High-resolution mapping of forest structure from integrated SAR and optical images using an enhanced U-net method

Michele Gazzea *, Adrian Solheim, Reza Arghandeh

Western Norway University of Applied Sciences, Inndalsveien 28, Bergen, 5063, Norway

A B S T R A C T

Forest structure is an essential part of biodiversity and ecological analysis and provides crucial insights to address challenges in these areas. Modern sensor technologies unlock new possibilities for more advanced vegetation monitoring. This study examines the potential of single high resolution X-band synthetic aperture radar (SAR) and optical images for pixel-wise mapping of four forest structure attributes (height, average height, fractional cover, and density) at a striking 0.5 m resolution. The study site is situated in Western Norway, hosting trees from flatlands to elevated mountainous areas and in-between. The proposed model architecture, called PSE-UNet, is a modified UNet incorporating key components from state-of-the-art deep learning from the field of forest structure monitoring. A comparative analysis involving state-of-the-art models shows promising results with *MAE%* between 21.5 and 24.7, depending on the variable.

1. Introduction

Forests are the most dominant terrestrial ecosystem on Earth and support most of the world's terrestrial species (Arroyo-Rodríguez et al., 2020). Moreover, forests hold 45% of the world's active carbon (Fischer et al., 2019), (Stovall et al., 2021). Conserving forest structure components enhances biodiversity and ecosystem functioning (Casula et al., 2021), (LaRue et al., 2020), (Camarretta et al., 2020). Forest structure attributes are biodiversity indicators and, thus, valuable metrics for monitoring biodiversity. These attributes can also aid in tracking the effects of management actions towards sustainable use (Oettel and Lapin, 2021), (Camarretta et al., 2020).

Remote sensing has an exceptional ability to monitor forest variables over time enabling solutions of unprecedented scale and low cost. As a result, local high-resolution surveys covering relatively large swaths are becoming possible at lower costs. Remote sensing exploits modern sensor technologies to monitor the Earth's surface from afar, thanks to advancements in computer vision. Passive optical sensors capture electromagnetic waves from visible light to near-infrared invisible light. Active SAR sensors transmit radio wave pulses, which make them operational without the sun needing to illuminate the surface. SAR operates with wavelengths up to 1 m enabling the sensor to work in adverse climatic and ground conditions, such as penetrating through cloud cover or accessing sub-canopy layers in forests. However, the active radar sensor is susceptible to speckles, often called noise. Speckle is not noise but a scattering phenomenon that arises because the sensor's resolution is insufficient to resolve individual scatterers. SAR is also a side-looking radar that opens it up to unwanted effects such as layover, shadowing, and foreshortening. More research on SAR imagery is needed to fully exploit its advantages and address some of its challenges.

The data fusion of SAR and optical sensor technologies improve predictions more than any single sensor technology alone, which speaks to the complementary information these unique sensor technologies have to offer (Ge et al., 2022; Chen et al., 2018; Laurin et al., 2018; Shao et al., 2017). SAR imagery is available during all weather conditions and can characterize the structural properties of objects. However, it does not contain spectral information, has poor interpretability, and is susceptible to speckle noise. On the other hand, optical images have rich spectral information (often composed of several spectral bands), which is crucial for many remote sensing applications. Thus, SAR and optical imagery offer complementary information about the imaged area. Fusing these images helps synthetically generate an image rich in spatial and spectral information (Kulkarni and Rege, 2020). Accurately merging SAR and optical data is a current research field, and many methods, including recently machine learning-based algorithms, have been proposed (Safari and Sohrabi, 2020), (Sommervold et al., 2023).

Various studies have experimented with optical images and machine learning approaches for describing forest structure, such as forest

---

canopy height (Potapov et al., 2021) and carbon stock (Lang et al., 2107). Fully connected deep neural networks (DNN) have been applied to predict forest structure attributes such as growing stock volume (GSV) (Astola et al., 2019, 2021; Chrysafis et al., 2017). Forest height is one of the most common forest structure attributes estimated using optical data (Lang et al., 2107; Potapov et al., 2021; Waldeland et al., 2022; Lang et al., 2019). A convolutional neural network (CNN) inspired by the Xception architecture achieved low error with an MAE between 1.7 m and 4.3 m in Gabon and Switzerland (Lang et al., 2019). UNet is a modern deep learning architecture that has recently been proposed to estimate forest height ($MAE = 4.6\ m$) over Congolian rainforests and forests in Tanzania (Waldeland et al., 2022). Generalized additive models were trained in (Korhonen et al., 2017) to estimate canopy cover and Leaf Area Index (LAI), resulting in a coefficient of determination R^2 = 0.687 and R^2 = 0.74, respectively. In a recent study, (Potapov et al., 2021), a bagged regression tree ensemble was implemented to predict forest height with $MAE = 6.36\ m$ using Landsat data.

In parallel, other studies have used SAR data for forest structure estimation. A Random Forest (RF) regression model used SAR time series for predicting stand level (20 m, 40 m, and 100 m) cells of forest height and fractional canopy cover with RMSE scores of 4.6 m and 0.08, respectively (Bruggisser et al., 2021). In (Ge et al.,), the authors use SAR Sentinel-1 time-series data to predict forest growing stock volume using support vector and random forests regression. The advantage of feeding SAR data to deep learning compared to traditional semi-empirical methods was demonstrated in (Ghosh and Behera, 2021) estimating above-ground biomass (AGB). A Bayesian ResNeXt model trained and examined the contribution of SAR inputs on estimating forest structure attributes height, average height, cover, density, and Gini coefficient in (Becker et al., 2111). Similarly, deep learning and machine learning models were trained on SAR inputs in (Waldeland et al., 2022) for forest height estimation. Recent sensor ablation studies generally agree that optical imagery contributes more to forest structure estimation than SAR (Becker et al., 2111; Waldeland et al., 2022). On the contrary, SAR time series has shown to be superior to single optical images (Waldeland et al., 2022) for mapping forest height. We see SAR time series superior to single SAR images in other forest applications such as mapping of forest harvesting (Zhao et al., 2022).

Very few studies tried to combine the two sensors for better forest characterization. Clerici et al. (2017) used a combination of Sentinel-1 (SAR) and Sentinel-2 (optical) for land cover mapping in Colombia. Ge et al. (2022) used the same Sentinel-1 and Sentinel-2 imagery for forest height mapping. In (Hirschmugl et al., 2020), authors used SAR and optical Sentinel time series data for tropical forest mapping.

The literature shows that the European Space Agency (ESA)'s Sentinel satellites are the most widely used imagery source for this application. However, the spatial resolution for both Sentinel-1 (SAR) and Sentinel-2 (optical) is 10 m/pixel, which is suitable for large-scale applications (forest stands). Still, it is not precise enough to estimate forest attributes at a smaller scale (e.g., at the tree or near-tree level). Nowadays, commercial operators can provide high-resolution images for both optical and SAR. These types of sensors, especially the high-resolution SAR, are very recent, and extensive research in this direction is still lacking. Based on our knowledge, no experiments have been performed yet for forest structure retrieval using single, high-resolution (50 cm) SAT and optical imagery.

As such, in this paper, we present a deep learning model to generate high-resolution (in this study, 50 cm per pixel) forest structure maps by regressing the following variables: forest height, average forest height, forest cover, and forest density. Our proposed framework relies on a UNet architecture as a baseline model, enhanced with *sensor-specific entry blocks*, *Squeeze-and-Excitation blocks*, and *partial convolutions* components. The framework is called *Partial Convolution and Squeeze-and-Excitation-based UNet (PSE-UNet)* for mapping forest structure. The proposed PSE-UNet-based forest structure modeling framework is trained and tested on a dataset consisting of pairs of high-resolution (50

cm per pixel) satellite images, namely multispectral optical and X-band SAR images. To our knowledge, combining X-band and optical images at this resolution has not been done before for mapping forest structures.

To summarize, the main contributions are as follows:

1. Using a single view and 50 cm per pixel resolution SAR and optical image for pixel-wise forest structure mapping.
2. Developing a deep learning model to simultaneously estimate forest height, average height, forest cover, and forest density. We show that our model achieves comparable, and for some forest attributes improvement, in the estimation predictions while being at the same time faster to train.

2. Data description

2.1. Study area

The study area is located in Askvoll, Western Norway, and covers roughly 22.5 km^2 (see Fig. 1). Norwegian spruce, pine, and deciduous trees constitute the majority of trees in the area (Norwegian institute of bioeconomy research (nibio), 2022). The terrain ranges from flat to mountainous regions of up to 230 m elevation. Trees are distributed across both low and high elevations but also reside in more challenging areas, such as residential areas and terrain with significant vertical elevation changes, such as steep slopes and cliffs.

2.2. LiDAR data

Forest structure reference maps are computed from LiDAR point clouds originally captured by airborne laser scanning (ALS). ALS data is available through a national program that provides nationwide point cloud coverage accessible to the public. Point clouds with 25 $points/m^2$ density from May 2018 was carried out by Terratec AS, covering the entire study area (see Fig. 3). Each point in the point cloud is measured as a height above sea level.

2.3. Optical data

The Pléiades 1A and Pléiades 1B satellites produce very-high-resolution 50 cm multispectral imagery products. The multispectral image used in this study was captured in July 2021 and includes four bands, i.e., RGB and near-infrared (NIR) (see Fig. 3). Optical images are prone to obstruction by cloud coverage, and thus the image is selected for being cloud-free across the entire area of interest.

2.4. SAR data

ICEYE-X8 was launched as recently as January 2021 and is the source of the SAR image used in this study (see Figs. 2 and 3). The image was captured in June 2021 in Spotlight mode and yields an unprecedented ground sampling distance (i.e., spatial resolution) of 0.5 m with the image already multi-looked (i.e., a preprocessing step to reduce the grainy effect of speckle). The X-band satellite image is downloaded as a Ground Range Detected (GRD) amplitude image processed from a complex image. In other words, in-phase and quadrature complex pixels are converted into amplitude-only values. Fig. 2 shows a side-by-side comparison of optical and SAR images for a portion of the study area.

3. Methodology

Our proposed PSE-UNet-based framework for mapping of forest structure has two major blocks: data generation and preprocessing (Block 1) and PSE-UNet-based mapping of forest structure (Block 2) (see Fig. 4). This section outlines the data generation and preprocessing steps of Block 1 3.1: Satellite imagery preprocessing, tree mask generation, LiDAR, and constructing the datasets. In Block 2, we closely examine the
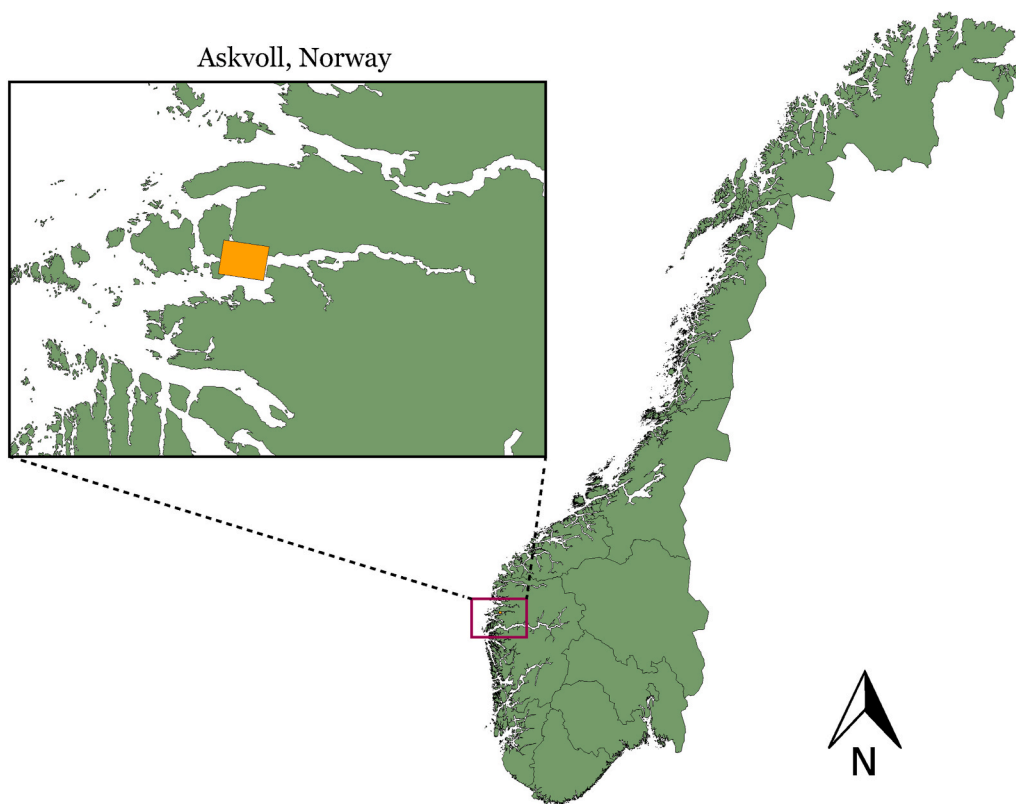
April 13, 2023

Fig. 1. The study site is located in Askvoll, Western Norway.
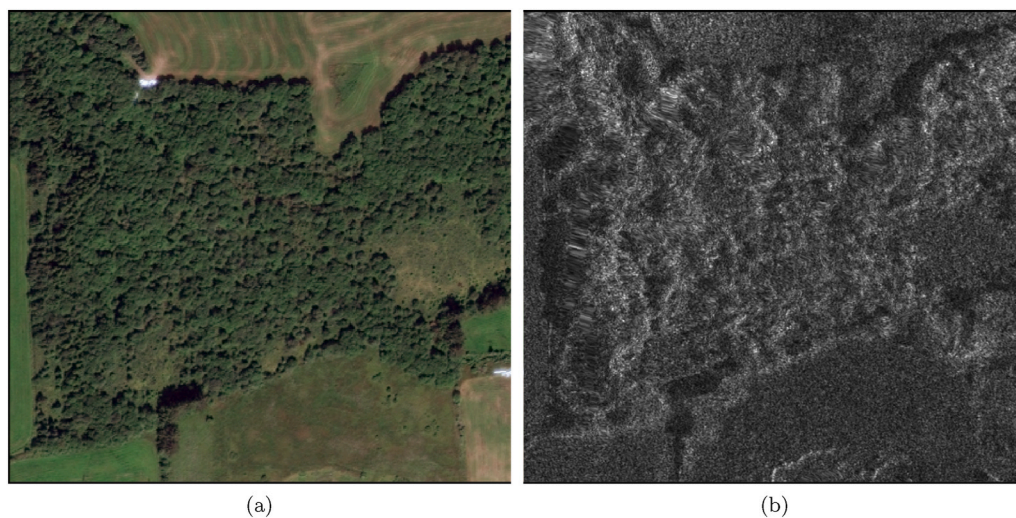

Fig. 2. Side-by-side comparison of two images up close by (a) Pléiades optical and (b) ICEYE SAR sensors for a portion of the study area.

deep learning model 3.2 by dissecting the fundamental components of the model architecture. Finally, also part of Block 2, we explain the training process 3.3 for all models in this study.

3.1. Block 1: data generation

The major tasks for Block 1 are preprocessing the input SAR and optical data and generating ground truth from the LiDAR point cloud (see Fig. 3). We explain the major components of Block 1 as follows:

3.1.1. Satellite imagery preprocessing

The ICEYE products are amplitude-only and come multi-looked, a technique to reduce speckle by processing the image in sections (*looks*) and later combining these sections back together (Vavriv and Bezvesilniy, 2013). More looks decrease speckle noise but lead to a decrease in the resolution and loss of information in the process. The SAR image in
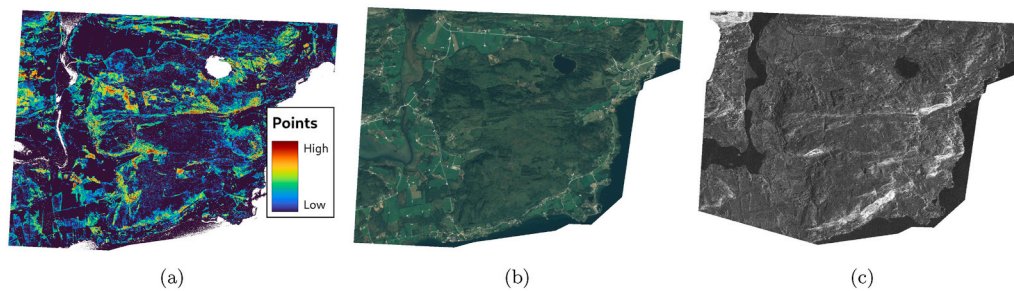
Fig. 3. (a) Normalized LiDAR point cloud, (b) Optical image, and (c) SAR image over the study area located in Askvoll, Western Norway.
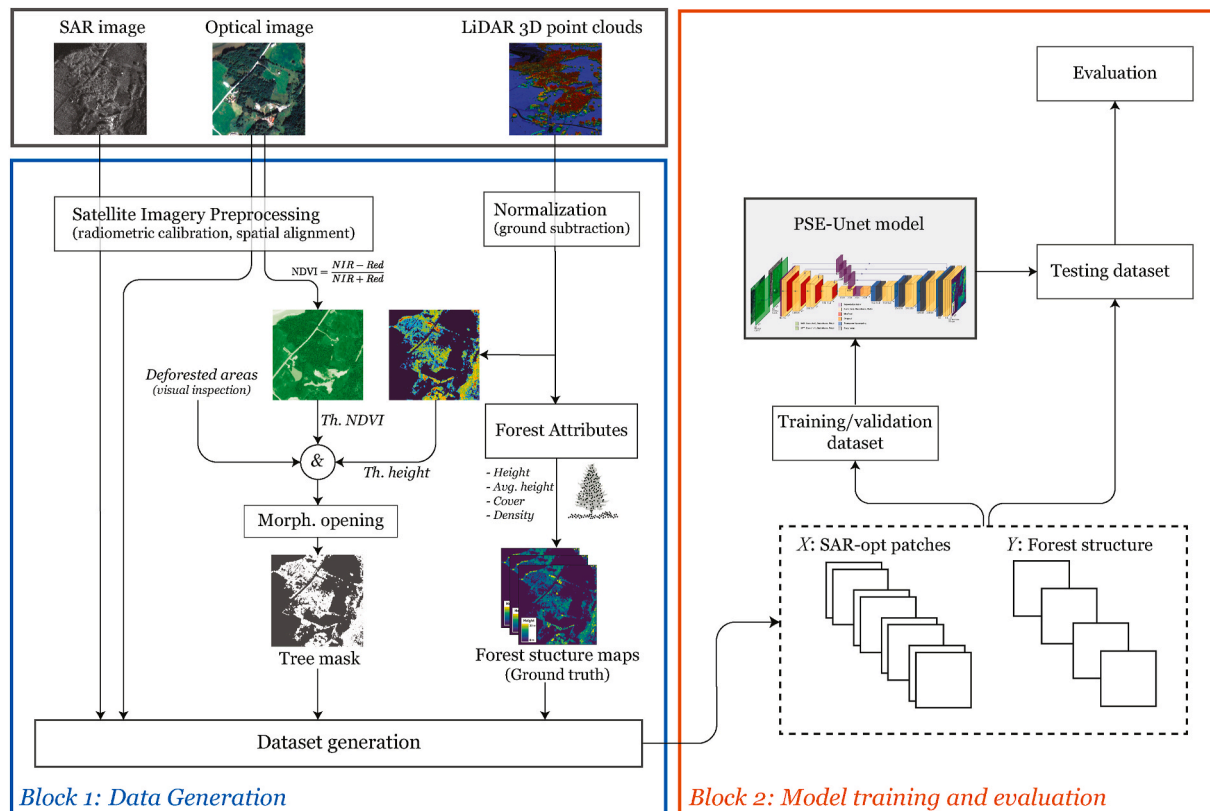

Fig. 4. Pipeline of our approach: in the first part (Block 1), we create the dataset from the available input as SAR, optical imagery, and LiDAR point clouds. In the second part (Block 2), we train a model to infer the forest structure attributes.

this study is formed from 4 independent looks. Speckle filtering is another technique that can be applied in conjunction with multi-looking. However, we refrained from denoising the SAR image beyond the preprocessing step already performed by ICEYE. Speckle has already been decreased, and speckle filtering will reduce resolution and lose information even further.

The downloaded image is already projected to the ground surface and has a natural range-azimuth orientation. However, SAR is a side-looking radar capturing information in varied, rugged terrain, so radiometric terrain correction is needed as a processing step. In rugged terrain, the changing local imaging geometry may result in backscatter changes up to ±5 dB (Loew and Mauser, 2007). Radiometric terrain correction corrects the backscatter intensity of pixels distorted by the local incidence angle. Range Doppler terrain correction was performed using the SNAP toolbox (Sentinel application platform, 2022) with a

high-resolution Digital Terrain Model (DTM) provided by the TerraTec ALS survey.

As the last step, both SAR and optical went through a manual affine transformation shifting pixel positions for both images to more closely overlap the ground truth images. The offset was calculated with respect to the ground truth (i.e., LiDAR), and it was 1 m in the north-south direction and 16 m in the west-east direction. In very high-resolution imagery, offsets become very noticeable and negatively affect the model when training.

3.1.2. LiDAR preprocessing

The ALS survey was carried out and processed by TerraTec AS. TerraTec classifies points through automatic labeling using their in-house software package TerraSolid and manual editing for classes such as bridges. Points in the raw point cloud are distinguished by

ground, noise, and unclassified labels. Using points classified as ground points, the entire point cloud is normalized to the height above ground by subtracting ground point height values from residual points' height values. Lastly, all ground points are set to zero.

3.1.3. Forest structure attributes

Finally, the following forest structure attributes are calculated per cell. A cell is a 2D square that groups 3D points from the LiDAR point clouds. Each cell matches the dimensions of the satellite spatial resolution (in our study, 0.5×0.5 m) and contains LiDAR 3D points measurements.

- Height (m): calculated from the highest points. In this case, the 95th percentile of the highest points is selected to counteract some of the inevitable noise captured by the LiDAR sensor, such as birds and power lines. Height is one of the most common forest characteristics to estimate and is an important component of forest structural complexity.
- Average Height (m): the average height of all vegetation points (i.e., *points > 1.37 m*) within each cell. Average Height gives insight into the vertical distribution of vegetation and can be used as a predictor of forest biomass (Astrup et al., 2019)
- Forest Cover (%): the % of first return vegetation points compared to all first returns. The fraction of ground covered by vegetation, often referred to as fractional cover, corresponds to the fraction of ground covered by vegetation and considers only the horizontal distribution of vegetation cover (i.e., a 2-D variable).
- Forest Density (%): the % of vegetation points compared to all points. Compared to forest cover, forest density include information about the vertical distribution of vegetation points within a cell.

A more visual representation of how these forest structure variables are calculated can be seen in Fig. 5.

3.1.4. Tree mask generation

This study is focused on estimating forest structure attributes, and hence only forested pixels are of interest. A binary mask denoting whether a pixel is forested or not is used to define valid pixels. Image patches without valid pixels are discarded. As such, only valid pixels count toward the loss during training.

The first component of the tree mask is the Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1973) calculated from near-infrared (NIR) and red band (see Eq. (1)).

$$NDVI = \frac{NIR - Red}{NIR + Red} \qquad (1)$$

The NDVI is an indicator ranging between -1 and 1 expressing the amount of chlorophyll in objects. Rocks, sand, water, or concrete show very low NDVI values (0.1 or less). Sparse vegetation, such as shrubs and grasslands or senescing crops results in moderate NDVI values (approximately 0.2–0.5). High NDVI values (approximately from 0.5 to 0.6 to 1) correspond to dense vegetation such as that found in temperate forests or crops at their peak growth stage. Therefore, for the NDVI, we selected a threshold of 0.55 as a value of "greenness" for the vegetation under investigation.

To exclude non-forested vegetation, such as fields, we set an additional threshold on the height value for LiDAR point clouds as 1.37 m for valid pixels. Such a threshold for trees is justified by the fact that one of the most common botanical measurements for trees is the "Diameter at breast height" (DBH), which is defined as the diameter of a tree trunk when at the height of an adult's breast, which is set practically at 1.37.
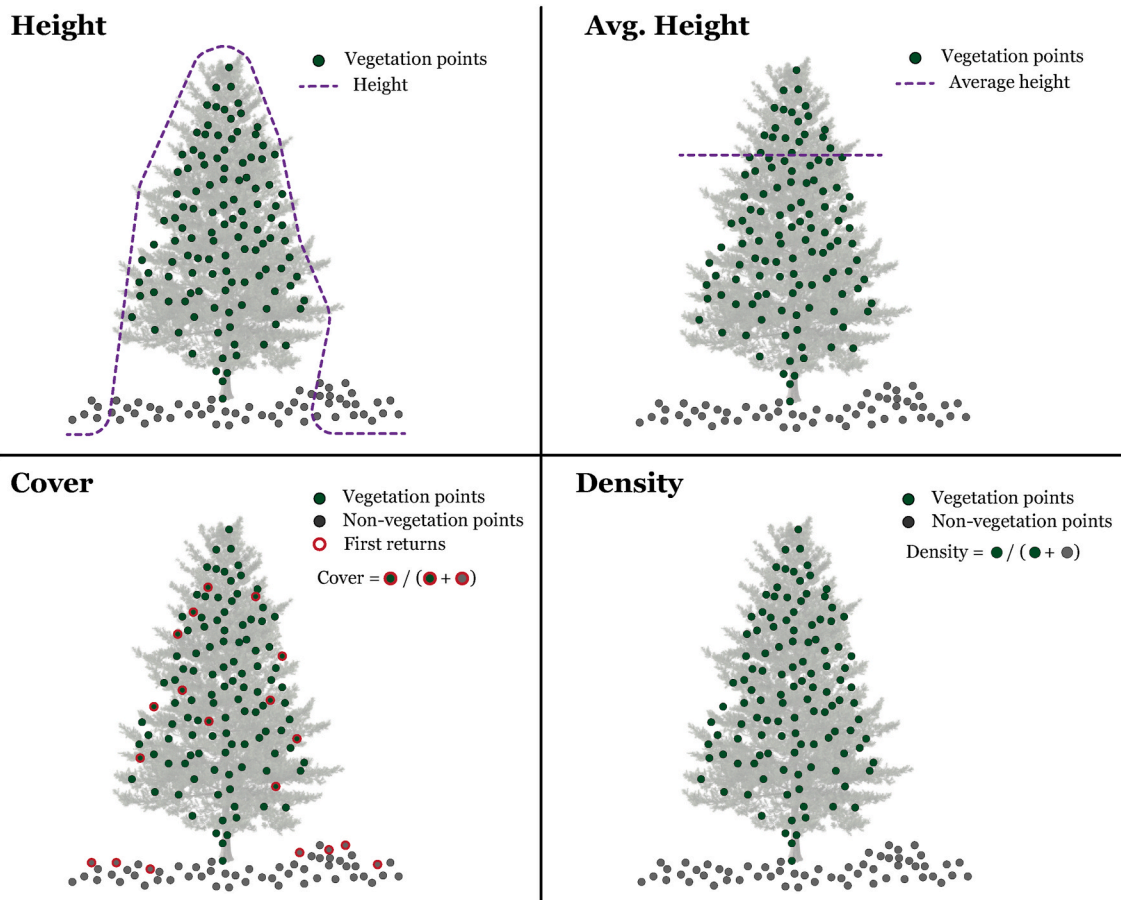

Fig. 5. Deriving forest metrics from LiDAR point cloud. Note that in the figure, the density and distribution of LiDAR points are used only for visualization purposes.

Although we were not interested in estimating the DBH in this study, the threshold of 1.37 m is often chosen to distinguish LiDAR points that can belong to trees from bushes or not yet developed trees (Tinkham et al., 2017).

The last component of the tree mask is, in particular for this use case, a consequence of the time discrepancy between the input satellite images and the LiDAR point cloud acquisition. Between 2018 and 2021, in certain areas, trees were cut down. Consequently, a deforestation mask is created by visual inspection to address most of these pixels. These three binary masks denote whether a pixel is forested, given all conditions are true for that pixel (see Eq. (2)).

$$\text{Tree mask} = (NDVI > 0.55) \wedge (Height > 1.37) \wedge \neg(Deforested) \tag{2}$$

Morphological image processing is performed on the mask through *erosion* and *dilation*. This technique removes unwanted artifacts and isolated points from an image that may still be present (Fig. 6).

Erosion is performed using a 2×2 structuring element, effectively shrinking image pixels. Dilation utilizes the same structuring element to restore and retain outlying vegetation pixels. The operation, usually referred to as *opening*, overall removes noise artifacts without changing the image boundaries.

3.1.5. Dataset generation

A polygon is generated as the intersection of all inputs and the LiDAR-derived ground truth (GT) defining the area of interest (AOI) in this study (see Eq. (3)). Subsequently, every raster is clipped to this polygon.

$$AOI = (OPT) \cap (SAR) \cap (GT) \tag{3}$$

The input and ground truth images have been clipped to the AOI polygon (Eq. (3)), creating tensors of matching size (i.e., matching height and width $H \times W$). A sliding window is used to generate patches from both the input image (i.e., SAR and optical) and the ground truth. We generate patches of size 64×64 pixels, with a slide of 32 pixels. As such, the algorithm extracts patches with some overlap. If the patch does not contain any trees or any of the input images (SAR and optical) contain pixels with NaN values, the patch is discarded. However, the percentage of discarded patches due to NaN values was very low (0.01%). The dataset is split as 70/10/20 percent for training, validation, and testing, respectively.

3.2. Block 2: PSE-UNet architecture

The second block in our framework includes our proposed PSE-UNet architecture (see Fig. 7). The PSE-UNet architecture is based on the following components:

3.2.1. Partial convolutions layer

Traditionally, when performing convolutions in CNNs, the input tensor is padded to retain the spatial dimensions. Padding extends the original input with a border filled w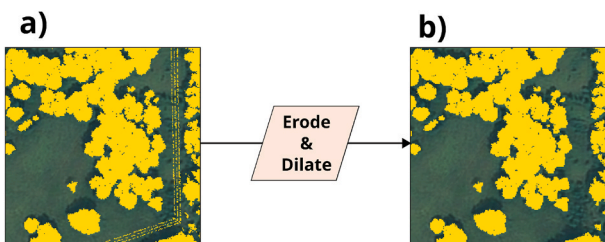ith fixed values using different padding schemes such as *zeros-padding, the most common, reflect*, or *replicate*. These methods add data that are either unrelated to input data (zero padding) or plausible at best (reflect or replicate). The added or repeated pixel values at the borders are considered valid and treated equally as the original input, which may confuse the network. The proposed network replaces regular convolutional layers with *partial convolutions*, originally proposed to handle images with missing data, such as images with holes (Liu et al., 2018). *Partial based convolutional padding* is a new padding scheme introduced in (Liu et al., 2023). The padded regions are treated as holes in the input data, and partial convolutional-based padding allows performing convolutions at the edge of the patch without training the network on unrelated pixels introduced by padding.

3.2.2. Entry blocks

Early convolutional layers aim to 1) learn basic features on very high spatial resolution inputs and 2) potential denoising. Optical inputs go through a basic 1×1 convolutional layer, batch normalization, and a rectified linear unit (ReLU). In contrast, the SAR entry block uses a 5×5 convolutional layer. The reason behind this is a larger filter considers a larger input field and can potentially smooth out some of the speckle noise inherent in SAR imagery as shown in (Becker et al., 2111). Batch normalization and ReLU follows. Entry blocks can be thought of as a sensor-specific preprocessing step and have previously been applied to deep learning on forest structure in (Becker et al., 2111) with SAR optical data fusion.

3.2.3. UNet block

The baseline model consists of a UNet architecture initially developed for biomedical image segmentation (Ronneberger et al., 2015). UNet has been applied to estimating and mapping forest structural parameters, such as in (Lang et al., 2019) with forest height. UNet can extract deeper features and retain spatial information by introducing skip connections to an encoder-decoder network.

The encoder, also called the contracting path, contains a sequence of 3×3 convolution, batch normalization, and ReLU that is repeated twice, followed by a 2×2 MaxPool at each level. For each level in the encoder, the spatial dimensions are halved while the channel dimensions are doubled.

The bottommost layer, sometimes referred to as the bottleneck, is similar to the preceding levels, except the MaxPool layer is replaced with Dropout with probability $p = 0.3$, which is only in effect during training in order to reduce overfitting.

The decoder, also called the expansive path, starts each level with a 2×2 transpose convolution followed by concatenating the feature map from the same-level skip-connection and ends with a double sequence of 3×3 convolution, batch normalization, and ReLU. UNet achieves feature reusability through concatenation from the skip connections. The decoder doubles the spatial dimensions and half the channel dimensions for each successive step through the decoder, ultimately restoring the feature map to its original size.

The last layer is a 1×1 convolution, reducing the number of channels down to the number of forest variables. For fractional variable outputs (cover & density), a final sigmoid activation function $\sigma = \frac{1}{1+e^{-1}}$ is applied to keep the predictions within the 0 to 1 range.

3.2.4. Channel attention

Convolution operations extract information by fusing spatial and channel-wise information within local receptive fields. The Squeeze-and-Excitation (SE) block is a lightweight gating mechanism introduced in (Hu et al., 2018) to focus on the channel relationship and explicitly model interdependencies between channels at a minimal additional computational cost. In this context, the model can learn to recognize sensor-specific characteristics in the input data sensitive to forest structure (i.e., wavelength and polarization).
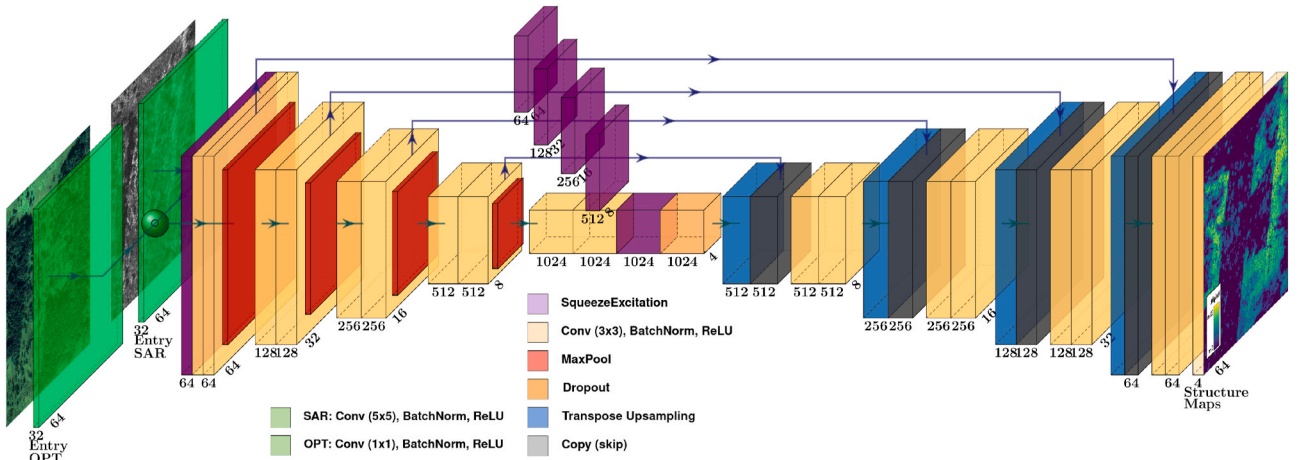
Fig. 6. The mask before **a)** and after **b)** eroding and dilating. The power lines spanning across the right border of the untreated mask **a)** are removed by erosion and dilation in the processed mask **b)**.

Fig. 7. Proposed UNet architecture with Squeeze-Excitation blocks and entry blocks for each sensor. Convolutional layers refer to partial convolutions.

The squeeze operation flattens the spatial dimensions of a tensor T from $C \times H \times W$ to $C \times 1 \times 1$. The squeezing process F_{sq} is achieved by channel-wise global average pooling (GAP) to generate the vector $\boldsymbol{z}$ with values z_c from each channel tensor $\boldsymbol{t}_c$ (see Eq. (4)). The result can be interpreted as a vector of local descriptors expressive for the whole image.

$$z_c = F_{sq}(t_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} t_c(i,j) \tag{4}$$

The excitation operation captures the nonlinear dependencies between channels through a gating mechanism F_{ex}. The vector after excitation, $\boldsymbol{s}$, is generated from the squeezed tensor $\boldsymbol{z}$ by two convolutions $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ followed by a ReLU function δ and sigmoid function σ. $\boldsymbol{W}_1$ reduce the number of channels to $\frac{C}{2}$ after which ReLU δ is applied. Subsequently, $\boldsymbol{W}_2$ restores the number of channels C followed by σ finalizing the weights for each channel (see Eq. (5)).

$$\boldsymbol{s} = F_{ex}(\boldsymbol{z}) = \sigma(\boldsymbol{W}_2 \delta(\boldsymbol{W}_1 \boldsymbol{z})) \tag{5}$$

Finally, the original tensor is effectively scaled by multiplication with the corresponding scalar for each channel from the excitation process (see Eq. (6)). The result is a new recalibrated tensor $\widetilde{T} = [\widetilde{t}_1, \widetilde{t}_2, \dots, \widetilde{t}_C]$ where (Fig. 8)

$$\widetilde{t}_c = F_{scale}(t_c, s_c) = s_c t_c \tag{6}$$

3.3. Block 2: PSE-UNet model training

Another major component of Block 2 is training the PSE-UNet model for mapping forest structures. The training and test dataset consists of 64×64 patches in batches of $B = 64$. In total, there are $C = 5$ input channels. Four optical bands (Red, Green, black, and Near-infrared) and a single VV SAR polarization form a tensor $T = B \times C \times 64 \times 64$ at runtime. Non-forested pixels and pixels without corresponding observations (i.e., missing values in the ground truth) are not considered when calculating the loss. During training, the dataset is shuffled after each epoch. Before jumping to the next epoch, the model is validated against the validation dataset. A collection of the best-performing and most trained models are stored for further testing and analysis. Model parameters θ are saved, given that the model performance on the validation set has improved since the previous best-performing model. Additionally, at the tail end of the training process, the model parameters of the last five models are saved. The model that performs best on the test dataset is ultimately selected as the best model θ^*. Given the regression task, the Mean Squared Error (MSE) is implemented as the loss function by calculating the average squared difference between predicted $\widehat{y}$ and observed y values.

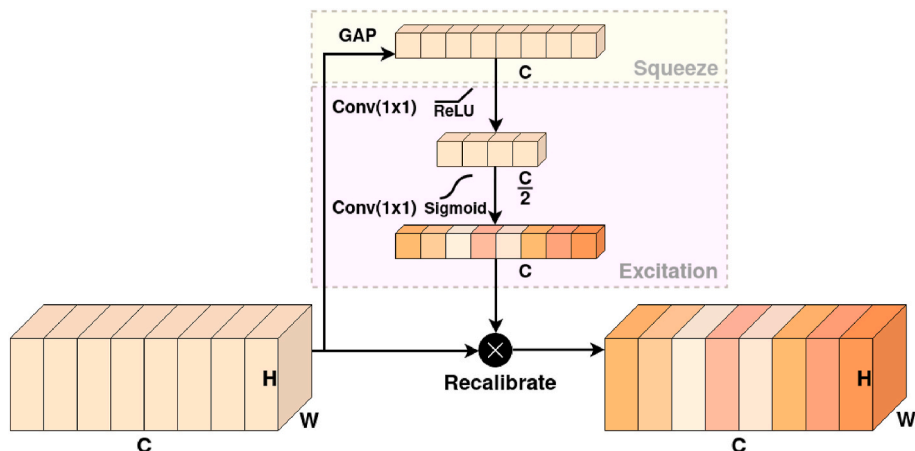$$MSE = \frac{\sum_i (\widehat{y}_i - y_i)^2}{n} \tag{4}$$

Fig. 8. Channel-wise recalibration of a tensor T through an SE block.

4. Results & discussion

In this section, we go over each model from the comparative analysis and how they are evaluated. Then we discuss the height and fractional variable results and analyze the sensor ablation study results.

4.1. Experimental setup

The training was performed on Linux with a single NVIDIA GeForce RTX 3060 and 32 GB RAM. The codebase was implemented in Python 3.10.4 using PyTorch 1.11.0. Adam was selected as the optimization method with learning rate $\gamma = 5 \times 10^{-4}$, weight decay $\lambda = 10^{-4}$, betas $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 1e - 08$. Additionally, the MultiStepLR scheduling technique was used to reduce the decay of the learning rate at fixed milestones. GradScaler was used to prevent gradient values from flushing to zero.

A test dataset consisting of unseen forested areas. Only the innermost 32×32 predicted pixels of the patch are preserved, resulting in no overlapping predictions. This mean ignoring a border of $\frac{Patch\,Size}{4}$ pixels. Discarding the predictions residing at the edge of the patch allows more spatial context for each patch prediction (Huang et al., 1805).

To evaluate the performance of the proposed network architecture, a set of state-of-the-art deep learning models is trained on the same data for comparison. These architectures have previously been tested on at least one forest structure attribute.

- UNet: The first model is a UNet proposed in (Waldeland et al., 2022) fitted to Sentinel-2 optical data in order to produce forest height maps with per-pixel forest height estimates. A few modifications are made to the original model such that it accepts SAR inputs as well.
- SeUNet: The second model is a UNet extended with SqueezeExcitation blocks proposed in (Ge et al., 2022) estimating forest height. The original paper implements a semi-supervised training strategy. In this study, SeUNet is trained on the entire dataset in a supervised manner.
- UNet3+: A new modified version of the UNet architecture improving on the also recent UNet++ architecture. UNet3+, with and without deep supervision, has shown great promise surpassing previous state-of-the-art approaches in segmentation (Huang et al., 2020). A simpler UNet3+ without deep supervision is implemented here and tested in forest structure estimation.
- ResNeXt: Lastly, a large ResNeXt architecture retaining spatial dimensions of feature maps throughout the entire network is implemented. The ResNeXt model was proposed in (Becker et al., 2111) with Bayesian deep learning predicting forest structure attributes along with their associated uncertainties. In this study, the ResNeXt model is trained using the MSE loss function instead of the negative log-likelihood of the original paper.

Given the size of the ResNeXt architecture, the spatial dimensions of the input are halved to address limited GPU memory. This yields input tensors $T = B \times C \times 32 \times 32$ with *stride* = 16 and retaining only the innermost 16×16 pixels accordingly.

The experimentation phase found most of the convergence to happen early, before or around 30 epochs for most of the models. 30 epochs were selected as the best middle ground between convergence and training time. Training time differed significantly between the fastest and slowest models. The number of trainable parameters also highlights differences between the models. Table 1 illustrates the disparity in training time for the same amount of epochs.

4.2. Performance metrics

Metrics for evaluating the performance of each model are Root Mean Squared Error (RMSE) (see Eq. (7)), MAE (see Eq. (9)), and their normalized counterparts (see Eq. (8) & Eq. (10)). Reiterating from previously, y and $\hat{y}$ denote predicted and observed values while $\bar{y}$ denotes the mean of observed values.

$$RMSE = \sqrt{\frac{\sum_i (\hat{y}_i - y_i)^2}{n}} \qquad (7)$$

$$RMSE\% = \frac{RMSE}{\bar{y}} \cdot 100\% \qquad (8)$$

$$MAE = \frac{\sum_i |\hat{y}_i - y_i|}{n} \qquad (9)$$

$$MAE\% = \frac{MAE}{\bar{y}} \cdot 100\% \qquad (10)$$

4.3. PSE-UNet performance evaluation

To evaluate the performance of the proposed PSE-UNet we first compare its performance in estimating all forest structure variables to the other models.

The results showcase the proposed model's ability to outperform several state-of-the-art DL models on forest height metrics (see Table 2) and perform similarly to other models on fractional variables (see Table 3). Forest height estimation performance is particularly interesting, considering most of these models were originally trained for estimating forest height. Here we achieve low mean absolute errors of 2.23 m and 1.99 m for forest height and forest average height, respectively (see Table 2). On forest height, our proposed PSE-UNet model performed 17.4%, 15.2%, 6.3% and 3.9% better than UNet3+, ResNeXt, SeUNet and UNet, respectively, in terms of RMSE. There are no particular stand-out performers in regard to the fractional variables (see Table 3). Most models achieve a mean absolute error of 0.18. Density mapping slightly favors the SeUNet architecture. UNet3+ edged out the other models on forest cover.

The scatter matrix, however, paints a slightly different picture (see Fig. 9). The proposed PSE-UNet model shows that the height predictions correlate well with the ground truth. There is observable uncertainty, and predictions skew slightly towards underestimating both height and average height. By contrast, the model seems to struggle with predicting fractional variables. While the model has been able to reduce prediction errors, it has not adequately established the relationship between input features and the fractional output variables. This is more true for forest cover than forest density. Density predictions show a stronger correlation with ground truth than the cover attribute. The cover and density ground truth pixels are mostly clustered at the higher value range. As a result, the model sees mostly high cover and high density during training. The inability of any model to adequately recognize the fractional features leaves more to be desired and motivates similar efforts on larger datasets. Looking more closely at the forest cover and forest density scatter plots, it is noticeable that there are gaps in the plots. Forest cover and forest density are continuous variables, but due to how these variables are calculated 5, they can practically become discrete variables when the LiDAR point density is low. This is a side-effect of deriving forest metrics from sub-meter cells isolating LiDAR points into smaller groups for each calculation. Since forest cover is calculated from first returns only, this effect is more prominent for forest cover as a

Table 1
Training time.

	UNet (Waldeland et al., 2022)	**SeUNet** (Ge et al., 2022)	**UNet3+** (Huang et al., 2020)	**ResNeXt** (Becker et al., 2111)	Proposed
Time	26 min	51 min	4 h 10 min	16 h 48 min	49 min
Params	7.2 M	7.8 M	27 M	27.4 M	32.5 M

Table 2
Forest height characterization error.

		MAE (m)	MAE%	RMSE (m)	RMSE%
Height	UNet	2.32	22.31	3.23	31.08
	SeUNet	2.38	22.86	3.37	32.44
	UNet3+	2.70	26.02	3.62	34.87
	ResNeXt	2.63	25.31	3.70	35.61
	Proposed	**2.23**	**21.49**	**3.11**	**29.96**
Avg. Height	UNet	2.06	23.08	2.89	32.46
	SeUNet	2.11	23.62	3.03	33.97
	UNet3+	2.37	26.61	3.17	35.55
	ResNeXt	2.31	25.96	3.29	36.98
	Proposed	**1.99**	**22.32**	**2.79**	**31.33**

Table 3
Cover & density characterization error.

		MAE	MAE%	RMSE	RMSE%
Cover	UNet	0.18	20.93	0.24	27.13
	SeUNet	0.18	20.82	0.24	27.02
	UNet3+	0.17	19.46	0.24	27.54
	ResNeXt	0.18	21.02	0.24	27.61
	Proposed	0.18	20.78	0.24	27.14
Density	UNet	0.18	24.71	0.22	30.94
	SeUNet	**0.17**	**24.12**	**0.21**	**30.18**
	UNet3+	0.18	24.79	0.22	31.19
	ResNeXt	0.18	25.12	0.22	31.32
	Proposed	0.18	24.68	0.22	30.98

consequence of further limiting the number of LiDAR points.

To better visualize the output of our PSE-UNet-based forest structure map, we create $80 \times 400\,m$ stripes of prediction value, ground truth (reference), and error (MAE) for the selected forest structure parameters under the study (see Fig. 10. These output stripes are selected to show an

uninterrupted contiguous community of trees and simultaneously illustrate the level of prediction detail achievable from high-resolution forest structure modeling. The predictions appear more smoothed out compared to the reference maps indicating that the model is unable to estimate the fine-grained texture of the canopy. Instead, the predictions are more continuous as of a more smooth canopy surface.

Forest cover is subject to gaps in the canopy layer where LiDAR pulses penetrate straight through to the ground layer. When looking at very high-resolution forest cover maps, these gaps appear as holes scattered across entire forest stands. Normally these gaps are averaged out when working with high to low-resolution mapping rendering them effectively nonexistent. In this case, however, these gaps become more prominent. The forest cover MAE error map 10 illustrates this with high error pixels (dark red spots) scattered across as a grainy effect throughout the entire MAE stripe. The forest cover scatter plot 9 highlights this at the bottom, where a cluster of forest pixels with 0 cover has not been identified by the model. The model is unable to resolve this level of detail within the canopy using SAR and optical. Furthermore, the time discrepancy between SAR and optical images and the derived maps from LiDAR make fine details like this nearly impossible to estimate regardless of the model.

4.4. SAR-optical ablation study

A SAR-optical ablation study is performed to analyze the contribution of each sensor technology to the model performance in estimating all forest structure variables. The results are summarized in Table 4.

Some studies suggest that optical imagery has better performance than SAR imagery for mapping forest structure [26, 20, 13]. Our SAR-optical ablation study confirms this statement. However, there is a notable improvement in height and average height predictions when we combine SAR and optical images. For Cover and Density parameters, combining SAR and optical does not lead to a noticeable improvement in
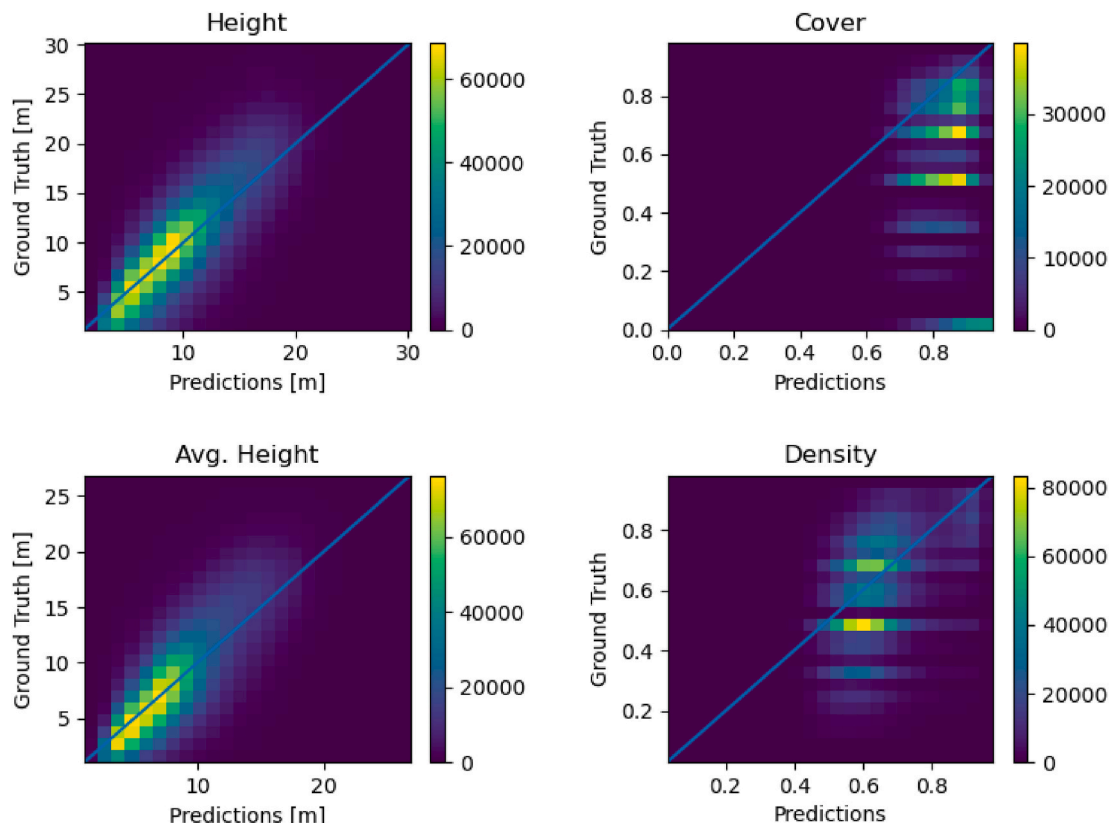

Fig. 9. Scatter plots from the proposed model predictions of all forest structure variables on the test dataset.
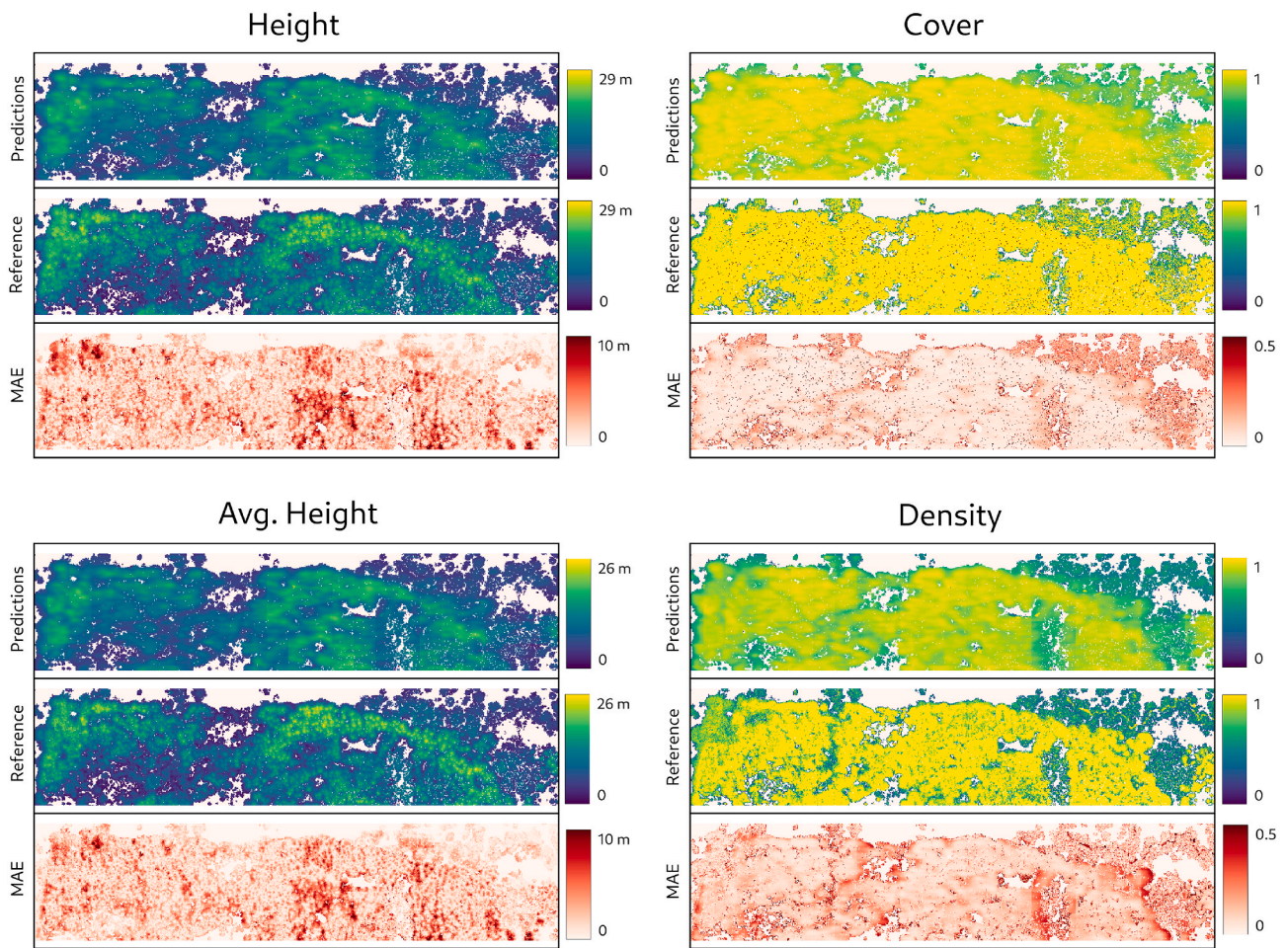
Fig. 10. 400 m stripes of predictions, reference and error of all forest variables.

Table 4
Opt & SAR sensitivity results of proposed model.

	Input	MAE (m)	MAE%	RMSE (m)	RMSE%
Height	Opt	2.32	22.30	3.23	31.10
	SAR	3.19	30.65	4.25	40.93
	Opt + SAR	**2.23**	**21.49**	**3.11**	**29.96**
Avg. Height	Opt	2.05	23.01	2.89	32.42
	SAR	2.77	31.14	3.78	42.37
	Opt + SAR	**1.99**	22.32	**2.79**	31.33
Cover	Opt	0.18	20.88	0.24	**27.11**
	SAR	0.19	21.50	0.24	27.57
	Opt + SAR	**0.18**	20.78	0.24	27.14
Density	Opt	**0.17**	24.76	0.22	**30.96**
	SAR	0.19	25.90	0.23	31.99
	Opt + SAR	0.18	**24.68**	**0.22**	30.98

comparison to the optical-only case. The lack of penetration capability of the X-band SAR sensor is likely a limiting factor for forest structure modeling. SAR sensors of longer wavelengths are able to penetrate deeper into the canopy and reveal more structural information. It should be noted that for cover and density, SAR-only performed relatively close to the opt-only case. Overall, combining SAR and optical still has advantages compared to SAR-only or optical-only cases.

4.5. Practical challenges and lessons learned

Considering the increase in high-resolution SAR data availability through commercial providers, it is crucial to study the unique advantages of SAR in different applications, including forest structure

modeling. However, at 0.5 m resolution, there are some challenges that become more prominent than those with lower resolution.

4.5.1. Challenges with SAR and opt data

With only a single SAR image, the model is more exposed to unwanted effects such as layover, foreshortening, and shadow. A SAR image in ascending orbit can be unable to capture accurate information behind a mountain ridge due to SAR being a side-looking radar system. A corresponding image from the descending orbit will provide complementary information filling much of what is missed in the original image. Most studies utilize multiple SAR images in deep learning, either as time series or ascending/descending pairs. SAR time series can exploit methods to reduce speckle. However, very high-resolution SAR imagery is scarce compared to open datasets such as Sentinel. Multiple SAR images should be explored for high-resolution forest structure monitoring with deep learning.

A considerable portion of sections with high elevation in the SAR image was negatively affected by the terrain correction such that the SAR pixels were unable to correctly overlap the corresponding ground truth pixels. In other words, there was a slight offset that could resolve to meters in ground distance. Thus, high-resolution DEM data is needed for localization registration. Furthermore, optical and SAR automatic co-registration could help in solving the possible mismatch issues (Sommervold et al., 2023). Longer wavelengths may be better suited for mapping forest structures due to increased penetration capabilities (Omar et al., 2017). However, they can generate also more complex signals from sub-canopy, trunks, and even ground coupling as shown in Fig. 11.
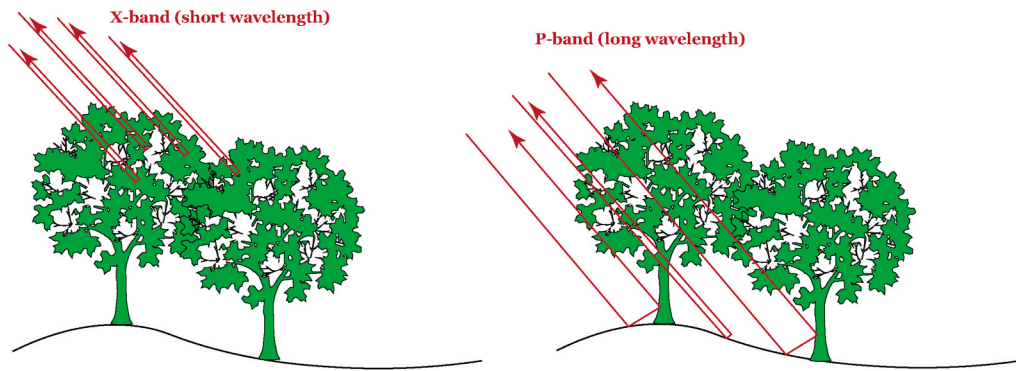
Fig. 11. Short wavelength (for example the X-band) signal reflects from canopy surface while long wavelength (for example the P-band) signal penetrates through canopy and reflects from stems and terrain surface.

Optical is an excellent technology for capturing information about the surface and can be further complemented with sub-canopy information from penetrating deeper into the canopy with SAR.

4.5.2. Challenges with LiDAR data

When deriving forest metrics from LiDAR at the sub-meter resolution, the number of points used for each calculation is significantly reduced. As mentioned in the evaluation section 4.3, the LiDAR-derived forest metrics become extremely sensitive to each LiDAR pulse when the point density is low. The forest cover low points are distributed across the forest canopies, and the model is unable to identify them. The sigmoid σ activation function was applied to the fractional variables to ensure the prediction were distributed within the appropriate range. Testing other activation functions, such as tanh did not improve the results. Future works should encourage more training on forest cover and forest density by ensuring a large enough dataset.

Normalizing a point cloud for the purpose of mapping tree height presents a few challenges. The method of using ground points to normalize non-ground points (e.g., vegetation points) to the height above ground is subject to extreme outliers in terrain with steep inclines creating towering high points (see Fig. 12a).

Norway is one of Europe's most mountainous countries, which makes creating accurate ground truth challenging in some areas. Extreme cases of erroneous height values are easily detectable due to their striking height and, in this case, appear mostly around mountains at the very edges of the point cloud. A more subtle case of erroneous terrain-induced high points happens in forested terrain with steep slopes or cliffs. The effect of sloped terrain on deriving forest metrics from LiDAR is known (Khosravipour et al., 2015). Vegetation points of trees can reside far enough out horizontally from the tree stem such that the ground point selected for normalizing is substantially lower compared to the true ground level of the tree. This can inflate the height value of the given vegetation point (see Fig. 12b). The more subtle case of artificial high points can occur inside a contiguous community of trees and be very hard to detect within a dense cloud of points. Thus, it is hard to estimate the scale of the issue at first glance.

5. Conclusions

The results showed that the proposed PSE-UNet model can perform very high-resolution mapping of forest structures using a single SAR and optical image. The proposed model performs well on mapping height, as
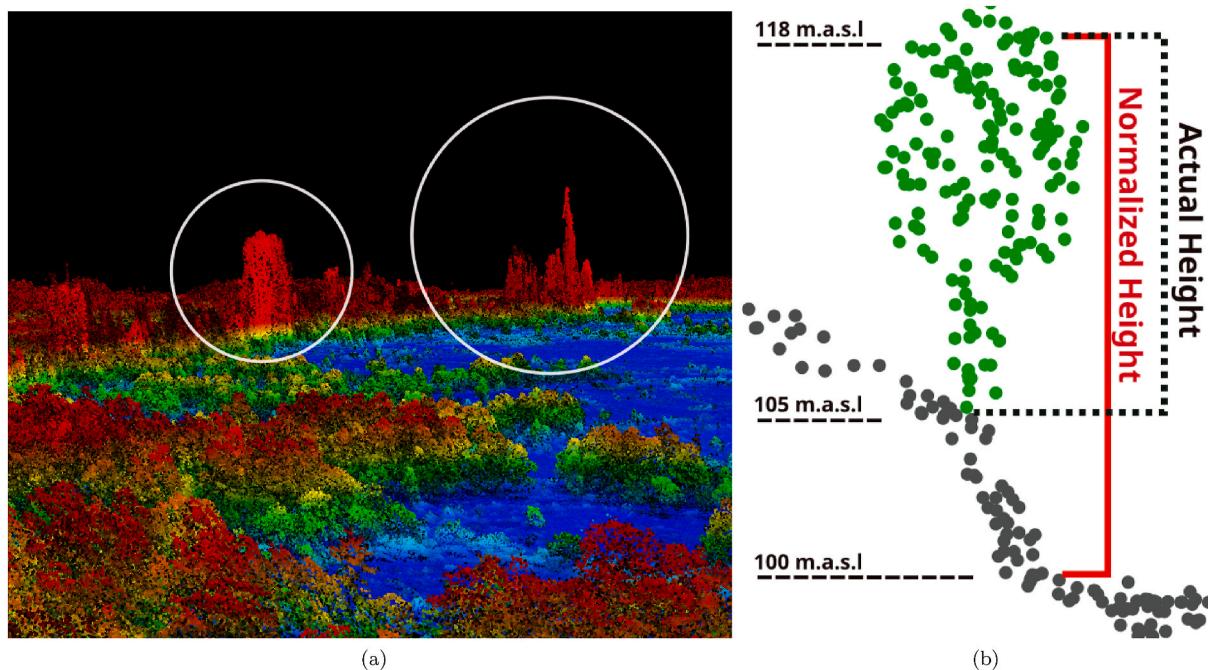

Fig. 12. (a): Showcasing terrain-induced false tree height after normalization. (b): LiDAR points at >100 m above sea level (m.a.s.l) showcasing terrain-induced tree height anomaly after normalization.

supported by the literature. Overall, the results were satisfactory, considering the challenges discussed. In particular, we got higher performances on forest height estimation compared to other state-of-the-art models. Longer wavelength SAR bands such as L- and P-band should be explored for this use case. Challenges related to very-high-resolution mapping of forest structure were highlighted, such as gaps in the LiDAR-derived ground truth due to low point density and the effect of rugged terrain in normalizing the point cloud. In regard to future works, we focus on other forest metrics to improve forest structure high-resolution mapping.URL https://doi.org/10.3390/rs15030850.

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.

Data availability

Data will be made available on request.

Acknowledgement

References

Arroyo-Rodríguez, V., Fahrig, L., Tabarelli, M., Watling, J.I., Tischendorf, L., Benchimol, M., Cazetta, E., Faria, D., Leal, I.R., Melo, F.P., et al., 2020. Designing optimal human-modified landscapes for forest biodiversity conservation. Ecol. Lett. 23 (9), 1404–1420.

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.

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. Rem. Sens. 13 (12), 2392.

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 (6), 484–496.

A. Becker, S. Russo, S. Puliti, N. Lang, K. Schindler, J. D. Wegner, Country-wide Retrieval of Forest Structure from Optical and Sar Satellite Imagery with Bayesian Deep Learning, arXiv preprint arXiv:2111.13154.

Bruggisser, M., Dorigo, W., Dostálová, A., Hollaus, M., Navacchi, C., Schlaffer, S., Pfeifer, N., 2021. Rem. Sens. 13 (4), 798. https://doi.org/10.3390/2Frs13040798, 10.3390/rs13040798.

Camarretta, N., Harrison, P.A., Bailey, T., Potts, B., Lucieer, A., Davidson, N., Hunt, M., 2020. Monitoring forest structure to guide adaptive management of forest restoration: a review of remote sensing approaches. N. For. 51 (4), 573–596.

Casula, P., Fantini, S., Fenu, G., Fois, M., Calvia, G., Bacchetta, G., 2021. Positive interactions between great longhorn beetles and forest structure. For. Ecol. Manag. 486, 118981.

Chen, L., Ren, C., Zhang, B., Wang, Z., Xi, Y., 2018. Estimation of forest above-ground biomass by geographically weighted regression and machine learning with sentinel imagery. Forests 9 (10), 582.

Chrysafis, I., Mallinis, G., Siachalou, S., Patias, P., 2017. Assessing the relationships between growing stock volume and sentinel-2 imagery in a mediterranean forest ecosystem. Remote Sensing Letters 8 (6), 508–517.

Clerici, N., Calderón, C.A.V., Posada, J.M., 2017. Fusion of sentinel-1a and sentinel-2a data for land cover mapping: a case study in the lower magdalena region, Colombia. J. Maps 13 (2), 718–726. https://doi.org/10.1080/17445647.2017.1372316.

Fischer, R., Knapp, N., Bohn, F., Shugart, H.H., Huth, A., 2019. The relevance of forest structure for biomass and productivity in temperate forests: new perspectives for remote sensing. Surv. Geophys. 40 (4), 709–734.

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," in. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 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., n.d.. n.d. Using Hypertemporal Sentinel-1 Data to Predict Forest Growing Stock Volume. bioRxiv. doi:10.1101/2021.09.02.458789.

Ghosh, S.M., Behera, M.D., 2021. Aboveground biomass estimates of tropical mangrove forest using sentinel-1 sar coherence data-the superiority of deep learning over a semi-empirical model. Comput. Geosci. 150, 104737.

Hirschmugl, M., Deutscher, J., Sobe, C., Bouvet, A., Mermoz, S., Schardt, M., 2020. Use of sar and optical time series for tropical forest disturbance mapping. Rem. Sens. 12 (4), 727. https://doi.org/10.3390/rs12040727, 10.3390/rs12040727.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141.

B. Huang, D. Reichman, L. M. Collins, K. Bradbury, J. M. Malof, Tiling and Stitching Segmentation Output for Remote Sensing: Basic Challenges and Recommendations, arXiv preprint arXiv:1805.12219.

Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., Wu, J., 2020. Unet 3+: a full-scale connected unet for medical image segmentation. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 1055–1059.

Khosravipour, A., Skidmore, A.K., Wang, T., Isenburg, M., Khoshelham, K., 2015. Effect of slope on treetop detection using a lidar canopy height model. ISPRS J. Photogrammetry Remote Sens. 104, 44–52.

Korhonen, L., Packalen, P., Rautiainen, M., et al., 2017. Comparison of sentinel-2 and landsat 8 in the estimation of boreal forest canopy cover and leaf area index. Remote Sens. Environ. 195, 259–274.

Kulkarni, S., Rege, P.P., 2020. Pixel level fusion techniques for sar and optical images: a review. Inf. Fusion 59, 13–29.

Lang, N., Schindler, K., Wegner, J.D., 2019. Country-wide high-resolution vegetation height mapping with sentinel-2. Remote Sens. Environ. 233, 111347.

N. Lang, K. Schindler, J. D. Wegner, High Carbon Stock Mapping at Large Scale with Optical Satellite Imagery and Spaceborne Lidar, arXiv preprint arXiv:2107.07431.

LaRue, E.A., Wagner, F.W., Fei, S., Atkins, J.W., Fahey, R.T., Gough, C.M., Hardiman, B. S., 2020. Compatibility of aerial and terrestrial lidar for quantifying forest structural diversity. Rem. Sens. 12 (9), 1407.

Laurin, G.V., Balling, J., Corona, P., Mattioli, W., Papale, D., Puletti, N., Rizzo, M., Truckenbrodt, J., Urban, M., 2018. Above-ground biomass prediction by sentinel-1 multitemporal data in central Italy with integration of alos2 and sentinel-2 data. J. Appl. Remote Sens. 12 (1), 016008.

Liu, G., Reda, F.A., Shih, K.J., Wang, T.-C., Tao, A., Catanzaro, B., 2018. Image inpainting for irregular holes using partial convolutions. In: Proceedings of the European Conference on Computer Vision. ECCV).

Liu, G., Dundar, A., Shih, K.J., Wang, T., Reda, F.A., Sapra, K., Yu, Z., Yang, X., Tao, A., Catanzaro, B., 2023. Partial convolution for padding, inpainting, and image synthesis. IEEE Transactions on Pattern Analysis 45 (5), 6096–6110. https://doi.org/10.1109/TPAMI.2022.3209702.

Loew, A., Mauser, W., 2007. Generation of geometrically and radiometrically terrain corrected sar image products. Remote Sens. Environ. 106 (3), 337–349.

Norwegian institute of bioeconomy research (nibio). https://kilden.nibio.no/?topic=skogportal. (Accessed 18 January 2022).

Oettel, J., Lapin, K., 2021. Linking forest management and biodiversity indicators to strengthen sustainable forest management in europe. Ecol. Indicat. 122, 107275.

Omar, H., Misman, M., Kassim, A., 2017. Synergetic of palsar-2 and sentinel-1a sar polarimetry for retrieving aboveground biomass in dipterocarp forest of Malaysia. Appl. Sci. 7 (7), 675. https://doi.org/10.3390/app7070675, 10.3390/ap.7070675.

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M.C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C.E., et al., 2021. Mapping global forest canopy height through integration of gedi and landsat data. Remote Sens. Environ. 253, 112165.

Ronneberger, O., Fischer, P., Brox, T., U-net, 2015. Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring Vegetation Systems in the Great Plains with Erts.

Safari, A., Sohrabi, H., 2020. Integration of synthetic aperture radar and multispectral data for aboveground biomass retrieval in zagros oak forests, Iran: an attempt on sentinel imagery. Int. J. Rem. Sens. 41 (20), 8069–8095.

Sentinel application platform (snap). https://step.esa.int/main/download/snap-download/. (Accessed 28 January 2022).

Shao, Z., Zhang, L., Wang, L., 2017. Stacked sparse autoencoder modeling using the synergy of airborne lidar and satellite optical and sar data to map forest above-ground biomass. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 10 (12), 5569–5582.

Sommervold, O., Gazzea, M., Arghandeh, R., 2023. A survey on sar and optical satellite image registration. Rem. Sens. 15 (3), 850. https://doi.org/10.3390/rs15030850.

Stovall, A.E., Fatoyinbo, T., Thomas, N.M., Armston, J., Ebanega, M.O., Simard, M., Trettin, C., Zogo, R.V.O., Aken, I.A., Debina, M., et al., 2021. Comprehensive comparison of airborne and spaceborne sar and lidar estimates of forest structure in the tallest mangrove forest on earth. Science of Remote Sensing 4, 100034.

Tinkham, W.T., Dickinson, Y., Hoffman, C.M., Battaglia, M.A., Ex, S., Underhill, J., 2017. Tech. rep. https://doi.org/10.2737/2Frmrs-gtr-365, 10.2737/rmrs-gtr-365.

Vavriv, D.M., Bezvesilniy, O.O., 2013. Advantages of multi-look sar processing. In: 2013 IX Internatioal Conference on Antenna Theory and Techniques, pp. 217–219. https://doi.org/10.1109/ICATT.2013.6650730.

Waldeland, A.U., Trier, Ø.D., Salberg, A.-B., 2022. Forest mapping and monitoring in africa using sentinel-2 data and deep learning. Int. J. Appl. Earth Obs. Geoinf. 111, 102840.

Zhao, F., Sun, R., Zhong, L., Meng, R., Huang, C., Zeng, X., Wang, M., Li, Y., Wang, Z., 2022. Monthly mapping of forest harvesting using dense time series sentinel-1 sar imagery and deep learning. Remote Sens. Environ. 269, 112822.