International Journal of Digital Earth

ISSN: 1753-8947 (Print) 1753-8955 (Online) Journal homepage: www.tandfonline.com/journals/tjde20

A modelling framework to estimate canopy height at local scale using optical and radar images paired with GEDI measurements in Mediterranean-like landscapes

João E. Pereira-Pires, Juan Guerra-Hernández, Adrián Pascual, Raffaella Guida, João M. N. Silva, José M. Fonseca & André Mora

To cite this article: João E. Pereira-Pires, Juan Guerra-Hernández, Adrián Pascual, Raffaella Guida, João M. N. Silva, José M. Fonseca & André Mora (2026) A modelling framework to estimate canopy height at local scale using optical and radar images paired with GEDI measurements in Mediterranean-like landscapes, International Journal of Digital Earth, 19:2, 2687201, DOI: 10.1080/17538947.2026.2687201

To link to this article: https://doi.org/10.1080/17538947.2026.2687201

Published online: 28 Jul 2026.

INTERNATIONAL JOURNAL OF DIGITAL EARTH
2026, VOL. 19, NO. 2, 2687201
https://doi.org/10.1080/17538947.2026.2687201

RESEARCH ARTICLE

OPEN ACCESS Check for updates

A modelling framework to estimate canopy height at local scale using optical and radar images paired with GEDI measurements in Mediterranean-like landscapes

João E. Pereira-Pires[a], Juan Guerra-Hernández[b], Adrián Pascual[c], Raffaella Guida[d], João M. N. Silva[b], José M. Fonseca[a] and André Mora[a]

[a]Centre of Technology and Systems/UNINOVA, and Associated Lab of Intelligent Systems, and the School of Science and Technology, NOVA University of Lisbon, Caparica, Portugal; [b]Forest Research Centre, Associate Laboratory TERRA, School of Agriculture, University of Lisbon, Lisbon, Portugal; [c]Department of Geographical Sciences, University of Maryland, College Park, MD, USA; [d]Surrey Space Centre, University of Surrey, Guildford, UK

ABSTRACT

The near-worldwide coverage of spaceborne LiDAR data from the GEDI offers unprecedented opportunities for mapping canopy height (CH). Notwithstanding the sensitivity of the GEDI to forests' vertical structure, it provides sparse sampling measurements, which hinder gap-free mapping. Several machine and deep learning models that resort to optical, radar, and GEDI have been tested to produce gap-free CH maps. Not all factors affecting the accuracy and consistency of these GEDI-fused products have been explored. Specifically, the sampling fraction coverage and the geolocation correction of footprints on-orbit positions have not been deeply studied. In this article, a collection of data from 15 study areas, characterised by Mediterranean landscapes, had their CH mapped resorting to Sentinel-1/2, ALOS-2, ancillary data, and a locally fitted extreme gradient boosting regressor. The produced maps had an average %RMSE of 32.41%, outperforming the other two global products in Mediterranean regions. Additionally, the geolocation correction of the GEDI footprints was limited to 1.23 percentage points. This experiment was able to demonstrate three key points: (1) the importance of including InSAR in the optical/SAR synergy; (2) the reduced impact of collocating GEDI footprints at the track level; and (3) the benefits of increasing GEDI-footprint coverage using multi-year data are hindered by temporal variation effects.

ARTICLE HISTORY
Received 29 December 2025
Revised 2 June 2026
Accepted 5 June 2026

KEYWORDS
Canopy height; optical; multifrequency SAR; spaceborne LiDAR; interferometric SAR

1. Introduction

Earth observation remote sensing plays a pivotal role in monitoring forest ecosystems (Xiao et al. 2019). An example of this is the Global Ecosystem Dynamics Investigation (GEDI), which was specifically designed to sense forest structure using small laser pulses capable of penetrating closed-canopy environments. GEDI has been used for mapping Canopy Height (CH) dynamics globally (Besic et al. 2025; Healey et al. 2020; Lang et al. 2022; Potapov et al. 2021) and regionally (Adam et al. 2020; Fayad et al. 2021; Li et al. 2023; Qi et al. 2025; Schwartz et al. 2022), by leveraging its spaceborne Light Detection and Ranging (LiDAR) sensor.

The sparse sampling nature of GEDI acquisitions and low availability of data hinder the creation of the gap-free maps (Pascual et al. 2025; Pereira-Pires et al. 2024; Qi et al. 2025). Therefore, many studies have proposed frameworks that combine GEDI, optical, and/or synthetic aperture radar (SAR) spaceborne sensors to produce spatially continuous CH maps (Healey et al. 2020; Lang et al. 2022; Pereira-Pires et al. 2024; Potapov et al. 2021; Schwartz et al. 2022). Potapov et al. (2021) and Turubanova et al. (2023) generated worldwide and European CH maps, respectively, with regression trees and Landsat observations. Alternatively, Lang et al. (2022) utilised convolutional neural networks and Sentinel-2 (S2) observations for global mapping. In both, each tile had a dedicated model with a slight difference in the calibration process: in Lang et al. (2022), the model fitting only resorted to footprints within the corresponding tile,

CONTACT João E. Pereira-Pires joaoepereirapires.work@gmail.com Centre of Technology and Systems/UNINOVA, and Associated Lab of Intelligent Systems, and the School of Science and Technology, NOVA University of Lisbon, Caparica 2829-516, Portugal

while in Potapov et al. (2021), footprints from the neighbouring tiles were also included. In turn, Healey et al. (2020) assessed three different spatial calibration scales while estimating CH globally with random forests. It was verified that smaller scales led to better results.

In the following years, Pereira-Pires et al. (2024) and Schwartz et al. (2022) explored alternative approaches for local CH estimation with GEDI, optical, and SAR images. Schwartz et al. (2022) used Sentinel-1 (S1) and S2, and (Pereira-Pires et al. 2024) further added Advanced Land Observing Satellite 2 (ALOS-2). The former estimated CH through a deep learning approach, while the latter through gradient boosting. Other authors resorted to ICESat-1/2 or ALS acquired data instead of GEDI, using regression trees (Hansen et al. 2016; Li et al. 2011; Peterson and Nelson 2014), random forests (Bolton et al. 2020; Cui et al. 2019; Jiang et al. 2021; Wang, Sun, and Xiao 2018) and neural networks (Astola et al. 2019; Radke, Radke, and Radke 2020). Qi et al. (2025) proposed a different GEDI-based approach, in which the CH was inverted from an interferometric SAR (InSAR) model derived from TanDEM-X.

A common hindrance to accurate CH estimation is the geolocation errors inherent to GEDI data (Dubayah et al. 2020; Hancock et al. 2019; Roy, Kashongwe, and Armston 2021; Tang et al. 2023; Wang et al. 2022). Existing methods to adjust GEDI footprints' geolocation can be leveraged for continuous CH mapping using data fusion. However, correcting the geolocation of thousands, or millions, of GEDI measurements requires consistent airborne LiDAR coverage to estimate horizontal and vertical offsets (Hancock et al. 2019). Tests on the impact of these on-orbit position corrections for CH estimation can be found in the literature (Roy, Kashongwe, and Armston 2021; Tang et al. 2023; Yang et al. 2024), but they focused exclusively on GEDI footprints, not considering the integration of different remote sensing sources. Additionally, the role of the GEDI data sampling fraction rate to calibrate CH modelling architectures remains underexplored. For instance, in Lang et al. (2022), Potapov et al. (2021), a fixed rate was utilised.

Considering the previous method, this article proposes a CH modelling architecture that combines optical and radar observations to assess the impact of using enhanced-geolocated GEDI footprints and contingent on ALS data, and if larger GEDI sampling fraction rates decrease CH estimation errors. Finally, most articles that resort radar data sources for CH mapping with machine learning do not explore their full potential. They focus on the use of backscatter products, overlooking the possible benefits of using interferometric variables. Therefore, in addition to the novelty of including a pre-processing method to adjust GEDI footprint geolocation, interferometric variables here were also derived for the CH estimation.

This research experiment covers 15 study areas across three countries and diverse Mediterranean landscapes in Portugal, Spain, and the US. It leverages GEDI's near-worldwide coverage to generate models harmonised with the conditions of the study area, through local calibration sets derived from CH measurements from the aforementioned sensor. For the assessment of the benefits of the proposed approach, the maps produced here were compared with two global products (Lang et al. 2022; Potapov et al. 2021), using ALS-derived Canopy Height Models (CHM) as a benchmark. An extreme gradient boosting was utilised as it offers a robust fitting process without requiring large datasets, which is typical in neural networks (Sun et al. 2023). It has also been demonstrated to be superior to a random forest for mapping CH in Mediterranean forests (Pereira-Pires et al. 2025; Pereira-Pires et al. 2023). The latter is justified by the sequential fitting process, in which each tree considers the residuals of the previous tree (instead of just randomly generating them).

2. Data and methods

2.1. Study areas

The study areas considered here have extensions ranging from 856 ha to 29,953 ha, and canopy height of up to 40 m (or 80 m in the case of California) across Mediterranean landscapes. Additionally, there were no identified major disturbances in any of these areas during the period of analysis. Of the selected areas, 6 were in Portugal, 8 were in Spain, and 1 was in the US state of California (Figure 1). In Portugal and California, they were selected based on data availability. In Spain, nine representative forest areas were identified. According to the respective Land Cover Land Use maps (LCLU), described in Section 2.2.4, the broadleaf species dominate (>75%) in 3 areas, while conifers do in 6 areas (see Table 1). The remaining

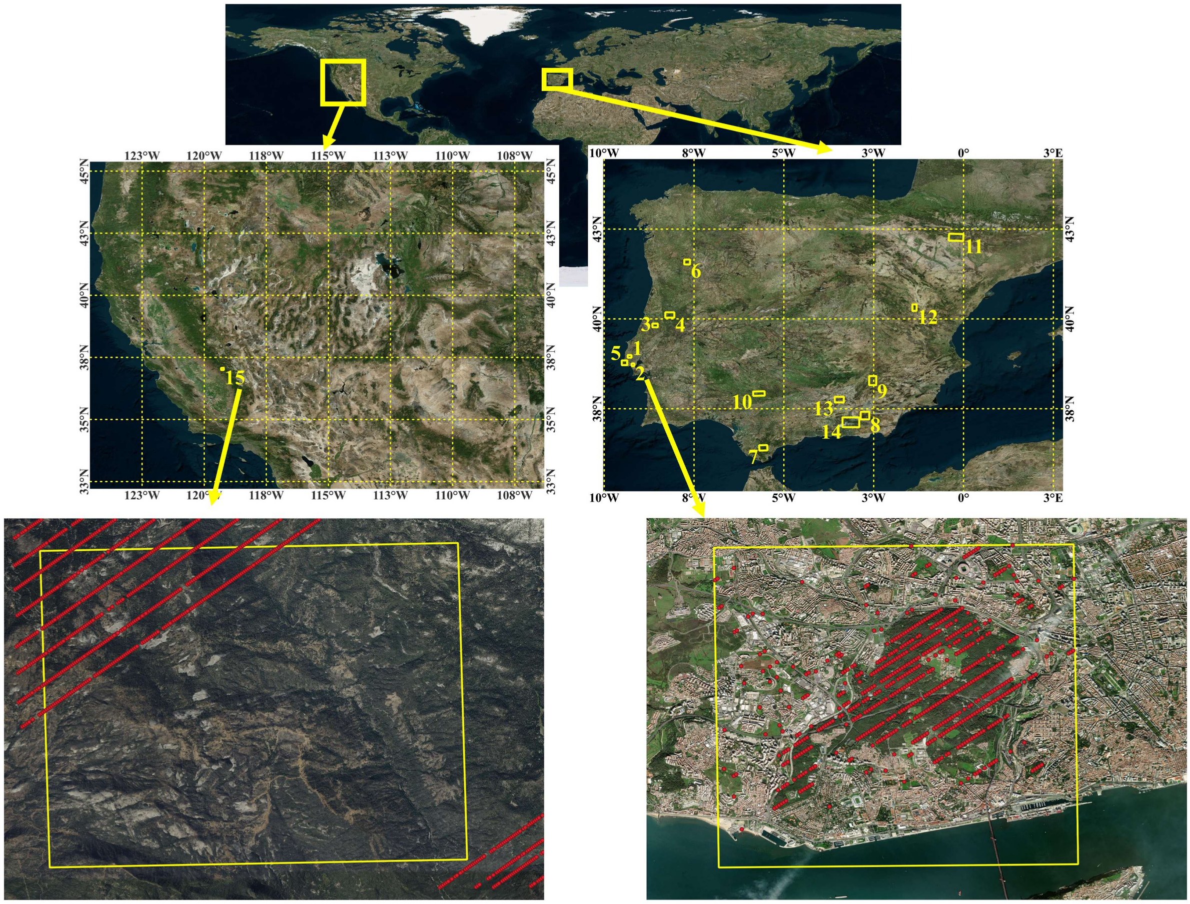

Figure 1. Geographic location of the study areas (WGS84), and GEDI footprints distribution for Soaproot Saddle, US and Monsanto, PT. The study areas corresponding to the identifiers are provided in Table 1.

samples were classified as mixed. Regarding their topography, they represent a diverse set, as shown by their elevation ranges presented in Table 1.

A reference year was assigned for each area based on the ALS acquisition dates. Specifically, the year 2019 for Soaproot Saddle, 2021 for Guara and Rodeno, and 2020 for the remaining.

2.2. Remote sensing data

Several remote sensing missions were used to estimate CH: LiDAR measurements from the GEDI and Airborne Laser Scanning campaigns (Table 1); optical images from S2; and SAR observations from ALOS-2 and S1. In addition to these, ancillary data on the land cover and topography already embedded in some remote sensing products were considered for the mapping.

2.2.1. Spaceborne LiDAR from GEDI

The GEDI tracks of the 2019–2023 collection of 25 m footprints were acquired using a spacing of 60 m along-track and 600 m across-tracks. In this article, four calendar years of the GEDI L2A product (Dubayah et al. 2020) from 2019 (April) to 2022 (December) were used. The data quality requirements were as follows (Pascual and Guerra-Hernández 2023; Pascual et al. 2025):

1) Waveform fidelity above 0.95 and greater than the Canopy Cover;
2) GEDI Shot needs to meet the L2A and L4A quality criteria, according to the respective product flags;
3) The difference between the GEDI elevation lowest mode and TanDEM-X terrain elevation cannot exceed 50 m.

Table 1. Study areas: tree species, dominant Plant Functional Type (PFT), elevation range, and Airborne Laser Scanning surveys specifications.

Study area (ID)	Country	Main tree species	Dominant GEDI PFT	Elevation range (m)	Year of ALS survey	ALS density (Points/m^2)	ALS instrument	ALS vertical error (m)
Mafra (1)	Portugal	Broadleaf	Evergreen Broadleaf	95–437	2020	10.02–13.88	Teledyne Galaxy PRIME	0.40
Monsanto (2)		Coniferous	Evergreen Needleleaf	0–277	2020			
Pombal (3)		Mixed	Evergreen Needleleaf	148–413	2020			
Sintra-Cascais (4)		Mixed	Evergreen Broadleaf & Closed Shrublands	0–568	2020			
Serra Lousã (5)		Mixed	Evergreen Needleleaf	122–1241	2020			
VPA (6)		Coniferous	Evergreen Needleleaf	266–1252	2020			
Alcornocales (7)	Spain	Broadleaf	Evergreen Broadleaf	70–910	2020	1.50	Optech Prime, ALS80HP	0.20
Baza (8)		Coniferous	Evergreen Needleleaf & Closed Shrublands	971–2273	2020	1.50	Optech Prime	
Cazorla (9)		Coniferous	Evergreen Needleleaf	708–1929	2020	1.50	Optech Prime	
Constantina (10)		Broadleaf	Evergreen Needleleaf	359–904	2020	1.50	Leica ALS80	
Guara (11)		Mixed	Evergreen Needleleaf	520–2100	2021	1.00	Optech Prime	
Rodeno (12)		Coniferous	Evergreen Needleleaf	1.057–1743	2021	1.00	Optech Prime	
Sierra Magina (13)		Mixed	Closed Shrublands	513–2204	2020/ 2021	1.50	Optech Prime, Optech T2000, ALS80HP	
Sierra Nevada (14)		Mixed	Closed Shrublands	701–3515	2020	1.50	Optech Prime	
Soaproot Saddle (15)	U.S.A.	Coniferous	Evergreen Needleleaf	537–2017	2019	2.00–6.00	ALTM Gemini LiDAR	0.36

A total of ~236, ~203 and ~239 thousand footprints met the defined criteria in the study areas in Spain, Portugal and California, respectively. Two LiDAR datasets were used per area: one including footprints acquired in the reference year, and the other including footprints acquired in the 4-year period. Finally, in this article, the relative height (RH) 98 retrieved from the GEDI footprints was adopted as the metric corresponding to the CH.

2.2.2. Optical satellite images from Sentinel-2

As optical images, S2 Level 2A products provided by the Copernicus Data Space Ecosystem were used. These correspond to bottom-of-atmosphere reflectance values across 13 spectral bands (all used with the exception of Band 10), covering the visible, near infrared, and shortwave infrared spectrums. The spatial resolution varies with the band from 10 to 60 m. All observations, without clouds covering the study area, in the 6-month period from July 1st to December 31st of the respective reference year, were considered. This provided a heterogeneous set of observations, making the models less dependent on the data acquisition period. Despite not considering entire calendar years for GEDI, this 6-month period accurately represents the annual data trends, requiring less computational effort and lowering possible errors from the temporal mismatch between GEDI and S2 images.

2.2.3. Synthetic aperture radar observations

The Global L-band mosaic (1.200 GHz) delivered by the ALOS Research Application Project was acquired in strip-map mode with dual polarisation (HH+HV). This product corresponds to the terrain-flattened gamma noise at a linear scale and a spatial resolution of 25 m, providing one observation per year. From S1 (which operates in the C-band, 5.405 GHz), two products provided by the Copernicus Data Space Ecosystem were used: single look complex (SLC) and ground range detected (GRD). Both were acquired in interferometric wide mode with dual polarisation (VV+VH). They were acquired with a 6-day frequency, and the GRD was provided with a 10 m resolution. Data from the second semester of the reference year were considered both for the SLC and GRD products. While for the SLC, a pair of observations was used for each month, for the GRD, two orbits (one for S1A and the other for S1B) were selected and all available observations were used.

2.2.4. Ancillary data

The CH mapping resorted to three LCLU depending on the study area: for Portugal, it was the Carta de Ocupação do Solo 2018 (Direção-Geral do Território 2018); for Spain, it was the Corine Land Cover 2018 (European Environment Agency 2019); and for California, it was the National Land Cover Database 2019 (Dewitz and U.S. Geological Survey 2021). COS 2018 is a vector product with ten forest classes and a minimum mapping unit of 1 ha; Corine 2018 is also a vector product, but with three forest classes and a minimum mapping unit of 25 ha; on the contrary, NLCD 2019 is a raster product with three forest classes and a spatial resolution of 30 m.

The digital elevation model (DEM) used here was acquired by the Shuttle Radar Topography Mission (SRTM) in February 2000 (OpenTopography 2013). The local incidence angles of the ALOS-2 and S1 observations were also included in this article as ancillary data.

2.2.5. Airborne Lidar data

This article used ALS data both for model validation and correction of the GEDI footprint geolocation. The data was acquired under the campaigns áGIL (ICNF 2022) for Portugal, Plan Nacional de Ortografía Aérea (PNOA 2020) for Spain, and NEON (NEON [National Ecological Observatory Network] 2023) for California. For the Portuguese campaign, data from six flights that took place in 2020 were used. The provided point clouds had densities ranging from 10.02 to 13.88 points/m^2. Spanish data was acquired by seven different flights in 2020 and 2021, with densities varying from 1.00 to 2.00 points/m^2. Finally, regarding California, only data from a single flight in 2019 was included. The point cloud had a density ranging from 2.00 to 6.00 points/m^2. Other specifications of the ALSs' campaigns are provided in Table 1.

The process used to generate the CHMs from point clouds resorted to both LASTools and the lidR package. It started with the creation of 500 × 500 m tile with 50 m buffer using the *lastile* function. Then, the height above the ground was computed with *lasheight*, and the previously introduced buffer was removed. The process concludes with the normalisation of the point clouds, resorting to the *readLAScatalog* and the creation of the CHMs at 25 and 10 m resolution, applying the pit-free algorithm from the *rasterize_canopy* function. Note that lower point densities will capture fewer height variations within the study area, generating more homogenous reference data. These products are accessible in Guerra-Hernández, Pereira-pires, and Pascual (2024) and were compared with the CH maps produced here. The number of ALS cells (i.e. pixels) corresponding to forests and the percentage of the study area's coverage by the GEDI are provided in Table 2. In addition, the ALS data was used in the gediSimulator for correcting the geolocation errors of the GEDI footprints (Hancock et al. 2019).

2.2.6. State-of-the-art canopy height maps used for comparison

A comparison between the maps produced here and the maps originated by two alternative approaches (Lang et al. 2022 and Potapov et al. 2021) was made, assessing the advantages of local calibration. The first produced a global map with a spatial resolution of 10 m for 2020, and herein will be referred to as Lang maps. The second produced a global map for 2019 with a spatial resolution of 30 m. However, in Turubanova et al. (2023), CH maps for Europe were produced for all years between 2001 and 2021, using the same approach. Therefore, for Soaproot Saddle, the 2019 map was used, while for Portugal and Spain, the maps produced for 2020 and 2021 were used accordingly. These maps will be referenced as Potapov maps.

Regarding the potential effects of the temporal mismatch between Lang maps (produced for 2020) and reference data for Guara, Rodeno, and Sierra Magina (produced for 2021), and Soaproot Saddle (produced for 2019), they are not expected to be substantial. This is concluded due to the slow growth dynamics of Mediterranean forests and the non-significant forest loss according to the maps from Hansen et al. (2013) in the Spanish areas. For Soaproot Saddle, some forest loss occurred; however, this is not reported for most

Table 2. Number of GEDI footprints and Airborne Laser Scanning cells, GEDI coverage of the study areas, and corresponding minimum Relative Height 98 value. For the study areas where the footprints geolocation correction was tested, the number of corrected footprints is also provided (see Section 3.2).

	No of GEDI footprints				
Study area	All	w/Corrected geolocation	No. of ALS Cells	% Area covered by GEDI	Minimum RH98
Mafra	416	25	34,213	1.18	2.60
Monsanto	432	435	13,667	3.16	4.78
Pombal	834	417	90,149	0.91	3.20
Serra Lousã	6,782	6,699	24,0726	2.77	1.86
Sintra Cascais	757	665	75,995	0.99	3.69
VPA	452	0	107,777	0.41	4.14
Alcornocales	4,252	-	362,038	1.17	3.27
Baza	4,183	-	237,621	1.74	3.09
Cazorla	13,361	-	350,697	3.81	3.08
Constantina	6,361	-	280,042	2.25	2.88
Guara	4,759	-	226,057	2.09	2.62
Rodeno	1,625	-	101,586	1.60	3.44
Sierra Magina	3553	-	100,608	3.45	2.41
Sierra Nevada	11,872	-	472,081	2.49	1.74
Soaproot Saddle	385	-	137,617	0.28	5.58

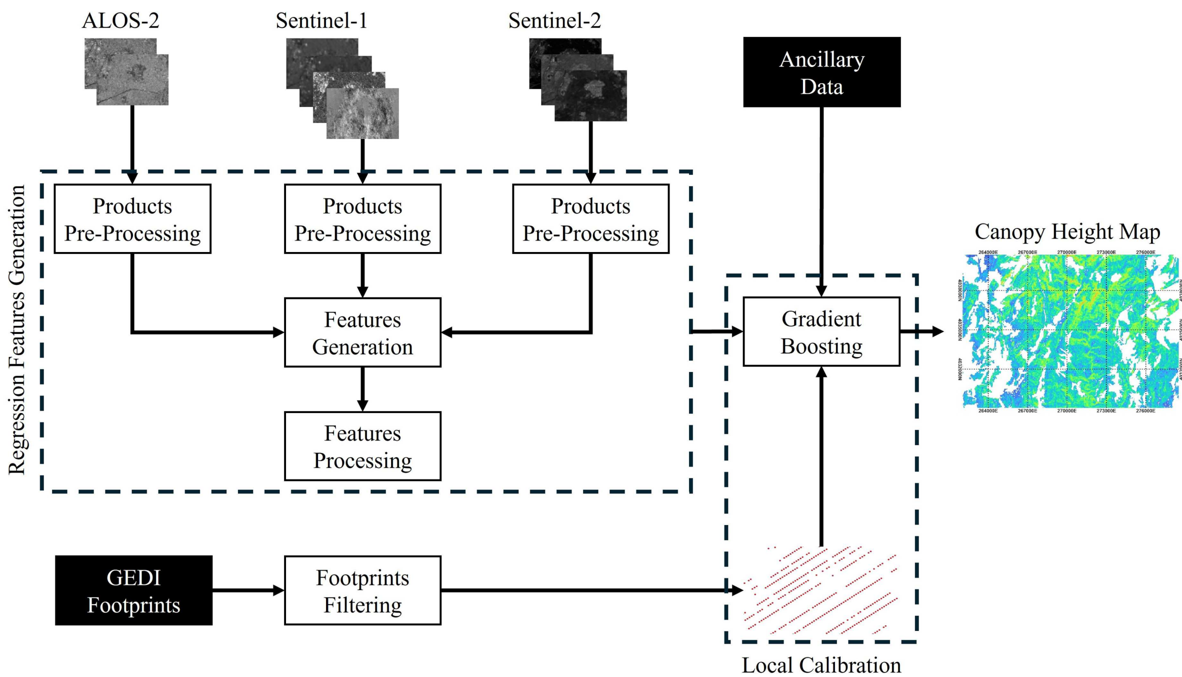

Figure 2. Overview of the Canopy height mapping model is proposed here.

of the study area. Relatively to Potapov maps, the different spatial resolution of the product (30 m) and reference (25 m) did not impact the comparison, as previously reported (Potapov et al. 2021).

2.3. Model architecture for CH estimation

The proposed CH mapping model, presented in Figure 2, has three main processes: pre-processing, feature generation and processing, and CH estimation, which are described in the following sections. An additional fourth process regarding the geolocation correction of the GEDI footprints was also assessed in this experiment, being also presented in this section.

2.3.1. Data pre-processing

During the first process, ALOS-2, S1 (GRD and SLC), and S2 observations were pre-processed independently. For ALOS-2 and S1-GRD, the methodology proposed in Pereira-Pires et al. (2025) was employed.

In the case of ALOS-2, it only consisted of the speckle filtering, while for S1-GRD, because it is at a lower processing level, a more complex approach was required. The following steps were employed: orbit file application, thermal noise removal, terrain flattening, speckle filtering, and terrain corrections. For the speckle filtering, a Lee filter with a 7×7 kernel was used. Regarding the terrain flattening and terrain correction, the utilised DEM was the SRTM. During the terrain correction, the observations were also resampled to 25 m using the bicubic interpolation.

For S1-SLC, the computation of the coherence and phase followed the steps proposed in Braun and Veci (2021). This process began with the selection of the sub-swaths covering the study areas, the application of the orbit files, and the coregistration of the observation pairs (using the back-geocoding operator from SNAP). Then, the interferogram was computed, debursted, and the Goldstein filter was applied. The pre-processing was concluded with the phase unwrapping and terrain correction (where the observations were also resampled to 25 m with the bicubic interpolation).

Finally, the S2 observations were only required to be resampled to 25 m (using bicubic interpolation for upsampling and the median for downsampling). The pre-processing process was conducted with Sentinel Application Platform (SNAP) software.

2.3.2. Features generation and processing

In the second process, texture features, principal components, polarisation ratios and differences, biophysical variables, and spectral indices were derived as previously described (Barreira et al. 2024; Pereira-Pires et al. 2024, 2025) for the reflectance bands and backscattering coefficients, resulting in a total of 277 features. Note that the texture features computation utilised the gray-level co-occurrence matrix, with the following settings: all directions, without displacement, and 64 quantisation levels. Then, they were re-projected to the defined coordinate reference system, and a 6-month median temporal composite was computed for the ones from S1 and S2. Finally, the 3×3 mean spatial filter proposed in Pereira-Pires et al. (2025) was applied to all of them.

2.3.3. Canopy height estimation

The last process consisted of producing the study areas' CH maps. For each area, an extreme gradient boosting regressor (using the XGBoost Python library implementation) was locally calibrated by resorting to the available local GEDI footprints (as shown in Figure 1). The purpose of this was to take advantage of the local patterns and spatial autocorrelations of each study area. Then, the resulting regressor estimated the CH at the pixel level, being all the non-forest pixels according to the study area LCLU filtered out, generating the respective area map.

2.3.4. GEDI footprints geolocation correction

In addition to these processes, the inclusion of a fourth one was assessed to investigate the impact of correcting the on-orbit positions of the GEDI footprints in the CH mapping, which geolocation errors are approximately 10.20 m (Dubayah et al. 2020). The correction was performed by matching the GEDI footprint's geolocation with the ALS position using gediSimulator's 'collocateWaves' function (Hancock et al. 2019; Hofton, Minster, and Blair 2000). First, for each footprint, a set of simulated waveforms, spaced at 1 m intervals within a 20×20 m window relative to its position centre, was computed from the available ALS. Then, the Pearson correlation between the on-orbit GEDI L1B and each simulated waveform was calculated. Since the correction was performed at the track level, the mean correlation value of all the track footprints was used to select the best candidate for the geolocation offset. The last step consisted of updating the track geolocation with the position of the best candidate. In this experiment, an additional step was included to filter out collocated tracks that were poorly correlated with the original GEDI L1B product. Therefore, for correlations below 0.80, the track was not collocated. The GEDI mission recommends only applying the gediSimulator's functions when ALS data with at least 4.00 points/m^2 are available. The number of footprints available after quality filtering and geolocation correction is provided in Table 2. Despite the inclusion of ALS data in pre-processing pipeline, it is only employed to reposition the GEDI's track. Therefore, the CH derived from these products do have an impact on the GEDI's height measurements, ensuring their independence from the ALS data.

2.4. Model building and testing

The model building included the hyperparameter optimisation and feature selection. First, the extreme gradient boosting number of estimators and maximum depth were selected within the ranges of 5–1000 and 1–8, respectively. For the second task, a sequential feature elimination process was employed to identify the most suitable ones for the model. It works as follows: starting with the top 270 features, at each iteration, the 10 least important features (according to the gain, weight, and cover using the XGBoost feature importance) are eliminated, the model is retrained, and its results are recomputed (reducing possible redundancies). Based on the %RMSE response to the decreasing number of features, the best ones were selected. The process was stopped when there were only 10 features remained. Regarding the study area's GEDI footprints, only the ones acquired during their reference year were used for the regressor calibration. During the model building, only the Portuguese study areas were considered, being evaluated individually through a five-fold cross-validation. The hyperparameters and features were selected based on the average results of all of them. Then, the regression approach was tested in Spain and California, assessing the applicability of the previously defined parameters for new areas.

After model building, an important factor in the generation of GEDI-fused products was tested: the benefit of using GEDI footprints with corrected on-orbit positions for calibrating CH models. Then, the potential benefit of using all available footprints (from 2019 to 2022) for mapping was subsequently assessed. Both of these tests consider only the Portuguese study areas, due to the ALS point density requirement of gediSimulator.

Finally, for evaluating the potential of the proposed approach regarding wall-to-wall mapping, the CH maps produced from it were compared with the two global products defined in Section 2.2.6, using the local ALS-derived CHMs as a reference. To ensure an unbiased comparison, the Potapov maps were resampled to 25 m using a nearest neighbour interpolation and for the Lang maps, the CHMs to be used as baselines were computed at 10 m (guaranteeing that the maps and references produced had a corresponding spatial resolution). Additionally, only the pixels that had a CH estimate for all these products were considered for the comparison.

The evaluation metrics utilised in this article were: Root Mean Squared Error (RMSE, (1a)), Bias ((2a)), Mean Absolute Error (MAE, Equation (3a)a), and the respective normalised values ((1a), (2b), and (3b)). Both the model calibration and validation made use of the GEDI-RH98 values. Additionally, the Coefficient of Variation (CV, Equation (4)) was computed to verify how accurately the regressor was with respect to the distribution and domain of the reference data. In Equations (1)–(4), N represents the number of samples, x_i represents the observed value GEDI-RH98, $\hat{x}_i$ represents the estimated value CH, μ represents the average, and σ represents the standard deviation of the reference. As can be seen in Table 1, the ALS data used in the scope of this experiment had sub-meter vertical accuracies. Therefore, these errors were considered negligible in the scope of this article, not being propagated for the final error estimation.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \widehat{x_i})^2}{N}} \tag{1a}$$

$$\%\mathrm{RMSE} = \frac{\mathrm{RMSE}}{\mu} \tag{1b}$$

$$\mathrm{Bias} = \frac{\sum_{i=1}^{N}(x_i - \widehat{x_i})}{N} \tag{2a}$$

$$\%\mathrm{Bias} = \frac{\mathrm{Bias}}{\mu} \tag{2b}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N}|x_i - \widehat{x_i}|}{N} \tag{3a}$$

$$\%\mathrm{MAE} = \frac{\mathrm{MAE}}{\mu} \tag{3b}$$

$$\mathrm{CV} = \frac{\sigma}{\mu} \tag{4}$$

3. Results

The results section is divided into four parts: model building and testing; impacts of correcting the GEDI footprints geolocation; impacts of using longer GEDI acquisition periods; and comparison with Lang and Potapov maps using an ALS-derived CHM as a reference.

3.1. Forest height mapping model

Considering that the model building results in extreme gradient boosting with 600 estimators, a maximum depth of 5, and the top 40 features were selected. Among the top features, there were four from ALOS-2, ten from S1-GRD, seven from S1-SLC, sixteen from S2, and three ancillary (Appendix A Figure A1, Table A1). Figure 3 shows that the gain was able to select the smallest number of features while achieving the lowest %RMSE (34.20%).

During testing, the model achieved an average %RMSE of 29.77%, demonstrating the applicability of the adjusted parameters to new areas. Globally, the results show an average %RMSE of 31.54%, with values ranging from 26.39% to 39.18% (Table 3). Figure 4 scatterplots are evidence that for short- and tall-tree conditions, overestimation, and underestimation occur, respectively. Notwithstanding the previous, a good estimation for the most-represented CH interval in the calibration set is verified (Figure 5).

3.2. Impacts of correcting GEDI footprints on-orbit position

Owing to the correlation threshold defined in Section 2.3, not all the GEDI tracks could be collocated. In Table A.2, it can be noticed that 3 of the 14 tracks across the Portuguese study areas did not meet the defined criterion (excluding all VPA's footprints). Regarding the geolocation corrections, the average

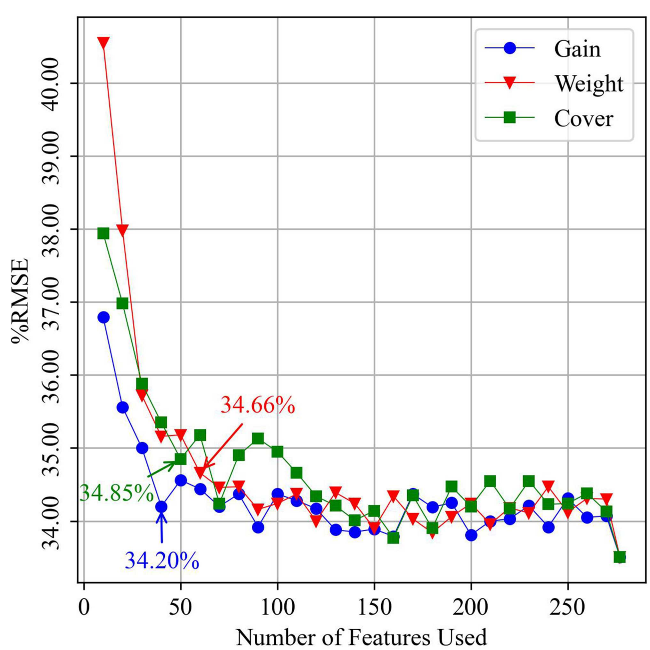

Figure 3. Progress of the relative mean square error (%RMSE) during the model fitting using different numbers of best features according to the areas of gain, weight, and cover.

Table 3. Estimation of forest Canopy height (CH) in the training areas using GEDI footprints for calibration.

	Five-fold cross-validation results					
Study area	%RMSE	RMSE (m)	%MAE	MAE (m)	%Bias	Bias (m)
Mafra	35.74	4.58	28.07	3.60	0.41	0.05
Monsanto	29.68	4.76	22.73	3.64	1.12	0.17
Pombal	36.89	4.99	28.78	3.89	0.45	0.05
Serra Lousã	32.51	5.32	25.52	4.18	0.30	0.05
Sintra Cascais	31.19	4.88	24.16	3.78	−0.08	−0.02
VPA	39.18	4.94	30.16	3.80	−0.24	−0.04
Alcornocales	31.89	3.97	24.55	3.06	0.53	0.06
Baza	26.39	2.65	19.09	1.92	0.04	0.00
Cazorla	27.63	4.93	21.32	3.80	0.09	0.02
Constantina	25.99	3.09	19.01	2.26	0.27	0.03
Guara	34.59	5.03	26.77	3.90	0.26	0.04
Rodeno	28.54	3.77	21.00	2.78	0.40	0.05
Sierra Magina	30.37	3.83	23.01	2.90	0.19	0.02
Sierra Nevada	28.64	3.63	20.81	2.64	−0.10	−0.01
Soaproot Saddle	33.86	8.92	26.56	7.00	−0.03	−0.02
Average	31.54	4.62	24.10	3.54	0.26	0.03

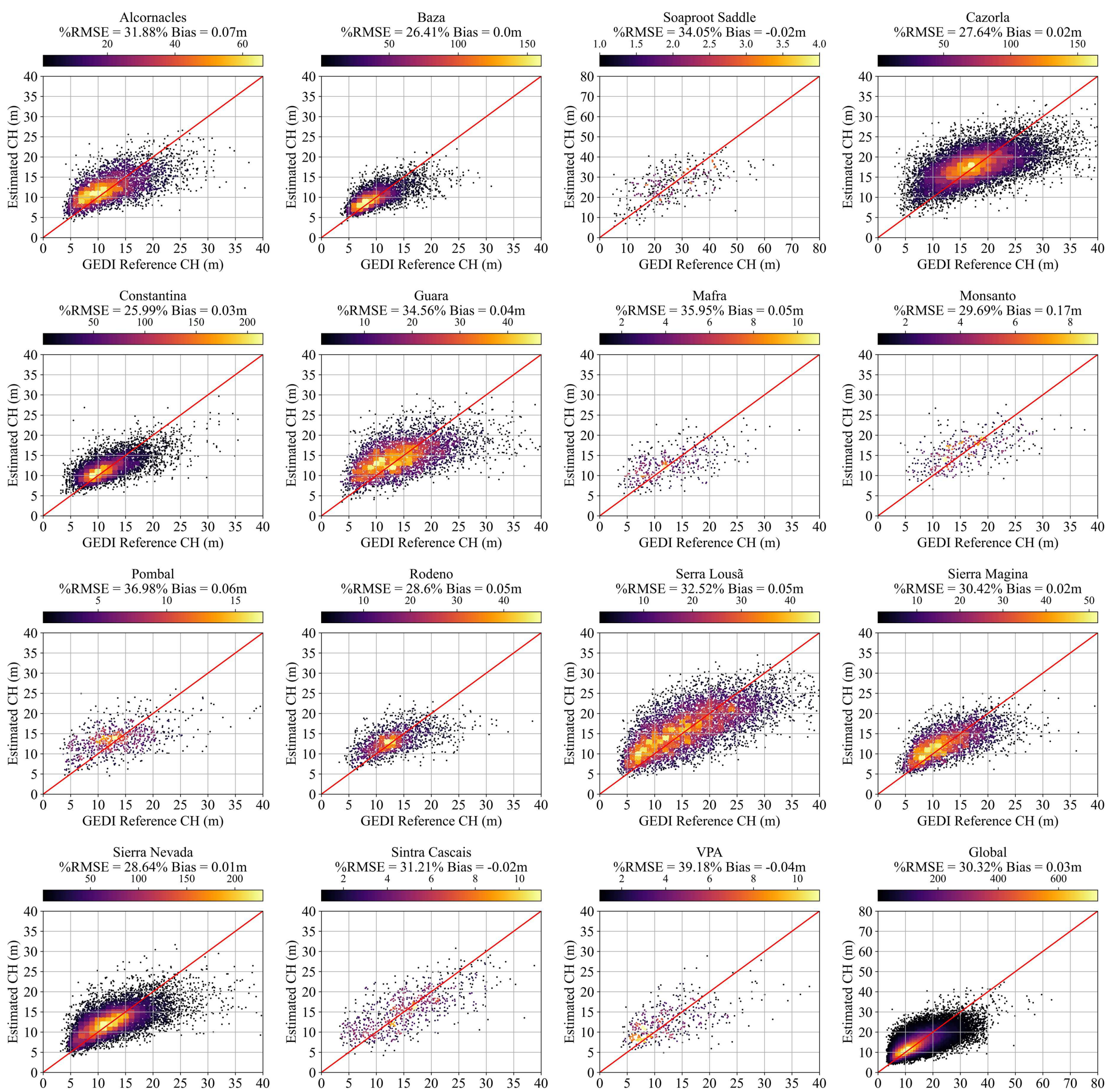

Figure 4. Scatterplots between the estimated Canopy height and respective GEDI reference, per study area and globally.

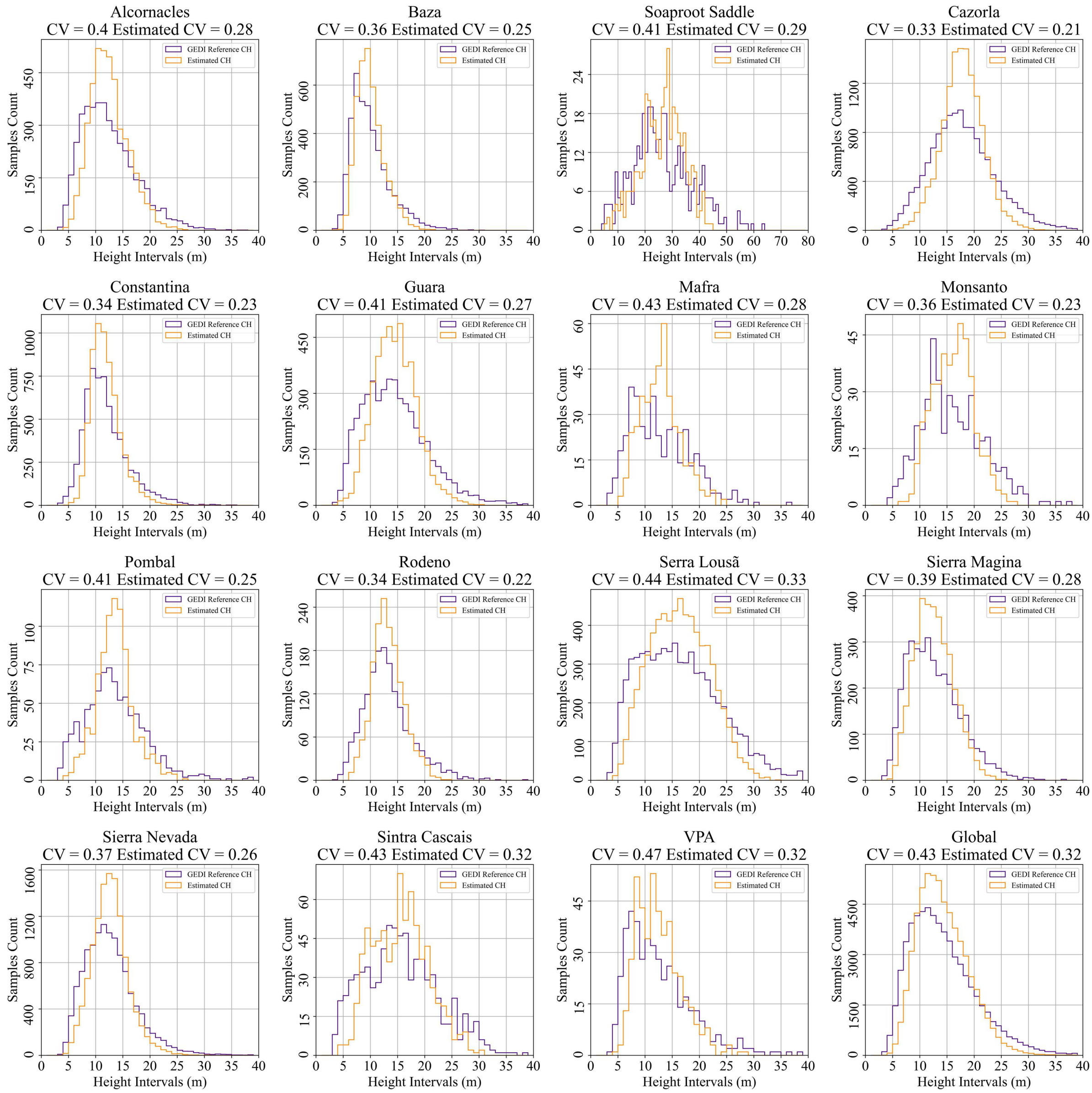

Figure 5. Histograms of the estimated Canopy height and GEDI reference, per study area and globally.

distances between the on-orbit and corrected position were 11.81 m (with a maximum value of 26.92 m) and 7.35 m (with a maximum value of 16.33 m) across the latitude and longitude directions.

To analyse the impacts of correcting the on-orbit position, two datasets were compared: (a) one including the footprints with corrected on-orbit position and (b) the other with the same footprints but with their on-orbit position (results presented in Table 4). Regarding the average %RMSE, it improved by 1.23 pp with the correction of the footprints on-orbit position. Furthermore, improvements were verified for all tested study areas (ranging from 0.29 to 2.30 pp).

To assess the impact of gap-free mapping, the CH maps produced using datasets (a) and (b) were compared with the reference CHMs. Both achieved similar results, with no verified advantage in including the collocation step. It is worth noting that not all footprints could be corrected, except for those of Monsanto. This is particularly evident for Mafra, where only 25 footprints out of 416 could be geolocated. The impact of this can be seen in the comparison of the results achieved with dataset (a) with those achieved when using all available footprints without any on-orbit position correction (Table 6). The use of dataset (a) led to an increase in the average %RMSE of 4.62 pp for the study areas considered, demonstrating the importance of the number of

Table 4. Results achieved when using the footprints with the corrected position (dataset a), and the footprints with those eligible for geolocation correction, but with the on-orbit position (dataset b). GEDI footprints and the ALS reference Canopy height model as reference. The Pearson correlation (ρ) between the respective dataset and ALS reference is also provided.

		Eligible for geolocation correction							
		(a) Corrected				(b) On-orbit			
Study areas		No of footprints	%RMSE	%MAE	ρ	No of footprints	%RMSE	%MAE	ρ
GEDI	Mafra	25	41.19	33.39	–	24	43.49	35.62	–
	Monsanto	435	29.38	22.27	–	432	30.84	23.06	–
	Pombal	417	33.37	26.23	–	412	35.14	27.48	–
	Serra Lousã	6699	32.47	25.45	–	6657	32.76	25.62	–
	Sintra Cascais	665	29.78	22.69	–	673	30.14	23.59	–
	Average	–	33.24	26.01	–	–	34.47	27.07	–
ALS	Mafra	–	56.13	44.35	0.47	–	53.30	41.89	0.06
	Monsanto	–	28.44	21.43	0.71	–	28.38	21.60	0.03
	Pombal	–	38.90	31.10	0.57	–	38.34	30.59	0.24
	Serra Lousã	–	31.56	24.21	0.66	–	31.43	24.15	0.33
	Sintra Cascais	–	29.46	22.53	0.77	–	29.06	22.31	0.32
	Average	–	36.90	28.72	–	–	36.10	28.11	–

Table 5. Results achieved using multi-year and single-year using the GEDI-relative height 98 footprints and ALS reference Canopy height model.

		GEDI footprints reference					
		Multi-year			Single-year		
Study areas		%RMSE	%MAE	%Bias	%RMSE	%MAE	%Bias
GEDI	Mafra	36.79	28.45	0.03	35.74	28.07	0.41
	Monsanto	29.88	22.66	1.01	29.68	22.73	1.12
	Pombal	36.56	27.83	0.25	36.89	28.78	0.45
	Serra Lousã	31.57	24.51	0.16	32.51	25.52	0.30
	Sintra Cascais	31.24	24.04	−0.14	31.19	24.16	−0.08
	VPA	32.18	24.73	0.42	39.18	30.16	−0.24
	Average	33.04	25.37	0.29	34.20	26.57	0.33
ALS	Mafra	37.94	29.07	1.64	38.24	29.22	3.88
	Monsanto	31.57	23.37	12.94	28.38	21.60	9.00
	Pombal	31.10	24.31	16.62	33.91	27.06	18.47
	Serra Lousã	29.98	22.88	−2.41	31.24	23.96	−3.55
	Sintra Cascais	28.83	21.74	7.66	29.63	22.59	7.55
	VPA	41.21	32.93	−18.87	46.46	36.85	−21.00
	Average	33.44	25.72	2.93	34.64	26.88	2.39

available footprints for calibration. Additionally, for Mafra and Pombal, where the results were significantly worse, the collocated footprints had lower correlation values with the corresponding ALS reference data. Finally, the fact that the geolocation correction in both directions was shorter than the 25 m spatial resolution hinders the potential improvement of the inclusion of this step in the model architecture (verified during the cross-validation phase).

3.3. Use of GEDI footprints acquired during a four-year period

The average %RMSE achieved for the three study areas with a GEDI coverage below 1.00% was 36.67%, which was considerably higher than that for all others that were more GEDI-populated (32.62%), as shown in Table 5. Adding 2019/22 GEDI footprints in the Portuguese study areas increased the spatial coverage to 3.77% for Mafra, 6.50% for Monsanto, 5.19% for Pombal, 6.45% for Serra Lousã, 3.83% for Sintra Cascais, and 4.66% for VPA. This improved the estimation by 1.16 pp, and for VPA, the improvement exceeded 5.00 pp. Regarding the results achieved when comparing with the reference CHMs, the average %RMSE decrease was nearly the same (1.20 pp). As before, VPA experienced the largest impact. However, for Monsanto (the area with the highest coverage in a single year), the estimation worsened by 3.19 pp. It is also interesting to note that the average %RMSE for the three study areas with a GEDI coverage below the 1.00% decreased by 2.95 pp, while for the others, it increased by 0.54 pp.

3.4. Comparison with CHMs and other CH products

The model used for producing the CH maps did not include the geolocation correction process and only resorted to footprints of the reference year for calibration. The comparison with the reference CHMs showed an average %RMSE of 32.41% for all the study areas, ranging from 20.17% to 38.94% (Table 6, Figure 6, and Appendix A Figure A2). Using the same CHMs, Lang and Potapov had higher %RMSEs (40.50% and 59.04%, respectively). Hence, the locally calibrated approach led to more accurate maps (exactly by 8.09 and 26.63 pp). Analysing the %Bias, while the Lang and Potapov maps underestimated CH for all study areas, the model proposed here does not have a clear tendency towards over- or under-estimation (Figure 6, Figure 7, and Appendix A Figure A2). Finally, the CH maps produced here provide a better representation of CH than the Lang maps, which are substantially smoother than the reference CHMs, as shown in Appendix A Figures A3–A5.

4. Discussion

This article demonstrated that the proposed modelling framework produced reliable CH maps across Mediterranean-like landscapes. Regarding the feature analysis, it was verified that data from all sources were relevant for the mapping, enhancing the importance of combining optical and SAR in CH mapping. This confirms the results previously achieved in García et al. (2018), Hyde et al. (2006); Pereira-Pires et al. (2024). Although most features within the top ones come from S2, the two most important were from ALOS-2, revealing the importance of the L-band in CH mapping. This was also verified in Guerra-Hernández et al. (2022), where the authors found ALOS-2 features to be the most sensitive to forest structure. In the opposite direction, in Fagua et al. (2019), the optical features (from Landsat-8) showed to be more sensitive to the CH than ALOS-2. However, as here, the authors also found that ALOS-2 was more important than S1, showing the greater importance of the L-band comparing to the C-band. This experiment also demonstrated the benefits of including S1-SLC in CH mapping. Finally, 30 out of the 40 selected features were derived during the feature generation, highlighting the importance of this process.

Regarding the feature selection, Astola et al. (2019) retained 59% of the available features for its final model with a forward selection approach. Here and in Wang et al. (2019) there were retained 14% and 19% with backward elimination. This may indicate greater robustness in the latter methods.

The potential improvement of correcting the on-orbit position of the GEDI footprints demonstrated in the cross-validation phase was not reflected in the produced CH maps. There are two main reasons that explain this fact. First, when only the tracks eligible for correction were utilised, the GEDI coverage of some study areas strongly decreased. This was evident for Mafra and Pombal, whose maps had substantially larger errors. For the other areas, where most footprints were successfully geolocated, there was no significant change in the results. Second, the average geolocation correction values in both directions were

Table 6. Validation of Canopy Height (CH) estimates using the local ALS reference Canopy height model derived from ALS. Lang and Potapov CH maps were also benchmarked against the maps here produced.

	Canopy height product								
	Proposed approach			Lang			Potapov		
Study areas	%RMSE	%MAE	%Bias	%RMSE	%MAE	%Bias	%RMSE	%MAE	%Bias
Mafra	38.19	29.21	−3.91	44.65	36.87	−30.28	70.82	61.87	−61.43
Monsanto	28.16	21.40	−9.32	30.85	26.00	−18.73	65.28	56.69	−56.62
Pombal	33.89	27.04	−18.54	35.16	27.68	−10.19	68.90	64.08	−64.02
Serra Lousã	31.17	23.91	3.51	38.23	31.28	−24.80	62.15	55.43	−55.21
Sintra Cascais	29.60	22.57	−7.52	34.74	28.14	−17.31	67.82	60.14	−59.83
VPA	46.42	36.84	21.03	44.39	38.92	−36.68	57.99	46.12	−41.15
Alcornocales	29.24	22.68	12.10	40.49	35.85	−34.67	53.71	47.61	−47.46
Baza	29.31	22.27	9.54	42.36	36.75	−31.63	46.59	38.02	−36.98
Cazorla	23.88	18.38	6.64	33.54	27.91	−25.23	59.73	54.45	−54.42
Constantina	22.39	15.81	7.46	38.59	32.37	−28.65	61.34	59.25	−59.28
Guara	48.65	38.54	29.00	49.85	42.98	−41.14	47.29	34.42	−29.66
Rodeno	20.17	15.43	−0.76	26.07	21.26	−16.72	58.39	53.88	−53.84
Sierra Magina	35.96	28.20	19.01	50.14	43.91	−41.73	58.75	49.48	−49.05
Sierra Nevada	30.19	22.74	15.51	45.76	40.87	−39.49	40.27	30.14	−27.20
Soaproot Saddle	38.94	31.25	−11.20	52.62	45.15	34.18	66.58	54.57	−52.27
Average	32.41	25.08	4.84	40.50	34.40	−24.20	59.04	51.08	−49.89

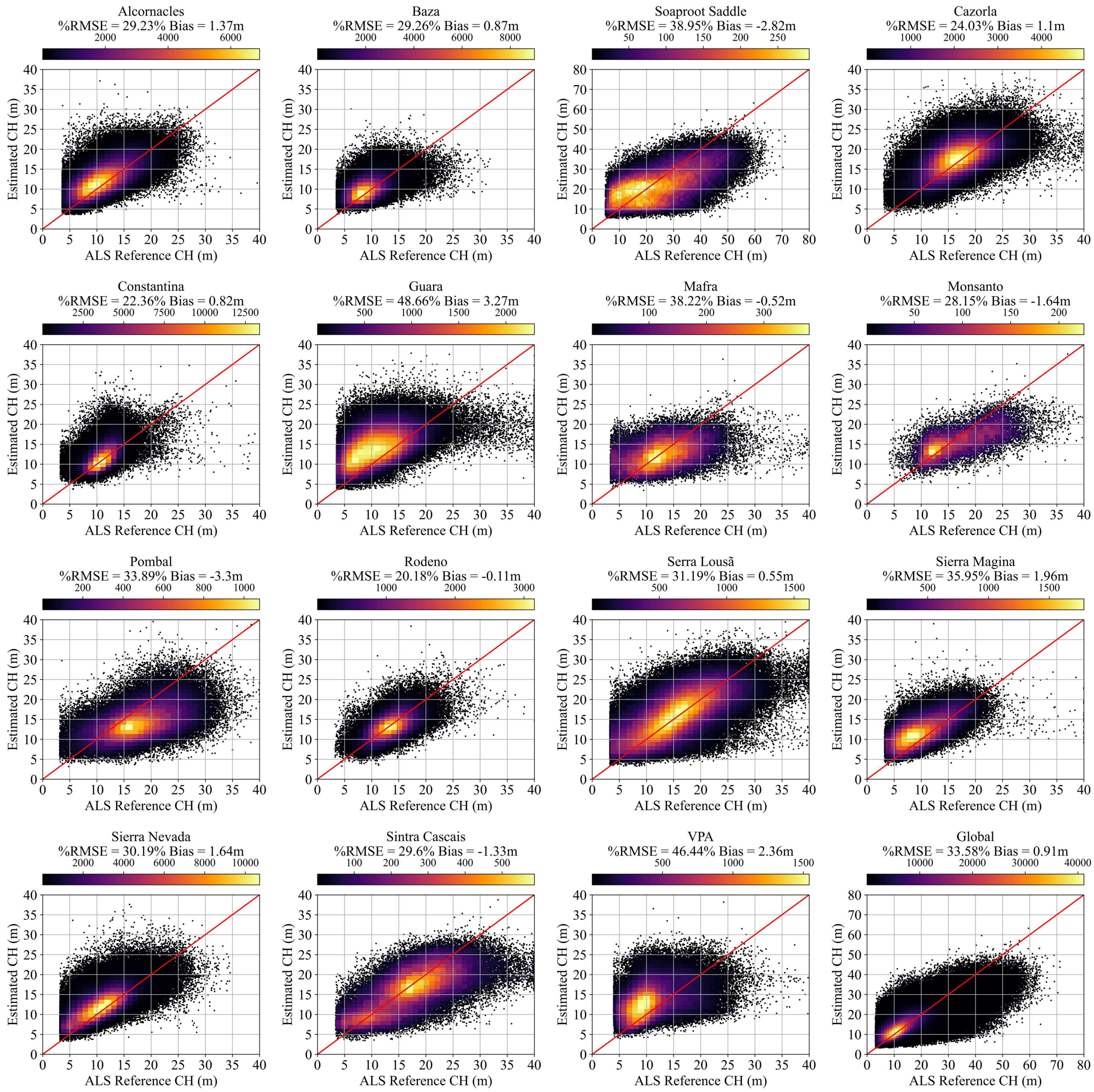

Figure 6. Scatterplots between the estimated Canopy height (CH) and respective airborne laser scanning (ALS) reference Canopy height model reference, per study area and globally.

smaller than the defined spatial resolution of this experiment. This led to corrections at the sub-pixel level, which have a less significant impact on the data. Furthermore, the inclusion of a mean spatial filter in the model architecture decreases the negative impacts of possible geolocation errors, reducing the need for this step. Another important fact, evidenced by the correlation values, geolocation correction did not significantly improve the calibration data quality. For instance, in Mafra and Pombal, it even led to calibration sets that were less correlated with the corresponding ALS reference. This shows that assuming the same geolocation offsets for all footprints in a given track may be effective for some of them, but it can introduce misalignments in others. Therefore, it is expected that more precise corrections may be achieved if the collocation is performed at the footprint level. However, this will require substantially higher processing time, hindering its application at larger scales. Additionally, the requirement of the overlapping between GEDI tracks and ALS point clouds with at least 4.00 points/m^2, restricts the worldwide applicability of this process (e.g. here it could only be performed in Portugal). Summarising, the set-up defined for this experiment is not optimal for taking advantage of the geolocation correction because: (1) the footprint spatial offsets do not exceed the product's resolution, hindering any improvements from this step; and (2)

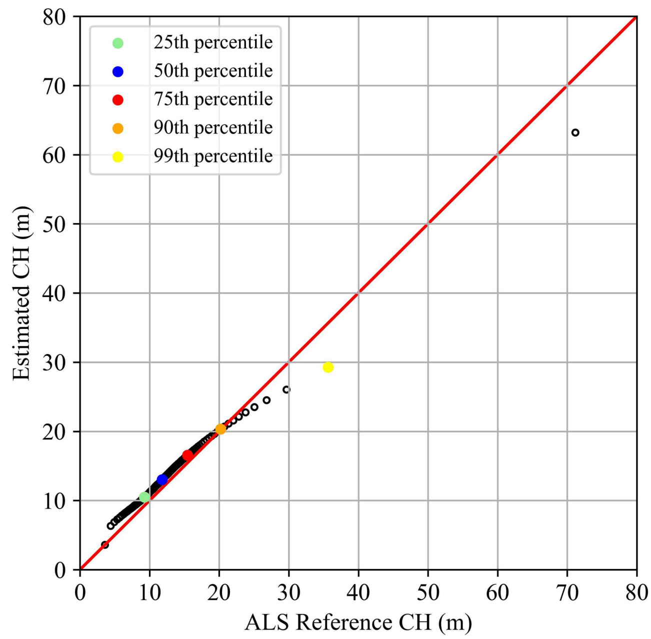

Figure 7. Comparison of distribution percentiles between the estimated Canopy height versus the airborne laser scanning reference Canopy height (CH).

the spatial averaging of the pre-processing smooths any sudden CH changes that would mislead the resulting models.

The results regarding the use of GEDI footprints acquired over a 4-year period were rather interesting. While on average they improved, this was more significant for areas with low footprint coverage, and not necessarily true for the others. Namely, for the most covered area, Monsanto, the estimation significantly worsened. This may be a consequence of the temporal variations' effects of utilising multi-year data for model calibration. However, it is worth noting that for study areas with more limited coverage, the increase of GEDI footprints for the calibration is more significant than the temporal effects.

Analysing the results by study areas, it was demonstrated that better mapping was achieved for Spain and California. There are apparently two reasons behind this fact: Portuguese study areas were more heterogeneous than the others, which can be verified by the average CV (0.42 for the Portuguese areas and 0.37 for the others), and they are steeper, making them more complex areas for the CH mapping. However, another reason that may justify the better mapping achieved for Spain and California seems to be the lower point cloud densities used for generating the reference CHMs. This leads to more homogeneous references, therefore easier to map. It would be expected that if the same point densities were available for all study areas, the resulting differences would be expected to be smaller. Regarding the previous results, it is expected that the model can produce reliable estimates for unseen areas (namely, those with different forest types), due to the local calibration. Furthermore, the accuracy of the produced maps is correlated with the CV, so the results may worsen for forest types with higher heterogeneity. Concerning different structural characteristics, it is also interesting to highlight the results achieved for Soaproot Saddle. Despite having significantly taller trees, the estimation quality did not decrease.

The main issue of the proposed modelling framework was the over- and underestimation at the lower and higher CH values, respectively. However, this is common in the literature when using machine learning regressors (Hansen et al. 2016; Healey et al. 2020; Lang et al. 2022; Min et al. 2024; Potapov et al. 2021; Schwartz et al. 2022). Notwithstanding the previous, the comparison with the Potapov and Lang maps showed that the ones here produced were significantly more reliable for Mediterranean landscapes. The main reason for this fact was the dedicated local calibration sets, opposing both the other maps that resorted to more heterogeneous calibration sets. Additionally, regarding the Potapov maps, it was expected once that the authors reported underestimation issues in their maps. They justified it by using Landsat data and the median as the aggregation metric of their random forests (Potapov et al. 2021). The underestimation of the methodology they proposed was also reported in Lang et al. (2022), Schwartz et al. (2022). Regarding the Lang maps, the worst results also seem to be linked to the fact that they only resort to S2 observations, and that for some of the study areas, there was a temporal misalignment of 1 year. In Schwartz et al. (2022), where the authors included S1-GRD data and local calibration sets, as in here, better results were also achieved. Another detail that may act as a disadvantage of Lang maps is their higher spatial resolution. On the one hand, it does not match the GEDI footprints size; and on the other hand, higher resolution usually leads to worse results.

Several steps can be taken to improve this model. First, to address the under- and overestimation, imbalanced learning methods may be applied to the data to generate more homogeneous calibration sets. Then, for the areas with a more limited GEDI coverage, the generation of synthetic data to artificially create larger calibration sets may be a solution for improving the reliability of their CH maps. In another line of research, it would be interesting to adapt this model to the estimation of biomass metrics. Other final tasks include the generation of CH maps at larger scales and the reuse of the model for mapping CH across multiple years, allowing time-series analysis as in Turubanova et al. (2023).

5. Conclusion

In this article, the benefits of combining optical, multifrequency SAR, InSAR, and ancillary data for the CH mapping through machine learning models were demonstrated. The potential of correcting the GEDI footprints on-orbit position was verified. However, its effectiveness in CH mapping was strongly hampered by the reduced coverage of the study area due to the calibration data, when only collocated tracks and the spatial resolution defined for this experiment were used. Regarding the calibration sets coverage, the model was also tested while using longer acquisition periods of GEDI footprints. Therefore, the use of multi-year periods of footprints should be reserved for study areas with coverages below 1.00%. The comparison of the maps produced here against existing global products showed the advantages of resorting to local calibration sets for mapping CH in Mediterranean-like study areas: they achieved lower errors and were less prone to CH underestimation.

Author contributions

CRediT: **João E. Pereira-Pires:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Juan Guerra-Hernández:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Adrián Pascual:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Raffaella Guida:** Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Writing – review & editing; **João M. N. Silva:** Conceptualization, Formal analysis, Investigation, Supervision, Writing – review & editing; **José M. Fonseca:** Conceptualization, Formal analysis, Investigation, Supervision, Writing – review & editing; **André Mora:** Conceptualization, Formal analysis, Investigation, Project administration, Supervision, Writing – review & editing.

Disclosure statement

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this study.

Funding

This research was supported by the Fundação para a Ciência Tecnologia under Grant UID/00066/2025 (awarded to Centre of Technology and Systems – Uninova), UID/00239/2025 (awarded to Forest Research Centre), LA/P/0092/2020 (awarded to Associate Laboratory TERRA), 2020.05015.BD (awarded to João E. Pereira-Pires for supporting his Ph.D.), and 2023.15225.TENURE.008 (awarded to Dr. Guerra-Hernández for supporting the FCT-Tenure). It was also supported by the European Space Agency under the grant 4000144713/24/I-DT-bgh (awarded Centre of Technology and Systems – Uninova and Forest Research Centre) and the NASA Fire Sense Grant #80NSSC24K1316 (for Adrián Pascual).

ORCID

João E. Pereira-Pires 0000-0003-2957-6372
Juan Guerra-Hernández 0000-0003-3518-2978

Adrián Pascual 0000-0002-2957-7810
Raffaella Guida 0000-0002-9041-4244
João M. N. Silva 0000-0001-5201-9836
José M. Fonseca 0000-0001-7173-7374
André Mora 0000-0003-1354-4739

Data availability statement

The data that support the findings of this research are available from the corresponding author upon reasonable request.

References

Adam, M., M. Urbazaev, C. Dubois, and C. Schmullius. 2020. "Accuracy Assessment of Gedi Terrain Elevation and Canopy Height Estimates in European Temperate Forests: Influence of Environmental and Acquisition Parameters." *Remote Sensing* 12 (23): 1–28. https://doi.org/10.3390/rs12233948.

Astola, H., T. Häme, L. Sirro, M. Molinier, and J. Kilpi. 2019. "Comparison of Sentinel-2 and Landsat 8 Imagery for Forest Variable Prediction in Boreal Region." *Remote Sensing of Environment* 223 (October 2017): 257–273. https://doi.org/10.1016/j.rse.2019.01.019.

Barreira, P., A. Mora, J. E. Pereira-Pires, J. M. Fonseca, and J. Guerra-Hernández. 2024. Forest Height Estimation Using Machine Learning Regressors with Sar Data. In *2024 8th International Young Engineers Forum on Electrical and Computer Engineer-ing (YEF-ECE)*8–13, Portugal, Almada: IEEE. https://doi.org/10.1109/YEF-ECE62614.2024.10624947

Besic, N., N. Picard, C. Vega, J.-D. Bontemps, L. Hertzog, J.-P. Renaud, F. Fogel, et al. 2025. "Remote-Sensing-Based Forest Canopy Height Mapping: Some Models Are Useful, but Might They Provide Us with Even More Insights When Combined?" *Geoscientific Model Development* 18 (2): 337–359. https://doi.org/10.5194/gmd-18-337-2025.

Bolton, D. K., P. Tompalski, N. C. Coops, J. C. White, M. A. Wulder, T. Hermosilla, M. Queinnec, et al. 2020. "Optimizing Landsat Time Series Length for Regional Mapping of Lidar-Derived Forest Structure." *Remote Sensing of Environment* 239 (June 2019): 111645. https://doi.org/10.1016/j.rse.2020.111645.

Braun, A., and Veci, L. 2021. Sentinel-1 Toolbox - TOPS Interferometry Tutorial. ESA. https://step.esa.int/docs/tutorials/S1TBX TOPSAR Interferometry with Sentinel-1 Tuto-rial_v2.pdf

Cui, L., Z. Jiao, Y. Dong, M. Sun, X. Zhang, S. Yin, A. Ding, Y. Chang, J. Guo, and R. Xie. 2019. "Estimating Forest Canopy Height Using MODIS BRDF Data Emphasizing Typical-Angle Reflectances." *Remote Sensing* 11 (19): 1–21. https://doi.org/10.3390/rs11192239.

Dewitz, J., and U.S. Geological Survey. 2021. *National Land Cover Database (NLCD) 2019 Products (ver. 2.0, June 2021)*. U.S. Geological Survey data release. https://doi.org/10.5066/P9KZCM54

Direção-Geral do Território. 2018. Carta de Ocupação do Solo 2018. https://snig.dgterritorio.gov.pt/rndg/srv/por/catalog.search#/metadata/b498e89c-1093-4793-ad22-63516062891b.

Dubayah, R., J. B. Blair, S. Goetz, L. Fatoyinbo, M. Hansen, S. Healey, M. Hofton, et al. 2020. "The Global Ecosystem Dynamics Investigation: High-Resolution Laser Ranging of the Earth'S Forests and Topography." *Science of Remote Sensing* 1: 100002. https://doi.org/10.1016/j.srs.2020.100002.

European Environment Agency. 2019. Corine Landcover 2018. https://doi.org/10.2909/960998c1-1870-4e82-8051-6485205ebbac

Fagua, J. C., P. Jantz, S. Rodriguez-Buritica, L. Duncanson, and S. J. Goetz. 2019. "Integrating LiDAR, Multispectral and Sar Data to Estimate and Map Canopy Height in Tropical Forests." *Remote Sensing* 11 (22): 2697. https://doi.org/10.3390/rs11222697.

Fayad, I., D. Ienco, N. Baghdadi, R. Gaetano, C. A. Alvares, J. L. Stape, H. Ferraço Scolforo, and G. Le Maire. 2021. "A CNN-based Approach for the Estimation of Canopy Heights and Wood Volume from GEDI Waveforms." *Remote Sensing of Environment* 265 (August): 112652. https://doi.org/10.1016/j.rse.2021.112652.

García, M., S. Saatchi, S. Ustin, and H. Balzter. 2018. "Modelling Forest Canopy Height By Integrating Airborne LiDAR Samples with Satellite Radar and Multispectral Imagery." *International Journal of Applied Earth Observation and Geoinformation* 66 (November 2017): 159–173. https://doi.org/10.1016/j.jag.2017.11.017.

Guerra-Hernández, J., J. E. Pereira-pires, and A. Pascual. 2024. "Canopy Height Models of Iberian Peninsula 2018-2021 at 10 M Resolution [Data Set]." *Zenodo*. https://doi.org/10.5281/zenodo.13946862.

Guerra-Hernández, J., L. L. Narine, A. Pascual, E. Gonzalez-Ferreiro, B. Botequim, L. Malambo, A. Neuenschwander, S. C. Popescu, and S. Godinho. 2022. "Aboveground Biomass Mapping By Integrating ICESat-2, SENTINEL-1, SENTINEL-2, ALOS2/PALSAR2, and Topographic Information in Mediterranean Forests." *GIScience & Remote Sensing* 59 (1): 1509–1533. https://doi.org/10.1080/15481603.2022.2115599.

Hancock, S., J. Armston, M. Hofton, X. Sun, H. Tang, L. I. Duncanson, J. R. Kellner, and R. Dubayah. 2019. "The GEDI Simulator: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of Spaceborne Missions." *Earth and Space Science* 6 (2): 294–310. https://doi.org/10.1029/2018EA000506.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, et al. 2013. "High-Resolution Global Maps of 21St-Century Forest Cover Change." *Science* 342(6160): 850–853. https://doi.org/10.1126/science.1244693.

Hansen, M. C., P. V. Potapov, S. J. Goetz, S. Turubanova, A. Tyukavina, A. Krylov, A. Kommareddy, and A. Egorov. 2016. "Mapping Tree Height Distributions in Sub-Saharan Africa Using Landsat 7 and 8 Data." *Remote Sensing of Environment* 185: 221–232. https://doi.org/10.1016/j.rse.2016.02.023.

Healey, S. P., Z. Yang, N. Gorelick, and S. Ilyushchenko. 2020. "Highly Local Model Calibration with a New GEDI LiDAR Asset on Google Earth Engine Reduces Landsat Forest Height Signal Saturation." *Remote Sensing* 12 (17): 2840. https://doi.org/10.3390/rs12172840.

Hofton, M. A., J. B. Minster, and J. B. Blair. 2000. "Decomposition of Laser Altimeter Waveforms." *IEEE Transactions on Geoscience and Remote Sensing* 38 (4): 1989–1996. https://doi.org/10.1109/36.851780.

Hyde, P., R. Dubayah, W. Walker, J. B. Blair, M. Hofton, and C. Hunsaker. 2006. "Mapping Forest Structure for Wildlife Habitat Analysis Using Multi-Sensor (LiDAR, Sar/InSAR, ETM+, Quickbird) Synergy." *Remote Sensing of Environment* 102 (1–2): 63–73. https://doi.org/10.1016/j.rse.2006.01.021.

ICNF. 2022. Projecto áGIL - Mafra, Monsanto, Pombal, Serra da Lousã, Sintra-Cascais, Vila Pouca de Aguiar. https://geocatalogo.icnf.pt/geovisualizador/agil.html

Jiang, F., F. Zhao, K. Ma, D. Li, and H. Sun. 2021. "Mapping the Forest Canopy Height in Northern China By Synergizing ICESat-2 with sentinel-2 Using a Stacking Algorithm." *Remote Sensing* 13 (8): 1535. https://doi.org/10.3390/rs13081535.

Lang, N., N. Kalischek, J. Armston, K. Schindler, R. Dubayah, and J. D. Wegner. 2022. "Global Canopy Height Regression and Uncertainty Estimation from GEDI LIDAR Waveforms with Deep Ensembles." *Remote Sensing of Environment* 268: 112760. https://doi.org/10.1016/j.rse.2021.112760.

Li, A., C. Huang, G. Sun, H. Shi, C. Toney, Z. Zhu, M. G. Rollins, S. N. Goward, and J. G. Masek. 2011. "Modeling the Height of Young Forests Regenerating from Recent Disturbances in Mississippi Using Landsat and ICESat Data." *Remote Sensing of Environment* 115 (8): 1837–1849. https://doi.org/10.1016/j.rse.2011.03.001.

Li, X., K. Wessels, J. Armston, S. Hancock, R. Mathieu, R. Main, L. Naidoo, B. Erasmus, and R. Scholes. 2023. "First Validation of GEDI Canopy Heights in African Savannas." *Remote Sensing of Environment* 285: 113402. https://doi.org/10.1016/j.rse.2022.113402.

Min, W., Y. Chen, W. Huang, J. P. Wilson, H. Tang, M. Guo, and R. Xu. 2024. "Incorporating of Spatial Effects in Forest Canopy Height Mapping Using Airborne, Spaceborne Lidar and Spatial Continuous Remote Sensing Data." *International Journal of Applied Earth Observation and Geoinformation* 133: 104123. https://doi.org/10.1016/j.jag.2024.104123.

NEON (National Ecological Observatory Network). 2023. Discrete return LiDAR point cloud (DP1.30003.001), RELEASE-2023. https://doi.org/10.48443/xxby-5a18

OpenTopography. 2013. Shuttle Radar Topography Mission (SRTM) Global. https://doi.org/10.5069/G9445JDF

Pascual, A., and J. Guerra-Hernández. 2023. "An Integrated Assessment of Carbon Emissions from Forest Fires Beyond Impacts on Aboveground Biomass. a Showcase Using Airborne Lidar and GEDI Data over a Megafire in Spain." *Journal of Environmental Management* 345 (April): 118709. https://doi.org/10.1016/j.jenvman.2023.118709.

Pascual, A., P. B. May, A. Cárdenas-Martínez, J. Guerra-Hernández, N. Hunka, J. M. Bruening, S. P. Healey, J. D. Armston, and R. O. Dubayah. 2025. "Calibration of GEDI Footprint Aboveground Biomass Models in Mediterranean Forests with NFI Plots: A Comparison of Approaches." *Journal of Environmental Management* 375: 124313. https://doi.org/10.1016/j.jenvman.2025.124313.

Pereira-Pires, J. E., R. Guida, J. M. N. Silva, A. Mora, and J. M. Fonseca. 2023. Forest Height Estimation Using Sentinel-1/2 and Alos-2. In *2023 8th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*1–6, Kuta, Bali, Indonesia: IEEE, https://doi.org/10.1109/APSAR58496.2023.10388740

Pereira-Pires, J. E., A. Mora, R. Guida, J. M. Fonseca, J. M. N. Silva, and P. Barreira. 2024. Mapping Forest Height with Multifrequency Sar, InSAR, and Multispectral Datasets In Technological Innovation for Human-Centric Systems (DoCEIS 2024). IFIP Advances in Information and Communication Technology. (vol. 716, p 309–321).https://doi.org/10.1007/978-3-031-63851-0_22

Pereira-Pires, J. E., J. Guerra-Hernández, J. M. N. Silva, J. M. Fonseca, R. Guida, and A. Mora. 2024. Forest Height Mapping Combining GEDI, ALOS-2, Sentinel-1/2, and Ancillary Data. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium* 5337–5340, Athens, Greece. https://doi.org/10.1109/IGARSS53475.2024.10640684

Pereira-Pires, J. E., J. Guerra-Hernández, J. M. N. Silva, J. M. Fonseca, R. Guida, and A. Mora. 2025. "Forest Height Mapping with Multifrequency Sar in Mediterranean Forests." *IEEE Transactions on Geoscience and Remote Sensing* 63: 1. https://doi.org/10.1109/TGRS.2025.3538216.

Peterson, B., and K. J. Nelson. 2014. "Mapping Forest Height in Alaska Using GLAS, Landsat Composites, and Airborne Lidar." *Remote Sensing* 6 (12): 12409–12426. https://doi.org/10.3390/rs61212409.

PNOA. 2020. LIDAR 2a Cobertura (2015-Actualidad). http://centrodedescargas.cnig.es/

Potapov, P., X. Li, A. Hernandez-Serna, A. Tyukavina, M. C. Hansen, A. Kommareddy, A. Pickens, et al. 2021. "Mapping Global Forest Canopy Height through Integration of GEDI and Landsat Data." *Remote Sensing of Environment* 253 (August): 112165. https://doi.org/10.1016/j.rse.2020.112165.

Qi, W., J. Armston, C. Choi, A. Stovall, S. Saarela, M. Pardini, L. Fatoyinbo, K. Papathanassiou, A. Pascual, and R. Dubayah. 2025. "Mapping Large-Scale Pantropical Forest Canopy Height By Integrating GEDI Lidar and TanDEM-X InSAR Data." *Remote Sensing of Environment* 318: 114534. https://doi.org/10.1016/j.rse.2024.114534.

Radke, D., D. Radke, and J. Radke. 2020. "Beyond Measurement: Extracting Vegetation Height from High Resolution Imagery with Deep Learning." *Remote Sensing* 12 (22): 1–22. https://doi.org/10.3390/rs12223797.

Roy, D. P., H. B. Kashongwe, and J. Armston. 2021. "The Impact of Geolocation Uncertainty on GEDI Tropical Forest Canopy Height Estimation and Change Monitoring." *Science of Remote Sensing* 4: 100024. https://doi.org/10.1016/j.srs.2021.100024.

Schwartz, M., P. Ciais, C. Ottlé, A. De Truchis, C. Vega, I. Fayad, and M. Brandt. 2024. "High-Resolution Canopy Height Map in the Landes Forest (France) Based on GEDI, Sen-Tinel-1, and Sentinel-2 Data with a Deep Learning Approach." International Journal of Applied Earth Observation and Geoinformation 128: 1–39. 103711. https://doi.org/10.1016/j.jag.2024.103711.

Sun, Y., J. Li, Y. Xu, T. Zhang, and X. Wang. 2023. "Deep Learning Versus Conventional Methods for Missing Data Imputation: A Review and Comparative Study." *Expert Systems with Applications* 227: 120201. https://doi.org/10.1016/j.eswa.2023.120201.

Tang, H., J. Stoker, S. Luthcke, J. Armston, K. Lee, B. Blair, and M. Hofton. 2023. "Evaluating and Mitigating the Impact of Systematic Geolocation Error on Canopy Height Measurement Performance of Gedi." *Remote Sensing of Environment* 291: 113571. https://doi.org/10.1016/j.rse.2023.113571.

Turubanova, S., P. Potapov, M. C. Hansen, X. Li, A. Tyukavina, A. H. Pickens, A. Hernandez-Serna, et al. 2023. "Tree Canopy Extent and Height Change in Europe, 2001–2021, Quantified Using Landsat Data Archive." *Remote Sensing of Environment* 298: 113797. https://doi.org/10.1016/j.rse.2023.113797.

Wang, M., R. Sun, and Z. Xiao. 2018. "Estimation of Forest Canopy Height and Aboveground Biomass from Spaceborne LiDAR and Landsat Imageries in Maryland." *Remote Sensing* 10 (2): 344. https://doi.org/10.3390/rs10020344.

Wang, D., B. Wan, P. Qiu, Z. Zuo, R. Wang, and X. Wu. 2019. "Mapping Height and Aboveground Biomass of Mangrove Forests on Hainan Island Using UAV-LiDAR Sampling." *Remote Sensing* 11 (18): 1–25. https://doi.org/10.3390/rs11182156.

Wang, C., A. J. Elmore, I. Numata, M. A. Cochrane, L. Shaogang, J. Huang, Y. Zhao, and Y. Li. 2022. "Factors Affecting Relative Height and Ground Elevation Estimations of GEDI Among Forest Types Across the Conterminous Usa." *GIScience & Remote Sensing* 59 (1): 975–999. https://doi.org/10.1080/15481603.2022.2085354.

Xiao, J., F. Chevallier, C. Gomez, L. Guanter, J. A. Hicke, A. R. Huete, K. Ichii, et al. 2019. "Remote Sensing of the Terrestrial Carbon Cycle: A Review of Advances over 50 Years." *Remote Sensing of Environment* 233 (January): 111383. https://doi.org/10.1016/j.rse.2019.111383.

Yang, C., D. Peng, K. Deng, L. Jiang, M. Zhao, W. Zeng, Y. Shao, and N. Wang. 2024. "Quantification of GEDI Geolocation Error and Its Influence on Elevation and Canopy Height." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17: 18012–18024. https://doi.org/10.1109/JSTARS.2024.3461843.

Appendix A Supplementary Data

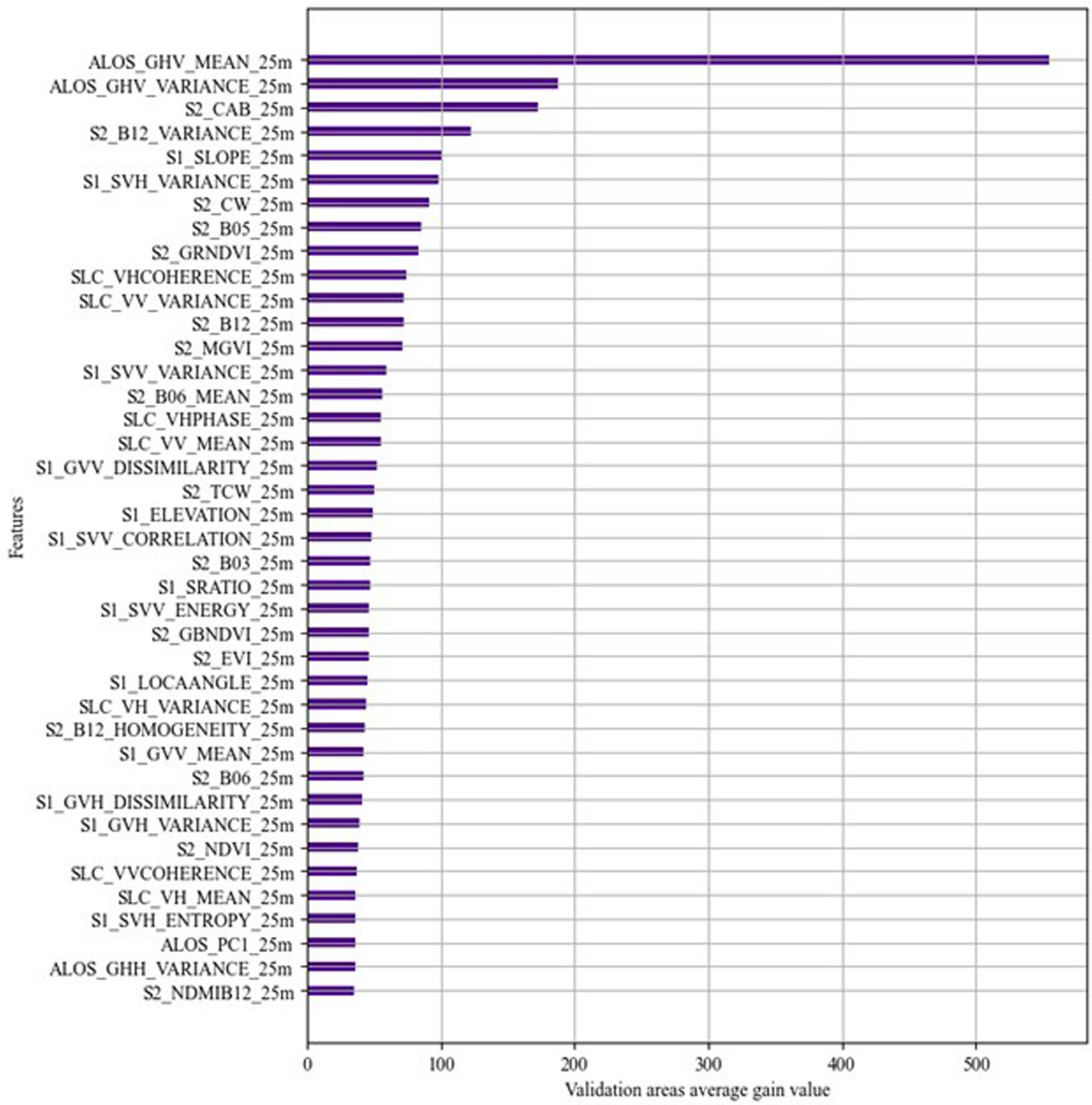

Figure A1. Best 40 features according to gain and respective score. The feature description is provided in Table A1.

Table A1. Description of the Top 40 variables. The spectral indices equations can be consulted at https://www.indexdatabase.de/.

Feature	Description
SLC_VV_VARIANCE_25m	Texture variance for the InSAR coherence with VV polarisation
SLC_VV_MEAN_25m	Texture mean for the InSAR coherence with VV polarisation
SLC_VVCOHERENCE_25m	InSAR coherence with VV polarisation
SLC_VH_VARIANCE_25m	Texture variance for the InSAR coherence with VH polarisation
SLC_VH_MEAN_25m	Texture mean for the InSAR coherence with VH polarisation
SLC_VHPHASE_25m	InSAR phase with VH polarisation
SLC_VHCOHERENCE_25m	InSAR coherence with VH polarisation
S2_TCW_25m	Reflectance Tasselled Cap Wetness
S2_NDVI_25m	Normalised Difference Vegetation Index (NDVI)
S2_NDMIB12_25m	Normalised Difference Moisture Index using Band 12
S2_MGVI_25m	Modified Green Vegetation Index
S2_GRNDVI_25m	Green-Red NDVI
S2_GBNDVI_25m	Green-Blue NDVI
S2_EVI_25m	Enhanced Vegetation Index
S2_CW_25m	Canopy Water Content (Biophysical variable)
S2_CAB_25m	Chlorophyll content in the leaf (Biophysical variable)
S2_B12_VARIANCE_25m	Texture variance for-reflectance Band 12
S2_B12_HOMOGENEITY_25m	Texture homogeneity for-reflectance Band 12
S2_B12_25m	Reflectance Band 12
S2_B06_MEAN_25m	Texture mean for-reflectance Band 6
S2_B06_25m	Reflectance Band 6
S2_B05_25m	Reflectance Band 5
S2_B03_25m	Reflectance Band 3
S1_SVV_VARIANCE_25m	Texture variance for Sentinel-1 Sigma Backscatter with VV polarisation
S1_SVV_ENERGY_25m	Texture energy for Sentinel-1 Sigma Backscatter with VV polarisation
S1_SVV_CORRELATION_25m	Texture correlation for Sentinel-1 Sigma Backscatter with VV polarisation
S1_SVH_VARIANCE_25m	Texture variance for Sentinel-1 Sigma Backscatter with VH polarisation
S1_SVH_ENTROPY_25m	Texture entropy for Sentinel-1 Sigma Backscatter with VH polarisation
S1_SRATIO_25m	Sentinel-1 Sigma Backscatter ratio between VH and VV polarisations
S1_SLOPE_25m	Terrain slope
S1_LOCAANGLE_25m	Sentinel-1 Local Incidence Angle
S1_GVV_MEAN_25m	Texture mean for Sentinel-1 Gamma Backscatter with VV polarisation
S1_GVV_DISSIMILARITY_25m	Texture dissimilarity for Sentinel-1 Gamma Backscatter with VV polarisation
S1_GVH_VARIANCE_25m	Texture variance for Sentinel-1 Gamma Backscatter with VH polarisation
S1_GVH_DISSIMILARITY_25m	Texture dissimilarity for Sentinel-1 Gamma Backscatter with VH polarisation
S1_ELEVATION_25m	Terrain Elevation
ALOS_PC1_25m	ALOS-2 first principal component
ALOS_GHV_VARIANCE_25m	Texture variance for ALOS-2 Gamma Backscatter with HV polarisation
ALOS_GHV_MEAN_25m	Texture mean for ALOS-2 Gamma Backscatter with HV polarisation
ALOS_GHH_VARIANCE_25m	Texture variance for ALOS-2 Gamma Backscatter with HH polarisation

Table A2. Correction values obtained by the gediSimulator's collocate Waves function, and respective correlation values achieved for the corrected track. The average of the correction values was calculated only considering the tracks used for calibration.

Study area	Latitude correction (m)	Longitude correction (m)	Collocation correlation	Used for calibration
Mafra	−26.92	11.02	0.93	Yes
	10.16	16.20	0.86	Yes
	−0.01	−11.92	0.35	No
Monsanto	8.04	2.05	0.93	Yes
	19.36	2.31	0.88	Yes
	11.27	6.04	0.82	Yes
	2.64	16.33	0.91	Yes
	3.00	19.00	0.04	No
Pombal	9.78	1.26	0.88	Yes
	14.90	−8.05	0.89	Yes
Serra Lousã	−12.52	-0.97	0.91	Yes
	11.36	−3.69	0.86	Yes
Sintra-Cascais	−2.96	−12.95	0.89	Yes
	86.19	−12.52	0.54	No
Average correction values	11.81	7.35	–	

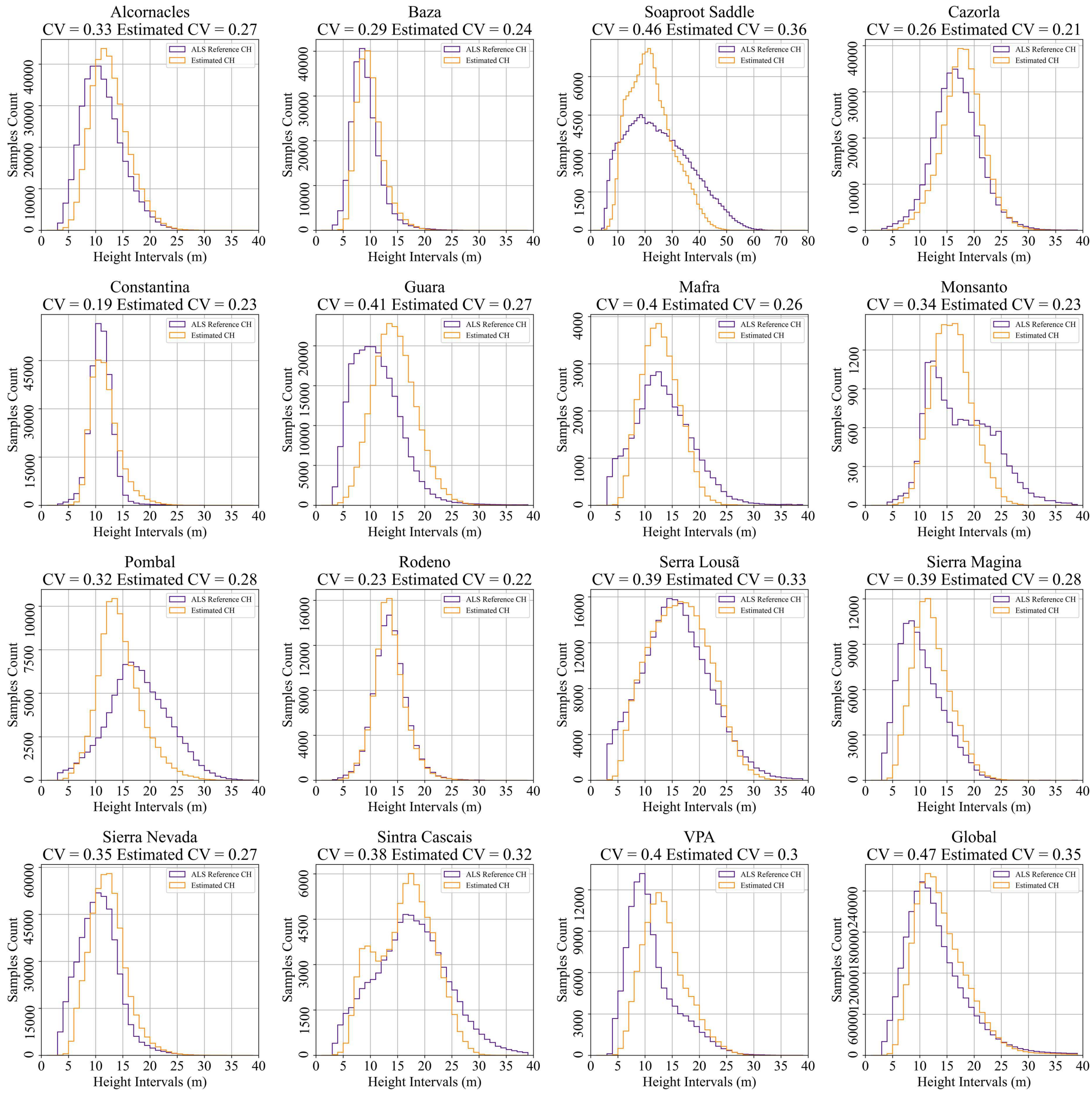

Figure A2. Histograms of the estimated Canopy height and the corresponding airborne laser scanning reference Canopy height model, per study area and combining all study areas.

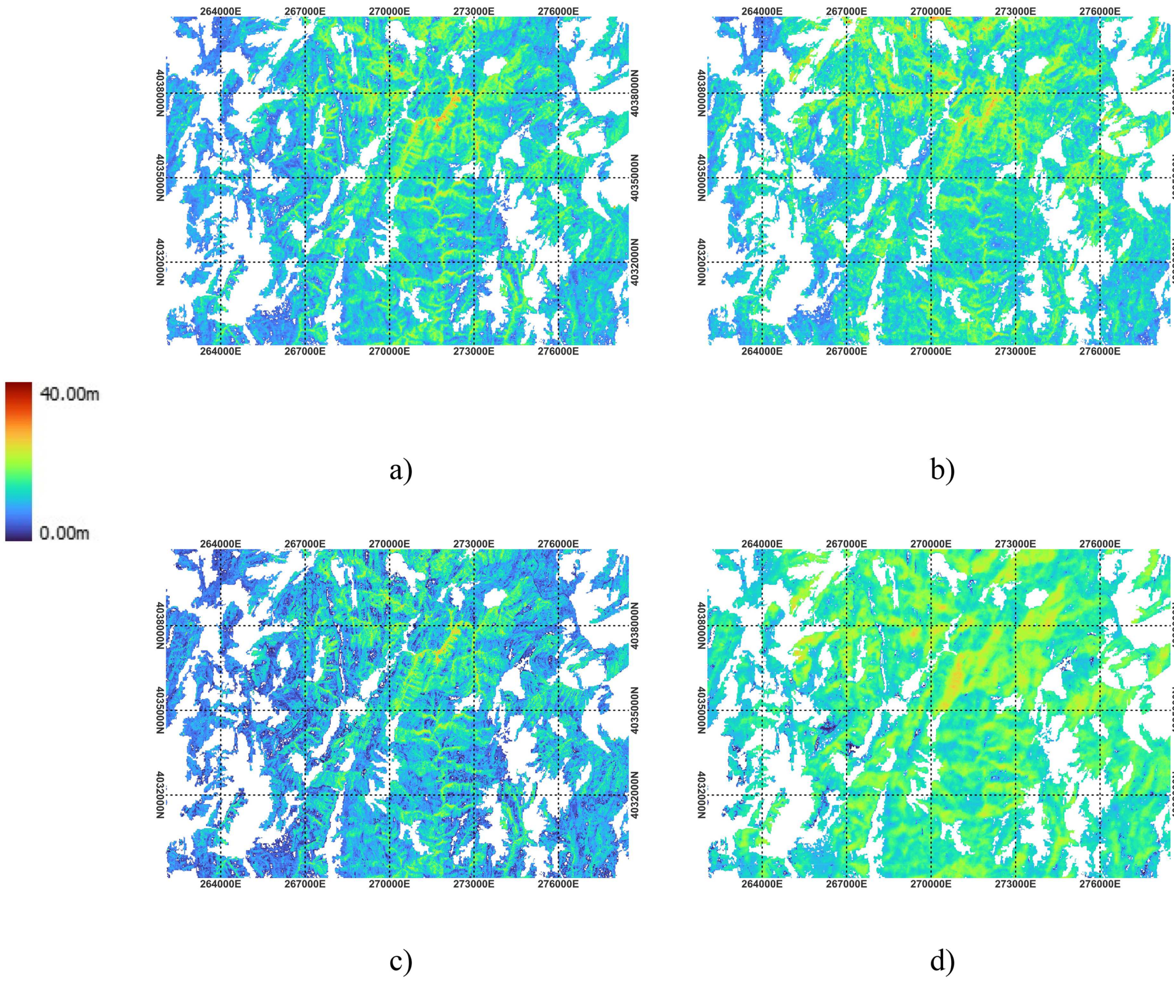

Figure A3. Canopy height map for Alcornacles (EPSG: 25830) study area: (a) the ALS reference at 25m; (b) our map; (c) the ALS reference at 10m; and (d) the Lang et al. global map.

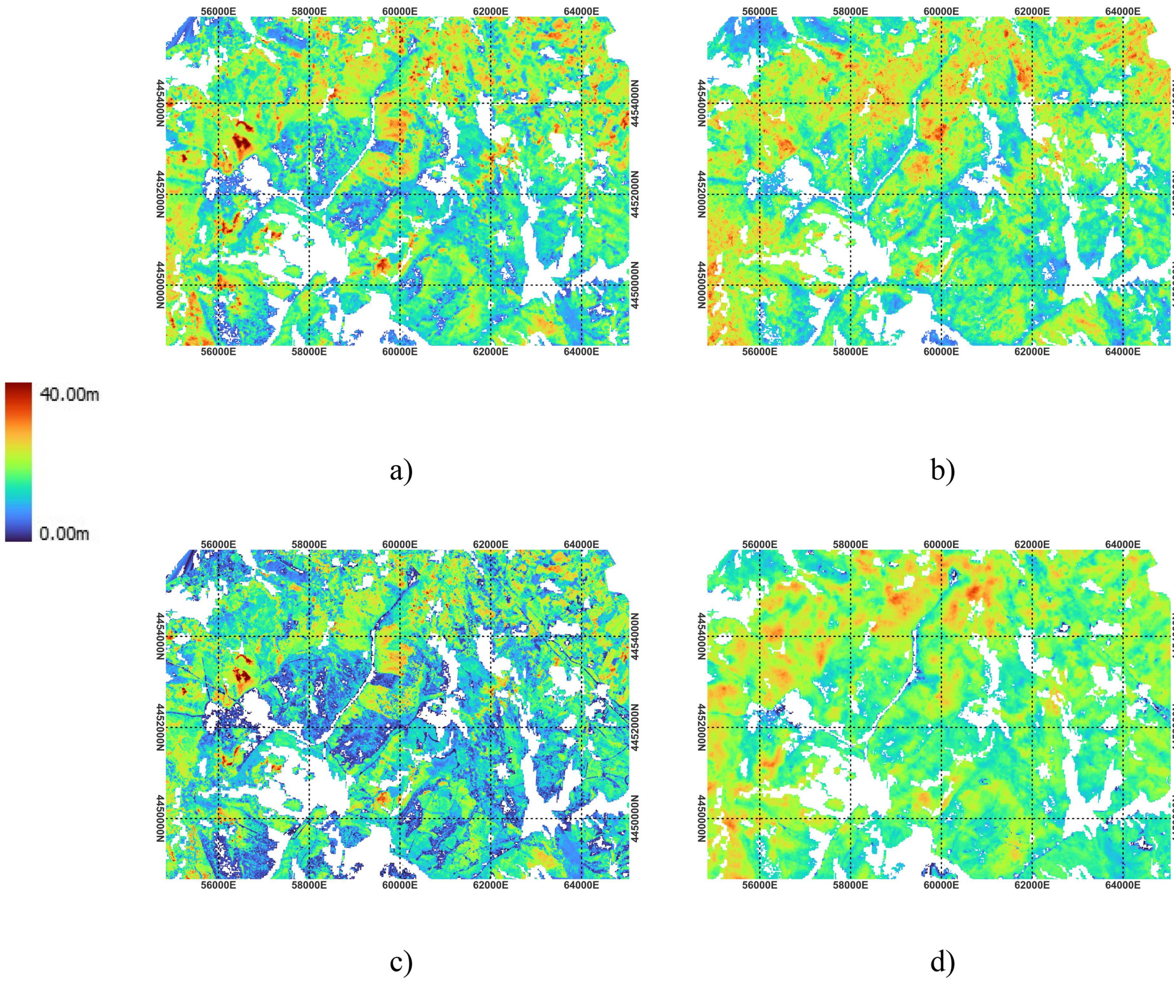

Figure A4. CH maps for Serra Lousã (EPSG: 3763): (a) the ALS reference at 25 m; (b) the map produced here; (c) the ALS reference at 10 m; and (d) the Lang et al. global map.

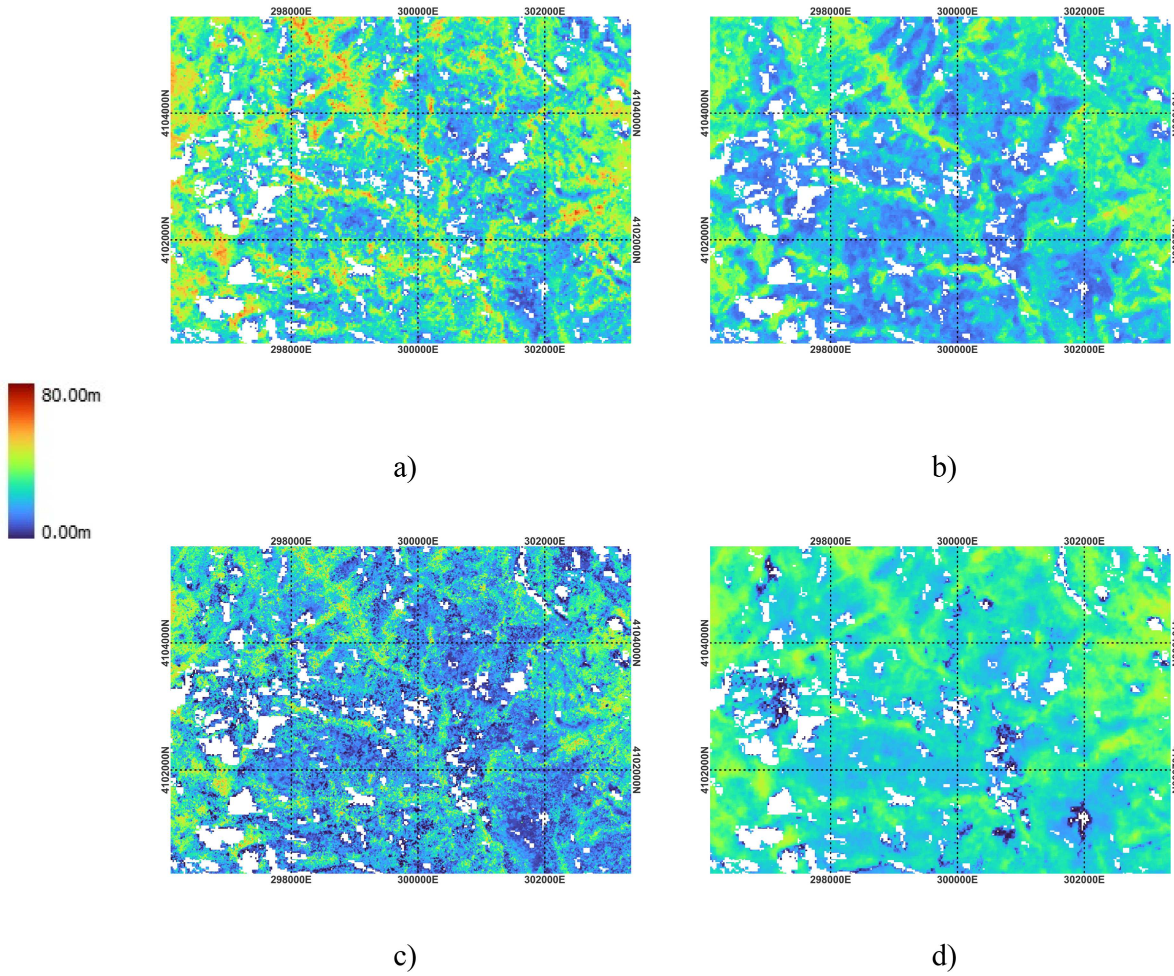

Figure A5. CH maps for Soaproot Saddle (EPSG: 32611): (a) the ALS reference at 25 m; (b) the map produced here; (c) the ALS reference at 10 m; and (d) the Lang global map.