International Journal of Applied Earth Observation and Geoinformation 153 (2026) 105527

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Deep learning models for estimating volume and Lorey's height across Nordic countries using optical and SAR satellite images

Zsofia Koma[a,*], Oleg Antropov[b], Oliver Cartus[c], Jukka Miettinen[b], Johannes Breidenbach[a]

[a] Norwegian Institute for Bioeconomy (NIBIO), Division of Forest and Forest Resources, National Forest Inventory, Høgskoleveien 7, 1433 Ås, Norway
[b] VTT Technical Research Centre of Finland, P.O. Box 1000, 02044 Espoo, Finland
[c] Gamma Remote Sensing, Worbstrasse 225, 3073 Gümligen, Switzerland

ARTICLE INFO

Keywords:
UNet
Deep learning
Forest resource mapping
Satellite imagery
SAR
Multisource

ABSTRACT

Spatially explicit information on forest resources and structure is essential for sustainable forest management and evidence-based policy-making. In the Nordic region, large-scale forest mapping often relies on integrating National Forest Inventory (NFI) field plots with airborne laser scanning (ALS) data. However, infrequent nationwide ALS campaign coverage limits their use for continuous monitoring. Satellite imagery, with its high temporal and spatial resolution, provides a promising alternative. We evaluate UNet-based deep learning models trained on wall-to-wall ALS-derived forest resource maps for predicting volume and Lorey's height in Norway using optical (Sentinel-2) and SAR (Sentinel-1, PALSAR-2) data. The UNet models, trained on both Finnish and Norwegian ALS maps, are benchmarked against extreme gradient boosting (XGB) models. Transfer learning is further explored by finetuning models using Norwegian NFI plots. Model accuracies are assessed using 541 reserved test NFI plots and 44 independent forest stands, representing high-volume mature boreal forests (>200 m^3 ha^{-1}). The UNet model trained on Norwegian ALS-based data achieved R2 values of 0.59 for both volume and Lorey's height when evaluated on NFI plots, and 0.70 and 0.59 for forest stands, respectively, outperforming the XGB models. Finetuning improved model transferability, yielding gains of up to 0.13 in R2 for volume and 0.46 for Lorey's height when adapting the Finnish model to Norwegian conditions. Utilizing SAR data alongside optical data enhanced model accuracy. Overall, our findings demonstrate the potential of UNet models trained on wall-to-wall ALS maps for forest resource mapping across Nordic countries.

1. Introduction

Quantifying the status and trends of forest resources and structure—such as timber volume and Lorey's height—is crucial for informing evidence-based policy and management decisions across local, national, and international levels (Tomppo et al., 2010; Vidal et al., 2016a). Sampling-based National Forest Inventories (NFIs) provide this information at national and regional scales, forming the foundation for strategic decision and policy-making regarding key ecosystem services, including bioenergy potential, biodiversity, and changes in carbon stock (Vidal et al., 2016b; Breidenbach et al., 2020). However, because NFIs are sampling-based surveys, their capacity to produce reliable estimates at local scales is restricted (Breidenbach and Astrup, 2012; Astrup et al., 2019). To bridge this gap, NFI field plots are often combined with remotely sensed data to model relationships between field-observed forest attributes and remotely sensedpredictor variables, enabling mapping between NFI plot locations (McRoberts and Tomppo, 2007; Kangas et al., 2018). This approach aligns with the requirements of several international policy frameworks—including the EU Forest Strategy for 2030, the EU Biodiversity Strategy, and the EU Deforestation Regulation—which emphasize the need for spatially explicit forest monitoring at both national and local scales. Consequently, integrating remotely sensed data has become essential for producing detailed forest resource maps and information.

Some of the earliest national forest maps were based on a combination of NFI and Landsat data (Gjertsen, 2007; McRoberts and Tomppo, 2007). In many countries, national forest resource maps are today based on integrating field plots from NFIs with country-wide Airborne Laser Scanning (ALS) data (Kangas et al., 2018). ALS provides detailed 3D information on forest structure, enabling accurate estimation of key

* Corresponding author at: Zsofia Koma, Norwegian Institute of Bioeconomy Research, Høgskoleveien 7, 1433 Ås, Norway.
E-mail address: zsofia.koma@nibio.no (Z. Koma).

https://doi.org/10.1016/j.jag.2026.105527
Received 25 October 2025; Received in revised form 24 July 2026; Accepted 8 August 2026
Available online 20 August 2026

attributes, such as timber volume and Lorey's height. However, national ALS campaigns lack the temporal frequency needed for effective forest monitoring. For example, Norway's nationwide ALS campaign required approximately ten years to achieve full coverage (Hauglin et al., 2021), compared to a seven-year timeframe in Sweden (Nilsson et al., 2017). In Finland, the future ALS campaigns are planned to be conducted in nine-year cycles (NLS, 2025). In contrast, Earth Observation (EO) datasets—such as optical imagery from Sentinel-2 (S2) and Synthetic Aperture Radar (SAR) data from Sentinel-1 (S1) or ALOS-2 PALSAR-2 (P2) —offer both high temporal and spatial coverage. Several studies have utilized optical remote sensing for forest resource mapping, such as S2 or LANDSAT (Astola et al., 2019; Miettinen et al., 2021) or have combined EO with ALS data (Esteban et al., 2019; Astola et al., 2021). SAR data with clear sensitivity to forest structure have also been used for forest mapping, primarily at stand-level or coarser mapping units and in wide-area studies (Santoro et al., 2011; Antropov et al., 2017; Ge et al., 2023b). Recent studies have shown that the combination of optical and SAR data improves the quantification of forest structural characteristics (Wittke et al., 2019; Persson et al., 2021; Ge et al., 2022; Ge et al., 2023a), especially if SAR operates in lower frequencies (e.g., L-band used on board ALOS-2), which opens new possibilities for improving the accuracy of maps for estimating forest attributes.

EO-based forest resource maps have been created using statistical methods such as linear regression, k-nearest neighbours (kNN), and ensemble-based machine learning (ML) techniques e.g. extreme gradient boosting (XGB). Studies using only S2 data with these approaches have reported RMSEs ranging from 27.2% to 59.3% for estimates of volume and Lorey's height in boreal forest contexts (Astola et al., 2019; Miettinen et al., 2021; Persson et al., 2021). When using SAR data alone, RMSEs for volume range from 32% to 54.3% (Wittke et al., 2019; Ge et al., 2023b; Persson et al., 2021). One study systematically tested the integration of optical and SAR datasets using kNN method and reported improved accuracies in Sweden, with volume RMSE reduced from 37.9% to 30.2% in boreal forests (Persson et al., 2021). Recent advances in deep learning (DL) show promise for improving EO-based forest resource mapping. Ge et al. (2023a) demonstrated that the prediction of Lorey's height using S2 data alone, as well as combined S2 and SAR data, improved by 0.3% to 1.6% RMSE when applying the UNet DL method compared to traditional approaches such as kNN, random forest, or linear regression. Similarly, Astola et al. (2021) reported RMSE improvements of 7.9% for volume and 2.5% for height when using DL models compared to reference methods. Nevertheless, DL remains underutilized in regression tasks for EO-based forest attribute estimation, especially when compared to its more widespread application in classification tasks such as disturbance mapping (Yuan et al., 2020).

One of the key challenges in applying DL to forest resource mapping is the limited availability and quality of suitable training data (Ball et al., 2017; Li et al., 2019). NFI datasets are often considered weak labels in the context of DL because they lack spatially continuous target information (e.g. within image patches) due to a relatively small plot size. To address this challenge, transfer learning has emerged as a promising strategy in various remote sensing tasks by pretraining models on datasets where spatially explicit information is available at a larger spatial scale (Wurm et al., 2019; Zhou et al., 2023; Kuzu et al., 2024). However, only a few studies have been used in forest resource mapping (Astola et al., 2021; Ge et al., 2023a). Instead of training a model solely on data from the target region, transfer learning allows the use of a pretrained model—developed on a related task —which can then be finetuned using a smaller set of local training data. This approach leverages the generalizable features learned during pretraining, enabling effective model adaptation even with limited labelled data. In addition, ALS-based forest resource maps can provide wall-to-wall training data that complement or substitute for traditional plot-based labels (Ge et al., 2022, 2023a). For example, Ge et al. (2023a) demonstrated the successful training of models in Finnish Lapland using ALS-derived maps, along with optical and SAR satellite imagery, followed by fine-tuning with plot-level measurement data in southern Finland. To further evaluate the effectiveness of this approach, applying transfer learning to neighbouring countries—such as Norway—provides an ideal test case.

In this study, we develop and evaluate UNet-based DL models to estimate timber volume and Lorey's height using optical (S2) and SAR (S1 and P2) satellite imagery. S1 and S2 were selected for their 10 m spatial resolution and high revisit frequency, enabling fine-scale and temporally consistent forest monitoring. P2 L-band SAR was additionally included to complement S1C-band data by providing deeper canopy penetration. The selection of EO products aimed to include datasets widely available to the EO community. Although ALS-based forest resource maps provide high accuracy, their limited temporal availability and acquisition costs restrict their direct use for temporal forest monitoring. To address this gap, we utilize ALS-based forest resource maps as wall-to-wall training data to support EO-based UNet modelling of timber volume and Lorey's height. We further analyze transfer learning by finetuning ALS-based pretrained models using NFI plots, including transferring a Finnish model to Norwegian conditions and exploring additional options to further improve the Norwegian model. Model accuracy is assessed using field-observed data based on reserved test NFI plots and independent forest stands. All UNet models are compared to XGB and kNN models trained on Norwegian NFI field plots. The XGB and kNN models served as widely used baselines that incorporated only spectral and polarization information. Although ALS-based forest resource maps are used during UNet pretraining, the predictive objective of the UNet model is to approximate field-measured volume and Lorey's height. Our research questions are: (1) Does proposed ALS-pretrained and NFI-finetuned UNet framework outperform established plot-based machine-learning baselines such as XGB and kNN? (2) How can transfer learning, where NFI field plots are incorporated in the model, improve the UNet accuracy, especially in the case of applying the Finnish model to Norwegian conditions? (3) Can the inclusion of SAR imagery-based predictors, such as Sentinel-1C-band and PALSAR-2 L-band backscatter, enhance the model's prediction accuracy?

2. Materials and methods

The overall workflow developed for this study (Fig. 1) consists of four main processing steps: (a) XGB modelling; (b) pretraining UNet models using ALS-based forest resource maps from Finland and Norway; (c) finetuning UNet models from Finland and Norway for the Norwegian study site; and (d) assessing the accuracy associated with each model type (XGB, pretrained UNet, and finetuned UNet). Additionally, we also carried out kNN modelling, which is described in detail in the Appendix (A.3.).

2.1. Study area

The study area comprises one Finnish and one Norwegian study site (Fig. 2). The Finnish study site encompasses central and southern Finland (Fig. 2) and has relatively small elevation differences, with the highest points generally less than 300 m above sea level. The forest in the reference areas has around 40% pine, more than 30% spruce, and the rest deciduous trees. The average forest volume across the study sites was 120 m^3 ha^{-1}, with an average height of 14 m.

The Norwegian study site covers approximately 210,000 km^2 of land, encompassing the South-Eastern, Central, Southern, and Western regions, extending from the county of Trøndelag southward (Fig. 2). It is located in the boreal climate zone and contains most of the productive forest of Norway. The boreal forests are dominated by three main species: Norway spruce, Scots pine, and deciduous trees, primarily birch, which are often mixed with coniferous species. The study site has diverse topographical variation.

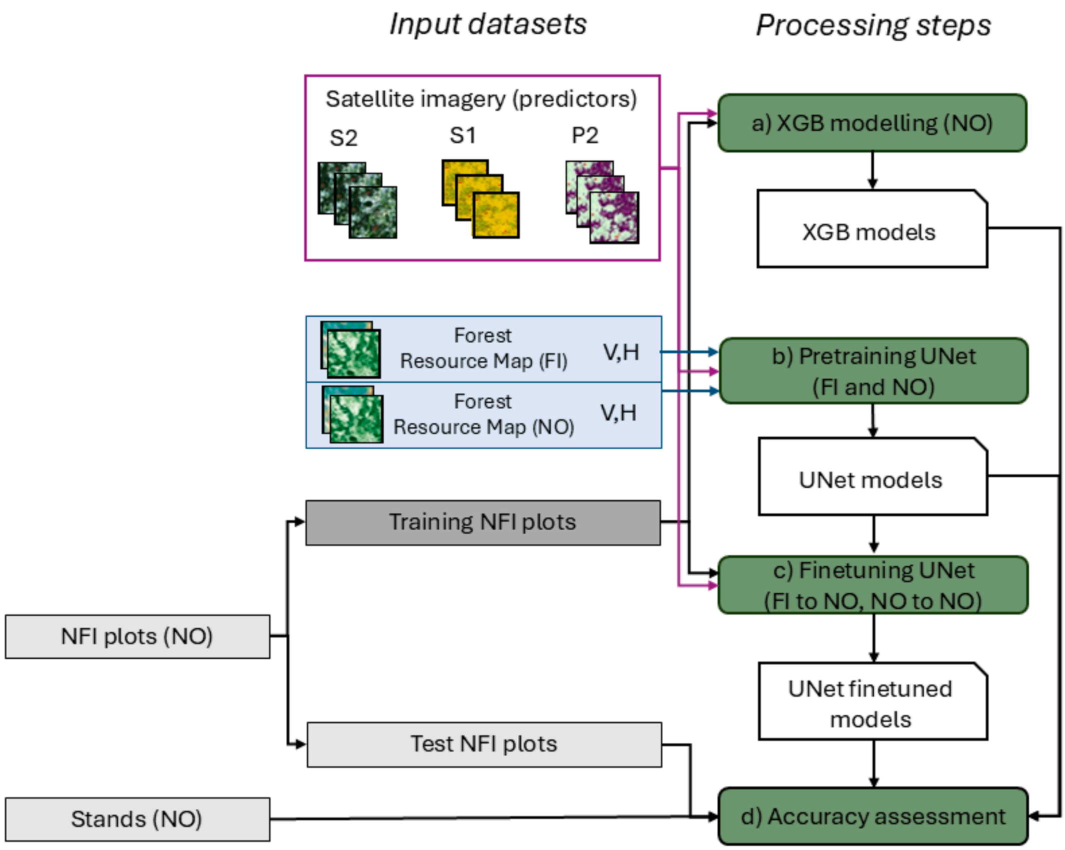

Fig. 1. Overview of the workflow used for assessing the accuracy of different models (XGB, UNet, finetuned UNet). The input datasets are marked with rectangles (satellite imagery, ALS-based forest resource maps, Norwegian NFI plots, and Norwegian validation stands), with dark grey indicating input for training the models. The intermediate models (XGB, UNet, and finetuned UNet) are marked with a snipped rectangle. The main processing steps are shown with green rounded rectangles (a-d corresponding to the main text). V is volume, and H is Lorey's height, S2 is Sentinel-2 optical images, S1 is Sentinel-1 SAR images, and P2 is PALSAR-2 SAR images. FI = Finland and NO = Norway.

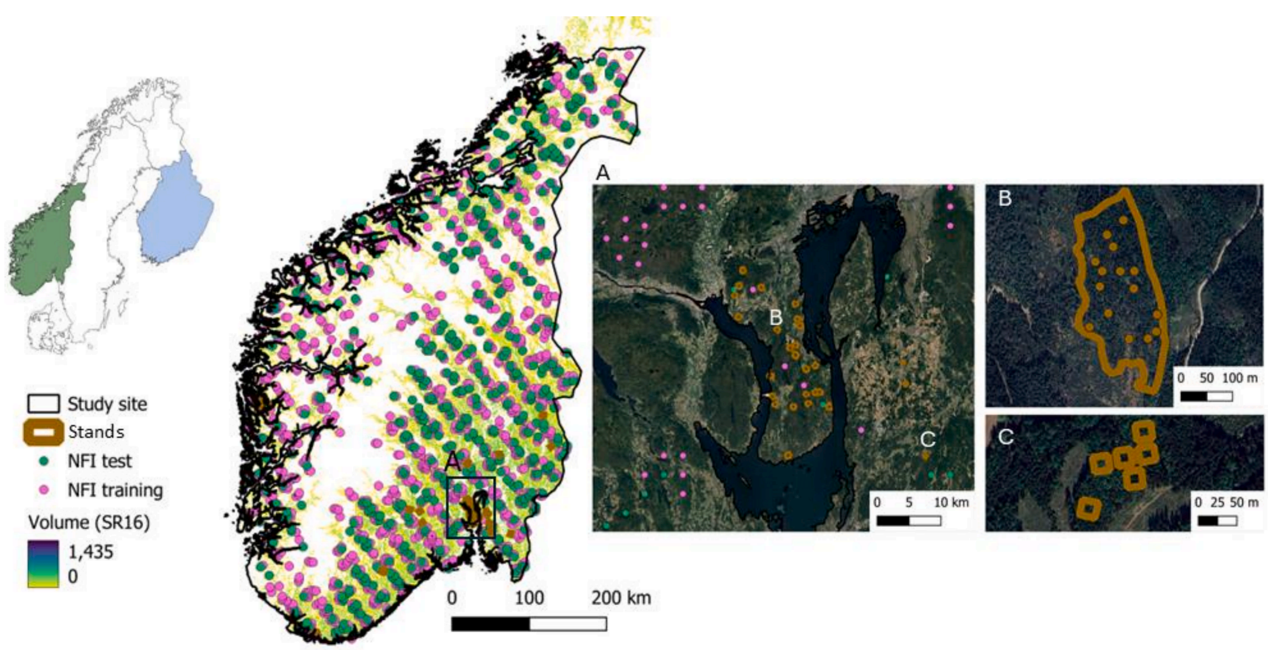

Fig. 2. The Norwegian study area (highlighted in green on the Nordic map) is where the different model types—XGB, UNet, and UNet finetuned —were evaluated. Green dots represent the NFI plots used for testing, while pink coloured dots indicate the training NFI plots used for the XGB model and UNet finetuning. Brown areas mark the independent forest stands. Panel A shows a zoomed-in view of the study area, with panels B and C highlighting two example stands used in the accuracy assessment. The background maps in panel A are the ALS-based timber volume map from Kilden and Google satellite imagery.

2.2. Datasets

2.2.1. Satellite imagery data and pre-processing

ESA Copernicus S2 satellites operating with the Multi-Spectral Instrument (MSI) provide optical imagery data in 13 spectral bands at a 2–3-day revisit frequency in the Nordic countries, covering the visible, near-infrared, and short-wave infrared parts of the electromagnetic spectrum. In this study, we used the Level-2 surface reflectance products provided by ESA in the visible and near-infrared bands (B2, B3, B4, B5, B8) and the short-wave infrared bands (B11, B12). The 20 m resolution bands (B5, B11, B12) were resampled to 10 m resolution using bilinear interpolation method to match the resolution of the other bands (10 m). The S2 data were further processed with the Terramonitor compositing script (Miettinen et al., 2021), producing growing season composites by weighting based on cloudiness, haze, and shadows. The compositing process also re-turned a quality flag used to filter cloud-affected or low-quality pixels.

In this study, we have used two types of SAR datasets. The first consisted of C-band backscatter imagery acquired by Sentinel-1A in Interferometric Wide (IW) swath mode, using VV and VH polarisation. The images in Ground-Range-Detected (GRD) format were pre-processed using the GAMMA Remote Sensing software, resulting in geocoded,radiometric and terrain-corrected g0 backscatter images. The backscatter images were compensated for the NESZ (noise equivalent sigma zero) and the terrain-flattening was done as in Frey et al. (2013) using the 1-arcsec Copernicus digital elevation model. The S1 data acquired over a year (30 to 31 images) were then used to calculate the annual mean backscatter per polarization. Maps depicting layover or shadow areas were applied to mask them. Mosaic channels were resampled to 10 m spacing to match S2 imagery.

The second type of SAR dataset was based on the Japan Aerospace Exploration Agency (JAXA), which releases free,open annual mosaics of the L-band data acquired by the Phased-Array type L-band SAR (P2) onboard the Advanced Land Observing Satellite-2 (ALOS-2). The backscatter mosaics, with a resolution of $\sim 25 \times 25$ m2, are produced from imagery acquired in Fine Beam Dual Polarization mode (HH and HV) using stripmap mode under JAXA's global basic observation scenario. The images used to generate the mosaics were radiometric, terrain-corrected, and geocoded to a geographic spatial reference system, as with Shimada and Ohtaki (2010). In this study, the HH and HV mosaics were resampled to a 10 m grid using bilinear interpolation and masked for layover and shadow using the layover/shadow maps produced by JAXA to harmonize all input datasets for UNet.

All satellite imagery datasets were processed for the study areas in Finland and Norway. In Finland, the datasets were processed for the year 2020, and in Norway for 2021, to align as closely as possible with the timing of the ALS-based forest resource maps.

2.2.2. Finnish reference datasets

The Finnish UNet model was trained based on the wall-to-wall Finnish ALS-based forest resource map produced by the Finnish Forest Centre using ALS acquisitions from 2020 and associated field inventory plots as training for kNN-based approaches (Maltamo and Packalen, 2014). The map has been produced at 16 m spatial resolution, and resampled using bilinear interpolation to align with the satellite imagery datasets and harmonize all input datasets.

2.2.3. Norwegian reference datasets

Different field observation datasets and remote sensing data products were used in the Norwegian study site (Fig. 2, Table 1). First, the Norwegian NFI is a continuously operating inventory based on a permanent sample grid, with each field plot remeasured every five years (Breidenbach et al., 2020). The NFI field measurements are carried out within circular 250 m^2 plots where the diameter at breast height (dbh) and species of all trees with a dbh $>=$ 5 cm are recorded, and the height of ten trees is measured to calculate timber volume and Lorey's height. The NFI plots represent average conditions across Norway, including all three main species groups: spruce, pine, and broadleaf species. Only NFI plots measured in 2021 within a valid remote sensing based forest mask were included in the study, resulting in 1566 NFI plots. 541 NFI plots were reserved for testing and not used for XGB or DL modelling, and the remaining 1025 NFI plots were used in model training (Table 1).

Additional stand-level field measurements (Fig. 2, Table 1), independent of the NFI, were used for validation (Koma and Breidenbach, 2025). The dataset includes two subsets: (a) mature stands in Asker (eastern region) and Alver (western region), where 10–15 plots per stand were sampled following NFI protocols. Mean timber volume and Lorey's height were calculated to characterise each stand. (b) Long-term field trials assessing forest management methods, consisting of 1–30 adjacent rectangular stands. Within each stand, dbh and species were recorded for all trees, and height was measured for every fourth tree. The stands represent high-volume, mature, late-stage development where only two species are present: spruce and pine. These stands are particularly important for forest management. Stand-level timber volume and Lorey's height were calculated and aggregated to trial-level averages. After applying a remote sensing-based forest mask and excluding cloud-covered stands from 2021, 44 validation stands covering 62 ha were available for accuracy assessment.

Third, the Norwegian Forest Resource Map (SR16) was used as a wall-to-wall training dataset for the UNet model. These spatially continuous maps provide estimates of timber volume, Lorey's height, and other forest attributes at a 16 m $\times$ 16 m resolution and are operationally updated each year following the workflow described by Hauglin et al. (2021). They are produced by combining NFI plot data with predictor variables derived from nationwide ALS acquisitions conducted by the Norwegian Mapping Agency. Species-specific univariate linear mixed-effects regression models are used to account for

Table 1

Summary statistics of training NFI, testing NFI, and validation stand datasets. The table includes the number of observations (N), minimum (Min), maximum (Max), mean, and standard deviation (SD) for each variable (Volume, Lorey's height, area of the plot, and the stands).

Dataset	Variable	Unit	N	Min	Max	Mean	SD
Training NFI plots	Volume	m^3ha^{-1}	1025	0.18	1157.59	118.13	127.99
	Lorey's height	m		2.40	28.18	11.96	5.11
	Area	m^2		250			
Testing NFI plots	Volume	m^3ha^{-1}	541	0.30	779.70	125.91	118.66
	Lorey's height	m		3.40	28.00	12.38	4.66
	Area	m^2		250			
Validation stands	Volume	m^3ha^{-1}	44	100.62	906.70	369.14	206.17
	Lorey's height	m		12.24	26.55	18.58	4.11
	Area	ha		0.11	4.32	1.40	1.06

systematic differences across ALS projects and regions. To ensure consistency, NFI plot values are projected to a common reference date, and a remote sensing-based forest mask is applied. The accuracy of SR16 maps is 40.5% RMSE and 1.64% bias for volume and 11.5% RMSE and 0.03% bias for Lorey's height, with corresponding R2 values of 0.86 and 0.87, respectively (Table S1). The full independence of the SR16-based pretraining labels from the NFI plots cannot be ensured, as SR16 was generated using all available NFI plots, including those from 2021 used in this study. However, additional analyses suggest that the impact of this overlap is negligible (Appendix A1). The resolution of the forest resource maps was aligned with the S2 satellite imagery datasets by resampling using bilinear interpolation, from the native spatial resolution of 16 × 16 m to 10 × 10 m to harmonize the data products for UNet.

2.3. Modelling

2.3.1. Extreme gradient boost modeling

The XGB method has been chosen as the baseline machine learning method for comparison with the UNet method. While traditional algorithms such as kNN have been widely used in forest inventory applications, recent studies have shown that gradient-boosting methods reach performance levels comparable to deep learning models (Grinsztajn et al., 2022, Björnberg et al., 2026). We have additionally conducted kNN modelling, which can be found in the Appendix (A.3.).

The modelling data tables containing the predictor variables (S2 and S2S1P2) were prepared by spatially intersecting NFI plot locations with the corresponding remote sensing datasets. For each plot, all pixels overlapping the 250 m^2 plot area were considered, and their contributions were weighted by the fraction of pixel area falling within the plot. Predictor values were extracted for both the training dataset (1025 NFI plots) and the reserved test dataset (541 NFI plots).

XGB regression was applied using the R package xgboost for each predictor configuration (S2 only and S2S1P2) and for both target variables (volume and Lorey's height). Model hyperparameters were optimized using k-fold cross-validation with early stopping, evaluating a randomized subset of candidate parameter combinations (learning rate, tree depth, subsampling, and L1/L2 regularization) and selecting the configuration that minimized cross-validated RMSE. Hyperparameters were determined separately for each predictor combination and target variable, resulting in shallow to moderate tree depths (max_depth = 5–7), low learning rates (η = 0.01–0.10), and explicit L1/L2 regularization. The final models were refit using the full training dataset (1025 NFI plots), and predictions were generated for the reserved test NFI plots (541 plots) and for independent forest stands used in the accuracy assessment.

2.3.2. UNet model

We employed a vanilla UNet deep learning architecture (Ronneberger et al., 2015) adapted for pixel-level regression of forest attributes. While UNet has itself a convolutional neural network (CNN) architecture, its encoder—decoder design with skip connections enables the utilization of spatial information, such as neighbourhood texture and spatial patterns, thereby exploiting information present in EO imagery that is not typically captured by simpler CNN models or pixel-based ML methods. This makes it well-suited for pixel-level regression tasks in EO-based forestry (Ge et al., 2022). The vanilla UNet architecture was selected as a balance between model capacity and robustness, given the limited amount of NFI data available for model finetuning within the transfer learning approach (Ge et al., 2023a).

We used a UNet encoder-decoder architecture (Fig. 3B) with an initial double-convolution block followed by three downsampling blocks (each starting with 2 × 2 max-pooling) and three upsampling blocks (using 2 × 2 transposed convolutions), connected via skip connections. Starting from an input patch size of 256 × 256, successive downsampling reduces the spatial resolution to 128 × 128, 64 × 64, and 32 × 32, while the decoder restores the resolution stepwise back to 64 × 64, 128 × 128, and 256 × 256. Each convolutional block contains two 3 × 3 convolutions with ReLU activation. The number of feature channels increases as 64 → 128 → 256 → 512 in the encoder and is mirrored in the decoder (512 → 256 → 128 → 64). A final 1 × 1 convolution projects the features to a single regression output channel per target variable.

Model training used a masked Mean Squared Error (MSE) loss function to ensure that only pixels with valid reference data contributed to the loss computation. During pretraining, non-forest pixels were excluded using a forest mask. During finetuning with NFI data, valid target values were available only for pixels intersecting NFI plots, while all remaining pixels within each patch were ignored. This way, ignored pixels did not influence the loss magnitude or gradient updates.

The Adam optimiser was used throughout training. A ReduceLROnPlateau scheduler reduced the learning rate by a factor of two when validation loss plateaued for ten epochs. Overfitting was monitored using masked MSE on an internal validation set, and model selection was based on the minimum validation loss. Training and validation losses were tracked throughout optimisation and used for learning-rate scheduling and model selection.

The same architecture, preprocessing steps, and loss formulation were applied across all predictor combinations (S2 and S2S1P2). The further exact parameter settings used during pretraining and finetuning are described in Section 2.3.4. No extensive hyperparameter optimisation was performed; instead, commonly used UNet training settings were adopted and kept fixed across all experiments to enable controlled comparisons among predictor configurations.

2.3.3. UNet model input

EO input data were prepared as fixed-size image patches of 2.56 × 2.56 km2 (256 × 256 pixels) prior to model training (Fig. 3). During pretraining, the Finnish and Norwegian wall-to-wall ALS-based forest resource maps, described in Sections 2.1.2–2.1.3, were used as reference datasets. The pretraining dataset comprised 1,068 square patches from Finland and 1,025 patches from Norway. In Finland, the patches were randomly distributed across the forested areas covered by the ALS-based forest resource maps. In Norway, the ALS-based patches used in pretraining were placed in proximity (<3 km) to the coordinates of the training NFI plots. During finetuning, the same 1,025 image patch locations were reused. Specifically, only the 10 × 10 m pixels intersecting the NFI field plots were assigned valid target values (i.e. observed volume and Lorey's height), with the plot-level measurements assigned uniformly to all intersecting pixels, including those with only partial overlap, while all remaining pixels were masked out during loss computation.

The pre-processed predictor variables (S2 and S2S1P2) were extracted for the selected image patches. In all image patches, non-forest pixels were masked. All EO patches were further normalised on a per-band basis using z-score normalization, whereby the mean of each band was subtracted and the result divided by its standard deviation. For independent validation, EO image patches related to the reserved test NFI were prepared using the same patch size. The test set consists of 541 field plots that were reserved at the beginning of the modelling process and not used during training. Additional patches were prepared for validation stands.

2.3.4. UNet model pretraining and finetuning

The UNet architecture did not use a pretrained backbone; instead, all weights were trained from scratch during pretraining using ALS–based forest resource maps and subsequently reused for finetuning. During UNet model pretraining, slightly different data-splitting strategies were applied for Finland and Norway. For Finland, the full dataset was randomly split into training (70%), internal validation (15%), and internal test (15%) sets, with 10 random splits. As no independent Finnish NFI plots were available for external model evaluation, the 15% test patches were used for internal model assessment. In total, pretraining in Finland used 744 training, 160 internal validation, and 160 internal test

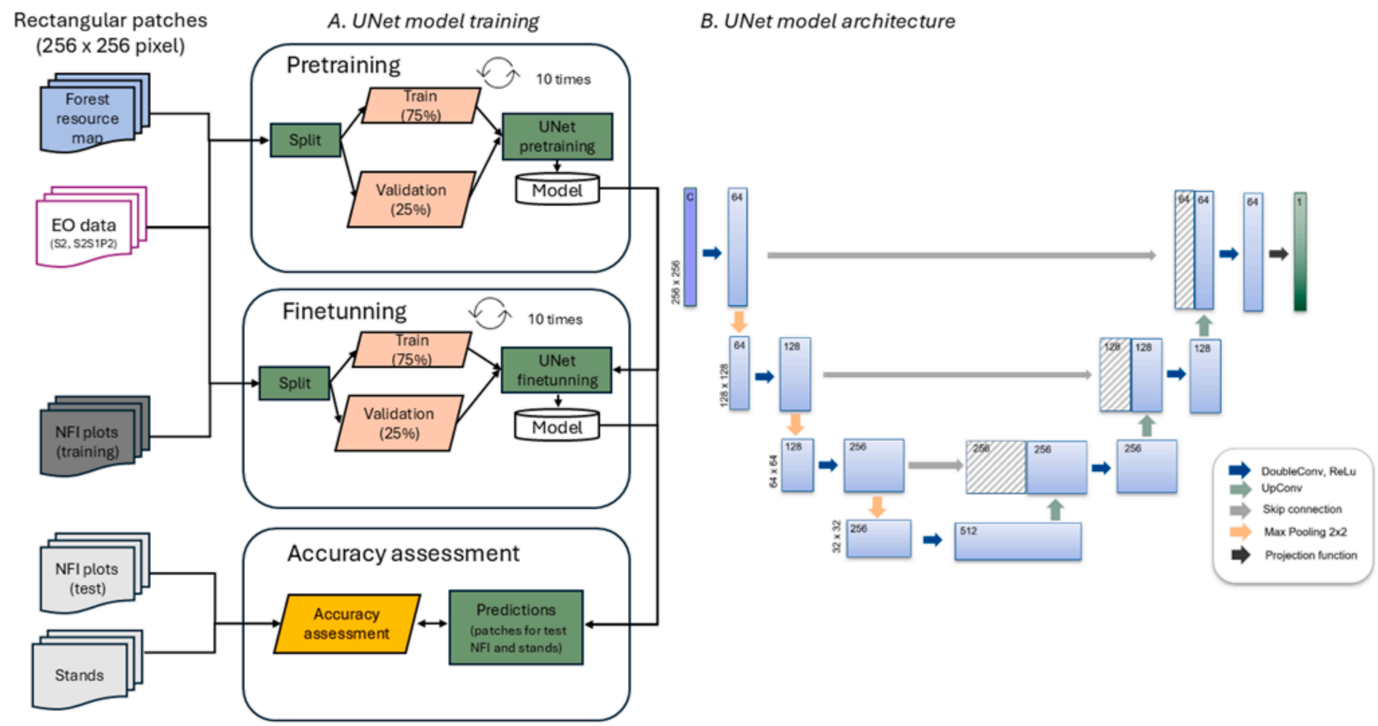

Fig. 3. Overview of the UNet modelling workflow for Norway. The left side (A) comprises the UNet training set-up for pre-training and finetunning. Independent NFI test plots were reserved before model training and used only for final accuracy assessment. The right side (B) shows the basic vanilla UNet architecture employed in this study. The C in the input block refers to the input EO data bands (S2 and S2S1P2).

patches, each covering 2.56 × 2.56 km2. For Norway (Fig. 3), the dataset was randomly split 10 times into 75% training (769 patches) and 25% internal validation (256 patches), as independent NFI plots were reserved prior UNet training. For both models, the initial learning rate was set to 1×10^{-4}. The model was trained for 200 epochs with a mild weight decay of 1×10^{-5}. The best-performing model was selected based on the lowest masked MSE-based validation loss, determined using the internal validation data, and the corresponding weights were saved for subsequent finetuning for each target variable. Training was conducted with a batch size of 16. The final model was used to predict EO patches at test NFI plot locations and validation stands in Norway.

For UNet finetuning, the pretrained models from Finland and Norway were used to initialize the network weights. The finetuning dataset, consisting of patches centered on training NFI plot locations, was randomly split into 75% for training (769 patches), 25% for internal validation (256 patches), and this procedure was repeated 10 times. During finetuning, all encoder layers of the pretrained UNet were frozen (including convolutional and downsampling blocks), while the decoder, skip-connection layers, and final output convolution were kept trainable and updated. As a result, only a subset of the network parameters was updated during finetuning. Freezing the encoder retains pretrained feature representations, while updating the decoder enables adaptation to sparse NFI data and reduces overfitting. The initial learning rate was set to 1×10^{-6}, the number of epochs to 30, and the batch size to 16. The best-performing finetuned model, based on masked MSE-based validation loss, was used to predict volume and Lorey's height values for the NFI and stand-level testing datasets.

2.4. Accuracy assessment

Accuracy assessment was carried out using the reserved test NFI plotsalong with independent forest stands. Model predictions are produced at the pixel level and subsequently aggregated to the NFI plot and stand level by computing an area-weighted mean of all pixels intersecting each unit, accounting for fractional pixel coverage. The accuracy was evaluated separately for the test NFI plot and for validation stands using weighted root mean square error (RMSE) and weighted bias, which was calculated as

$$\{RMSE, bias\} = \left[\frac{1}{n}\sum_{i=1}^{n} w_i(\widehat{y}_i - y_i)^{\varphi}\right]^{\omega}$$

with $\varphi = 2$ and $\omega = 0.5$ for RMSE and $\varphi = 1$ and $\omega = 1$ for bias, where y_i and $\widehat{y}_i$ represent the observed and estimated forest attributes at the stand level, and $\boldsymbol{w_i}$ is the area of each NFI plot and validation stand; $i = 1,\ldots,n$. Both RMSE and bias were additionally normalized by the area-weighted mean of the observed values and reported as percentages.

We further calculated the weighted coefficient of determination (R^2) as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} w_i(\widehat{y}_i - y_i)^2}{\sum_{i=1}^{n} w_i(y_i - \overline{y}_i)^2}$$

where $\overline{y}_i$ is the mean of the observed values.

3. Results

Prediction accuracies were assessed for the different types of models (XGB, UNet, finetuned UNet) using only optical and optical and SAR-based imagery data as predictors. Furthermore, the UNet model was trained on Finnish and Norwegian forest resource maps, and the effect of finetuning was evaluated. The results are presented separately for timber volume and Lorey's height.

3.1. Timber volume

The UNet models pretrained on the Norwegian forest resource map, including both the base and finetuned variants—consistently outperformed XGB (Table 2, Figs. 4 and 5, Table S3) and kNN models (Table S5). The RMSE was reduced by up to 13.3% for NFI plots. For the

Table 2
Overall accuracy assessment for timber volume comparing XGB, UNet, and finetuned UNet models trained on Finnish (FI) and Norwegian (NO) data. Results are shown for NFI plots, forest stands, and two types of predictor variable combinations (Pred. type): S2 (Sentinel-2) and S2S1P2 (Sentinel-2 with Sentinel-1 and PALSAR-2). Finetuned models were pretrained on ALS-based forest resource maps from Finland (FI) or Norway (NO) and further finetuned using Norwegian training NFI plots. Area-weighted RMSE and bias are reported as percentages, normalized by the area-weighted mean of the observed values. Bold formatting indicates the best-performing model within groups that share the same test data and predictor configuration.

Test data type	Pred. type	Training data	XGB	UNet	Fine- tuned UNet	XGB	UNet	Fine- tuned UNet	XGB	UNet	Fine- tuned UNet
			RMSE [%]			BIAS [%]			R^2		
NFI	S2	FI		74.5	72.6		−16.2	−7.9		0.37	0.41
NFI	S2	NO	75.1	64.4	**64.1**	6.5	5.1	**3.4**	0.36	0.53	**0.54**
NFI	S2S1P2	FI		71.8	71.7		−4.1	−2.5		0.42	0.42
NFI	S2S1P2	NO	74.0	61.3	**60.7**	8.1	1.9	**−0.6**	0.38	0.58	**0.59**
Stand	S2	FI		62.0	57.7		−38.1	−32.2		0.04	0.17
Stand	S2	NO	40.4	**37.4**	37.5	−15.8	**−7.4**	−8.2	0.59	**0.65**	0.65
Stand	S2S1P2	FI		55.7	55.1		−24.2	−23.1		0.23	0.25
Stand	S2S1P2	NO	38.6	**34.8**	35.9	−11.8	**−3.3**	−6.6	0.63	**0.70**	0.68

independent forest stands, UNet outperformed XGB, reducing the RMSE by up to 3.8%. The UNet models consistently reduced bias up to 8.7% for NFI plots and 15.1% for stands. In contrast, the UNet model based on the Finnish ALS-based forest resource map showed only marginal or no improvements over XGB. Regarding bias, the Finnish UNet achieved a reduction of up to 5.6% in NFI plots for S2S1P2 predictors, while XGB showed a reduction of up to 16.4% in bias for stands.

Finetuning effects were most beneficial for UNet models using the Finnish ALS-based forest resource map (Table 2, Fig. 4). For these models, improvements ranged from 0.1% to 4.3% in RMSE, 0.02% to 0.13% in R^2, and 1.1% to 8.3% in bias. In contrast, finetuning the Norwegian forest resource map-based models (Table 2, Fig. 5) yielded limited gains for NFI plots, with improvements of up to 0.6% in RMSE, 0.01 in R^2, and 1.7% in bias. At the stand level, finetuning the Norwegian forest resource map-based model often led to reduced model performance, with RMSE increasing by 0.1% to 1.1%, R^2 decreasing by up to 0.02, and underestimation (bias) increasing by 0.8% to 3.3%.

Combining optical and SAR predictors generally enhanced UNet performance compared to using S2 data alone (Table 2, Fig. 4). For the best-performing UNet model, switching from S2 to S2S1P2 predictors reduced RMSE by 3.4% for NFI plots and 2.6% for stands, while R^2 increased by 0.05 for both validation types. In terms of bias, compared to the S2-only model, the S2S1P2 model showed improvements ranging from 3.2% to 4.0% for NFI plots and 1.6% to 4.1% for stands. For XGB, the impact of adding SAR data was aligned with the UNet results: accuracy was enhanced, reducing RMSE by up to 1.8%, and increasing R^2 by 0.04. The bias is slightly reduced for stands with 4% but higher with 1.6% in the case of NFI plots.

3.2. Lorey's height

Similarly to volume, UNet models outperformed kNN and XGB in most cases (Table 3, Figs. 5 and 6, Tables S4 and S5). RMSE improvements for best performing UNet models compared to XGB ranged from 4.3% to 5.1% for NFI plots and 3.3% to 3.7% for stands, and R^2 increases by 0.17 and 0.19 for NFI plots and by 0.23 to 0.26 for stands. Bias was improved for XGB models with 0.2% in the case of NFI plots but best-performing UNet models reduced bias by up 7.5% to for stands.

The effects of finetuning followed a similar overall pattern to those observed for volume. RMSE improvements (Table 3, Fig. 6) ranged from 0.1% to 7.5%, while R^2 changes varied from 0.01 to 0.46. The most substantial gains were achieved by finetuning the pretrained Finnish model with Norwegian NFI plot data, where R^2 increased from 0.03 to 0.38 and from –0.17 to 0.29, for NFI plots and stands, respectively, improving poor model performance to moderate. However, for the UNet model pretrained in Norway, the RMSE improvements in NFI plots were small (0.1%), the stand-level results showed less consistency, including instances of worsening performance with finetuning—evidenced by slightly increased RMSE and decreased R^2 (Table 3, Fig. 6).

Similar to volume, combining optical and SAR predictors generally improved Norwegian-pretrained UNet performance for Lorey's height estimation on NFI plots (Table 3, Fig. 6), reducing RMSE by 1.6% and increasing R2 by 0.06. For stand-level predictions, RMSE decreased up to 2.4%, and R2 increased by up to 0.15 by adding SAR predictors. For the Finnish pretrained model, the S2-only model shows improved or comparable performance to the S2S1P2 model for NFI plots and stands (Table 3). Regarding bias, S2-only predictors yielded up to 3.6% improvement for NFI plots, while the S2S1P2 combination for the Norwegian pretrained model reduced stand-level bias by up to 2.5%. For XGB, differences between S2 and S2S1P2 were minimal: RMSE changed by up to 0.8% for NFI plots, while RMSE decreased by 0.6% for stands. Bias was slightly increased by 0.7% for NFI plots using combined optical and SAR predictors and reduced by 0.2% for stands.

4. Discussion

We evaluated the accuracies of kNN, XGB, UNet, and finetuned UNet models for estimating forest volume and Lorey's height across a large area of Norway, totalling 76 ha. This study provides an evaluation of S2 and SAR imagery using yearly composites for large-scale forest resource mapping in Norway, where earlier efforts have relied mainly on Landsat data (Gjertsen, 2007), and current operational mapping of volume and Lorey's height is based exclusively on ALS predictors (Hauglin et al., 2021, Koma and Breidenbach, 2025). The best overall accuracies ranged from 34.8 to 60.7% RMSE for volume and 15.6 to 24.1% RMSE for Lorey's height in the case of two different validation datasets. These results were comparable with other studies using only satellite-based imagery datasets from a single time in the boreal forest region, RMSE ranging from 27.2 to 59.3% for volume and 15.4 to 33.5% for Lorey's height (Wittke et al., 2019; Persson et al., 2021; Miettinen et al., 2021; Ge et al., 2022, 2023a). The accuracy in this study was evaluated using two different types of field-observed datasets. Reserved test NFI plots, which represent average boreal forest conditions across Norway, and validation stands correspond to high-volume, mature, late development stages, which are particularly important for forest management.

The UNet models trained on wall-to-wall forest resource maps—particularly those derived from Norwegian ALS data—outperformed both XGB and kNN— underscoring the advantages of applying them in UNet-based DL methods for satellite-based forest attribute mapping (Tables 2 and 3 and S5). The effectiveness of ALS-based forest resource maps as training data for DL has been confirmed by a few previous studies (Becker et al., 2023; Ge et al., 2022, 2023a). It has been further demonstrated that using a sorrounding area around the plot (e.g. 3x3 window) with a DL method improved accuracy compared to a pixel-wise approach, supporting our finding that spatial contextual information (neighborhood) enhances DL training (Astola et al., 2021, Björnberg et al., 2026). As illustrated in Fig. 5, UNet predictions trained using 2.5 km × 2.5 km image patches, thereby exploiting a larger

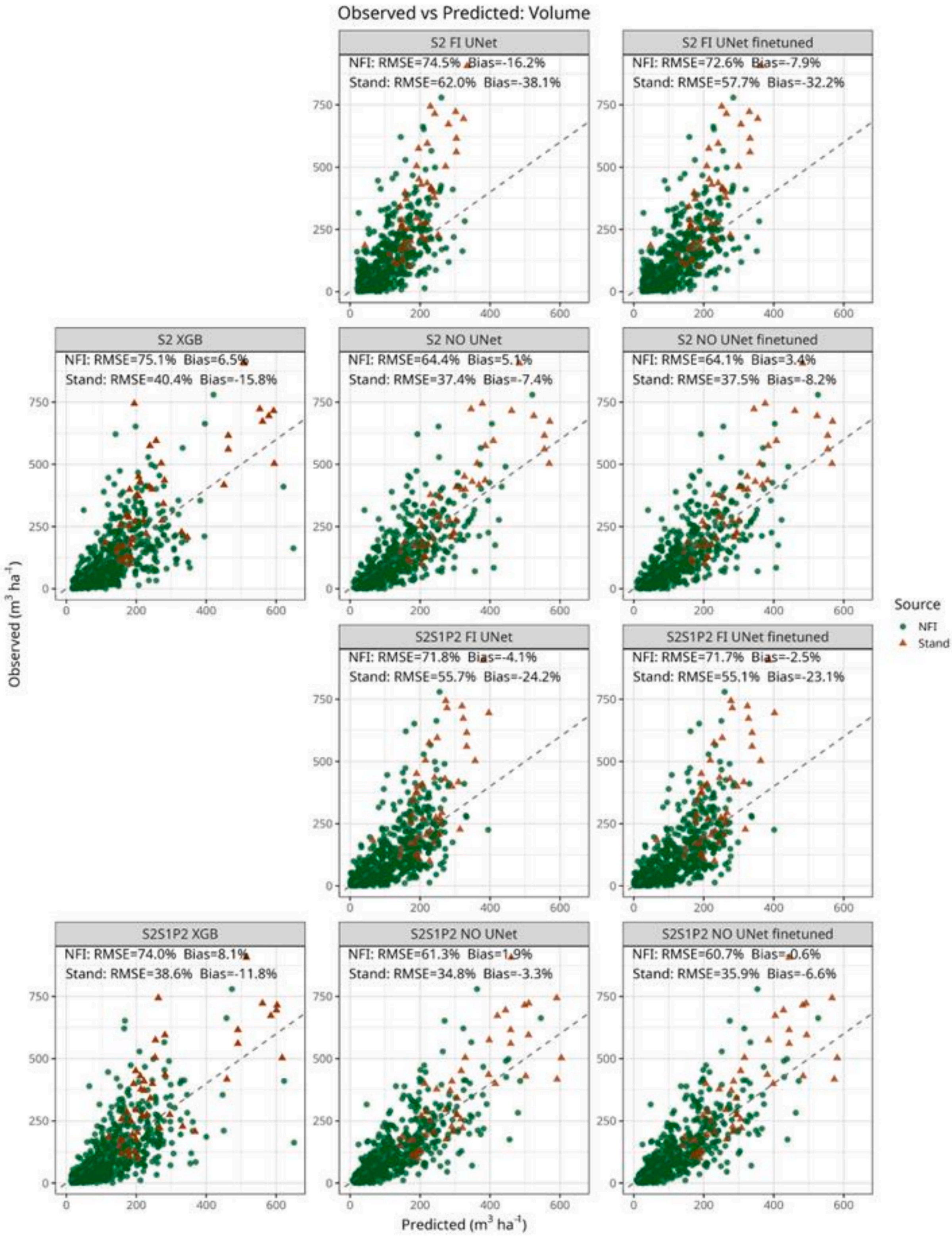

Fig. 4. Observed vs. predicted timber volume for different predictor types: S2 (Sentinel-2 only) and S2S1P2 (Sentinel-2 with Sentinel-1 and PALSAR-2). Models include XGB, UNet, and finetuned UNet trained on Finnish (FI) or Norwegian (NO) datasets. Finetuned models were first pretrained on ALS-based forest resource maps (FI or NO) and then updated using Norwegian NFI training plots. Area-weighted RMSE and bias are reported as percentages, normalized by the area-weighted mean of the observed values.

spatial neighborhood, exhibit spatially coherent but smoother patterns. In some cases, this learning strategy enables a more robust representation of forest structure than XGB predictions, particularly for stands with volumes below 400 m^3 ha^{-1}. However, for the highest-volume stands (>400 m^3 ha^{-1}), this smoothing can reduce local variability and may lead to underestimation of timber volume compared to the XGB method. One way to sharpen UNet predictions is to incorporate higher-resolution satellite inputs (e.g., PlanetScope) to better preserve fine-scale

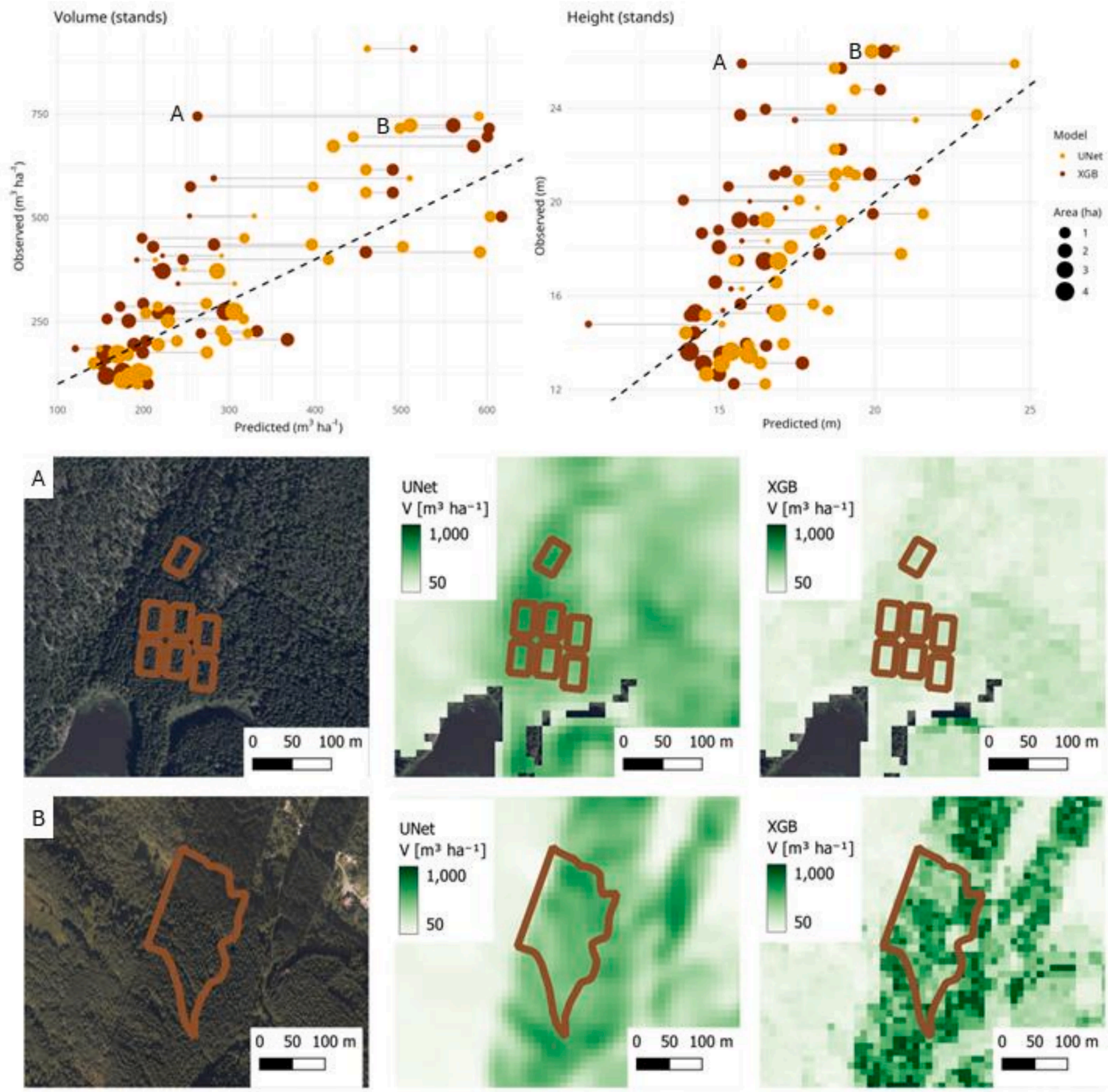

Fig. 5. Comparison of the UNet and XGB models for volume and Lorey's height estimation across the 44 independent forest stands. The upper panel shows cross-plots of predicted versus observed values for volume and Lorey's height, with paired predictions from the two models connected by lines to illustrate model-specific differences at the stand level. The lower panels present spatial prediction maps for volume only. Panels A and B highlight two selected high-volume examples, where A illustrates better performance of UNet and B illustrates better performance of XGB. For both sites, stand boundaries are shown in brown, and high-resolution orthophotos are included as background for spatial context.

variability, or to combine convolutional feature extraction with downstream models (e.g., using learned embeddings as inputs to XGB).

Similar patterns in achieved accuracies have been reported in other DL-based boreal forest studies. For example, Ge et al. (2022) demonstrated improved accuracy in Lorey's height prediction in Finland, with RMSE decreasing from 30.1% to 24.1% at the plot level and from 19.7% to 15.4% at the stand level. Astola et al. (2021) reported reductions in RMSE for both volume (50.3% to 42.4%) and height (30.7% to 28.2%) using deep learning. In our study, the improvements were of comparable magnitude, with RMSE of volume decreasing from 74.0% to 60.7% and Lorey's height from 29.2% to 24.1% at the plot level. While Ge et al. (2022) evaluated stand-level accuracy only for Lorey's height, our results show that the advantage of UNet over XGB also extends to volume.

In the terms of bias, UNet reduced overestimation of volume for low-volume NFI plots (<200 m3 ha^{-}1), consistent with previous studies (Astola et al., 2021; Ge et al., 2022, Ge et al., 2023a), although a slight increase in bias was observed for Lorey's height compared to XGB. At the stand level, UNet was generally less biased than XGB, particularly in lower volume stands (<400 m3 ha^{-}1 in Fig. 5.). It is important to note that the ALS-based map used as training is a modelled product itself with its own error structure and potential biases, which can propagate in the UNet models. Overall, the accuracy comparison between XGB and UNet should be interpreted as a practical use case, since the amount and type of training data differ between the ML and DL methods. Additionally, stratified analysis did not reveal clear or systematic differences in model accuracy across slope, region, species, or forest development classes

Table 3
Overall accuracy assessment for Lorey's height comparing XGB, UNet, and finetuned UNet models trained on Finnish (FI) and Norwegian (NO) data. Results are shown for NFI plots, validation stands, and two types of predictor variable combinations (Pred. type): S2 (Sentinel-2) and S2S1P2 (Sentinel-2 with Sentinel-1 and PALSAR-2). Finetuned models were pretrained on ALS-based forest resource maps from Finland (FI) or Norway (NO) and finetuned further using Norwegian training NFI plots. Area-weighted RMSE and BIAS are reported as percentages, normalized by the area-weighted mean of the observed values. Bold formatting indicates the best-performing model within groups that share the same test data and predictor configuration.

Test data type	Pred. type	Training data	XGB	UNet	Fine- Tuned UNet	XGB	UNet	Fine- Tuned UNet	XGB	UNet	Fine- Tuned UNet
			RMSE [%]			BIAS [%]			R^2		
NFI	S2	FI		30.2	29.6		2.5	−1.7		0.36	0.38
NFI	S2	NO	30.0	25.7	**25.6**	**1.1**	1.9	1.3	0.37	0.53	**0.54**
NFI	S2S1P2	FI		37.1	29.6		6.1	3.5		0.03	0.38
NFI	S2S1P2	NO	29.2	**24.1**	**24.1**	**1.8**	2.0	2.4	0.40	**0.59**	**0.59**
Stand	S2	FI		**16.6**	17.0		**0.8**	−3.2		**0.48**	0.46
Stand	S2	NO	19.9	17.9	18.0	−8.3	−2.7	−3.4	0.25	0.40	0.39
Stand	S2S1P2	FI		24.9	19.4		15.7	3.3		−0.17	0.29
Stand	S2S1P2	NO	19.3	15.7	**15.6**	−8.1	−1.1	**−0.9**	0.28	0.54	**0.54**

(Appendix A5).

Our study also examined the geographical transferability of pretrained models from Finland to Norwegian conditions. Cross-country transfers are rarely explored; most previous research has focused on transfers within Finland (Astola et al., 2021; Ge et al., 2023a), while Schumacher et al. (2026) evaluated local and central models for volume estimation across Norway, Sweden, Finland and Italy based on satellite time series data and CNNs.. In our case, Finnish pretrained models applied without finetunning achieved only poor to moderate performance (Tables 2 and 3), with particularly low accuracy for Lorey's height, including negative R^2 values when SAR predictors were used. These results indicate substantial domain shifts between Finland and Norway, arising from differences in phenology, satellite acquisition timing (2020 in Finland versus 2021 in Norway), cloud conditions, viewing geometry, NFI plot definitions, ALS-based forest resource mapping practices, and higher topographic variability in Norway. Such differences are likely to affect both the transferability of learned representations and the consistency of the predictor, particularly for SAR-based backscatter–structure relationships. Our results show that Finnish pretrained UNet models, when finetuned with Norwegian NFI plots, achieved accuracies comparable to XGB models in most cases (Tables 2 and 3). This is consistent with Astola et al. (2021), who found that transfer learning yielded equal or better performance, and with Ge et al. (2023a), who showed that finetuning improved accuracy relative to models applied without adaptation. Finetuning helps mitigate the domain shift by adapting the model to local conditions while preserving transferable spatial and structural patterns learned during pretraining, which also explains the generally reduced bias observed in the finetuned models. Nevertheless, for stand-level volume estimation using both S2 and S2S1P2 predictors, overall accuracies after fine-tuning Finnish pretrained models remained lower than those of XGB (Table 2), indicating that transfer learning does not fully compensate for all domain-shift effects. At the same time, models trained directly on Norwegian data achieved higher accuracies than finetuned Finnish models, consistent with results from other DL-based studies (Kattenborn et al., 2021, Schumacher et al., 2026). By explicitly evaluating performance, including high-volume mature forest stands (>200 m^3 ha^{-1}), we further found that finetuning could, in some cases, slightly underperform models without finetuning, highlighting a limitation when applied to regions where the model was not trained. Nevertheless, the finetuning results also showed that in most cases bias was reduced, whereas models without finetunning often exhibited higher bias, which aligns with Ge et al., (2023a). Future work should focus on exploring alternative UNet DL architectures, including different attention mechanisms and loss functions, to improve the incorporation of NFI information during fine-tuning, as well as on expanding and evaluating model transferability beyond the boreal forest region.

Our results indicate that combining optical and SAR predictors generally improved model performance compared to optical-only inputs for models trained on Norwegian data (Tables 2 and 3). Similar trends have been reported in other boreal forest studies. For example, Persson et al. (2021) demonstrated that including SAR data reduced the RMSE for volume estimates in Sweden from 37.2% to 30.2%. Likewise, Ge et al. (2022, 2023a) reported improvements in Lorey's height prediction in Finland, with RMSE decreasing from 30.3% to 18.1%. In our study, the magnitude of improvement was smaller: the maximum reduction in RMSE was from 62.0% to 55.7% for volume and from 18.0% to 15.6% for Lorey's height. Additionally, the inclusion of SAR has reduced overall accuracy in Finnish pretrained Lorey's height models (Table 3). Two main factors can explain these results. First, Norway's more complex topography compared to Sweden or Finland can influence the SAR backscatter signal. Second, Persson et al. (2021) used TanDEM-X data, which offers a finer spatial resolution (10 m), a more frequent acquisition cycle (11-day repeat cycle), and, most importantly, interferometric coherence that is inherently sensitive to the vertical structure of forests (Olesk et al., 2016, Telli et al., 2025). In our study, the L-band P2 data, which offer deeper penetration than S1 C-band, are constrained by their relatively coarse spatial resolution (25 m) and are provided by an openly available annual mosaic product, a single image from the year, which limits sensitivity to height variations. Additionally, SAR information was incorporated through a composite of polarizations rather than interferometric variables. Consequently, SAR contributed only modestly to overall prediction accuracy (Appendix A.4), with no clear advantage of L-band SAR data, whereas multitemporal composites from S1 data showed a slightly more consistent benefit. Future work should therefore explore SAR data with greater penetration depth, multitemporal composites, finer spatial resolution, and interferometric information to better exploit SAR sensitivity to forest vertical structure, particularly for improving the transferability of DL models across different regions.

5. Conclusion

Our study demonstrates the potential of using spatially explicit ALS-based forest resource maps as training data for UNet-based DL models. The highest accuracies were obtained when pretraining the UNet model on Norwegian forest resource maps, resulting in an overall accuracy that surpassed those of both XGB and kNN baseline machine-learning models. In addition, a model pretrained on Finnish data and finetuned with Norwegian NFI plots achieved performance in most cases comparable to that of XGB models, trained exclusively on Norwegian data. Overall, this approach can be applied to produce spatially explicit forest structure information that supports model-assisted estimation frameworks to complement sampling-based NFIs, with relevance for carbon stock reporting, forest management, and policy-related applications. With the development of more comprehensive foundation models trained on multi-country datasets, this approach may enable robust, large-scale, cross-country forest resource mapping.

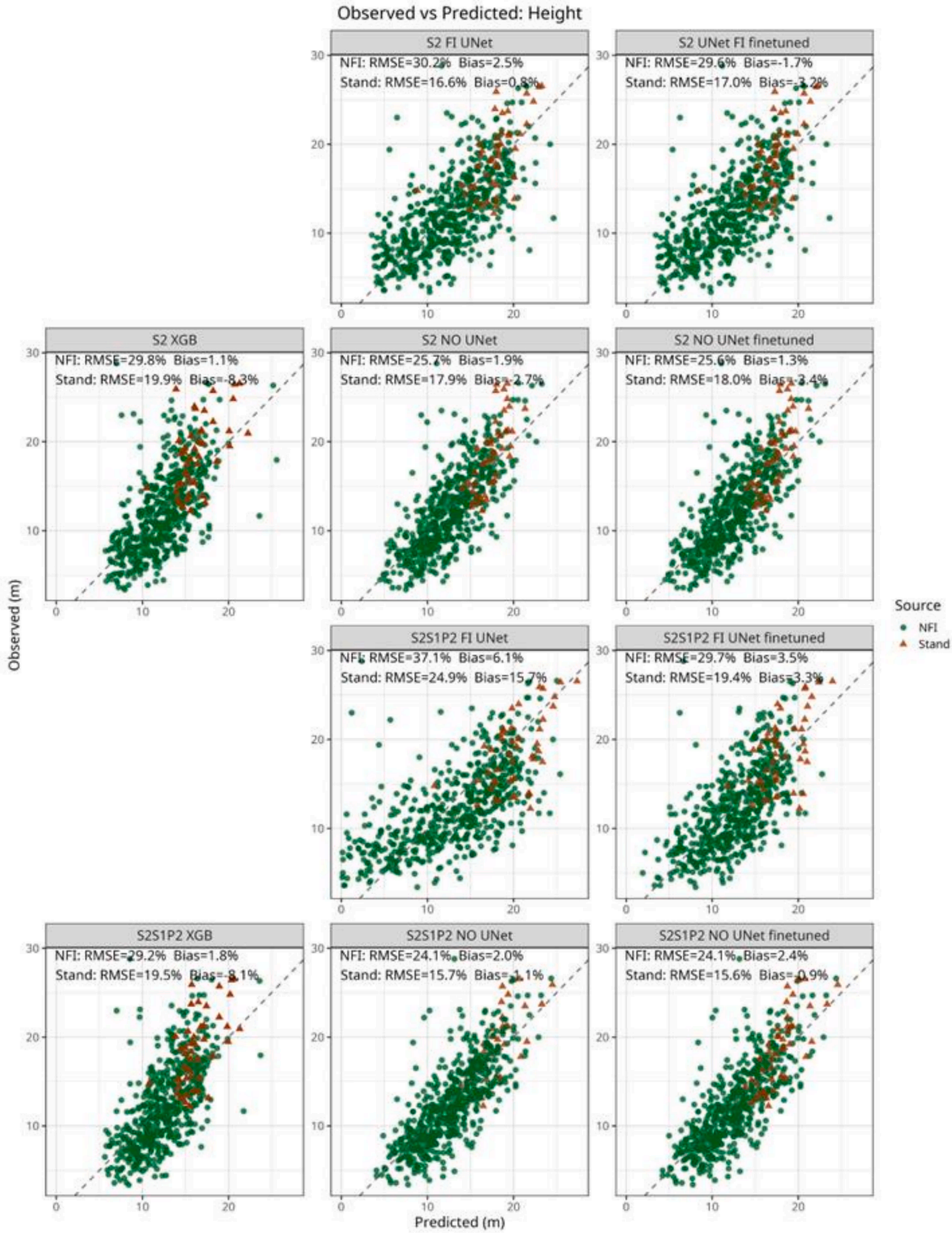

Fig. 6. Observed vs. predicted Lorey's height for different predictor types: S2 (Sentinel-2 only) and S2S1P2 (Sentinel-2 with Sentinel-1 and PALSAR-2). Models include XGB, UNet, and finetuned UNet trained on Finnish (FI) or Norwegian (NO) datasets. Finetuned models were first pretrained on ALS-based forest resource maps (FI or NO) and then updated using Norwegian NFI training plots. Area-weighted RMSE and BIAS are reported as percentages, normalized by the area-weighted mean of the observed values.

Author contributions

ZK: Conceptualization, Data curation, Formal analysis, Validation, Writing original draft, OA: Conceptualization, Data curation, Formal analysis, Methodology, Writing review and editing, OC: Data curation, Writing review and editing, JM: Formal analysis, Project administration, Funding acquisition, Writing review and editing, JB: Project administration, Funding acquisition, Writing review and editing.

CRediT authorship contribution statement

Zsofia Koma: Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Oleg Antropov:** Writing – review & editing, Methodology, Formal analysis, Data curation, Conceptualization. **Oliver Cartus:** Writing – review & editing, Data curation. **Jukka Miettinen:** Writing – review & editing, Project administration, Funding acquisition, Formal analysis. **Johannes Breidenbach:** Writing – review & editing, Project administration, Funding acquisition.

Declaration of competing interest

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

Acknowledgments

This research was funded by the European Space Agency (ESA) project on Forest Carbon Monitoring under contract number 4000135015/21/I-NB—Forest Carbon Monitoring, and by the EU under GA101056907 (PathFinder). The satellite data processing was supported by the ESA Network of Resources Initiative. Microsoft Copilot assisted in enhancing the clarity and correctness of the English text.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2026.105527.

Data availability

field observation datasets used in this study contain confidential information, including field plot locations, and therefore cannot be made publicly available. The processed datasets and trained model weights required to reproduce the analyses have been deposited in Zenodo (https://doi.org/10.5281/zenodo.21511637). The used preprocessing workflow and modelling methods used in the study are available within the Forestry-TEP platform (https://f-tep.com/) developed by VTT.

References

Antropov, O., Rauste, Y., Häme, T., Praks, J., 2017. Polarimetric ALOS PALSAR time series in mapping biomass of boreal forests. Remote Sens. (Basel) 9, 999. https://doi.org/10.3390/rs9100999.

Astola, H., Häme, T., Sirro, L., Molinier, M., Kilpi, J., 2019. Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. Remote Sens. Environ. 223, 257–273. https://doi.org/10.1016/j.rse.2019.01.019.

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. (Basel) 13, 2392. https://doi.org/10.3390/rs13122392.

Astrup, R., Rahlf, J., Bjørkelo, K., Debella-Gilo, M., Gjertsen, A.-K., Breidenbach, J., 2019. Forest information at multiple scales: development, evaluation and application of the Norwegian forest resources map SR16. Scand. J. For. Res. 34, 484–496. https://doi.org/10.1080/02827581.2019.1588989.

Ball, J.E., Anderson, D.T., Chan Sr, C.S., 2017. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. J. Appl. Remote Sens. 11, 042609. https://doi.org/10.1117/1.JRS.11.042609.

Becker, A., Russo, S., Puliti, S., Lang, N., Schindler, K., Wegner, J.D., 2023. Country-wide retrieval of forest structure from optical and SAR satellite imagery with deep ensembles. ISPRS J. Photogramm. Remote Sens. 195, 269–286. https://doi.org/10.1016/j.isprsjprs.2022.11.011.

Björnberg, D., Ericsson, M., Lindeberg, J., Löwe, W., Nordqvist, J., Wallerman, J., Fransson, J.E.S., 2026. Improving national forest attribute maps of Sweden with machine learning. Sci. Remote Sens. 13, 100395. https://doi.org/10.1016/j.srs.2026.100395.

Breidenbach, J., Astrup, R., 2012. Small area estimation of forest attributes in the Norwegian National Forest Inventory. Eur. J. For. Res. 131, 1255–1267. https://doi.org/10.1007/s10342-012-0596-7.

Breidenbach, J., Granhus, A., Hylen, G., Eriksen, R., Astrup, R., 2020. A century of National Forest Inventory in Norway – informing past, present, and future decisions. For. Ecosyst. 7, 46. https://doi.org/10.1186/s40663-020-00261-0.

Esteban, J., McRoberts, R.E., Fernández-Landa, A., Tomé, J.L., Næsset, E., 2019. Estimating forest volume and biomass and their changes using random forests and remotely sensed data. Remote Sens. (Basel) 11, 1944. https://doi.org/10.3390/rs11161944.

Frey, O., Santoro, M., Werner, C.L., Wegmüller, U., 2013. DEM-based SAR pixel-area estimation for enhanced geocoding refinement and radiometric normalization. IEEE Geosci. Remote Sens. Lett. 10, 48–52. https://doi.org/10.1109/LGRS.2012.2192093.

Ge, S., Antropov, O., Häme, T., McRoberts, R.E., Miettinen, J., 2023a. Deep learning model transfer in forest mapping using multi-source satellite SAR and optical images. Remote Sens. (Basel) 15, 5152. https://doi.org/10.3390/rs15215152.

Ge, S., Gu, H., Su, W., Praks, J., Antropov, O., 2022. Improved semisupervised UNet deep learning model for forest height mapping with satellite SAR and optical data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 5776–5787. https://doi.org/10.1109/JSTARS.2022.3188201.

Ge, S., Tomppo, E., Rauste, Y., McRoberts, R.E., Praks, J., Gu, H., Su, W., Antropov, O., 2023b. Sentinel-1 time series for predicting growing stock volume of boreal forest: multitemporal analysis and feature selection. Remote Sens. (Basel) 15, 3489. https://doi.org/10.3390/rs15143489.

Gjertsen, A.K., 2007. Accuracy of forest mapping based on Landsat TM data and a kNN-based method. Remote Sens. Environ. 110, 420–430. https://doi.org/10.1016/j.rse.2006.08.018.

Grinsztajn, L., Oyallon, E., Varoquaux, G., 2022. Why do tree-based models still outperform deep learning on typical tabular data? Adv. Neural Inf. Proces. Syst. 35, 507–520.

Hauglin, M., Rahlf, J., Schumacher, J., Astrup, R., Breidenbach, J., 2021. Large scale mapping of forest attributes using heterogeneous sets of airborne laser scanning and National Forest Inventory data. For. Ecosyst. 8, 65. https://doi.org/10.1186/s40663-021-00338-4.

Kangas, A., et al., 2018. Remote sensing and forest inventories in Nordic countries – roadmap for the future. Scand. J. For. Res. 33, 397–412. https://doi.org/10.1080/02827581.2017.1416666.

Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS J. Photogramm. Remote Sens. 173, 24–49. https://doi.org/10.1016/j.isprsjprs.2020.12.010.

Koma Z, Breidenbach J. 2025. Large-scale validation of forest attribute maps across different spatial resolutions. Silva Fennica **59** : Article 24061. DOI: 10.14214/sf.24061.

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 J. Sel. Top. Appl. Earth Obs. Remote Sens. 17, 4751–4767. https://doi.org/10.1109/JSTARS.2024.3361183.

Li, J., Huang, X., Gong, J., 2019. Deep neural network for remote-sensing image interpretation: status and perspectives. Natl. Sci. Rev. 6, 1082–1086. https://doi.org/10.1093/nsr/nwz058.

Maltamo M, Packalen P. 2014. Species-Specific Management Inventory in Finland. In: Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies, In: Maltamo M, Næsset E, Vauhkonen J (eds). Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies. Springer Netherlands: Dordrecht. 241–252. DOI: 10.1007/978-94-017-8663-8_12.

McRoberts, R.E., Tomppo, E.O., 2007. Remote sensing support for national forest inventories. Remote Sens. Environ. 110, 412–419. https://doi.org/10.1016/j.rse.2006.09.034.

Miettinen, J., et al., 2021. Demonstration of large area forest volume and primary production estimation approach based on Sentinel-2 imagery and process based ecosystem modelling. Int. J. Remote Sens. 42, 9467–9489. https://doi.org/10.1080/01431161.2021.1998715.

Nilsson, M., et al., 2017. A nationwide forest attribute map of Sweden predicted using airborne laser scanning data and field data from the National Forest Inventory. Remote Sens. Environ. 194, 447–454. https://doi.org/10.1016/j.rse.2016.10.022.

NLS. 2025. National Land Survey (NLS) Finland. 2025. Kansalliset ilmakuvaus- ja laserkeilausohjelmat saavat jatkoa – luvassa entistä tarkempaa kolmiulotteista tietoa Suomesta. Available at: https://www.maanmittauslaitos.fi/ajankohtaista/kansallis et-ilmakuvaus-ja-laserkeilausohjelmat-saavat-jatkoa-luvassa-entista (accessed 19 Sep 2025). [In Finnish and Swedish].

Olesk, A., Praks, J., Antropov, O., Zalite, K., Arumäe, T., Voormansik, K., 2016. Interferometric SAR coherence models for characterization of hemiboreal forests using TanDEM-X data. Remote Sens. (Basel) 8, 700. https://doi.org/10.3390/rs8090700.

Persson, H.J., Jonzén, J., Nilsson, M., 2021. Combining TanDEM-X and Sentinel-2 for large-area species-wise prediction of forest biomass and volume. Int. J. Appl. Earth Obs. Geoinf. 96, 102275. https://doi.org/10.1016/j.jag.2020.102275.

Ronneberger O, Fischer P, Brox T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi A, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, vol. 9351. Cham: Springer. pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

Santoro, M., Beer, C., Cartus, O., Schmullius, C., Shvidenko, A., McCallum, I., Wegmüller, U., Wiesmann, A., 2011. Retrieval of growing stock volume in boreal forest using hyper-temporal series of Envisat ASAR ScanSAR backscatter measurements. Remote Sens. Environ. 115, 490–507. https://doi.org/10.1016/j.rse.2010.09.018.

Schumacher, J., Cescatti, A., Chirici, G., D'Amico, G., Francini, S., Hertzler, J., Mehtätalo, L., Nabuurs, G.-J., Nilsson, M., Pitkänen, J., Breidenbach, J., 2026. Federated learning in forest resource modelling and monitoring: bridging data confidentiality and collaborative research. Int. J. Appl. Earth Obs. Geoinf. 152, 105452. https://doi.org/10.1016/j.jag.2026.105452.

Shimada, M., Ohtaki, T., 2010. Generating large-scale high-quality SAR mosaic datasets: application to PALSAR data for global monitoring. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 3, 637–656. https://doi.org/10.1109/JSTARS.2010.2077619.

Telli, C., Lavalle, M., Pierdicca, N., 2025. Vegetation height from L-band SAR backscatter and interferometric temporal coherence measurements. Remote Sens. Environ. 328, 114879. https://doi.org/10.1016/j.rse.2025.114879.

Tomppo E, Gschwantner T, Lawrence M, McRoberts RE (eds). 2010. National Forest Inventories: Pathways for Common Reporting. Springer Netherlands: Dordrecht. DOI: 10.1007/978-90-481-3233-1.

Vidal, C., Alberdi, I., Redmond, J., Vestman, M., Lanz, A., Schadauer, K., 2016a. The role of European National Forest Inventories for international forestry reporting. Ann. For. Sci. 73, 793–806. https://doi.org/10.1007/s13595-016-0545-6.

Vidal C, Alberdi IA, Hernández Mateo L, Redmond JJ (eds). 2016b. National Forest Inventories. Springer International Publishing: Cham. DOI: 10.1007/978-3-319-44015-6.

Wittke, S., Yu, X., Karjalainen, M., Hyyppä, J., Puttonen, E., 2019. Comparison of two-dimensional multitemporal Sentinel-2 data with three-dimensional remote sensing data sources for forest inventory parameter estimation over a boreal forest. Int. J. Appl. Earth Obs. Geoinf. 76, 167–178. https://doi.org/10.1016/j.jag.2018.11.009.

Wurm, M., Stark, T., Zhu, X.X., Weigand, M., Taubenböck, H., 2019. Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. ISPRS J. Photogramm. Remote Sens. 150, 59–69. https://doi.org/10.1016/j.isprsjprs.2019.02.006.

Yuan, Q., et al., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens. Environ. 241, 111716. https://doi.org/10.1016/j.rse.2020.111716.

Zhou, J., et al., 2023. A deep transfer learning framework for mapping high spatiotemporal resolution LAI. ISPRS J. Photogramm. Remote Sens. 206, 30–48. https://doi.org/10.1016/j.isprsjprs.2023.10.017.