net modeling approaches and a hybrid strategy that integrates model-based parameter retrieval with deep learning modeling.

1.3. Our study goals and objectives

The overarching objective of this work is to develop and evaluate a robust methodology for exploiting time series of interferometric L-band data dominated by temporal decorrelation to maximize the accuracy of forest height prediction, thereby supporting operational forestry practices and preparing for the uptake of NISAR mission data. The study relies on ALOS-2 PALSAR-2 datasets acquired over highly productive forests in Northern Spain, an area characterized by challenging topography and diverse forest structure.

For the deep learning framework, we rely on a family of U-Net models, which have consistently demonstrated high performance in predicting forest structural variables and are still competitive to the latest advances in geospatial foundation modeling. Three variants are considered: the classical U-Net (Ronneberger et al., 2015), the nested U-Net (Zhou et al., 2018), and the Squeeze-and-Excitation U-Net (SeU-Net) (Ge et al., 2022a; Hu et al., 2020).

The specific objectives of this study are to:

1. Evaluate the potential of advanced deep learning architectures, including nested and attention-enhanced U-Net variants, for quantitative forest-structure mapping from multi-temporal PALSAR-2 interferometric data, and benchmark their performance against conventional machine-learning and model-based approaches;
2. Develop a hybrid UNet model based approach that combines input polarimetric observables and physics based inverted parameters within CNN modeling approach to improve forest height prediction performance;
3. Quantify the contribution of different SAR observables to forest height retrieval accuracy within hybrid UNet model based prediction, and assess the added value of integrating physics-model derived parameters and geometric features such as incidence angle and DEM layers in forest height retrievals; and
4. Examine the impact of spatial resolution and interferometric data processing on model accuracy and spatial fidelity, determining optimal trade-offs between detail and stability for PALSAR-2 based forest height retrieval.

To achieve these objectives, we combine dual-polarization ALOS-2 backscatter and interferometric time-series data with model-based features derived from a physics-based exponential decay model of temporal coherence (Sica et al., 2019), and UNet family based deep learning framework using airborne LiDAR measurements as reference. The approach includes a multi-resolution analysis to assess the robustness of the model across different spatial scales.

The paper is organized as follows: Section 2 describes the study area and the datasets used in this work, Section 3 details the deep learning architectures and the overall methodological framework, and Section 4 presents the experimental results and performance assessment. Finally, Section 5 summarizes the main findings and outlines directions for future research.

2. Data

2.1. Study site

The study area is located in northern Spain (longitude: -6.33°, -5.33°; latitude: 42.98°, 43.68°) within the Principality of Asturias. It is characterized by a Temperate Broadleaf and Mixed Forests biome, and it is predominantly covered by vegetation. In the area, the most common species are sweet chestnut (45.8%), beech (25.9%), blue gum (19.5%), and oaks (4.5%) (MITECO, 2020). The site is characterized by rough topography, being relatively flat in the north and mountainous toward the south, with elevations ranging from 100 m to 2500 m. The climate is mild, with an average annual temperature of 10°C. The mean annual precipitation is approximately 1000 mm, with the highest rainfall occurring mainly during the winter months. Due to its proximity to the ocean, the region is frequently affected by strong winds. The study site is shown in Figure 1.

2.2. SAR and auxiliary data

The SAR dataset consists of a time series of 9 dual-pol (HH, HV) ALOS-2 SLC images acquired at 14-day intervals between June and September 2020. Data were collected in Fine Beam Dual (FBD) mode along an ascending orbit. The local acquisition time is approximately midnight, and the off-nadir angle is 28°.

The ALOS-2 data processing was performed using JPL’s ISCE3 (InSAR Scientific Computing Environment) software (Rosen et al., 2018). All SAR images were orthorectified and radiometrically normalized with respect to the projected incidence angle area to obtain gamma-naught (γ^0) images using the ISCE3 Area-Based Projection Algorithm (Shiroma et al., 2022), which averages the backscatter measurements using an adaptive multi-looking window. Orthorectification was performed using Copernicus DEM. Three adaptive multilooking configurations were applied to generate backscatter time series for each polarization with resolutions of 20 m, 40 m, and 60 m, respectively.

Interferometric processing produced 36 coherence maps per polarization, with temporal baseline ranging from 14- days up to a maximum of 112 days. To ensure consistency with the three backscatter resolutions, coherence was estimated using the following number of looks:

- 21 looks (3 range $\times$ 7 azimuth),
- 60 looks (4 range $\times$ 10 azimuth), and
- 119 looks (7 range $\times$ 17 azimuth)

The resulting coherence maps were resampled to 20 m, 40 m, and 60 m resolutions, respectively, using bilinear interpolation, to perfectly align with the backscatter grid.

Using multiple resolutions enables testing the ability of deep learning models to provide reliable high-resolution tree-height estimates while coping with speckle and interferometric coherence biases.

In addition to pairwise coherences, the median temporal coherence was computed for each time lag and polarization, and

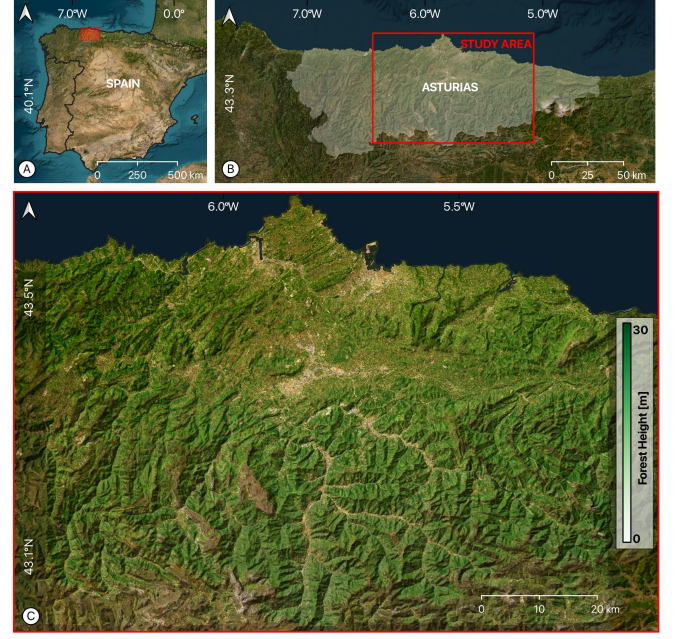


Figure 1: Location of the study site in Asturias, Spain. Panel B shows the extent of the study area within the Asturias region, and Panel C presents the detailed study area along with the reference tree height map (CNIG, 2021). Coordinate reference system: EPSG:4326, WGS84 (geodetic latitude/longitude).

the temporal median backscatter maps were also derived. Geometry layers obtained during the SAR processing, including the local incidence angle and layover–shadow masks, were also generated and used as ancillary information in the analysis. All intensity and coherence layers were geocoded to WGS84 (EPSG:4326) reference system and re-sampled to a regular grid with 20 m pixel spacing using bilinear interpolation, while nearest neighbor interpolation was used for resampling mask layers.

2.2.1. Model-derived features

Given the interferometric geometry ($k_z^{max} = 0.02 \text{ rad m}^{-1}$), the contribution of volume decorrelation to the measured coherence is negligible (Kugler et al., 2015). Therefore, the dependence of coherence on tree height is driven primarily by temporal decorrelation.

As reported in the literature (Lavalley et al., 2011), the impact of temporal decorrelation is expected to vary with tree height, as taller canopies are more generally affected by wind-induced motion.

To capture this behavior, the SAR dataset includes parameters estimated from an exponential decay model of temporal coherence as a function of repeat interval (Sica et al., 2019)

$$\gamma_i(t) = (1 - \rho_\infty)e^{-t/\tau} + \rho_\infty \quad (1)$$

where $\gamma_i(t)$ is the temporal coherence, and ρ_∞ and τ are the long-term coherence and the coherence rate of decay, respectively. The resulting maps, ρ_∞ and τ , are expected to capture the temporal and spatio-temporal variability of the interferometric coherence at the pixel level, thus to be sensitive to forest height. Model fitting was performed using median-aggregated

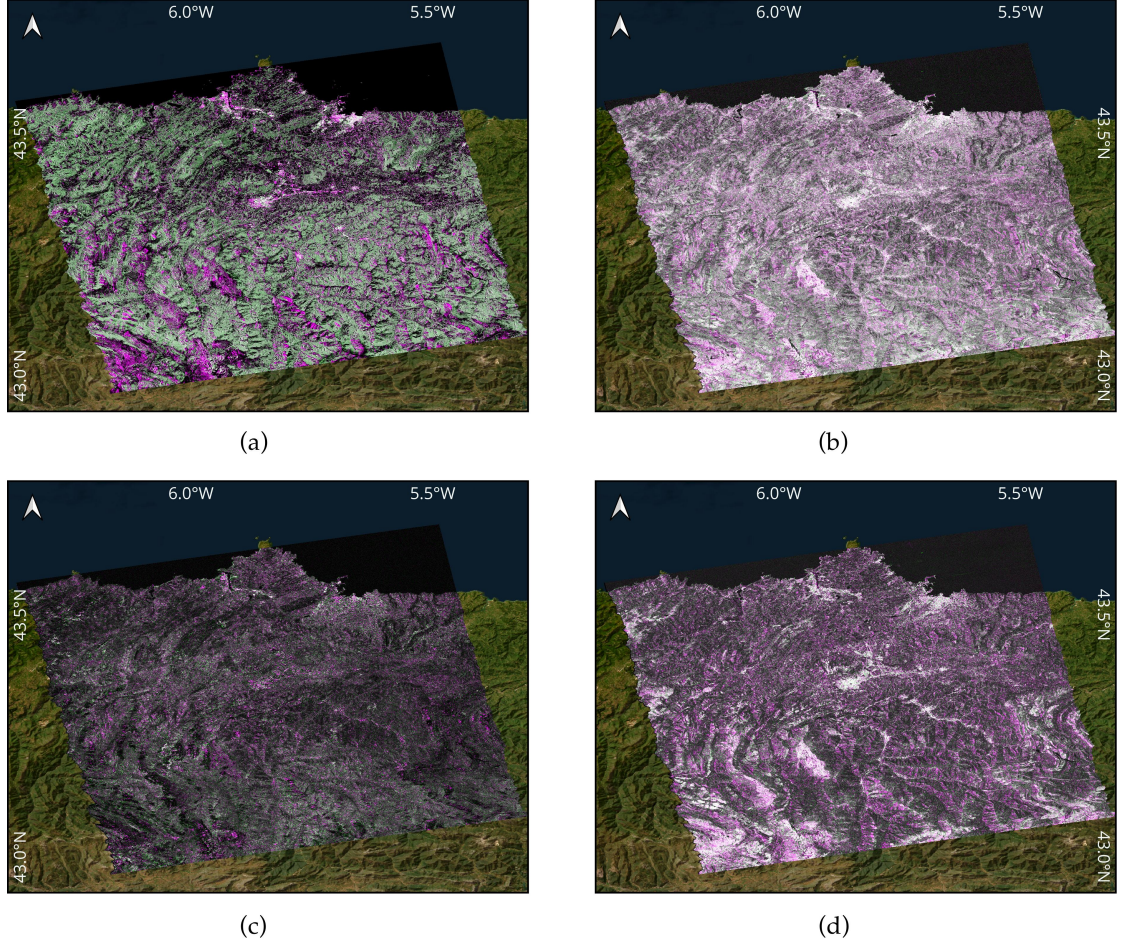


Figure 2: RGB composites of the 20 m resolution InSAR dataset and model derived layers used in the analysis (R: HH, G: HV, B: HH). (a) Median backscatter maps in dB scale, with co- and cross-polarization intensities ranging from -14 to -4 dB and -20 to -10 dB, respectively. (b) Median 14-day coherence maps, with values between 0.2 and 0.8. (c) Temporal decay parameters, τ , spanning 1–45 days. (d) Long-term coherence, ρ_{∞} , ranging 0.2–0.7.

coherence estimates at the different spatial resolutions (20 m, 40 m, and 60 m) through a trust-region reflective regression algorithm.

The InSAR features, including backscatter intensities, interferometric coherence layers, and model-derived parameters, are illustrated in Fig. 2. Figure 3 presents the sensitivity of the 20 m resolution median HV backscatter to reference tree height (Figure 3,a), the sensitivity of the 14-day temporal median coherence to tree height (Figure 3,b), and the temporal decay of coherence over increasing baselines (Figure 3-c,d). Figure 3-d provides an example of a pixel-level fit of the temporal coherence decay model.

2.3. Reference data

Reference tree height data are provided by airborne LiDAR-derived forest height measurements. These data come from the Normalized Digital Surface Model of Vegetation Classes of Spain (MDSnV), generated from the second national LiDAR coverage acquired between 2015 and 2021, with an original spatial resolution of 2.5 m (CNIG, 2021). The MDSnV was obtained by interpolating vegetation height above ground from LiDAR returns classified into low, medium, and tall vegetation

classes (C3, C4, and C5, respectively) as part of the Spanish National Plan for Aerial Orthophotography (PNOA). For the study site, LiDAR data were collected in 2020, ensuring temporal consistency with SAR acquisitions from the same year.

The mean tree height at the study site was 7.80 m. The reference data were aggregated into 20 m, 40 m, and 60 m mapping units to match the spatial resolutions of the input features.

3. Methods

3.1. General approach

This section presents the proposed methodology, describing the adopted deep learning framework, the training and validation strategies, and the final inference and reliability assessment steps.

The input feature set comprises interferometric and ancillary layers derived from ALOS-2 data, which were used either individually or in combination to predict forest height. Specifically, the considered features include:

- Median backscattering coefficients in HH and HV polarizations

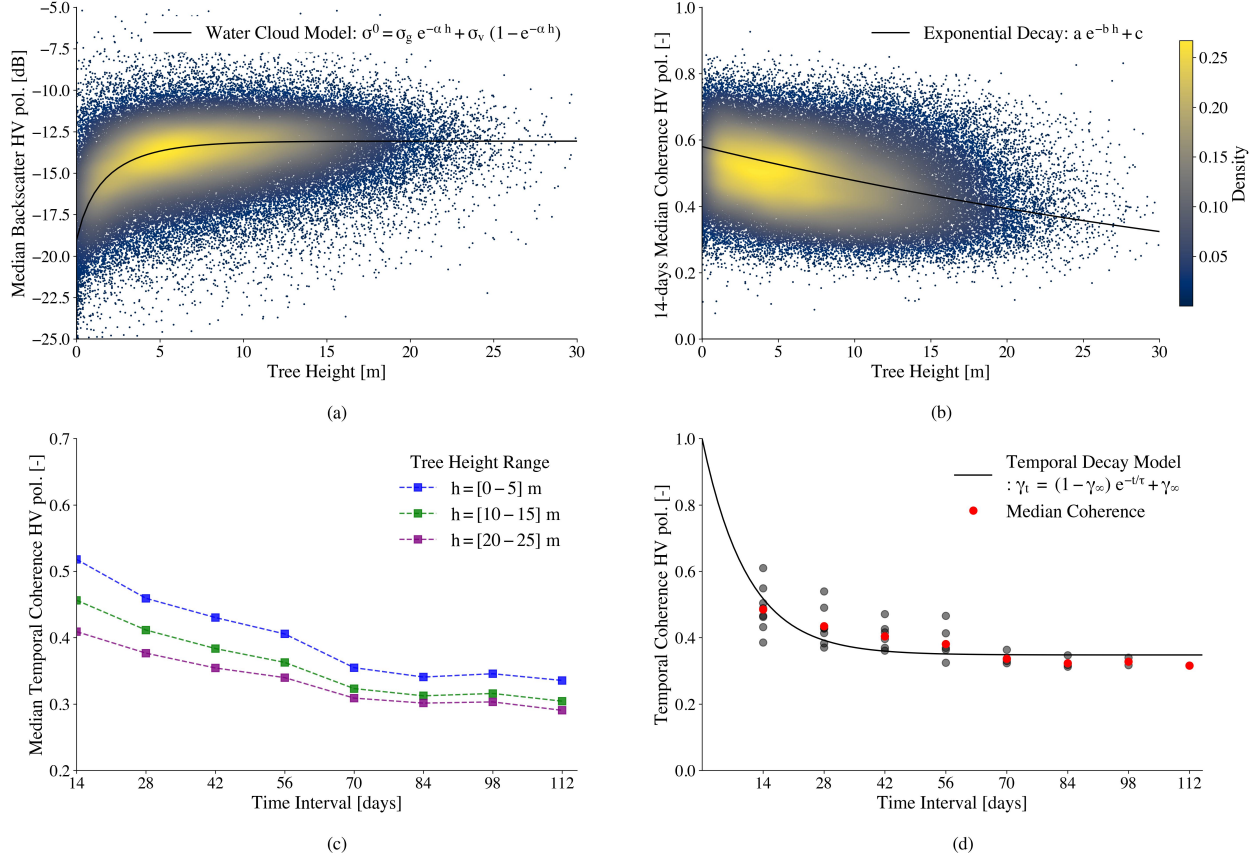


Figure 3: Top row: Sensitivity of median HV backscatter (left) and 14-day HV temporal coherence (right) to reference tree height. The backscatter is fitted using the Water Cloud Model (WCM), while the coherence is fitted with an exponential decay function, highlighting its dependence on forest structure. Bottom row: Sensitivity of ALOS-2 HV temporal coherence to temporal baseline. The left panel shows the median HV coherence as a function of repeat interval for different tree-height intervals, while the right panel presents a pixel-level example of temporal coherence decay, showing individual HV interferometric pairs (gray dots), the median stack (red dots), and the fitted exponential decay model (solid line) following Sica et al. (2019). Only pixels classified as vegetated were included in the analysis.

- Median HV and HH temporal coherence time series, with temporal baselines ranging from 14 to 112 days;
- Model-derived parameters τ and ρ_∞ ;
- Local incidence angle;
- Copernicus DEM at 30 m spatial resolution.

A family of U-Net-based architectures was adopted as the backbone for the proposed SAR-time-series based forest height mapping framework. The tested models include the original U-Net model (Ronneberger et al., 2015), Nested U-Net (Zhou et al., 2018), and SeU-Net (Ge et al., 2022a), which incorporates squeeze-excitation blocks (Hu et al., 2020). These architectures were selected for their capability to effectively model spatially structured patterns from limited reference data, a common constraint in remote sensing forest inventory mapping tasks.

To evaluate the performance of the proposed models, we conducted experiments using different combinations of input features and spatial resolutions. Each configuration was initially trained and evaluated at 20 m resolution, while additional experiments at 40 m and 60 m were performed for selected representative cases. For these cases, conventional machine

learning algorithms, namely Random Forest (RF) and k-Nearest Neighbors (kNN), were also implemented to provide a baseline comparison against the proposed deep learning models.

The overall workflow of the study is shown in Fig. 4.

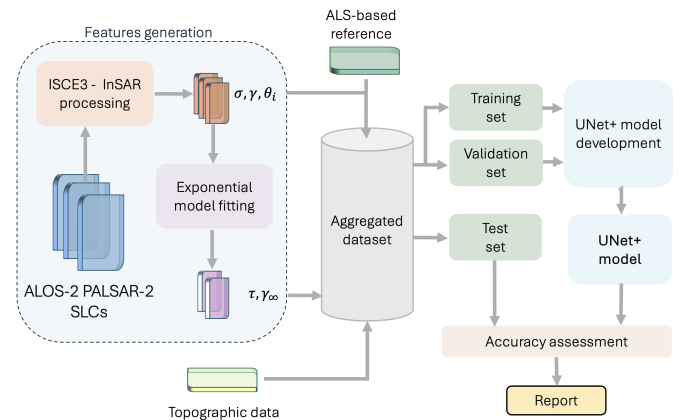


Figure 4: General workflow of the proposed hybrid deep learning approach for forest height mapping from multi-temporal PALSAR-2 interferometric and backscatter features.

3.2. UNet modeling

The U-Net architecture is essentially a convolutional encoder-decoder network originally developed for biomedical image segmentation (Ronneberger et al., 2015). It has since been widely applied to EO-based semantic segmentation and regression tasks (e.g., Scepanović et al., 2021; Ge et al., 2022a, 2023; Illarionova et al., 2022; Bolyn et al., 2022). Its encoder-decoder design, reinforced by skip connections, enables the preservation of fine spatial details and the reconstruction of high-resolution output maps, essential for predicting forest attributes such as forest height, growing stock volume and above ground biomass density.

The Vanilla U-Net consists of a symmetrical encoder-decoder structure composed of double convolutional blocks, each including a 2-D convolution, batch normalization, and ReLU activation. This design allows the model to extract deep hierarchical features while maintaining spatial correspondence between input and output layers, making it well suited for pixel-level regression and classification tasks (Figure 5a).

Building on this foundation, several U-Net family variants were explored to enhance the representation of forest structural heterogeneity in spatially explicit EO data:

- Nested U-Net introduces densely connected skip pathways that improve multi-scale feature fusion and gradient propagation, enhancing robustness in heterogeneous forest environments where spectral and structural complexity may lead to ambiguous signals (Fig.5,b).
- SeU-Net extends the U-Net design by incorporating squeeze-and-excitation blocks, which enable the network to adaptively recalibrate channel-wise feature responses. This mechanism is particularly beneficial for multi-source input images, where the relative importance of different SAR and ancillary inputs can vary across forest types and ecological and seasonal conditions. The adoption of SE blocks has been shown to yield superior accuracy in boreal forest height mapping using Sentinel-1 and Sentinel-2 time series (Ge et al., 2022a) and in other studies employing multi-feature datasets (Gazzea et al., 2023) (Fig. 5,c).

3.3. Experimental design and setup

This section outlines the organization of the experiments performed with the preprocessed SAR feature stack and reference data set described in Sections 2.2, 2.2.1 and 2.3.

The whole study area was divided into image patches of size 128×128 px, with pixel size of 20 m, thus size of each patch was $2.56 \times 2.56 \text{ km}^2$. To train, validate, and test the studied models, the input dataset was divided into three subsets: 60% for training, 20% for validation, and 20% for testing, using random cropping. Before training, all input bands were normalized using Z-score based normalization using band-wise mean and standard deviation.

The models were trained using a masked mean squared error (masked-MSE) loss function, which computes the prediction error only over valid reference pixels. Invalid or missing-

data regions were excluded, such as radar layover/shadow areas. The loss function is defined as

$$\mathcal{L}(\theta) = \frac{\sum_{i=1}^n (f_{\theta}(x_i) - y_i)^2}{n} \quad (2)$$

where x_i represents the input features for valid pixel i , y_i is the corresponding reference tree height, $f_{\theta}(x_i)$ is the model prediction parameterized by the network weights θ .

Experiments were run on a Google Colab platform with Tesla T4 GPUs. The proposed model was built by PyTorch 2.8 with CUDA 12.6 and Python 3.12. During the training, Adam (Kingma and Ba, 2017) was chosen as the optimizer and OneCycleLR as the learning rate scheduler, the initial learning rate was set to 10^{-3} , and patience parameter was set to 5. We trained the model for 50 epochs, with batch size of 16, the weight decay factor was set to 10^{-4} . The checkpoints were saved and updated according to the validation loss, with the best checkpoint used in the testing stage. An example of training and validation loss functions is shown in Figure 6.

3.4. Accuracy metrics

To evaluate model performance, we employed the Mean Error (ME), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). Let y_i denote the reference values, $\hat{y}_i$ the predicted values, and n the total number of samples. Mean Error represents the average deviation between predicted and reference values:

$$\text{ME} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (3)$$

RMSE measures the overall magnitude of prediction errors, with larger errors contributing more strongly to the metric due to the squared residuals:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (4)$$

Coefficient of Determination R^2 quantifies the proportion of variance in the reference values explained by the predictions:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

where $\bar{y}$ is the mean of the reference values.

4. Results and Discussion

This section presents the results obtained from the proposed methodology, including the performance of the deep learning architectures across different input configurations and spatial resolutions. The tree height estimation results for the 20 m resolution experiments are reported in Table 1, while Table 2 summarizes the outcomes obtained at 20 m, 40 m, and 60 m resolutions for selected input feature configurations. Examples of forest patch predictions are shown in Fig. 7, and selected scatterplots illustrating model performance are presented in Fig. 8.

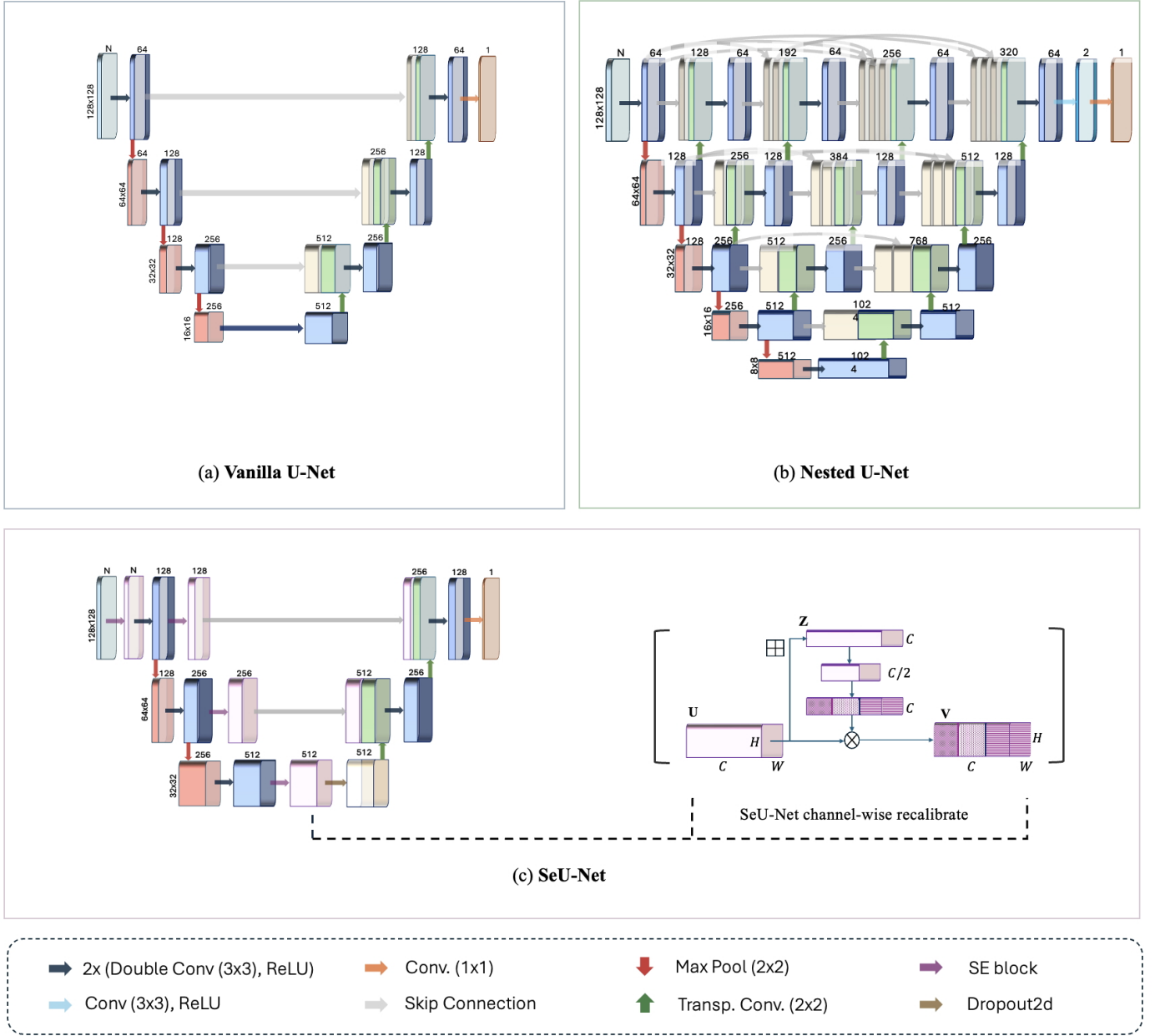


Figure 5: Structure of studied U-Net models, where the Vanilla U-Net (a) represents the baseline design, SeU-Net (c) incorporates squeeze-and-excitation blocks for channel-wise attention, and Nested U-Net (b) introduces dense skip connections

A map of the retrieved tree height obtained using the SeU-Net model at 20 m resolution and all available input features is shown in Fig. 9.

4.1. General prediction performance observations

Overall, examined deep learning models demonstrated consistent capability to predict the target variable from multi-temporal PALSAR-2 datasets with high accuracy. The baseline U-Net provided a consistent reference for comparison, while two modified variants, i.e., Nested U-Net, featuring nested skip connections for improved multiscale feature fusion, and SeU-Net enhanced with Squeeze-and-Excitation attention blocks, both

achieved higher predictive accuracy. Across all feature configurations, inclusion of additional polarimetric and geometry-derived features/predictors improved prediction accuracies, confirming that compiled combination of SAR polarimetric observables and interferometric coherence features provides the richest information for learning. Root-mean-square errors (RMSE) typically ranged between 2.8–3.9 m, and coefficients of determination (R^2) reached up to 0.67 for the best configurations at 20 m mapping units, indicating robust model generalization. Results further improved to less than 2 m RMSE for larger mapping units (20 m–40 m) when processing was performed in coarser resolution.

Performance Metrics										
Input	Model	ME [m]			RMSE [m]			R2 [-]		
		HH	HV	HH,HV	HH	HV	HH,HV	HH	HV	HH,HV
single-features										
σ	UNet	0.25	1.36	0.20	3.77	3.82	3.49	0.39	0.38	0.47
	Nested UNet	0.18	0.21	0.05	3.48	3.29	3.17	0.48	0.53	0.57
	SeU-Net	0.12	0.23	0.24	3.54	3.25	3.25	0.46	0.55	0.55
γ_{14}	UNet	0.23	0.18	-0.04	3.29	4.26	3.90	0.54	0.22	0.35
	Nested UNet	0.15	0.00	0.26	3.97	4.23	3.81	0.33	0.23	0.38
	SeU-Net	0.08	0.29	0.27	3.97	4.19	3.85	0.33	0.25	0.37
τ, ρ_{∞}	UNet	-0.02	0.10	0.17	4.02	4.14	4.01	0.31	0.29	0.31
	Nested UNet	0.11	-0.12	0.04	3.91	4.05	3.80	0.35	0.30	0.38
	SeU-Net	-0.01	0.08	0.05	3.91	4.05	3.81	0.35	0.30	0.39
$\gamma_{14}, \gamma_{28}, \gamma_{42}, \gamma_{56},$ $\gamma_{70}, \gamma_{84}, \gamma_{90}, \gamma_{112}$	UNet	0.20	-0.02	0.08	3.89	3.98	3.79	0.35	0.32	0.39
	Nested UNet	0.21	-0.15	0.13	3.75	3.92	3.62	0.40	0.34	0.44
	SeU-Net	0.08	0.19	0.18	3.70	3.91	3.66	0.41	0.34	0.43
combined-features										
σ, γ_{14}	UNet	0.24	0.13	0.16	3.29	3.20	3.19	0.54	0.56	0.56
	Nested UNet	0.08	-0.01	0.16	3.13	3.00	3.00	0.58	0.61	0.61
	SeU-Net	-0.15	0.07	0.00	3.15	3.09	2.94	0.57	0.59	0.63
$\sigma, \tau, \rho_{\infty}$	UNet	0.12	0.46	0.28	3.29	3.34	3.11	0.54	0.52	0.59
	Nested UNet	0.05	0.00	0.00	3.13	3.04	2.94	0.58	0.60	0.63
	SeU-Net	0.00	0.04	0.04	3.12	2.96	2.96	0.58	0.62	0.62
$\sigma, \gamma_{14}, \tau, \rho_{\infty}$	UNet	-0.02	-0.10	-0.03	3.30	3.16	3.03	0.53	0.57	0.61
	Nested UNet	0.26	0.17	0.10	3.13	2.94	2.89	0.58	0.63	0.64
	SeU-Net	0.05	-0.04	0.06	3.04	2.96	2.92	0.60	0.62	0.63
$\sigma, \gamma_{14}, \gamma_{28}, \gamma_{42}, \gamma_{56},$ $\gamma_{70}, \gamma_{84}, \gamma_{90}, \gamma_{112}$	UNet	0.16	0.06	0.00	3.22	3.07	3.03	0.56	0.60	0.61
	Nested UNet	-0.02	-0.09	0.08	3.06	2.94	2.90	0.60	0.63	0.64
	SeU-Net	0.10	0.16	-0.05	3.00	2.93	2.92	0.61	0.63	0.63
combined-features + geometry layers										
σ, θ_i	UNet	0.05	0.24	0.25	3.60	3.31	3.25	0.44	0.53	0.55
	Nested UNet	-0.09	-0.08	0.09	3.15	3.02	2.95	0.57	0.61	0.63
	SeU-Net	0.14	0.11	0.08	3.22	3.11	3.02	0.56	0.59	0.61
$\sigma, \theta_i, \text{dem}$	UNet	0.01	-0.12	0.15	3.28	3.15	3.13	0.54	0.58	0.58
	Nested UNet	-0.06	0.11	0.11	3.15	2.98	2.93	0.54	0.62	0.63
	SeU-Net	0.00	0.01	0.18	3.11	2.98	2.94	0.58	0.62	0.63
$\sigma, \theta_i, \text{dem}, \gamma_{14}$	UNet	-0.08	0.13	0.00	3.13	3.01	2.84	0.58	0.61	0.66
	Nested UNet	0.24	0.27	0.01	2.86	2.87	2.78	0.62	0.65	0.67
	SeU-Net	0.03	0.05	0.03	3.00	2.88	2.82	0.61	0.64	0.66
$\sigma, \theta_i, \text{dem}, \gamma_{14}, \tau, \rho_{\infty}$	UNet	0.05	0.10	-0.09	2.98	2.90	2.93	0.62	0.64	0.63
	Nested UNet	0.05	0.01	0.04	2.92	2.83	2.77	0.63	0.66	0.67
	SeU-Net	0.05	0.01	0.06	2.92	2.79	2.79	0.64	0.67	0.67
$\sigma, \theta_i, \tau, \rho_{\infty}$	UNet	0.09	0.14	0.04	3.36	3.12	3.10	0.52	0.58	0.57
	Nested UNet	-0.02	0.00	0.07	2.95	2.88	2.83	0.63	0.64	0.66
	SeU-Net	0.12	0.06	0.00	2.98	2.86	2.82	0.62	0.65	0.66
$\sigma, \theta_i, \text{dem}$ $\gamma_{14}, \gamma_{28}, \gamma_{42}, \gamma_{56},$ $\gamma_{70}, \gamma_{84}, \gamma_{90}, \gamma_{112}$	UNet	0.21	0.23	0.14	3.16	3.09	2.83	0.57	0.59	0.66
	Nested UNet	0.21	-0.09	0.04	2.93	2.83	2.79	0.63	0.66	0.67
	SeU-Net	0.16	0.00	0.17	2.88	2.82	2.78	0.64	0.66	0.67

Table 1: Prediction performance of all evaluated deep learning models across different SAR feature modalities at 20 m spatial resolution. Best achieved performance for each input modality is highlight in bold

4.2. SAR backscatter data only

When using only SAR backscatter intensities (σ^0), the models achieved moderate prediction performance, with RMSE between 3.1–3.9 m and R^2 values ranging from 0.45 to 0.55. Among the three architectures, Nested U-Net generally yielded the lowest errors and highest R^2 values, indicating that its dense skip connections effectively enhanced feature propagation and mul-

tiscale representation even from limited intensity-based information. SeU-Net achieved comparable accuracy in several cases, suggesting that the channel-wise recalibration introduced by the SE blocks also supports improved discrimination of backscatter variations. The baseline U-Net consistently provided lower accuracy than its two enhanced variants. Across all models, performance gains were clear when dual-pol (HH,HV) inputs were

Performance Metrics										
Input	Model	ME [m]			RMSE [m]			R2 [-]		
		20m	40m	60m	20m	40m	60m	20m	40m	60m
$\sigma, \theta_i, \text{dem}$	RF	0.12	0.08	0.08	3.90	2.40	2.40	0.29	0.66	0.66
	MLR	0.20	0.35	0.35	3.32	3.32	3.32	0.34	0.34	0.34
	kNN	0.07	0.05	0.05	3.62	2.60	2.60	0.59	0.58	0.59
	UNet	0.15	0.13	0.06	3.13	2.36	2.38	0.58	0.69	0.69
	Nested UNet	0.11	0.07	0.10	2.93	2.19	2.19	0.63	0.73	0.74
	SeU-Net	0.18	0.21	0.07	2.94	2.24	2.23	0.63	0.72	0.73
$\sigma, \gamma_{14}, \theta_i, \text{dem}$	RF	0.08	0.04	0.04	3.10	2.17	2.16	0.55	0.71	0.72
	MLR	0.04	0.06	0.05	3.50	2.76	2.74	0.43	0.54	0.55
	kNN	0.03	0.02	0.01	3.36	2.38	2.38	0.47	0.66	0.66
	UNet	0.00	0.02	0.01	2.84	2.12	2.06	0.66	0.75	0.76
	Nested UNet	0.01	0.06	-0.01	2.78	2.03	2.03	0.67	0.77	0.77
	SeU-Net	0.03	0.10	0.09	2.82	2.07	2.07	0.66	0.76	0.76
$\sigma, \theta_i, \text{dem}, \gamma_{14}, \tau, \rho_\infty$	RF	0.07	0.06	0.07	3.17	2.12	2.12	0.54	0.72	0.73
	MLR	0.04	0.09	0.07	3.48	2.74	2.73	0.44	0.55	0.56
	kNN	0.01	0.05	0.05	3.46	2.37	2.36	0.45	0.66	0.67
	UNet	-0.09	-0.03	-0.03	2.93	2.14	2.14	0.63	0.75	0.75
	Nested UNet	0.04	0.03	0.02	2.77	1.97	1.98	0.67	0.78	0.78
	SeU-Net	0.06	0.10	0.09	2.79	1.96	1.95	0.67	0.78	0.79
$\sigma, \theta_i, \text{dem}$ $\gamma_{14}, \gamma_{28}, \gamma_{42}, \gamma_{56},$ $\gamma_{70}, \gamma_{84}, \gamma_{90}, \gamma_{112}$	RF	0.12	0.07	0.07	3.06	2.12	2.11	0.56	0.73	0.73
	MLR	0.06	0.07	0.07	3.48	2.75	2.73	0.44	0.55	0.55
	kNN	0.07	0.06	0.06	3.44	2.47	2.43	0.45	0.63	0.64
	UNet	0.14	0.07	0.12	2.83	2.02	2.04	0.66	0.77	0.77
	Nested UNet	0.04	0.10	0.02	2.79	2.03	1.99	0.67	0.77	0.78
	SeU-Net	0.17	0.14	0.11	2.78	1.98	2.00	0.67	0.77	0.79

Table 2: Prediction performance of all evaluated deep learning models at 20m, 40m and 60m resolution spatial resolutions using dual pol (HH, HV) input features. Best achieved performance for each input modality is highlight in bold.

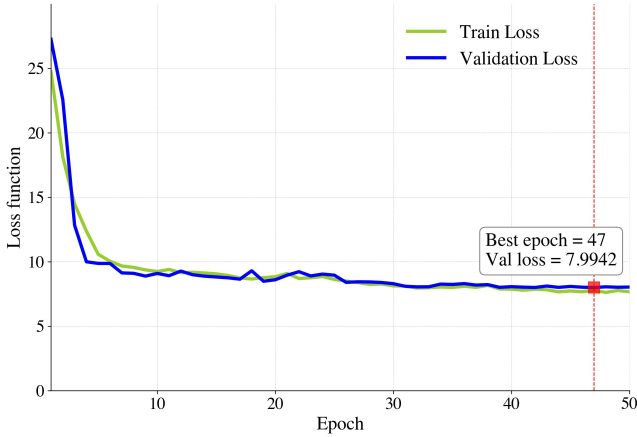


Figure 6: Accuracy curves during Nested UNet model training and development using all available input features.

used, confirming that the joint use of both polarizations provides a more complete scattering characterization than single-polarization inputs.

4.3. Intensity and coherence data

Incorporating temporal coherence (γ) in addition to backscatter significantly improved predictive accuracy. The inclusion of γ led to a notable reduction in RMSE (to approximately 3.0–3.3 m) and an increase in R^2 to 0.60-0.63, reflecting the

added sensitivity of temporal coherence to structural characteristics of the forests. The Nested U-Net architecture showed the most consistent improvement, as its nested skip connections enhanced the fusion of multiscale spatial and temporal features from amplitude and coherence inputs. The SeU-Net model had similar performance in prediction accuracy, occasionally even achieving slightly higher R^2 values, demonstrating that spatial attention via SE blocks complements the temporal information embedded in coherence. Overall, both architectures substantially outperformed the baseline U-Net.

While the combined use of backscatter intensity, geometric layers and insar coherence improved performance, the 14-day coherence used as a single input feature yielded comparatively low accuracy, particularly when generated with a limited number of looks. This behavior is due to the high spatial variability of the coherence measurements as well as uncompensated biases in the coherence estimation.

4.4. Including additional variables

Including model-derived parameters such as the InSAR coherence-based inverted parameters (τ, ρ_∞) yielded performance comparable to that obtained using the full temporal stack of interferometric coherence. This indicates both that the DL model is capable of directly learning the underlying physical relationships from the temporal coherence series, and that τ and ρ_∞ represent meaningful descriptors of the temporal coherence sensitivity to forest structure. The main advantage of employing these two model-derived parameters lies in the reduced di-

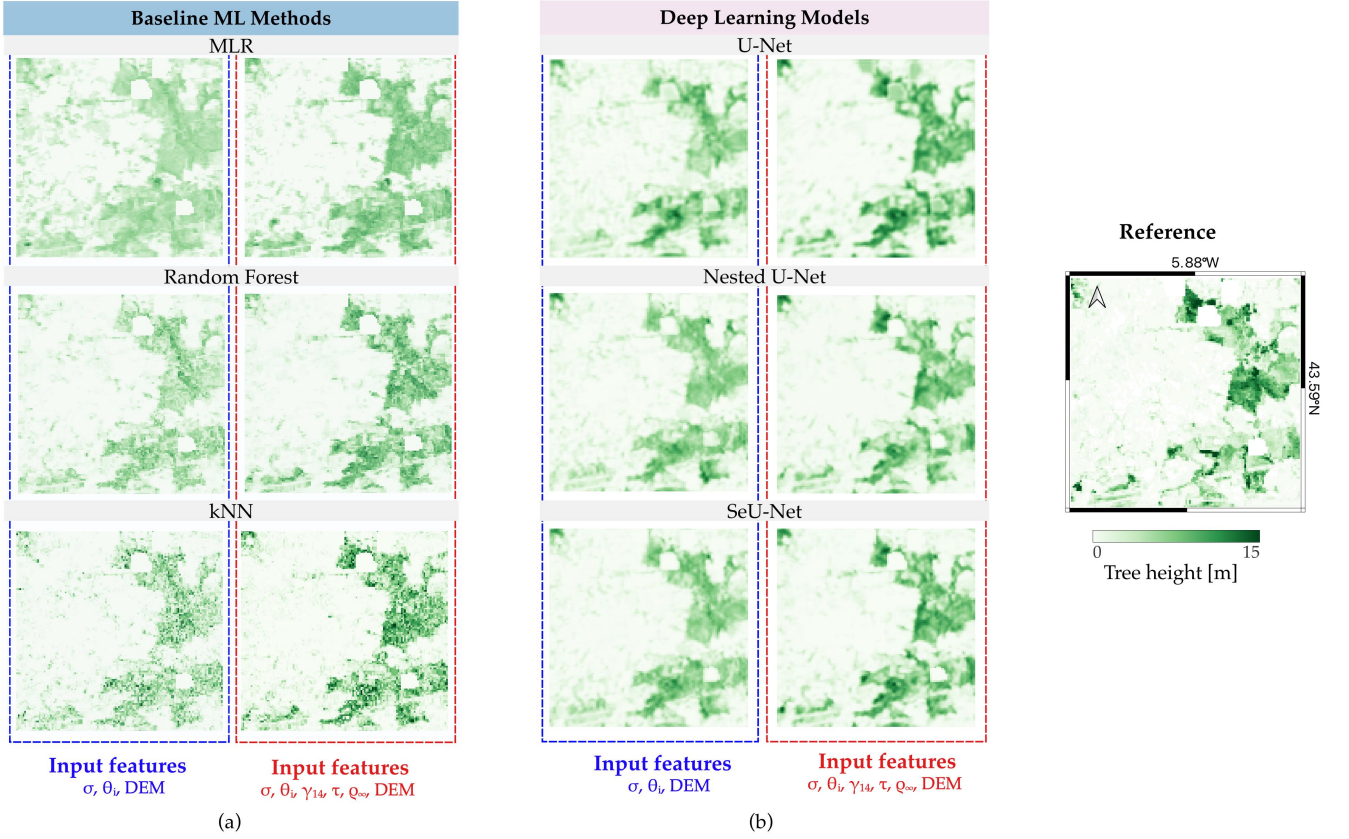


Figure 7: Example of tree height estimates over a forest patch obtained with baseline machine learning (a) and deep learning (b) models using two input feature configurations at 20 m resolution.

dimensionality of the input features, which lowers computational costs.

When adding incidence angle (θ_i) and digital elevation model (DEM) layers, model performance improved further. These features provided geometric and physical context, allowing the networks to better account for topographic and incidence-angle induced variations in backscatter and coherence. The Nested U-Net achieved the lowest RMSE values (approximately 2.8–2.9 m) and highest R^2 (up to 0.67), closely followed by SeU-Net. Overall, this indicates that combining SAR observables with auxiliary physical variables enables the extraction of more meaningful representations. The improvement was especially evident for areas with complex terrain, where geometry-driven variations are pronounced and cannot be captured by σ^0 and γ alone.

4.5. Coarser resolution mapping

The comparison of performances at 20 m, 40 m, and 60 m resolutions showed that all models maintained stable accuracy across scales, with improved prediction accuracy at coarser mapping units, as illustrated in Table 2 and Fig. 8-b,c,d.

From physics of radio wave scattering based perspective, larger mapping units increase the ENL and SNR and provide

more reliable coherence estimates, thereby enabling better predictions with less noisy inputs. Transition from 20 m to 60 m resolution resulted in an average increase in R^2 of approximately 0.1 and a reduction in RMSE by about 0.4–0.6 m across all input feature configurations. These improvements reflect the stabilizing effect of spatial aggregation, which reduces local speckle-induced variance and enhances the coherence of learned spatial patterns. Overall, these findings suggest that moderate spatial aggregation (40–60 m) provides the best trade-off between spatial detail and predictive stability for PALSAR-2 based deep learning retrievals.

At the finest resolution (20 m), the studied models achieved the highest spatial detail but also slightly larger noise-induced errors, with RMSE typically between 3.0–3.9 m and R^2 values around 0.63–0.67. Aggregation to 40 m led to modest improvements in accuracy for most configurations (RMSE $\approx$ 2.7–3.3 m; R^2 up to 0.73–0.76), suggesting that spatial averaging helps suppress speckle and local variability effects. Further coarsening to 60 m produced very limited additional gains, with RMSE as low as 1.9–2.1 m and R^2 values reaching up to 0.78–0.79 for the best-performing deep-learning models, indicating that likely 40m resolution can provide an optimal trade off between prediction accuracy and spatial detail.

Across all resolutions, both enhanced architectures—Nested

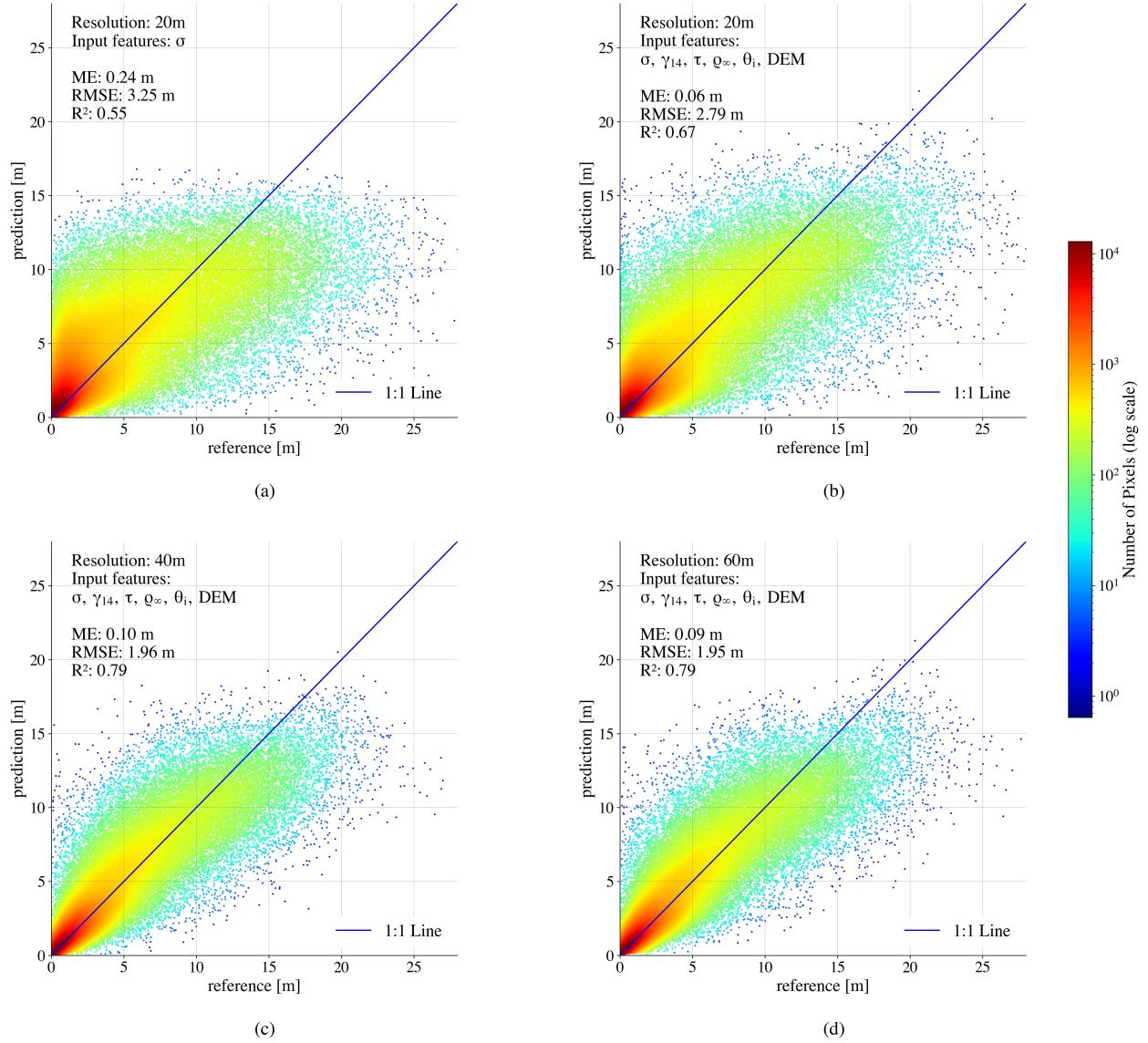


Figure 8: Scatterplots showing the effect of input feature configuration and spatial resolution on SeU-Net tree-height predictions using HH and HV polarizations. (a) 20 m resolution using backscatter (σ) only; (b) 20 m resolution using all features ($\sigma, \gamma_{14}, \tau, Q_{\infty}, \theta_i, DEM$); (c) 40 m resolution using all features; (d) 60 m resolution using all features. The 1:1 line is shown in blue for reference.

U-Net and SeU-Net—outperformed the baseline U-Net. The Nested U-Net demonstrated more consistent performance across feature combinations and scales, reflecting the benefit of dense skip connections for multiscale feature integration. Meanwhile, SeU-Net often achieved slightly higher accuracy at coarser resolutions (particularly 60 m) and for complex input configurations, likely due to its channel-wise recalibration mechanism that adaptively weights informative features. The baseline U-Net consistently lagged behind these variants, though still exceeded the accuracy of conventional machine learning methods summarized in Section 4.6.

Consequently, resolutions between 40 m and 60 m appeared to provide the best balance between spatial fidelity and predictive accuracy for PALSAR-2 based deep learning retrievals.

4.6. Comparison with baseline ML methods

Traditional machine-learning methods exhibited predictable scaling trends: Random Forest (RF) and kNN generally outperformed MLR, with RF achieving RMSE ≈ 3.1 m at 20 m and ≈ 2.2 m at 60 m. However, examined UNet-family models consistently surpassed them by 0.3–0.5 m across resolutions, demonstrating superior capacity to generalize nonlinear dependencies among backscatter intensity, interferometric coherence, and geometry-derived features.

Visual inspection of Figure 7 indicates that UNet models reproduce the reference spatial patterns more precisely than the baseline methods. The Nested U-Net and SeU-Net architectures capture both the fine-scale heterogeneity within forest stands and the broader gradients in canopy height across the scene. Their outputs show close correspondence with the reference, particularly in densely forested areas, where local maxima and

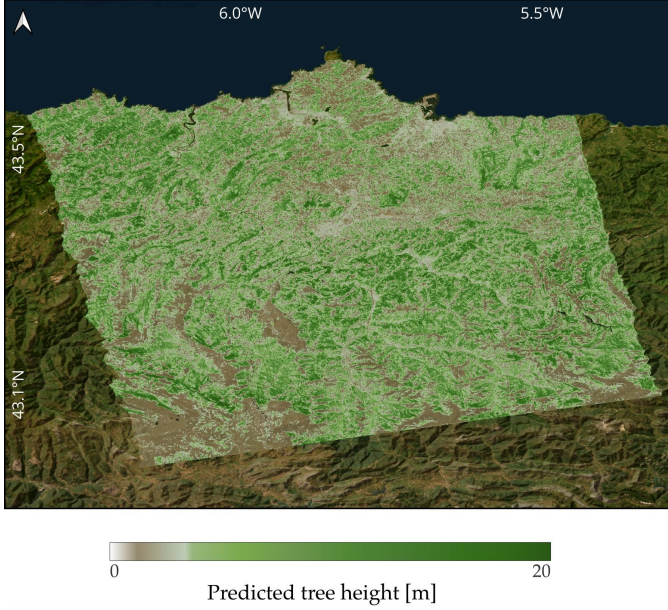


Figure 9: Tree height predictions from the SeU-Net model at 20 m resolution using median backscatter, 14-day coherence, and model-derived layers (τ , ρ_∞) at co- and cross-polarizations.

low-height clearings are clearly delineated. In contrast, the baseline methods exhibit stronger smoothing effects (especially MLR) and reduced dynamic range (all methods), leading to underestimation of tall canopies and less distinct stand boundaries. For Random Forests and kNN, some nonstationarity is evident in adjacent pixels even in homogeneous areas explained by pixel-level prediction approach.

The addition of interferometric coherence and polarimetric variables (γ_{14} , τ , ρ_∞) improves structural detail and contrast across all models, but the benefit is more apparent for the UNet model predictions. While Random Forest partly recovers large-scale gradients, the deep models produce maps whose textural and radiometric patterns align closely with the ALS based reference, allowing to produce spatially consistent height retrievals.

4.7. Comparison with prior literature

Results reported in our paper for deep learning architectures appear more competitive when compared to recent state-of-the-art studies employing both traditional model-based and modern deep learning approaches across different sensors and forest biomes.

Model-based approaches using spaceborne L-band time series have reported RMSE values typically ranging from ~ 4 to 6 m, depending on the study site and dataset. Seppi et al. (2022) achieved RMSEs of 4–5 m from SAOCOM L-band InSAR coherence over managed forests in Argentina. Using PALSAR-2 time series, Telli et al. (2025) showed that combining ALOS-2 backscatter and interferometric temporal coherence reduced RMSE to ~ 4.2 m over mixed evergreen and deciduous forests, with class-specific errors between 2.0 and 6.2 m. Comparable accuracies were obtained by Zhang et al. (2022), who reported RMSEs of 4.4–4.8 m using single-baseline L-band PolInSAR inversion, and by Lei and Siqueira (2014) who achieved RMSE

below 4 m using repeat-pass InSAR coherence over Maine, USA. Earlier airborne L-band PolInSAR work by Praks et al. (2012) similarly reported RMSE ≈ 3.7 m using different versions of simplified LIDAR-DEM aided RVoG inversion. In a study over Traunstein site in Germany, a similar accuracy with RMSE of 3.71 m was achieved when using RVoG-model based retrievals with L-band PolInSAR images.

Finally, a regionally focused study over Asturias region in Spain using multi-source Sentinel-1 and PALSAR-2 mosaic layers and optical Sentinel-2 images and wide range of machine-learning methods (RF, GPR, and kNN) obtained RMSE values of approximately 3–5 m for forest height mapping (Tejido-Murias et al., 2025). These results fall within the same accuracy range as the present study for pixel-level machine learning predictions, yet here the use of UNet models and interferometric features allowed to obtain superior results using PALSAR-2 time series in contrast to multi-source dataset used in prior paper.

Overall, our studied models, especially the attention-augmented SeU-Net and Nested UNet variants, achieved lower RMSE values than physics model-based and classical machine-learning approaches relying on similar L-band data, or even more advanced PolInSAR datasets or multi-source SAR and optical imagery, confirming the potential of the proposed deep learning framework for tree height estimation from PALSAR-2 time series.

4.8. Model applicability and limitations

Described approach is suitable for regional mapping under similar weather and seasonal conditions when there is a representative set of reference ALS data (hundreds of image patches) available for end-to-end training.

Similar performance can be expected in adjacent regions with comparable forest composition. However, applications in remote areas or different forest biomes would require significant model fine-tuning, such as domain adaptation or model transfer, supported by at least a small sample of reference ALS patches or a representative set of forest plots. The potential of pretrained model transfer in the forest context has been studied to a limited extent under simplified scenarios (Ge et al., 2023; Kuzu et al., 2024). Further research is needed to investigate more complex cases, as radar scattering signatures depend not only on forest structure but also on seasonal and weather conditions influencing forest transmissivity and ground conditions (e.g., rainfall, soil moisture, or frozen conditions), as well as on imaging geometry and topography.

5. Conclusions and Outlook

This study presented a deep learning framework to estimate tree height from time series of L-band ALOS-2 PALSAR-2 interferometric acquisitions. Three DL models from the U-Net family, i.e., U-Net, Nested U-Net, and SeU-Net, were used to assess the impact of architecture design on retrieval accuracy. Nested skip connections as well as channel-wise recalibration significantly improved performance compared to basic U-Net.

Among the tested models, Nested U-Net and SeU-Net achieved comparable accuracy, with the former yielding the best overall results.

Several combinations of SAR and InSAR input data were tested to evaluate the contribution of multi-polarization and multi-temporal features. Integrating temporal coherence with backscatter significantly reduced estimation errors, achieving the best results when the full temporal coherence stack was exploited. Similar accuracy was obtained when using model-derived parameters (τ , ρ_∞), demonstrating that these features effectively capture the temporal and spatial sensitivity of coherence to forest structure while reducing input dimensionality. The inclusion of geometry layers, such as incidence angle and DEM, further enhanced performance by allowing the network to better account for topography-induced variations in backscatter and coherence. The best accuracy at 20 m resolution was achieved using the full feature stack, with an RMSE of approximately 2.8 m.

The comparison across mapping scales reported higher estimation accuracies at 40 m and 60 m, achieving RMSE $\sim$ 1.95 m at coarser resolution. The achieved accuracies outperform those reported in recent studies based on model-based and traditional machine learning approaches, confirming the strong potential of deep learning for SAR-based forest structure retrieval.

Future developments will include the application of the proposed framework to upcoming NISAR data, enabling large-scale assessments across diverse forest biomes. Additional efforts will focus on integrating complementary datasets, such as optical and C-band SAR observations, to further enhance retrieval accuracy and generalization. Finally, testing alternative deep learning architectures and incorporating explainable AI techniques will be essential to improve model interpretability and advance the physical understanding of SAR-based forest parameter retrieval.

Acknowledgments

The authors would like to thank Japan Aerospace Exploration Agency - JAXA for providing ALOS-2 data and NASA JPL for providing the ISCE3 software.

References

- Antropov, O., Rauste, Y., Ahola, H., Hame, T., 2013. Stand-level stem volume of boreal forests from spaceborne SAR imagery at L-band. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6, 35–44. doi:10.1109/JSTARS.2013.2241018.
- Antropov, O., Rauste, Y., Häme, T., Praks, J., 2017. Polarimetric ALOS PALSAR time series in mapping biomass of boreal forests. *Remote Sensing* 9. URL: <https://www.mdpi.com/2072-4292/9/10/999>, doi:10.3390/rs9100999.
- Astola, H., Seitsonen, L., Halme, E., Molinier, M., Lönnqvist, A., 2021. Deep neural networks with transfer learning for forest variable estimation using Sentinel-2 imagery in boreal forest. *Remote Sens.* 13.
- Berger, K., Foerster, S., Szantoi, Z., Hostert, P., Foerster, M., Van De Kerchove, R., Vancutsem, C., Schweitzer, C., Masolele, R., Reiche, J., Dowell, M., Enssle, F., Suarez, D.R., Nepomshina, O., Herold, M., 2025. Evolving earth observation capabilities for recent land-related eu policies. *Land Use Policy* 158, 107749. doi:<https://doi.org/10.1016/j.landusepol.2025.107749>.
- Berninger, A., Lohberger, S., Stängel, M., Siegert, F., 2018. Sar-based estimation of above-ground biomass and its changes in tropical forests of kalimantan using l- and c-band. *Remote Sensing* 10. doi:10.3390/rs10060831.
- Bhogapurapu, N., Siqueira, P., Armston, J., Urbazhev, M., Wessels, K., Duncanson, L., 2024. Canopy height estimation using c- and l-band insar coherence over savannas and dry forests, in: *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2290–2293. doi:10.1109/IGARSS53475.2024.10640881.
- Bolyn, C., Lejeune, P., Michez, A., Latte, N., 2022. Mapping tree species proportions from satellite imagery using spectral-spatial deep learning. *Remote Sens. Env.* 280. doi:10.1016/j.rse.2022.113205.
- Carceneri, D., Rizzoli, P., Ienco, D., Bruzzone, L., 2023. A deep learning framework for the estimation of forest height from bistatic tandem-x data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
- Cartus, O., Santoro, M., Kelldorfer, J., 2012. Mapping forest aboveground biomass in the Northeastern United States with ALOS PALSAR dual-polarization L-band. *Remote Sensing of Environment* 124, 466–478. doi:<https://doi.org/10.1016/j.rse.2012.05.029>.
- Cartus, O., Santoro, M., Wegmuller, U., Labriere, N., Chave, J., 2022. Sentinel-1 coherence for mapping above-ground biomass in semiarid forest areas. *IEEE Geoscience and Remote Sensing Letters* 19. doi:10.1109/LGRS.2021.3071949.
- Cloude, S.R., Papathanassiou, K.P., 1998. Polarimetric sar interferometry. *IEEE Transactions on geoscience and remote sensing* 36, 1551–1565.
- CNIG, 2021. Obra derivada de mdsnvegetación2,5 2015–2021. Centro Nacional de Información Geográfica (CNIG). CC-BY 4.0. Available at: <https://scne.es>.
- Dobson, M.C., Ulaby, F.T., LeToan, T., Beaudoin, A., Kasischke, E.S., Christensen, N., 1992. Dependence of radar backscatter on coniferous forest biomass. *IEEE Transactions on Geoscience and remote Sensing* 30, 412–415.
- Englhart, S., Keuck, V., Siegert, F., 2012. Modeling above-ground biomass in tropical forests using multi-frequency sar data—a comparison of methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5, 298–306. doi:10.1109/JSTARS.2011.2176720.

- Gazzea, M., Solheim, A., Arghandeh, R., 2023. High-resolution mapping of forest structure from integrated sar and optical images using an enhanced u-net method. *Science of Remote Sensing* 8, 100093. doi:<https://doi.org/10.1016/j.srs.2023.100093>.
- Ge, S., Antropov, O., Häme, T., McRoberts, R.E., Miittinen, J., 2023. Deep learning model transfer in forest mapping using multi-source satellite sar and optical images. *Remote Sensing* 15. URL: <https://www.mdpi.com/2072-4292/15/21/5152>, doi:10.3390/rs15215152.
- Ge, S., Gu, H., Su, W., Praks, J., Antropov, O., 2022a. Improved semisupervised UNet deep learning model for forest height mapping with satellite SAR and optical data. *IEEE J. Sel. Topics Appl. Remote Sens.* 15, 5776–5787.
- Ge, S., Su, W., Gu, H., Rauste, Y., Praks, J., Antropov, O., 2022b. Improved LSTM model for boreal forest height mapping using Sentinel-1 time series. *Remote Sens.* 14. doi:10.3390/rs14215560.
- GFOI, 2014. Integrating remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and Guidance from the Global Forest Observations Initiative. Group on Earth Observations, Geneva, Switzerland.
- Hu, H., Qi, S., Zhang, Y., Chen, X., Zhang, Z., 2020. SEUNet: Squeeze-and-excitation u-net for medical image segmentation. *Applied Sciences* 10, 713. doi:10.3390/app10020713.
- Hunka, N., May, P., Babcock, C., de la Rosa, J.A.A., de los Angeles Soriano-Luna, M., Saucedo, R.M., Armston, J., Santoro, M., Suarez, D.R., Herold, M., Málaga, N., Healey, S.P., Kennedy, R.E., Hudak, A.T., Duncanson, L., 2025. A geostatistical approach to enhancing national forest biomass assessments with earth observation to aid climate policy needs. *Remote Sensing of Environment* 318, 114557. doi:<https://doi.org/10.1016/j.rse.2024.114557>.
- Illarionova, S., Shadrin, D., Ignatiev, V., Shayakhmetov, S., Trekin, A., Oseledets, I., 2022. Estimation of the canopy height model from multispectral satellite imagery with convolutional neural networks. *IEEE Access* 10, 34116–34132. doi:10.1109/ACCESS.2022.3161568.
- Kingma, D.P., Ba, J., 2017. Adam: A method for stochastic optimization. URL: <https://arxiv.org/abs/1412.6980>, arXiv:1412.6980.
- Kugler, F., Lee, S.K., Hajnsek, I., Papathanassiou, K.P., 2015. Forest height estimation by means of pol-insar data inversion: The role of the vertical wavenumber. *IEEE Transactions on Geoscience and Remote Sensing* 53, 5294–5311.
- Kurvonen, L., Pulliainen, J., Hallikainen, M., 1999. Retrieval of biomass in boreal forests from multitemporal ERS-1 and JERS-1 sar images. *IEEE Transactions on Geoscience and Remote Sensing* 37, 198–205. doi:10.1109/36.739154.
- Kuzu, R.S., Antropov, O., Molinier, M., Dumitru, C.O., Saha, S., Zhu, X.X., 2024. Forest disturbance detection via self-supervised and transfer learning with sentinel-1&2 images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17, 4751–4767. doi:10.1109/JSTARS.2024.3361183.
- Latterini, F., Camarretta, N., Watt, M.S., 2025. Remote sensing for planning harvesting operations and monitoring their effects on the forest ecosystem: State of the art and future perspectives. *Forest Ecology and Management* 597, 123175. doi:<https://doi.org/10.1016/j.foreco.2025.123175>.
- Lavalle, M., Simard, M., Hensley, S., 2011. A temporal decorrelation model for polarimetric radar interferometers. *IEEE Transactions on Geoscience and Remote Sensing* 50, 2880–2888.
- Lavalle, M., Simard, M., Hensley, S., 2012. A temporal decorrelation model for polarimetric radar interferometers. *IEEE Transactions on Geoscience and Remote Sensing* 50, 2880–2888. doi:10.1109/TGRS.2011.2174367.
- Lavalle, M., Telli, C., Pierdicca, N., Khati, U., Cartus, O., Kelldorfer, J., 2023. Model-based retrieval of forest parameters from sentinel-1 coherence and backscatter time series. *IEEE Geoscience and Remote Sensing Letters* 20. doi:10.1109/LGRS.2023.3239825.
- Le Toan, T., Quegan, S., Woodward, I., Lomas, M., Delbart, N., Picard, G., 2004. Relating radar remote sensing of biomass to modelling of forest carbon budgets. *Climatic Change* 67, 379–402. doi:10.1007/s10584-004-3155-5.
- Lei, Y., Siqueira, P., 2014. Estimation of forest height using spaceborne repeat-pass l-band insar correlation magnitude over the us state of maine. *Remote Sensing* 6, 10252–10285. URL: <https://www.mdpi.com/2072-4292/6/11/10252>, doi:10.3390/rs61110252.
- Mazza, A., Sica, F., Rizzoli, P., Scarpa, G., 2019. Tandem-x forest mapping using convolutional neural networks. *Remote Sensing* 11. URL: <https://www.mdpi.com/2072-4292/11/24/2980>, doi:10.3390/rs11242980.
- McRoberts, R.E., Tomppo, E.O., 2007. Remote sensing support for national forest inventories. *Remote Sensing of Environment* 110, 412–419. doi:10.1016/j.rse.2006.09.034.
- Mermoz, S., Le Toan, T., Villard, L., Réjou-Méchain, M., Seifert-Granzin, J., 2014. Biomass assessment in the cameroon savanna using ALOS PALSAR data. *Remote sensing of environment* 155, 109–119.
- Mitchard, E.T., Saatchi, S.S., Woodhouse, I.H., Nangendo, G., Ribeiro, N., Williams, M., Ryan, C.M., Lewis, S.L., Feldpausch, T., Meir, P., 2009. Using satellite radar backscatter to

- predict above-ground woody biomass: A consistent relationship across four different african landscapes. *Geophysical Research Letters* 36.
- MITECO, 2020. Mapa forestal de españa (mfe) de máxima actualidad.
- Moersberger, H., Valdez, J., Martin, J.G.C., Junker, J., Georgieva, I., Bauer, S., Beja, P., Breeze, T.D., Fernandez, M., Fernández, N., Brotons, L., Jandt, U., Bruehlheide, H., Kissling, W.D., Langer, C., Lique, C., Lumbierres, M., Solheim, A.L., Maes, J., Morán-Ordóñez, A., Moreira, F., Pe'er, G., Santana, J., Shamoun-Baranes, J., Smets, B., Capinha, C., McCallum, I., Pereira, H.M., Bonn, A., 2024. Biodiversity monitoring in europe: User and policy needs. *Conservation Letters* 17, e13038. doi:<https://doi.org/10.1111/conl.13038>.
- Neumann, M., Moreno, A., Mues, V., Härkönen, S., Mura, M., Bouriaud, O., Lang, M., Achten, W.M., Thivolle-Cazat, A., Bronisz, K., Merganič, J., Decuyper, M., Alberdi, I., Astrup, R., Mohren, F., Hasenauer, H., 2016. Comparison of carbon estimation methods for european forests. *Forest Ecology and Management* 361, 397–420. doi:<https://doi.org/10.1016/j.foreco.2015.11.016>.
- Penman, J., Gytarsky, M., Hiraishi, T., Krug, T., Kruger, D., Pippatti, R., Buendia, L., Miwa, K., Ngara, T., Tanabe, K., Wagner, F. (Eds.), 2003. Good Practice Guidance for Land Use, Land-Use Change and Forestry. Institute for Global Environmental Strategies (IGES) for the Intergovernmental Panel on Climate Change (IPCC), Kanagawa, Japan.
- Persello, C., Wegner, J.D., Hänsch, R., Tuia, D., Ghamisi, P., Koeva, M., Camps-Valls, G., 2022. Deep learning and earth observation to support the sustainable development goals: Current approaches, open challenges, and future opportunities. *IEEE Geosci. Remote Sens. Mag.* 10, 172–200. doi:[10.1109/MGRS.2021.3136100](https://doi.org/10.1109/MGRS.2021.3136100).
- Praks, J., Antropov, O., Hallikainen, M.T., 2012. Lidar-aided sar interferometry studies in boreal forest: Scattering phase center and extinction coefficient at x- and l-band. *IEEE Transactions on Geoscience and Remote Sensing* 50, 3831–3843. doi:[10.1109/TGRS.2012.2185803](https://doi.org/10.1109/TGRS.2012.2185803).
- Pulliainen, J., Kurvonen, L., Hallikainen, M., 1999. Multi-temporal behavior of L- and C-band sar observations of boreal forests. *IEEE Transactions on Geoscience and Remote Sensing* 37, 927–937.
- Rauste, Y., 2005. Multi-temporal JERS SAR data in boreal forest biomass mapping. *Remote Sensing of Environment* 97, 263–275. doi:<https://doi.org/10.1016/j.rse.2005.05.002>.
- Rodríguez-Veiga, P., Quegan, S., Carreiras, J., Persson, H.J., Fransson, J.E., Hoscilo, A., Ziolkowski, D., Stereńczak, K., Lohberger, S., Stängel, M., Berninger, A., Siegert, F., Avitabile, V., Herold, M., Mermoz, S., Bouvet, A., Le Toan, T., Carvalhais, N., Santoro, M., Cartus, O., Rauste, Y., Mathieu, R., Asner, G.P., Thiel, C., Pathe, C., Schmullius, C., Seifert, F.M., Tansey, K., Balzter, H., 2019. Forest biomass retrieval approaches from earth observation in different biomes. *International Journal of Applied Earth Observation and Geoinformation* 77, 53 – 68. doi:[10.1016/j.jag.2018.12.008](https://doi.org/10.1016/j.jag.2018.12.008).
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. URL: <https://arxiv.org/abs/1505.04597>, arXiv:1505.04597.
- Rosen, P.A., Gurrola, E.M., Agram, P., Cohen, J., Laval, M., Riel, B.V., Fattahi, H., Aivazis, M.A., Simons, M., Buckley, S.M., 2018. The InSAR scientific computing environment 3.0: a flexible framework for NISAR operational and user-led science processing, in: IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, IEEE. pp. 4897–4900.
- Santoro, M., 2, L.E., Askne, J., and, C.S., 2006. Assessment of stand-wise stem volume retrieval in boreal forest from JERS-1 L-band SAR backscatter. *International Journal of Remote Sensing* 27, 3425–3454. doi:[10.1080/01431160600646037](https://doi.org/10.1080/01431160600646037).
- Santoro, M., Cartus, O., Antropov, O., Miettinen, J., 2024a. Estimation of forest growing stock volume with synthetic aperture radar: A comparison of model-fitting methods. *Remote Sensing* 16. URL: <https://www.mdpi.com/2072-4292/16/21/4079>, doi:[10.3390/rs16214079](https://doi.org/10.3390/rs16214079).
- Santoro, M., Cartus, O., Quegan, S., Kay, H., Lucas, R.M., Araza, A., Herold, M., Labrière, N., Chave, J., Åke Rosenqvist, Tadono, T., Kobayashi, K., Kelndorfer, J., Avitabile, V., Brown, H., Carreiras, J., Campbell, M.J., Cavlovic, J., da Conceição Bispo, P., Gilani, H., Khan, M.L., Kumar, A., Lewis, S.L., Liang, J., Mitchard, E.T., Pacheco-Pascagaza, A.M., Phillips, O.L., Ryan, C.M., Saikia, P., Schepaschenko, D., Sukhdeo, H., Verbeeck, H., Vieilledent, G., Wijaya, A., Willcock, S., Seifert, F.M., 2024b. Design and performance of the climate change initiative biomass global retrieval algorithm. *Science of Remote Sensing* 10, 100169. doi:<https://doi.org/10.1016/j.srs.2024.100169>.
- Scepanović, S., Antropov, O., Laurila, P., Rauste, Y., Ignatenko, V., Praks, J., 2021. Wide-area land cover mapping with sentinel-1 imagery using deep learning semantic segmentation models. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, 10357–10374. doi:[10.1109/JSTARS.2021.3116094](https://doi.org/10.1109/JSTARS.2021.3116094).
- Schmullius, C., Thiel, F., Pathe, C., Santoro, M., 2015. Radar time series for land cover and forest mapping, in: Kuenzer, C., Dech, S., Wagner, W. (Eds.), *Remote Sensing Time Series: Revealing Land Surface Dynamics*. Springer, Cham. volume 22 of *Remote Sensing and Digital Image Processing*, pp. 231–250. doi:[10.1007/978-3-319-15967-6-12](https://doi.org/10.1007/978-3-319-15967-6-12).

- Seppi, S.A., López-Martínez, C., Joseau, M.J., 2022. Assessment of L-Band SAOCOM InSAR coherence and its comparison with C-band: A case study over managed forests in Argentina. *Remote Sensing* 14. doi:10.3390/rs14225652.
- Shiroma, G.H., Lavalle, M., Buckley, S.M., 2022. An area-based projection algorithm for sar radiometric terrain correction and geocoding. *IEEE Transactions on Geoscience and Remote Sensing* 60, 1–23.
- Sica, F., Pulella, A., Nannini, M., Pinheiro, M., Rizzoli, P., 2019. Repeat-pass sar interferometry for land cover classification: A methodology using sentinel-1 short-time-series. *Remote Sensing of Environment* 232, 111277.
- Teijido-Murias, I., Antropov, O., López-Sánchez, C.A., Barrio-Anta, M., Miettinen, J., 2025. Forest height and volume mapping in northern Spain with multi-source earth observation data: Method and data comparison. *Forests* 16. URL: <https://www.mdpi.com/1999-4907/16/4/563>, doi:10.3390/f16040563.
- Telli, C., Lavalle, M., Pierdicca, N., 2025. Vegetation height from l-band sar backscatter and interferometric temporal coherence measurements. *Remote Sensing of Environment* 328, 114879. doi:10.1016/j.rse.2025.114879.
- Treuhaft, R.N., Siqueira, P.R., 2000. Vertical structure of vegetated land surfaces from interferometric and polarimetric radar. *Radio science* 35, 141–177.
- Ulaby, F.T., Long, D.G., 2014. *Microwave Radar and Radiometric Remote Sensing*. University of Michigan Press, Ann Arbor, MI.
- Zhang, J., Zhang, Y., Fan, W., He, L., Yu, Y., Mao, X., 2022. A modified two-steps three-stage inversion algorithm for forest height inversion using single-baseline l-band polinsar data. *Remote Sensing* 14. URL: <https://www.mdpi.com/2072-4292/14/9/1986>, doi:10.3390/rs14091986.
- Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2018. Unet++: A nested u-net architecture for medical image segmentation. *CoRR* abs/1807.10165. URL: <http://arxiv.org/abs/1807.10165>, arXiv:1807.10165.
- Zhu, X.X., Montazeri, S., Ali, M., Hua, Y., Wang, Y., Mou, L., Shi, Y., Xu, F., Bamler, R., 2021. Deep learning meets SAR: Concepts, models, pitfalls, and perspectives. *IEEE Geosci. Remote Sens. Mag.* 9, 143–172. doi:10.1109/MGRS.2020.3046356.
- Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Magazine* 5:4, 8–36. doi:10.1109/MGRS.2017.2762307.