https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

Mapping forest canopy height over Europe by integrating Sentinel-1, Sentinel-2, GEDI, and ICESat-2 data

Wanda De Keersmaecker[1], Astrid Verhegghen[1], Luc Bertels[1], Cornelius Senf[2], Alba Viana-Soto[2], Pieter Johannes Verkerk[3], Daniele Zanaga[1], Ruben Van De Kerchove[1]

[1]Vlaamse Instelling voor Technologisch Onderzoek (VITO), Mol, 2400, Belgium
[2]Technical University of Munich, School of Life Sciences, Earth Observation for Ecosystem Management, Freising, 85354, Germany
[3]European Forest Institute, Joensuu, 80100, Finland

Correspondence to: Wanda De Keersmaecker (wanda.dekeersmaecker@vito.be)

Abstract. Timely, accurate, and spatial explicit information on forest structure, such as canopy height, is important to understand and respond to ongoing changes in forests and to support the mapping of habitat structure. The availability of spaceborne LiDAR data, such as those from GEDI, has stimulated the development of continental to global canopy height maps. Yet, while GEDI data are often used to train canopy height models, these data are lacking in northern areas. In this study, we mapped canopy height over Europe at 10 m resolution by combining Sentinel-1 and Sentinel-2 data and integrating training data from GEDI and ICESat-2. The integration of ICESat-2 and GEDI data mostly enhanced the model performance in the north of Europe, where GEDI data are lacking. The model reached a RMSE of 5.77 m and a MAE of 4.09 m based on an independent validation with ALS data over about 3,700 patches across Europe. The resulting canopy height map and validation dataset have been made publicly available at https://doi.org/10.5281/zenodo.13324731 and https://doi.org/10.5281/zenodo.18471620, respectively.

1 Introduction

Forests cover about 39% of the European land area and provide jobs, income, and many other benefits to society, including a wide range of ecosystem services (Felipe-Lucia et al., 2018; Lecina-Diaz et al., 2024). Europe's forests also play an important role in the carbon cycle and climate regulation (Taye et al., 2021), and have acted as a net carbon sink for decades (Ciais et al., 2008; Pan et al., 2011). Moreover, they are crucial for Europe's ambition to reduce greenhouse gas emissions by 2030 and reach net-zero emissions by 2050 (Verkerk et al., 2022). Finally, forests host a wide range of terrestrial biodiversity and are crucial in preventing further biodiversity loss (Muys et al. 2022). Yet, European forests are faced with increasing timber harvest rates and changing natural disturbance regimes, induced by climate change and facilitated by past land use decisions (Patacca et al., 2023; Seidl et al., 2014; Breidenbach et al., 2022; Seidl and Senf, 2024; Mcdowell et al., 2020). To understand and respond to these complex changes, timely, accurate, and spatially explicit information on the forest state and structure is needed. Forest canopy height is an important structural metric, as it is closely related to aboveground biomass and carbon stock

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

(Duncanson et al., 2022), and forms an important descriptor of habitat structure (Duncanson et al., 2022; Moudrý et al., 2023). Hence, spatial explicit information on forest structure can contribute to forest monitoring and support international reporting on forests.

Traditionally, forest canopy height has been monitored using field inventories, and -more recently- by using Airborne Laser Scanning (ALS) data (White et al., 2016). Although these data are highly valuable, they typically cover a limited geographic extent and have a limited update frequency, due to practical and economic reasons. To enable wall-to-wall mapping of canopy height over continental to global extents, canopy structure has been estimated from satellite imagery using machine learning approaches, which were often calibrated with ALS, or spaceborne LiDAR data (Tolan et al., 2024; Liu et al., 2023; Lang et al., 2023; Pauls et al., 2025; Potapov et al., 2021). In fact, the launch of spaceborne LiDAR data, such as acquired by the Global Ecosystem Dynamics Investigation (GEDI) instrument or the Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2) mission, facilitated the continuous, large-scale mapping and monitoring of canopy height. GEDI is a lidar instrument that is mounted on the International Space Station (ISS), quantifying the vertical distribution of the vegetation using a full-waveform system, from which canopy height measures can be derived. The instrument covers the Earth surface between 51.6°N and 51.6°S using footprints of about 25m in diameter (Dubayah et al., 2020). In contrast to GEDI, the Advanced Topographic Laser Altimeter System (ATLAS) instrument onboard ICEsat-2 is a photon-counting laser altimeter that was mainly optimized to monitor glaciers, sea ice, and ice sheets and covers the Earth surface between 88°N and 88°S (Abdalati et al., 2010).

These spaceborne LiDAR datasets have been intensively used to train classical machine learning and deep learning models to map forest structure from satellite imagery. For example, Potapov et al. (2021) made a global canopy height map at 30m resolution by modelling the 95th percentile canopy height from GEDI using annual Landsat metrics and random forest models. Over Europe, Turubanova et al. (2023) mapped changes in canopy height and extent between 2001 and 2021 from the Landsat archive. This model was calibrated using a combination of GEDI and ALS data. Several studies used deep learning approaches to map canopy height. Lang et al. (2023), for instance, trained a U-Net model using the 98th percentile height from GEDI and Sentinel-2 data, resulting in a global canopy height map at 10 m resolution. Pauls et al. (2024) also trained a U-Net architecture to model and map canopy height globally at 10m resolution from Sentinel-1 and Sentinel-2 data using the RH100 metric from GEDI. Their study focused on optimizing preprocessing techniques and employed a novel loss function to account for geolocation inaccuracies. Over Europe, Pauls et al. (2025) additionally mapped canopy height and canopy height changes using a three-dimensional U-Net architecture. Finally, several studies focused on deep learning approaches applied to high-resolution imagery. Liu et al. (2023) trained a deep learning model on ALS data and PlanetScope imagery to produce 3m resolution canopy height maps across Europe. Meta and World Resources Institute launched a global canopy height map at one meter resolution using natural color imagery from Maxar Technologies. The map is the result of the combination of low resolution predictions from a convolutional network trained on GEDI RH95 data and high resolution predictions from an DPT decoder trained with aerial LiDAR data (Tolan et al., 2024). Finally, Wagner et al. (2025) mapped canopy height over the

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

Amazon forest at about 4.78 m spatial resolution using a U-Net model on Planet NICFI data. Given that ICESat-2 was mainly designed to monitor glaciers, sea ice, and ice sheets, most studies focus on GEDI and ALS data to train canopy height models. Yet, since these data do not cover the full extent of the European land area, many canopy height maps either extrapolate beyond the area covered by the training data or provide estimates over a restricted area.

The existing canopy height maps were often validated by splitting off a subset of the training dataset for validation. This subset is not used for training (i.e., a test set) (Pauls et al., 2024; Pauls et al., 2026). Although this approach is correct, many countries provide freely available ALS data over Europe, offering an opportunity to build a large-scale canopy height validation dataset that is fully independent of the training data (Moudrý et al., 2026). Yet, while several studies validated their maps using an independent ALS dataset, this often covered a restricted area over Europe (Moudrý et al., 2024; Potapov et al., 2021; Lang et al., 2023).

To address these limitations, our overall aim was to model canopy height of woody vegetation at 10 m resolution for the year 2020 using Copernicus Sentinel-1 and Sentinel-2 data across Europe. Specifically, our objectives were to:

(i) Compare models trained using GEDI, ICESat-2, and the combination of GEDI and ICESat-2 data,
(ii) Assess the contribution of Sentinel-1 and Sentinel-2 features, and
(iii) Evaluate the models across Europe using ALS data

2 Materials and methods

2.1. Area of interest

This study focused on mapping canopy height of woody vegetation, i.e., trees and shrubs, over Europe. Our study included the 27 members of the European Union, Norway, the United Kingdom, Switzerland, Bosnia and Herzegovina, Serbia, Montenegro, Albania, Kosovo, and North Macedonia. Forests in Europe cover a variety of ecological zones, ranging from north to south over boreal forests, temperate, and Mediterranean forests, as well as mountain systems (De Rigo et al., 2016). Europe's forests are mostly even-aged and semi-natural, i.e. their structure, composition, and processes are being managed or disturbed by humans (Eea, 2025).

2.1 Spaceborne LiDAR data

2.1.1 GEDI

To train the canopy height model, we collected spaceborne LiDAR data acquired by Global Ecosystem Dynamics Investigation (GEDI) (Dubayah et al., 2020). GEDI L2A relative height metrics (RH98) were collected over the months April until October

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

and of the years 2019-2021. The data were filtered to keep high-quality observations (based on their quality flag) with a sensitivity larger than 0.95 and without degradation, acquired during nighttime, during the leaf-on season (i.e., based on the leaf-on flag), and using a power beam.

Since LiDAR observations close to buildings may introduce noise to the dataset, these observations were removed using a 50m by 50m window based on OpenStreetMap building data from the year 2020. Moreover, since LiDAR observations are known to be affected by topography (Adam et al., 2020; Fayad et al., 2021), leading to height overestimations, we further excluded observations with a slope larger than 10% using the 30m resolution Copernicus Global DEM (Copernicus, 2019). Finally, noisy observations were further reduced by excluding observations with a height larger than 60m.

2.1.1 ICESat-2

Since the GEDI data do not cover the northern part of Europe, we additionally collected relative height metrics derived from Ice, Cloud and land Elevation Satellite-2 (ICESat-2) observatory data (ATL08 product v005) to train a canopy height model. The ICESat-2 observatory and Advanced Topographic Laser Altimeter System (ATLAS) instrument utilize a photon-counting lidar that covers the Earth between a latitude of approximately 88° N to 88° S. The instrument covers three pairs of spots on the ground, forming six ground tracks that are typically about 14m wide. The beams have different transmit energies (i.e. weak and strong beams), with an energy ratio between the weak and strong beams of about one against four (Neuenschwander et al., 2019).

We used the 98% height of all the individual relative canopy heights (height above terrain) for the 20m geosegments of the ATL08 product as a measure of canopy height. Heights are only provided for segments that have sufficient photons, i.e., at least 10 signal photons and 3 canopy photons are required. Similar to GEDI L2A data, we collected the ICESat-2 observations over Europe for the months April – September over the years 2019 – 2021. Moreover, we only selected observations from strong beams, that were not located close to buildings, were not located on slopes higher than 10%, and had a height lower than 60m.

2.2 Predictor variables

2.2.1 Sentinel-1 and -2 features

Yearly temporal statistics derived from Copernicus Sentinel-1 and Sentinel-2 data were used as predictor variables in the canopy height models. The Sentinel-2 features were derived from Sentinel 2 Level 2A products for 2020, which were first selected and filtered for cloud cover. We excluded clouds, cloud shadows, cirrus, and saturated pixels from the Sentinel-2 bands based on the scene classification layer of the L2A product. To further reduce residual noise in the time series, ten days median composites with a moving window of twenty days were subsequently computed. The remaining missing values were filled using a linear interpolation of neighboring available pixels in the time series. After pre-processing the Sentinel-2 data, a

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

set of vegetation indices was calculated, including the NDVI, EVI, NIRV, NDWI, NDGI, NDMI, NBR, NBR2, REP, ANIR, NDRE2, and NDRE3 (see supplement S.1). Next, the time series of pre-processed Sentinel-2 data and vegetation indices were used to derive a set of descriptive statistics, i.e. features. These include the 10^{th}, 50^{th}, 90^{th}, and interquantile range (difference between 75^{th} and 25^{th} percentile) for each band (B02, B03, B04, B05, B06, B07, B08, B11, and B12) and vegetation index. In addition, NDVI time series that were downsampled to 6 timestamps using Fourier methods were added as feature.

The Sentinel-1 features were derived in a similar fashion to the Sentinel-2 features. The Sentinel-1 Ground Range Detected (GRD) products were first corrected to Gamma0 backscatter and geocoded to the Sentinel-2 grid. Maximum one observation per day and per tile was processed to reduce data redundancy. A multitemporal speckle filter was then applied to reduce noise, after which the time series were aggregated into ten daily composites. Finally, the percentiles and interquartile range were computed on the Sentinel-1 VV and VH backscatter, the difference between the VH and VV backscatter, and the radar vegetation index (RVI) time series.

2.2.1 Localization features

Next to the Sentinel-1 and Sentinel-2 data, several features were used to enable the model to localize in space, including the terrain height from the 30m resolution Copernicus GLO-30 Digital Elevation model (DEM) (Copernicus, 2019) and the latitude and longitude of the observation. The inclusion of such localizing features facilitates the usage of a single model instead of multiple regional models and helps to improve the prediction accuracy (Lang et al., 2023).

2.3 Model calibration

Separate canopy height models were trained for each of the spaceborne LiDAR datasets, i.e., using GEDI, ICESat-2, and the combination of GEDI and ICESat-2 data (further called ‘combined model’). Training, validation, and test data were sampled using a fixed 100x100 km grid. Since the spaceborne LiDAR data may show high values for the relative height metric outside areas covered by woody vegetation (e.g., due to the presence of buildings, tall crops, or water bodies) and hence do not always represent canopy height in these areas, we used a separate sampling approach inside and outside areas covered by woody vegetation. Per grid cell, we sampled up to 2500 observations from woody areas and targeted a 0.4 ratio of locations outside to within woody areas. Woody areas were delineated based on tree and shrub areas within the ESA WorldCover land cover map and the Copernicus High Resolution forest type layer. 80% of the grid cells were then assigned to training, 10% to validation, and 10% to testing.

Areas covered with woody vegetation were delineated based on the intersection of the Copernicus High Resolution forest type layer of 2018 and the tree and shrub areas within the ESA WorldCover land cover map of 2020 and 2021. Over these areas, relative height observations were sampled from locations with GEDI or ICESat-2 data. For each of the samples, the

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

independent variables (i.e., the features) were extracted by taking the mean over a 3x3 10m pixel window to represent the 25m GEDI observations and a single 10m pixel to represent the ICESat-2 observations.

Areas outside woody vegetation were delineated according to the intersection of the no-forest areas within the Copernicus High Resolution layer of 2018 and the areas without woody vegetation (trees or shrubs) within the ESA WorldCover land cover map of 2020 and 2021 (Zanaga et al., 2021; Zanaga et al., 2022; Eea, 2020). Over areas without woody vegetation, the relative height metric was set to zero. To ensure that the samples outside woody vegetation represent different land cover types without oversampling dominant classes, the sampling was performed using the ESA WorldCover land cover map of 2020. We sampled 10% of the pixels per land cover type, with a maximum of 50 samples per land cover type within a 10x10 km area.

Next, a multi-quantile CatBoost regression model was trained using the set of Sentinel-1 and Sentinel-2 temporal features of the year 2020, DEM, and location (i.e. latitude and longitude) as independent variables. To avoid model overfitting on the location, we added noise to the latitude and longitude that was randomly sampled between -1 and 1 degrees. Moreover, since the number of data samples was not equally distributed over the range of the canopy height metric, we computed a histogram and assigned weights inversely proportional to the square root of the number of observations in each bin. To avoid that height bins with very few observations would result in very high weights and lead to overfitting, we saturated the weight for heights higher than 35 m.

A multi-quantile loss function was applied to enable the model to predict multiple target quantiles, allowing to gain information about the conditional target distribution and providing an uncertainty estimate. Similar to Poggio et al. (2021), three quantiles were used: 0.05 ($q_{0.05}$), 0.5 (median, $q_{0.5}$), and 0.95 ($q_{0.95}$). The 0.5 quantile ($q_{0.5}$) predictions were used to estimate canopy height, where negative estimation values were set to zero. Uncertainty was mapped using the 90th prediction interval (PI_{90}) and prediction interval ratio (PIR) as indictor (Eq. 1 and 2).

$$PI_{90} = q_{0.95} - q_{0.05} \qquad \text{(Eq 1)}$$

$$PIR = \frac{q_{0.95} - q_{0.05}}{q_{0.5}} \qquad \text{(Eq 2)}$$

To assess the impact of the input features on the accuracy of the canopy height map, we trained different models using the combination of GEDI and ICESat-2 data and varying sets of input features (Table 1). The canopy height models were evaluated over the test set using the root mean squared error (RMSE), coefficient of determination (R2), Mean Absolute Error (MAE), and Mean Bias Estimate (MBE).

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

Table 1: Sets of input features used for benchmarking. B02 (blue), B03 (green), B04 (red), B05, B06, B07 (red edge), B08 (NIR), B11, and B12 (SWIR) represent the different Sentinel-2 bands.

Name	Input features
S1	Localizing and S-1 features
S2-4B	Localizing and S-2 features for 4 bands: B02, B03, B04, and B08
S2-6B	Localizing and S-2 features for 6 bands: B02, B03, B04, B08, B11, and B12
S2-9B	Localizing and S-2 features for 9 bands: B02, B03, B04, B08, B11, B12, B05, B06, and B07
S1S2-9B	Localizing, S-1, and S-2 features for 9 bands: B02, B03, B04, B08, B11, B12, B05, B06, and B07
All	All input features

2.4 Validation with ALS data

To independently validate the canopy height datasets, a validation dataset was prepared using freely available Airborne Laser Scanning (ALS) data over a set of 640 by 640 m patches across Europe (see supplement S.2 – S.4 for more information). The patches were selected to (i) be equally distributed across Europe, and (ii) cover areas that are partially covered by trees or has been covered by trees in the past. To ensure an equal distribution of patches across Europe, we defined a grid of 100 by 100 km cells over Europe and tried to sample ten patches within each grid cell. We selected patch locations at least 50% covered by trees as defined by the union of the Copernicus High Resolution Forest Type layers of 2012, 2015, and 2018. Yet, since the ALS data over the patches are collected over different dates, the actual tree cover over the patch at the flight date might be lower than 50%. In addition, we noticed that for a minority of the patches also urban areas with high tree cover were included.

ALS data were collected over each of the selected patch locations, provided that freely available data in the form of LAZ-files were available. Since each of the selected patches can be covered by one or more LAZ-files, we first collected the LAZ files per patch, clipped the data to the dimensions of the patch, selected the first returns, and merged them into a single LAZ file. The point cloud data per patch were subsequently normalized, resulting in the above-ground height of each LiDAR point. The normalization uses the a-priori classification and identification of ground points within the downloaded point clouds. These ground points were triangulated into a ground triangular irregular network (TIN) and the elevation of each point with respect to this TIN was then computed. Since the LAZ-files are commonly in a local coordinate system, the LiDAR point cloud data were reprojected to the Universal Transverse Mercator (UTM) projection, aligning with the Copernicus Sentinel-2 grid. The resulting LiDAR point cloud may still contain noise, such as isolated points due to power lines or birds. To reduce the effect of these noisy, isolated points on the vegetation structure metrics, a denoising operation was applied. Here, points are eliminated if they have less than f_d other points in their surrounding 3 by 3 by 3 grid of cells, each having a size of 2 by 2 by 1 meter. The most suitable denoising parameter f_d depends on the characteristics of the data (such as point density) as was thus defined per country. Finally, the data was filtered by flight date to maximize dataset consistency. The LiDAR point cloud data

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

of a patch may originate from different flight lines gathered on various dates. To reduce this effect, points from flights with a time difference exceeding 30 days since the most recent flight date were excluded, utilizing only the data from the most recent flight unless the difference is less than or equal to 30 days.

The height at the 98th percentile, i.e. the height at which 98 percent of the first returns above a height cutoff of zero meter has been recorded, was then derived from the LiDAR point cloud as a measure of canopy height. The canopy height metric was calculated on a grid of 10 by 10 meter, aligning with the grid and spatial resolution of Copernicus Sentinel-2 data and resulting in a raster size of 64 by 64 pixels. Additionally, per-pixel metadata such as point density and flight date were recorded alongside the vegetation structure metrics. This resulted in about 3700 patches across 18 countries in Europe. The patches cover a gradient from the north of Europe (Scandinavia) to the south of Europe (Spain) but show a gap in the Balkan countries and Italy (Figure 1).

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

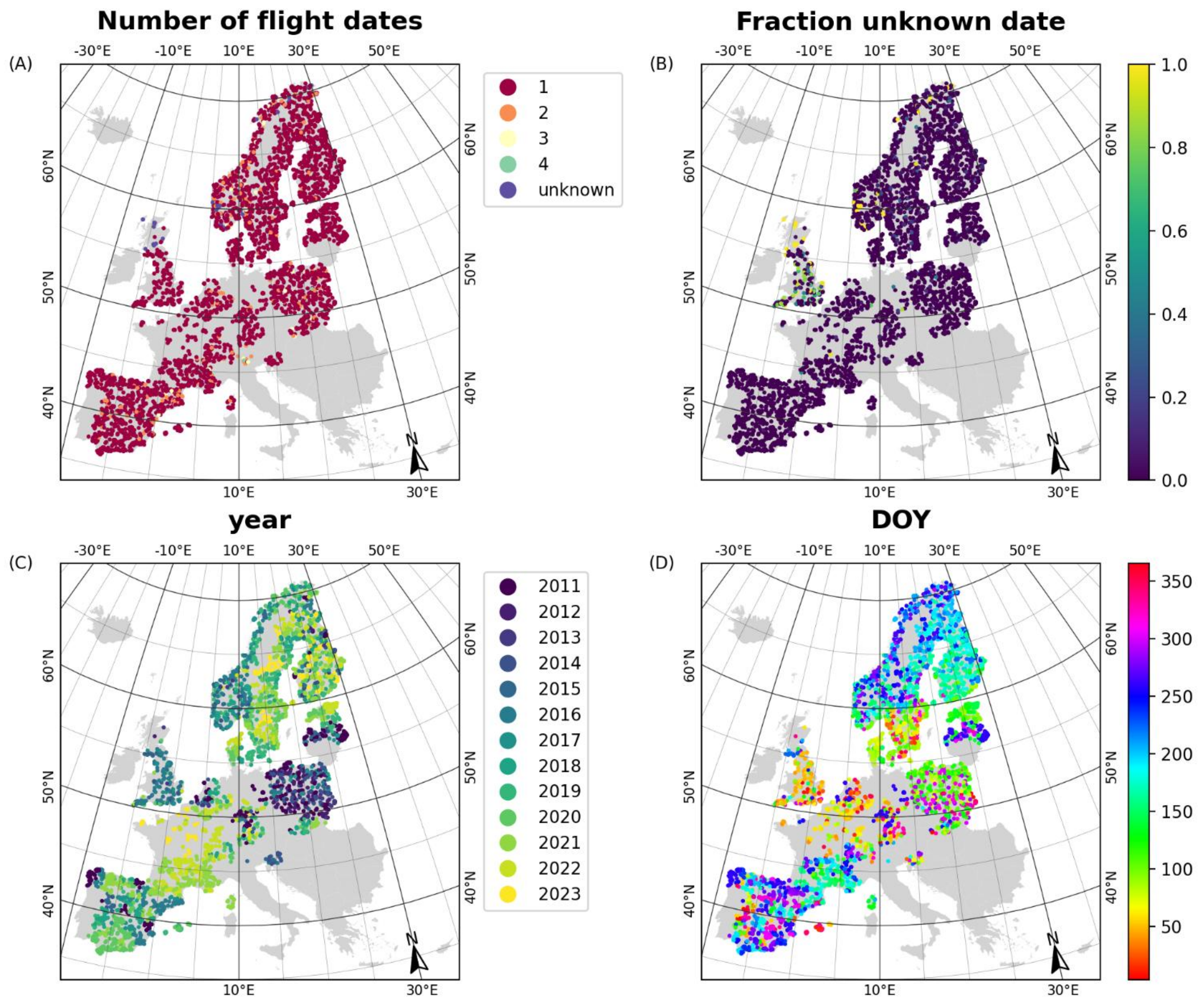

Figure 1: Flight days for the ALS patches. The number of flight days within each patch (A), the fraction of pixels within each patch with an unknown date (B), the flight year for patches with a single flight date (C), and the DOY for patches with a single flight date (D).

The validation dataset was constructed by sampling ten pixels from each ALS patch. Since the relative height metric might be impacted by the presence of buildings, water, or steep slopes, we excluded ALS data that was flagged as built-up areas or water in the WorldCover V200 dataset (Zanaga et al., 2022) and slopes larger than 20%. As the GEDI data only covers an area

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

up to 51.6° N, we evaluated the canopy height models using (i) all ALS patches, (ii) the patches north of 51.6°N (i.e. outside the area covered by GEDI training data), and (iii) the patches south of 51.6°N.

3 Results

3.1 Comparing ICESat-2 and GEDI observations

Comparing the ICESat-2 and GEDI observations with the ALS data shows a strong agreement between the datasets, reaching an R^2 between 0.47 and 0.61 for the GEDI and ICESat-2 data, respectively. The two spaceborne lidar datasets show a MAE between 3.5 and 4 m and a RMSE between 5.5 and 6.5 m.

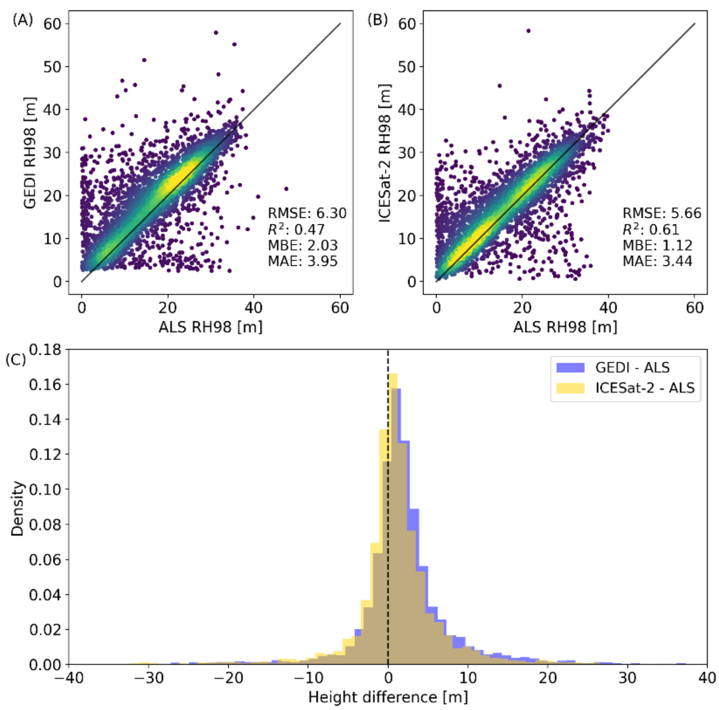

Figure 2: Comparison of the GEDI and ICESat-2 observations: scatterplots of the GEDI and ALS-based canopy height (A), ICESat-2 and ALS canopy height (B), and density plots of the canopy height difference between the spaceborne and the ALS data (C). Only observations south of 51.6°N with a canopy height less than 60 m have been considered.

3.2 Evaluation of the canopy height maps with ALS data

Independent validation of the canopy height maps with ALS data shows that the maps reach a coefficient of determination around 0.55, a RMSE of 5.8 m and MAE of 4.1 m (Figure 3). The combined model performed slightly better than the models using GEDI or ICESat-2 data only. The model based on solely GEDI data shows an overestimation in the northern part of

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

Europe (above 51.6 degrees latitude), which is absent in the other models. In the southern part of Europe, the three models perform similarly.

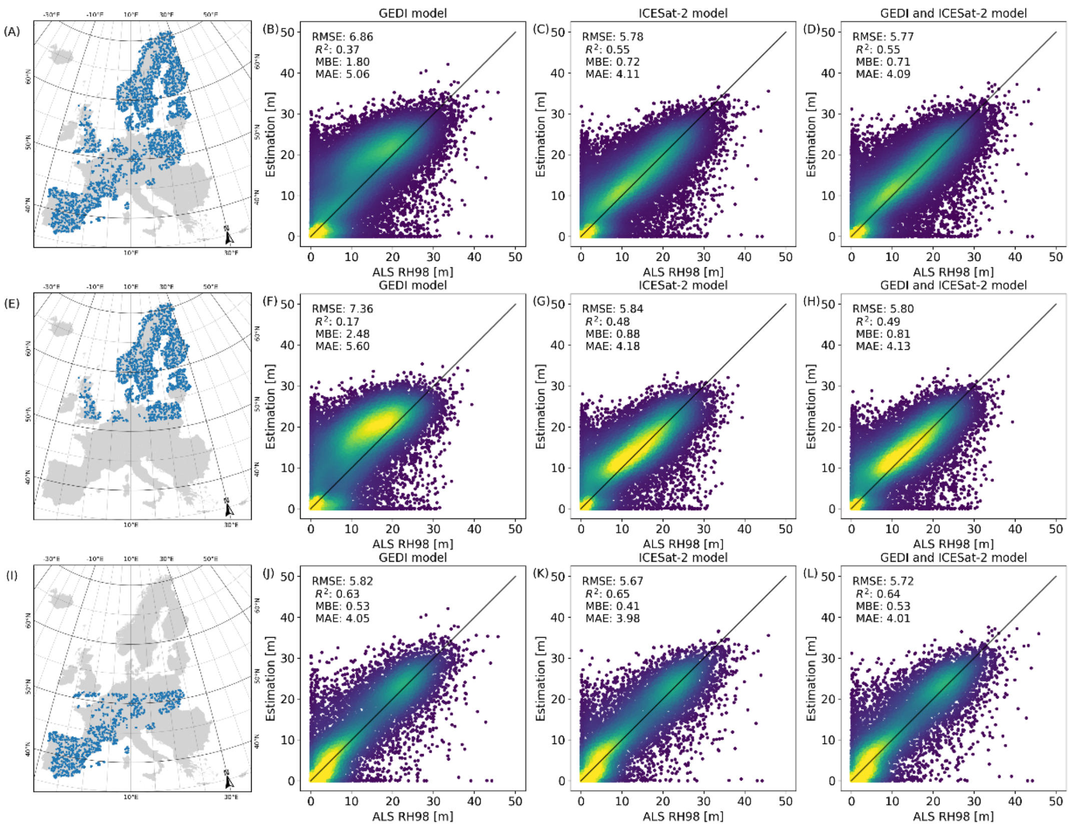

Figure 3: validation of the canopy height maps with ALS data over Europe (A-D), northern Europe (E-H), and southern Europe (I-L). Three maps were validated: using GEDI data only (B, F, J), ICESat-2 data only (C, G, K) and both (D, H, L).

In addition, the validation with ALS data indicated slightly better performance of our maps compared to existing canopy height maps (Figure 4 and Figure 5), such as those of Lang et al. (2023); Liu et al. (2023); Potapov et al. (2021). The GEDI and ICESat-2 model for instance reaches an RMSE around 5.77 m, while the RMSE of the other maps range between 6.11 and 7.21 m.

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

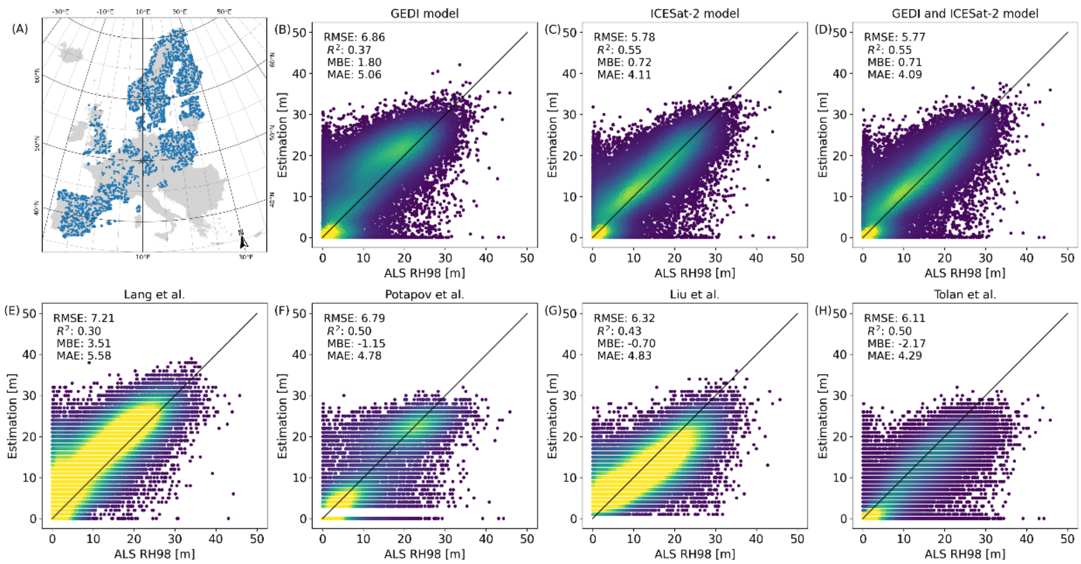

Figure 4: Evaluation of the canopy height maps with ALS data: based on GEDI data (A), ICESat-2 data (B), both GEDI and ICESat-2 data (C), and the canopy height maps of Lang et al. (2023) (D), Potapov et al. (2021) (E), Liu et al. (2023) (F), and Tolan et al. (2024) (H). A maximum value resampling was applied to the map of Tolan et al. (2024) to compare it with ALS data.

In general, our models show a lower error in the low canopy height ranges (0 – 10 m) than the maps of Lang et al. (2023) and Liu et al. (2023), which tend to overestimate canopy height. Within the higher canopy height ranges (30 – 40 m), all maps tend to underestimate canopy height. The underestimation of our maps is higher than those of Lang et al. (2023), but slightly lower than those of Tolan et al. (2024) and Liu et al. (2023).

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

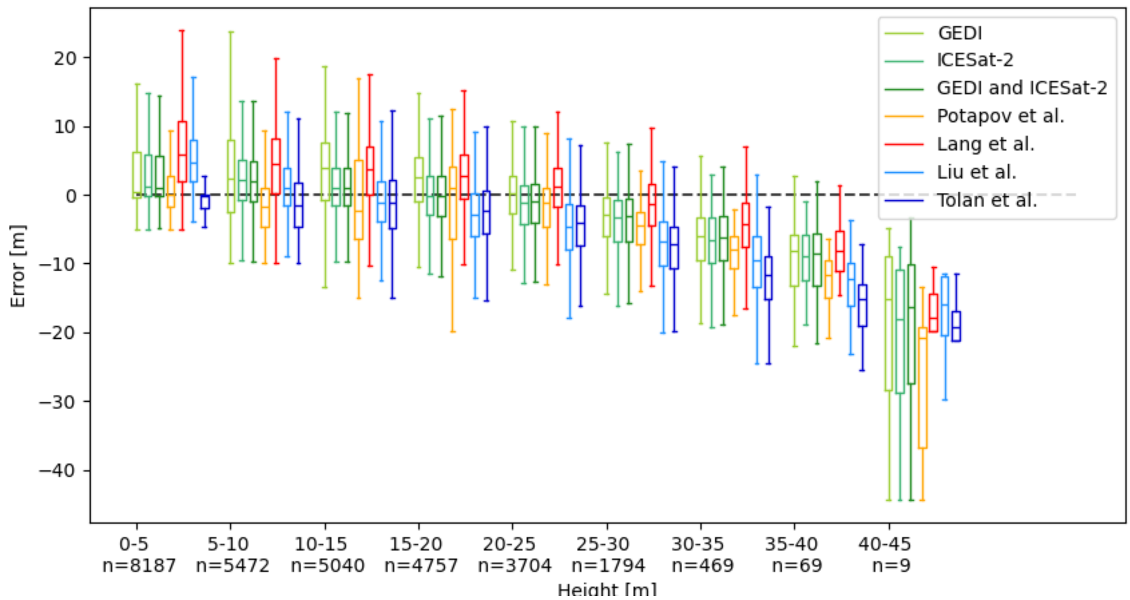

Figure 5: Error (estimated height – ALS height) per height interval. The number of observations is given along the x axis. For some datasets (e.g. those from Potapov et al.) this number might be smaller.

3.2 Canopy height maps

Training CatBoost models on the three datasets, i.e., (i) GEDI data only, (ii) ICESat-2 data only, and (iii) the combination results in one height prediction map for each of the datasets (Figure 6). Comparing these height maps shows that the model trained on GEDI data had higher height estimates over most of northern Europe (above 55 degrees latitude) than the model trained on both spaceborne LiDAR datasets. The difference between height estimates below 55 degrees latitude is smaller and mixed. The difference in canopy height estimates between the model trained on solely ICESat-2 data and the combined datasets is much smaller. This is consistent with validation using ALS data, which shows a height overestimation in northern areas when the model is trained exclusively on GEDI data—an effect that disappears when ICESat-2 data is included.

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

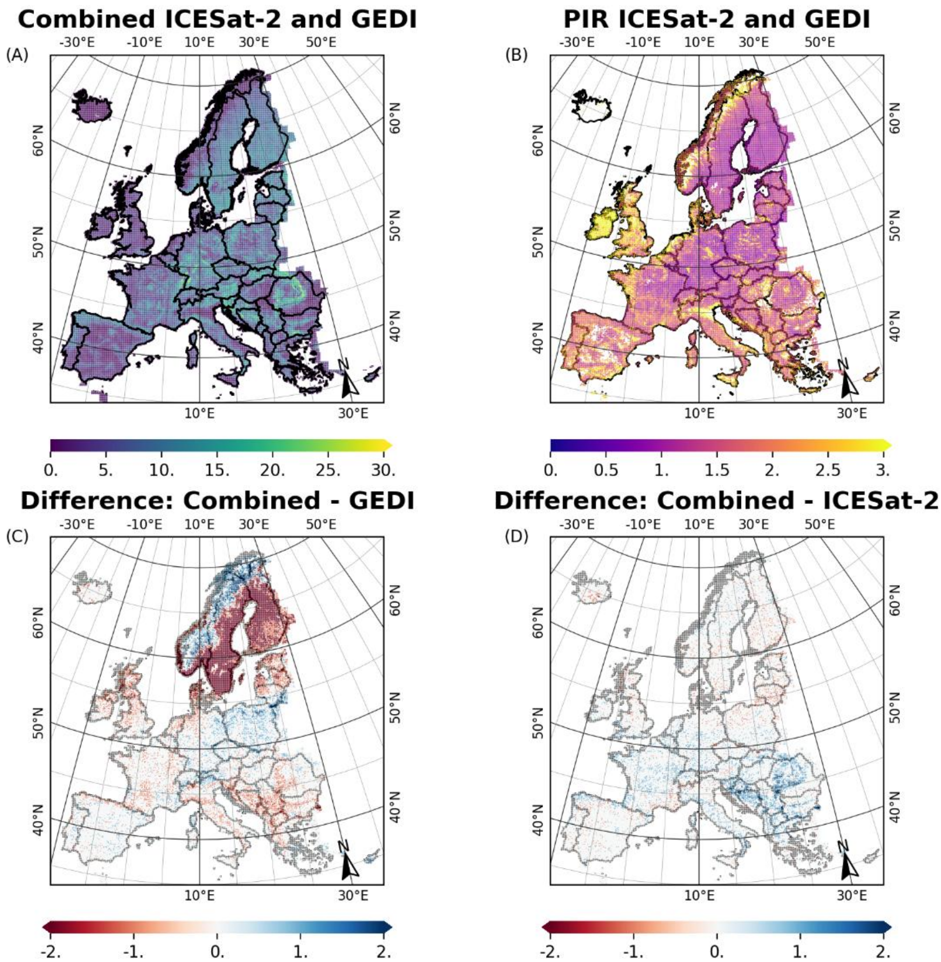

Figure 6: Canopy height map based on the combination of ICESat-2 and GEDI data (A), the prediction interval ratio (PIR) of the canopy height estimation using the combined data (B), the difference between the canopy height estimations using the combined data and GEDI data (C) the difference between the canopy height estimations produced by the model using the combined data and ICESat-2 data (D)

3.3 Effect of the input features on the accuracy metrics

The accuracy metrics (coefficient of determination and RMSE) show a rather gradual increase in accuracy with addition of features (Figure 7). The model using only Sentinel-1 and localizing features reached the lowest accuracy (R^2 and RMSE equal 0.63 and 5.69 m, respectively). The different groups of Sentinel-2 bands (RGB + NIR, SWIR, and red edge) and Sentinel-1 bands increased the accuracy of the model. Adding vegetation indices (model with all features) on top of the auxiliary, Sentinel-1 and nine Sentinel-2 bands (S2S2-B9 model) had a rather minor impact on accuracy.

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

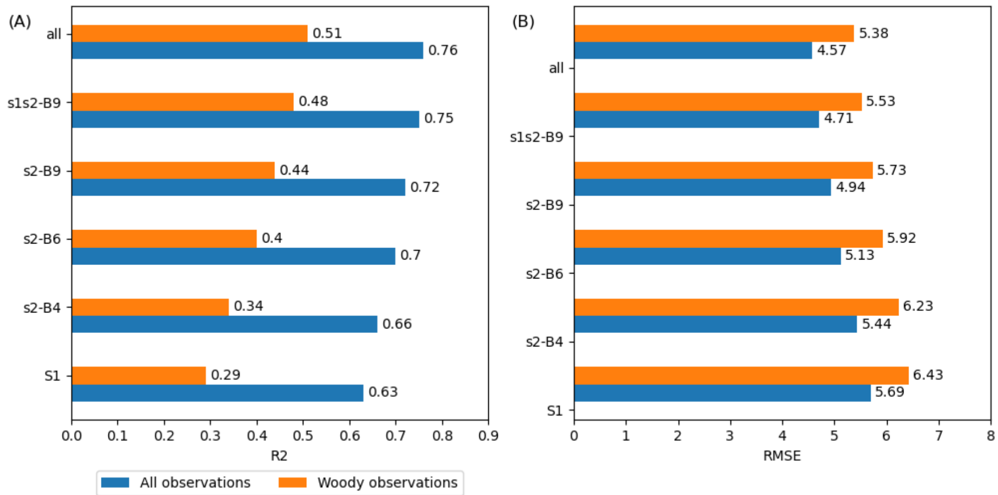

Figure 7: R^2 (A) and RMSE (B) for models trained with different sets of input features. All models use localizing features. S1: Sentinel-1 features, S2-B4: features of RGB, and NIR bands, S2-B6: features of RGB, NIR, and SWIR bands, s2-B9: RGB, NIR, SWIR, and red edge bands, all: all input features.

3.4 Visual comparison of canopy height maps

Visual inspection of the canopy height maps over a set of patches (Figure 8) indicates that our map is able to delineate fine spatial patterns (e.g. Figure 8B, C, and E). Yet, sparse woody vegetation or rows of single trees are better captured by the high-resolution map of Tolan et al. (2024). Second, the examples show that our map has a reasonable capacity to map tall trees, in line with the validation results shown in Figure 5. The map of Lang et al. (2023) shows the largest number of tall trees, followed by our map.

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

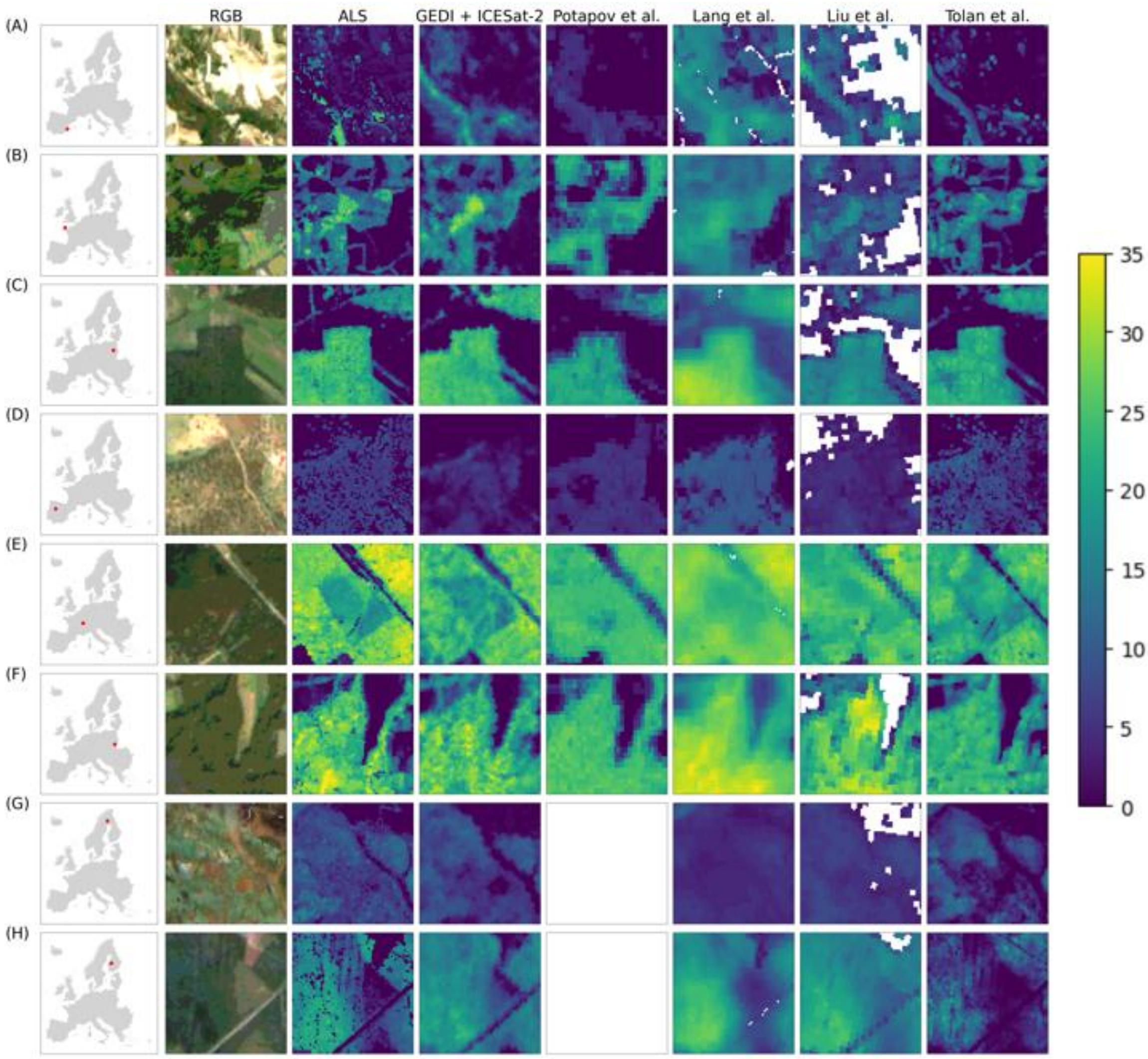

Figure 8: Examples of canopy height maps over patches across Europe. From left to right: patch location, annual median RGB Sentinel-2 composite, ALS-based height, canopy height estimate of the model trained with GEDI and ICESat-2, canopy height estimates of Potapov et al. (2021), Lang et al. (2023), Liu et al. (2023), and Tolan et al. (2024).

4 Discussion and conclusion

Our study shows the potential of combining GEDI and ICESat-2 data when mapping canopy height at continental scale covering a wide range of forest ecosystems. Although there might be differences in height estimates between the GEDI and ICESat-2 datasets, the combined use of spaceborne lidar datasets expanded the geographic coverage of training data, resulting in better height estimates over northern Europe. Liu et al. (2021) also compared GEDI and ICESat-2 observations with ALS data across multiple eco-climatic zones and found a RMSE around 3.93 and 3.56 m. Although they found that both datasets

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

showed a high correspondence with ALS data, GEDI performed better than ICESat-2, with a larger potential bias for ICESat-2.

The evaluation of canopy height maps over Europe with the independent ALS-based validation dataset indicated that our map trained with both GEDI and ICESat-2 data showed competitive or even better accuracies to other continental and global products (e.g. (Lang et al., 2023; Potapov et al., 2021; Tolan et al., 2024)). In addition, errors across height bins were relatively balanced, suggesting that the model accurately estimates canopy heights across low- to medium-high forest canopies. All existing canopy height maps show increasing errors with increasing tree height. Yet, the errors of our model are smaller than those of the other maps for high canopy height values, except for those of the map of Lang et al. (2023). The map of Lang et al. (2023), however, tends to overestimate height for low canopy height values, whereas the errors of our map are minimal within this height range.

Several factors may have contributed to the balanced error distribution. First, our canopy height models were specifically trained over Europe, while other canopy height maps have a global extent (e.g., those of Potapov et al. (2021) or Lang et al. (2023). The focus on a single continent decreases the data complexity and may have facilitated the canopy height mapping. Second, besides the quality flags associated with the spaceborne lidar data, an additional screening of observations over woody areas was performed to avoid errors due to e.g., steep slopes, neighboring buildings, or large errors around forest edges due to spatial inaccuracies. Over areas without woody vegetation, height overestimations from the spaceborne lidar datasets were avoided by setting the height to zero. Third, similar to Lang et al. (2023), we used localizing features and model weights dependent on the number of observations within the height bins. Whereas the former allows the model to localize and account for varying bio-climatic regions, the latter promotes the model to improve the height estimation for tall trees. Finally, the combination of seasonal Sentinel-1 and Sentinel-2 metrics, i.e. percentiles of the Sentinel data over the temporal profile, showed to promote the overall model accuracy.

It should, however, be noted that slightly different canopy height metrics were used in the maps. While we used the 98th percentile height as a canopy height metric to produce and evaluate the canopy height maps, the map of Potapov et al. (2021), for example, is based on the 95th percentile height. In addition, spatial resolution varies between the canopy height maps. Our 10m resolution map tends to better delineate fine spatial patterns than those of the 10 resolution map of Lang et al. (2023) or the 30m resolution map of Potapov et al. (2021). The map of Tolan et al. (2024) was produced at about 1.2 m, allowing to better detect single trees or lines of trees, which is particularly interesting in sparsely treed landscapes or urban areas. Liu et al. (2023) also produced a canopy height map at a higher resolution; however, their map was made publicly available only at 30 m resolution and was therefore evaluated in our study at that resolution.

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

In addition, an extensive ALS dataset, containing about 3700 patches across Europe, was collected to validate our canopy height map. The data covers a large share of Europe, except for Italy and the Balkan countries and spans diverse climate and forest zones. An important challenge in preparing the dataset was data acquisition. Since the ALS data were stored and maintained at regional or national levels, information about the data was sometimes only available in local languages, hampering to find the correct access points to the data. Moreover, information about the tiling grid was not always available or was provided in a non-georeferenced format. The means for accessing the data also varied significantly. In some countries, the data could easily be downloaded using a script, while in others, it requires manually clicking on download links. Hence, we anticipate that creating an openly available dataset of processed canopy height over Europe will facilitate the use of the European ALS data by scientists.

In summary, our study introduced a canopy height map for Europe trained using GEDI and ICESat-2 data. In addition, an extensive validation dataset of about 3700 640 by 640 m patches was constructed over Europe using freely available ALS data. According to our validation results, the addition of ICESat-2 data notably increased mapping accuracy in northern Europe, where GEDI data are lacking. Moreover, our map showed slightly better validation results compared to existing map products.

5 Code and data availability

The European canopy height map is available on Zenodo (https://doi.org/10.5281/zenodo.13324731) (De Keersmaecker et al., 2024). The validation dataset, i.e., containing 640 by 640 m patches with the canopy height at 98th percentile, point density, and the flight date, can also be accessed on Zenodo (https://doi.org/10.5281/zenodo.18471620) (Bertels et al., 2026).

Author contribution

WDK, RVDK, AVS, and CS contributed to the conceptualization. Data curation was performed by WDK, LB, and DZ. WDK did the formal analysis. AV, CS, AVS, PJV, and WDK contributed to the writing.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We would like to acknowledge the ForestPaths project (Co-designing Holistic Forest-based Policy Pathways for Climate Change Mitigation), that receives funding from the European Union's Horizon Europe Research and Innovation Programme

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

(ID No 101056755), as well as from the United Kingdom Research and Innovation Council (UKRI). In addition, we would like to thank the countries and institutions for sharing the ALS data.

References

Abdalati, W., Zwally, H. J., Bindschadler, R., Csatho, B., Farrell, S. L., Fricker, H. A., Harding, D., Kwok, R., Lefsky, M., and Markus, T.: The ICESat-2 laser altimetry mission, Proceedings of the IEEE, 98, 735–751, 2010.

Adam, M., Urbazaev, M., Dubois, C., and Schmullius, C.: Accuracy assessment of GEDI terrain elevation and canopy height estimates in European temperate forests: Influence of environmental and acquisition parameters, Remote Sensing, 12, 3948, 2020.

Bertels, L., De Keersmaecker, W., Verhegghen, A., Senf, C., Viana-Soto, A., Verkerk, P. J., Zanaga, D., and Van De Kerchove, R.: ALS-based canopy height across Europe [dataset], 10.5281/zenodo.18471620, 2026.

Breidenbach, J., Ellison, D., Petersson, H., Korhonen, K. T., Henttonen, H. M., Wallerman, J., Fridman, J., Gobakken, T., Astrup, R., and Næsset, E.: Harvested area did not increase abruptly—how advancements in satellite-based mapping led to erroneous conclusions, Annals of forest science, 79, 2, 2022.

Ciais, P., Schelhaas, M.-J., Zaehle, S., Piao, S., Cescatti, A., Liski, J., Luyssaert, S., Le-Maire, G., Schulze, E.-D., and Bouriaud, O.: Carbon accumulation in European forests, Nature geoscience, 1, 425–429, 2008.

Copernicus: Copernicus DEM [dataset], https://doi.org/10.5270/ESA-c5d3d65, 2019.

De Keersmaecker, W., Bertels, L., Zanaga, D., Viana-Soto, A., Senf, C., and Van De Kerchove, R.: ForestPaths: European canopy height map (v0.0.1) [dataset], 10.5281/zenodo.13324731, 2024.

De Rigo, D., Caudullo, G., Houston Durrant, T., and San-Miguel-Ayanz, J.: The European Atlas of Forest Tree Species: modelling, data and information on forest tree species, European atlas of forest tree species, 40–45, 2016.

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., and Luthcke, S.: The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography, Science of remote sensing, 1, 100002, 2020.

Duncanson, L., Kellner, J. R., Armston, J., Dubayah, R., Minor, D. M., Hancock, S., Healey, S. P., Patterson, P. L., Saarela, S., and Marselis, S.: Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission, Remote Sensing of Environment, 270, 112845, 2022.

EEA: Forest Type 2018 (raster 10 m), Europe, 3-yearly, Oct. 2020 [dataset], https://doi.org/10.2909/59b0620c-7bb4-4c82-b3ce-f16715573137, 2020.

Europe's forests: https://forest.eea.europa.eu/topics, last access: 12-12-2025.

Fayad, I., Baghdadi, N., Alcarde Alvares, C., Stape, J. L., Bailly, J. S., Scolforo, H. F., Cegatta, I. R., Zribi, M., and Le Maire, G.: Terrain slope effect on forest height and wood volume estimation from GEDI data, Remote Sensing, 13, 2136, 2021.

Felipe-Lucia, M. R., Soliveres, S., Penone, C., Manning, P., van der Plas, F., Boch, S., Prati, D., Ammer, C., Schall, P., and Gossner, M. M.: Multiple forest attributes underpin the supply of multiple ecosystem services, Nature communications, 9, 4839, 2018.

Lang, N., Jetz, W., Schindler, K., and Wegner, J. D.: A high-resolution canopy height model of the Earth, Nature Ecology & Evolution, 7, 1778–1789, 2023.

Lecina-Diaz, J., Senf, C., Grünig, M., and Seidl, R.: Ecosystem services at risk from disturbance in Europe's forests, Global change biology, 30, e17242, 2024.

Liu, A., Cheng, X., and Chen, Z.: Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals, Remote Sensing of Environment, 264, 112571, 2021.

Liu, S., Brandt, M., Nord-Larsen, T., Chave, J., Reiner, F., Lang, N., Tong, X., Ciais, P., Igel, C., and Li, S.: The overlooked contribution of trees outside forests to tree cover and woody biomass across Europe, Science Advances, 9, 2023.

McDowell, N. G., Allen, C. D., Anderson-Teixeira, K., Aukema, B. H., Bond-Lamberty, B., Chini, L., Clark, J. S., Dietze, M., Grossiord, C., and Hanbury-Brown, A.: Pervasive shifts in forest dynamics in a changing world, Science, 368, eaaz9463, 2020.

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

Moudrý, V., Cord, A. F., Gábor, L., Laurin, G. V., Barták, V., Gdulová, K., Malavasi, M., Rocchini, D., Stereńczak, K., and Prošek, J.: Vegetation structure derived from airborne laser scanning to assess species distribution and habitat suitability: The way forward, Diversity and Distributions, 29, 39–50, 2023.

Moudrý, V., Gábor, L., Marselis, S., Pracná, P., Barták, V., Prošek, J., Navrátilová, B., Novotný, J., Potůčková, M., and Gdulová, K.: Comparison of three global canopy height maps and their applicability to biodiversity modeling: Accuracy issues revealed, Ecosphere, 15, e70026, 2024.

Moudrý, V., Remelgado, R., Forkel, M., Torresani, M., Laurin, G. V., Šárovcová, E., Garcia Millan, V. E., Fischer, F. J., Jucker, T., and Gallay, M.: Spaceborne canopy height products should be complemented with airborne laser scanning data: Toward a European canopy height model, Earth and Space Science, 13, e2025EA004544, 2026.

Neuenschwander, A., Pitts, K., Jelley, B., Robbins, J., Markel, J., Popescu, S., Nelson, R., Harding, D., Pederson, D., and Klotz, B.: Ice, Cloud, and Land Elevation Satellite 2 (ICESat-2) algorithm theoretical basis document (ATBD) for land-vegetation along-track products (ATL08), National Aeronautics and Space Administration: Washington, DC, USA, 2019.

Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., Phillips, O. L., Shvidenko, A., Lewis, S. L., and Canadell, J. G.: A large and persistent carbon sink in the world's forests, science, 333, 988–993, 2011.

Patacca, M., Lindner, M., Lucas-Borja, M. E., Cordonnier, T., Fidej, G., Gardiner, B., Hauf, Y., Jasinevičius, G., Labonne, S., and Linkevičius, E.: Significant increase in natural disturbance impacts on European forests since 1950, Global change biology, 29, 1359–1376, 2023.

Pauls, J., Zimmer, M., Turan, B., Saatchi, S., Ciais, P., Pokutta, S., and Gieseke, F.: Capturing Temporal Dynamics in Large-Scale Canopy Tree Height Estimation, arXiv preprint arXiv:2501.19328, 2025.

Pauls, J., Zimmer, M., Kelly, U. M., Schwartz, M., Saatchi, S., Ciais, P., Pokutta, S., Brandt, M., and Gieseke, F.: Estimating Canopy Height at Scale, arXiv preprint arXiv:2406.01076, 2024.

Pauls, J., Schrödter, K., Ligensa, S., Schwartz, M., Turan, B., Zimmer, M., Saatchi, S., Pokutta, S., Ciais, P., and Gieseke, F.: ECHOSAT: Estimating Canopy Height Over Space And Time, arXiv preprint arXiv:2602.21421, 2026.

Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G. B., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, Soil, 7, 217–240, 2021.

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., and Silva, C. E.: Mapping global forest canopy height through integration of GEDI and Landsat data, Remote Sensing of Environment, 253, 112165, 2021.

Seidl, R. and Senf, C.: Changes in planned and unplanned canopy openings are linked in Europe's forests, Nature Communications, 15, 4741, 2024.

Seidl, R., Schelhaas, M.-J., Rammer, W., and Verkerk, P. J.: Increasing forest disturbances in Europe and their impact on carbon storage, Nature climate change, 4, 806–810, 2014.

Taye, F. A., Folkersen, M. V., Fleming, C. M., Buckwell, A., Mackey, B., Diwakar, K., Le, D., Hasan, S., and Saint Ange, C.: The economic values of global forest ecosystem services: A meta-analysis, Ecological Economics, 189, 107145, 2021.

Tolan, J., Yang, H.-I., Nosarzewski, B., Couairon, G., Vo, H. V., Brandt, J., Spore, J., Majumdar, S., Haziza, D., and Vamaraju, J.: Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar, Remote Sensing of Environment, 300, 113888, 2024.

Turubanova, S., Potapov, P., Hansen, M. C., Li, X., Tyukavina, A., Pickens, A. H., Hernandez-Serna, A., Arranz, A. P., Guerra-Hernandez, J., and Senf, C.: Tree canopy extent and height change in Europe, 2001–2021, quantified using Landsat data archive, Remote Sensing of Environment, 298, 113797, 2023.

Verkerk, H., Delacote, P., Hurmekoski, E., Kunttu, J., Matthews, R., Mäkipää, R., Mosley, F., Perugini, L., Reyer, C., and Roe, S.: Forest-based climate change mitigation and adaptation in Europe, 2022.

Wagner, F. H., Dalagnol, R., Carter, G., Hirye, M., Gill, S., Takougoum, L. B. S., Favrichon, S., Keller, M., Ometto, J. P., and Alves, L.: High Resolution Tree Height Mapping of the Amazon Forest using Planet NICFI Images and LiDAR-Informed U-Net Model, arXiv preprint arXiv:2501.10600, 2025.

White, J. C., Coops, N. C., Wulder, M. A., Vastaranta, M., Hilker, T., and Tompalski, P.: Remote sensing technologies for enhancing forest inventories: A review, Canadian Journal of Remote Sensing, 42, 619–641, 2016.

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.-E., Xu, P., Ramoino, F., and Arino, O.: ESA WorldCover 10 m 2021 v200 (Versie v200) [dataset], https://doi.org/10.5281/zenodo.7254221, 2022.

https://doi.org/10.5194/essd-2026-329
Preprint. Discussion started: 4 June 2026

Earth System Science Data
Open Access
Discussions

Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., Tsendbazar, N.-E., Ramoino, F., and Arino, O.: ESA WorldCover 10 m 2020 v100 (Versie v100) [dataset], 2021.