

Biomass Prediction Combining Radar, Multispectral and Field Inventory Data

João E. Pereira-Pires¹, Juan Guerra-Hernández², Amal Anil-Lathika², Svetlana Saarela³, João M. N. Silva², José M. Fonseca¹, André Mora¹

¹Centre of Technology and Systems/UNINOVA, Associate Lab of Intelligent Systems, NOVA School of Science and Technology, Portugal

²Forest Research Centre, Associate Laboratory TERRA, School of Agriculture, University of Lisbon, Portugal

³Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Norway

Keywords: Biomass, Synthetic Aperture Radar, Multispectral, Field Inventory, Extreme Gradient boosting

Abstract

Information on forest biomass is crucial for a better management and understanding of wildfires effects (pre- and post-fire analyses), as well as for a more accurate monitoring of carbon cycle dynamics. The current literature adopts two main approaches for the retrieving of biomass values: 1) predicting the biomass through allometric equations based on field measurements or Airborne Laser Scanning (ALS) metrics; and 2) using machine learning models based on spaceborne data sources (radar, optical, and laser), trained with field or ALS data. While approach 1) is usually spatially limited and has high costs; approach 2) mostly focuses on the Aboveground Biomass Density (AGBD) (i.e.: leaves, branches, trunk), disregarding the other components of the forest ecosystems as root biomass (BGBD), litter biomass (LBD) and shrubs biomass (SBD)) which leads to biomass underestimation.

This article proposes a novel methodology capable of producing accurate predictions of the four previously mentioned biomass components, that integrates multi-source remote sensing observations acquired by synthetic aperture radars (ALOS-2 and Sentinel-1), multispectral sensors (Sentinel-2), and ancillary data (on topography and land cover). The methodology resorts to an extreme gradient boosting regressor trained with field inventory data for the AGBD and BGBD, and ALS data for the LBD and SBD. It was demonstrated the importance of including all the proposed datasets and that reliable predictions were achieved. Namely, %RMSEs of 63.29%, 62.90%, 41.07%, 28.34% were achieved for the AGBD, BGBD, LBD, and SBD, respectively.

1. Introduction

The current scientific landscape concerning the prediction of biomass has focused on its aboveground components (over forests), overlooking others such as the belowground, litter, and shrub. The impact of disregarding these components when quantifying biomass have been analysed by several authors. Namely, (Pascual et al., 2023) reported that belowground and litter accounted for 38.1% of the total biomass in Extremadura, Spain; and (Y. Chen et al., 2023) estimate that the belowground are responsible for 18.1% of biomass growth in the last 20 years across China. Additionally, (Pascual & Guerra-Hernández, 2023) analysed this in the perspective of wildfires' carbon emissions, showing that for a megafire occurred in Northwestern Spain aboveground corresponds only to 63% of the emissions.

The importance of these additional biomass components also results from the changing landscape, specifically, the expansion of shrubs (A. Chen et al., 2024; Enes et al., 2020; Greaves et al., 2016). This phenomenon has been fuelled by the increasing extension of burnt areas resulting from wildfires, in which the previously existent vegetation is substituted by shrubs. For instance, (Enes et al., 2020) verified that *Pinus Pinasters* forests are systematically succeed by shrublands in post-fire landscapes. Other important factor for this landscape modification is the shrub encroachment (i.e., the transition from herbaceous to shrubs). This has been verified in temperate regions, where due to climate change and anthropogenic activities, grasslands are decreasing in favour of shrubs (A. Chen et al., 2024). Finally, the Arctic thaw, consequence of

global warming, is also leading to the growth of tundra shrubs (Greaves et al., 2016).

The direct measurement through field campaigns or the inventory of traditional forest variables (as diameter at breast height and canopy height) to be used in allometric models, represent a reliable approach for quantifying biomass. However, the associated financial costs and extensive work, hinder its use for mapping larger scales. Therefore, alternatives have been proposed that take advantage of remote sensing data.

As said previously, the literature has been focused on proposing alternative ways for mapping the Aboveground Biomass Density (AGBD). (Duncanson et al., 2022) have produced a global AGBD product resorting to the Global Ecosystem Dynamics Investigation (GEDI), a spaceborne Light Detection and Ranging (LiDAR) sensor. The authors used an ordinary least squares regression and GEDI's Relative Height (RH) metrics for the AGBD prediction. (Silva et al., 2025) resorted to the same predictors, however for predicting biomass in selected study areas in Portugal and Spain. An extensive comparison between different models was performed, and it was verified that a multilayer perceptron and a voting regressor (combining a multilayer perceptron and a random forest) outperformed all the other assessed models. Alternatively, (Guerra-Hernández et al., 2022) made use of the Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2), Airborne Laser Scanning (ALS), Sentinel-1/2, and the Advanced Land Observing Satellite 2 (ALOS-2) for predicting AGBD in a Spanish study area. More recently, (Pascual & Guerra-Hernández, 2026) have demonstrated the possibility of predicting AGBD from Google's Alpha Earth Embeddings.

As outlined in (Dlamini et al., 2025) there are three main platform have been employed for Belowground Biomass Density (BGBD) prediction: proximal, airborne, and spaceborne. While ground penetrating radar was already used for retrieving biomass directly (Hardiman et al., 2017), most approaches try to predict BGBD from the AGBD and other ancillary variables using parametric and machine learning models. For instance, (Pascual et al., 2023) proposed an exponential model with AGBD as the independent variable; (Y. Chen et al., 2023) and (Pereira-Pires et al., 2025) employed machine learning regressors (random forests and extreme gradient boosting, respectively) that combined optical and radar predictors. Finally, (Silva et al., 2025) in their experiment compared different regressors, verifying that random forests and voting regressors achieved the best results for local prediction while using GEDI's RH metrics for Portuguese and Spanish study areas.

The Litter Biomass Density (LBD) prediction has been ensured by LIDAR allometric models (Pascual et al., 2023); Moderate Resolution Imaging Spectroradiometer (MODIS) derived soil-adjusted spectral indices (Ren et al., 2018); and machine learning regressors that resort to multi-source spaceborne remote sensing observations (Pereira-Pires et al., 2025; Silva et al., 2025). Finally, for Shrub Biomass Density (SBD) prediction, (Liu et al., 2024) proposed a methodology that integrates ground measurements, high-resolution imagery acquired by unnamed aerial vehicles, Landsat, and Sentinel-1/2; resorting to a random forest for mapping SBD at a regional-scale. In (Greaves et al., 2016) verified the expansion of shrubs in the Arctic Tundra also resorting to a random forest and multi-source remote sensing datasets. Finally, in (Pereira-Pires et al., 2025; Silva et al., 2025) the use of extreme gradient boosting regressors was also explored for SBD prediction.

In this article a methodology capable of mapping biomass in four dimensions (AGBD, BGBD, LBD, and SBD) is proposed. It makes use of multi-source remote sensing datasets including multi-frequency radar backscatter from ALOS-2 and Sentinel-1, interferometry from Sentinel-1, multispectral from Sentinel-2, and ancillary features on the land cover and topography. The biomass prediction is ensured by extreme gradient boosting

regressors (one per component).

2. Study Areas

The study areas considered in this article encompass six sites Portugal and three in Spain, displayed in Figure 1. The Portuguese areas have a limited spatial extent, covering a total of 34,109 ha (from the 850 in Monsanto to the 15,569 ha in Serra Lousã). These areas are characterised by Resin pine (*Pinus Pinaster*), Eucalyptus (*Eucalyptus spp.*), Stone pine (*Pinus pinea*), Cork Oak (*Quercus suber*), Oaks (*Quercus spp.*) and other Broadleaves (such as *Castanea sativa*, *Quercus faginea*, *Olea europaea* var. *sylvestris*, *Betula celtiberica*, *Fraxinus spp.*).

The Spanish areas are substantially larger, covering a total of 2,741,281 ha. Extremadura, which has an area of 822,623 ha, is dominated by *Quercus spp.*, being found other species such as: *Quercus ilex*, *Pinus pinaster*, *Quercus suber*, and *Pinus pinea*. Leon province, with an area of 353,980 ha, is mostly composed of *Quercus pyrenaica* Wild. Additionally, large areas of *Pinus sylvestris* L., *Pinus pinaster*, *Pinus nigra* subsp. *Salzmannii*, *Fagus sylvatica* Wild, *Quercus robur*, *petraea*, *rubra* and *humilis*. Finally, Aragon, the largest area (1,564,678 ha), is covered by *Pinus halepensis*, *Pinus sylvestris*, *Quercus ilex*, mixed forest of conifers and broadleaved, mixed forest of conifers and *Pinus nigra*.

These areas together comprise a diverse dataset within the Mediterranean forest of the Iberian Peninsula, providing a reliable baseline for the training and validating the proposed methodology in this region of the world.

3. Datasets

In the scope of this article, field inventory data for training and validation was used. Regarding the biomass prediction there were utilised Synthetic Aperture Radar (SAR), Multispectral, and ancillary variables.



Figure 1: Study areas location within the Iberian Peninsula. The red outlined areas from South to North are Monsanto, Sintra-Cascais, Pombal, Serra Lousã, and Vila Pouca de Aguiar. The orange, green, and blue outlined are Extremadura, Aragon, and Leon.

Table 1: Summary of the Airborne Laser Scanning campaigns over Portugal and Spain.

Mission	áGIL	PNOA
Sensors	Teledyne Galaxy PRIME™ Airborne Lidar Terrain Mapper with SwathTRAK™	RIEGL LMS-Q1560, LEICA ALS80
Point Densities	10.02 to 13.88 points/m ²	1.00 to 2.00 points/m ²
Vertical RMSE	<10 cm	<20 cm
CRS	ETRS89/Portugal TM06	ETRS89/29-30N
Availability	ICNF Geocatálogo	Centro Nacional de Información Geográfica

3.1 Field Inventory and Airborne Laser Scanning

The study areas considered in the scope of this article were surveyed during two different field campaigns: the Portuguese ICNF campaign and the 4th Spanish National Forest Inventory (SNFI-4). In the former a total of 411 plots with a radius of 12.62 m (~500 m²) were measured, while for the latter there were 4,754 plots with a 25 m radius (~1960 m²). In addition, ALS data was also used in the scope of this article to generate the reference data. The Portuguese study areas were covered by the áGIL campaign (ICNF, 2022) and the Spanish by Plan Nacional de Ortofotografía Aérea (PNOA, 2020). A summary of the specifications of each of these campaigns is provided in Table 1. The retrieved measurements were then used for computing the reference biomass values through allometric models as discussed in the Data Preparation section, being the field data used for the AGBD and BGBD, and ALS data for the LBD and SBD.

3.2 Synthetic Aperture Radar

The proposed methodology resorts to two types of SAR data: multi-frequency radar and interferometry. Specifically, the first includes L-band backscatter from ALOS-2 and C-band backscatter Sentinel-1. The second, includes interferometric coherence and phase derived from Sentinel-1.

3.2.1 Radar Backscattering

Beginning with the ALOS-2 observations, the Global L-band mosaic provided by the ALOS Research and Application Project was utilised. This is a freely worldwide product, that contains a yearly observation per tile, with Horizontal/Horizontal (HH) and Horizontal/Vertical (HV) polarisations. This product is already pre-processed corresponding to terrain-flattened Gamma Nought (Gamma0), being delivered in linear scale with a 25 m spatial resolution.

From Sentinel-1, a lower processing level product was used, being it, the Ground Range Detected (GRD) available in the Copernicus Data Space Ecosystem. This corresponds to the returned C-band amplitude measurements at Vertical/Vertical (VV) and Vertical/Horizontal (VH) polarisations. These observations are already multilooked to a square pixel resolution of 10 m. Contrary to ALOS-2, Sentinel-1 provides a new observation every six days, being used two monthly observations between May and October for the study area reference year, acquired in ascending orbit.

3.2.2 Differential Interferometry

The derivation of the Differential-Interferometric SAR (D-InSAR) predictors required phase measurements of the returning SAR signal. Therefore, the Sentinel-1 Single Look Complex (SLC) product was obtained from the Copernicus Data Space Ecosystem. These are lower-level products, which have not yet been multilooked and it is in slant-range geometry, but still contain the phase information. Additionally, the computation of both the interferometric coherence and phase require an observations pair, being used a pair of them per month between May and October for the study area reference year (also acquired in ascending orbit as Sentinel-1 GRD observations). Note that for avoiding the inclusion of instrument noise in the computation of these predictors, only Sentinel-1A observations were considered, which imposes a temporal baseline of 12 days.

3.3 Multispectral

The multispectral dataset was built from Sentinel-2 observations. From this platform, the atmospherically corrected bottom-of-atmosphere reflectance measurements on the level 2A product made available in Copernicus Data Space Ecosystem were utilised. Sentinel-2 provides 13 reflectance bands spanning from the visible to shortwave infrared regions of the spectrum (including three in red edge), at spatial resolutions of 10, 20, and 60 m. In this article all of them were included, with exception of band 10 (Cirrus). Sentinel-2 constellation provides a new observation every 5 days, however some of them are not usable due to the presence of clouds or the respective shadows.

3.4 Ancillary Variables

In addition to the backscatter, interferometric, and reflectance predictors, ancillary variables concerning topography, land cover, and acquisition geometry of the radar data were also used by the proposed methodology.

The topographic predictors were obtained from the Shuttle Radar Topography Mission-Digital Elevation Model (SRTM-DEM) providing data on the Elevation and Slope measurements at pixel level with a 30 m spatial resolution. For the land cover, the Plant Functional Type (PFT) derived from the Moderate Resolution Imaging Radiometer (MODIS) was utilised. This product provides the PFT classification at pixel level with a 500 m spatial resolution. Both these products have worldwide coverage. In addition to these, the local incidence angles for both ALOS-2 and Sentinel-1 observations were also included in the predictors pool of proposed methodology.

4. Methodology

The methodology proposed in this article can be described in three steps: 1) data preparation, 2) features engineering, and 3) biomass prediction. Each step will be presented throughout this section, which is concluded model validation and testing approach employed.

4.1 Data Preparation

An independent data preparation pipeline was defined for each dataset. This was required to ensure that all products were brought to the same processing level, making them consistent for the biomass prediction.

4.1.1 Reference Biomass

The computation of the biomass reference datasets required the application of allometric models to the national inventory plots

measurements and ALS point clouds. The reference AGBD and BGBD were computed resorting to the field inventory data and the allometric equations proposed by (Ruiz-Peinado Gertrudix et al., 2012; Ruiz-Peinado et al., 2011; Tomé et al., 2007). Note that these models were specifically developed for the Iberian Peninsula. In its turn, the LBD and SBD references were obtained through allometric models based on the mean height and canopy cover metrics derived from ALS point clouds (Montero et al., 2020).

The large range of biomass values found for both the AGBD and BGBD led to the need of applying some filtering to the reference data to guarantee the possibility of achieving reliable models for a defined range of biomass. Considering this, a quartile range filter with a dynamic factor a was applied to these datasets was applied. The upper and lower limits can be found in equation (1) and (2), respectively.

$$Upper_{Limit} = Q3 + a \cdot IQR \quad (1)$$

$$Lower_{Limit} = Q1 - a \cdot IQR \quad (2)$$

Where $Q3$ and $Q1$ represent third and first quartile values, and IQR the Interquartile range. Factor a was optimised during the model building with the objective of being less restrictive.

4.1.2 ALOS-2 and Sentinel-1 GRD

For the ALOS-2 pre-processing a simple pipeline is employed since it corresponds to an already terrain-flattened product. It starts with the speckle filtering using the Lee filter (Lee et al., 1994) with a 7×7 kernel. Then, the trigonometric relationships presented in equation (3) and (4), where θ corresponds to the ALOS-2 incidence angle are used to obtain the additional backscattering coefficients (Sigma and Beta Nought). Then, these values were converted to the dB scale.

$$Beta0 = \frac{Gamma0}{\tan \theta} \quad (3)$$

$$Sigma0 = Gamma0 \cdot \cos \theta \quad (4)$$

The Sentinel-1 GRD pipeline was responsible for computing the Γ_0 , Σ_0 , and β_0 . It followed the approach proposed in (Pereira-Pires, Guerra-Hernández, Silva, et al., 2025), being applied the steps below:

- i) Application of the orbit files
- ii) Thermal noise removal
- iii) Terrain flattening (not required for β_0)
- iv) Speckle filtering (use the previous filter)
- v) Range-Doppler terrain correction
- vi) Conversion to a dB scale

At the end of these pipelines the C-/L-band multi-frequency backscatter is generated at a spatial resolution of 25 m.

4.1.3 Sentinel-1 SLC

The computation of the interferometric coherence and unwrapped phase follows the approach proposed in the Sentinel-1 Toolbox - TOPS Interferometry Tutorial (https://step.esa.int/docs/tutorials/S1TBX_TOPSAR_Interferometry_with_Sentinel-1_Tutorial_v2.pdf) (with the addition of a multilooking step, as seen below:

- i) Sub-swath selection
- ii) Application of orbit files
- iii) Coregistration of the observation pair
- iv) Interferogram formation and debursting
- v) Multilooking to a square pixel of 25 m
- vi) Goldstein filtering
- vii) Phase unwrapping

viii) Range-Doppler terrain correction

As for the multi-frequency backscatter dataset, this has a spatial resolution 25 m.

4.1.4 Sentinel-2

The pre-processing of Sentinel-2, due to high processing level at which this product is delivered, essentially consists of the resampling of all bands to a spatial resolution of 25 m using a bicubic interpolation for upsampling and the median for the downsampling.

All the aforementioned datasets are also reprojected for the defined coordinate reference system (ETRS89-29N for Portugal and ETRS89-30N for Spain). Additionally, all the available observations per study area, with exception of ALOS-2, are collocated and have the May-to-October mean temporal composite computed. All the data preparation steps were executed resorting to the Sentinel Application Platform (SNAP) and the Geospatial Data Abstraction Library (GDAL).

4.2 Features Engineering

After the data preparation, an extensive set of predictors is derived from the initial features, including: texture features, normalised polarisation differences and ratios (equations (3) and (4)), spectral indices, and five biophysical variables. The texture features were computed using SNAP Grey-level Co-occurrence Matrix (GLCM) being derived the ten available metrics for the backscatter coefficients, interferometric coherences, and reflectance bands, using the parameters presented in Table 2. Then, the twenty spectral indices and the five Tasseled Cap Transformations (TCT) presented in Table 3 were computed. The predictors set was concluded with the derivation of the five biophysical variables available on SNAP Biophysical processor (Leaf Area Index, Fraction of Absorbed Photosynthetically Active Radiation, Fraction of vegetation cover, Chlorophyll content in the leaf, Canopy Water Content). This makes a predictor set composed by a total of 301 features.

4.3 Model Building and Biomass Prediction

The biomass prediction was ensured by four independently developed extreme gradient boosting regressors (one per biomass component). These regressors were optimised through a five-fold cross-validation grid search over the hyperparameters presented in Table 4. Then, a sequential backwards feature elimination process was performed as follows:

- i) Model fitting with non-eliminated features
- ii) Computation the defined error metrics with a five-fold cross-validation

Table 2: Parameters defined for computing the texture features.

Parameter	Value
Window Size	5x5
Angle	All
Quantiser	Probabilistic Quantiser
Quantization Levels	64
Displacement	1
Output Texture Features	Contrast, Dissimilarity, Homogeneity, Angular Second Moment, Energy, Maximum Probability, Entropy, GLCM Mean, GLCM Variance, GLCM Correlation

Table 3: Spectral indices equations. The following abbreviations were used: ATSAVI - Adjusted Transformed Soil-Adjusted Vegetation Index, NDVI - Normalized Difference Vegetation Index; SLAVI - Specific Leaf Area Vegetation Index; MSAVI - Modified Soil Adjusted Vegetation Index; GSAVI - Green Soil Adjusted Vegetation Index; GARI - Green Atmospherically Resistant Vegetation Index, GCVI - Green Chlorophyll Vegetation Index.

Index	Equation	
Normalised Difference (ND)	$\frac{B_a - B_b}{B_a + B_b}$	Using the following B_a/B_b combinations: B08/B04 (NDVI); B08/B05; B08/B06; B08/B07; B08/B11 (NDMI); B08/B12; B03/B04 (NGRDI).
Normalised Difference (using 3 bands, ND3)	$\frac{B_a - (B_b + B_c)}{B_a + (B_b + B_c)}$	Using the following $B_a/B_b/B_c$ combinations: B08/B03/B02; B08/B03/B04; B08/B04/B02.
Misra Green Vegetation	$-0.386B03 - 0.530B04 + 0.535B06 + 0.532B08$	-
Misra Soil Brightness	$0.406B03 + 0.600B04 + 0.645B06 + 0.243B08$	-
ATSAVI	$1.22 \frac{B08 - 1.22B04 - 0.03}{1.22B08 + B04 - 1.22 \cdot 0.03 + 0.08(1 + 1.22^2)}$	-
Enhanced Vegetation Index	$2.5 \frac{B08 - B04}{(B08 + 6.0 \cdot B04 - 7.5 \cdot B02) + 1}$	-
Transformed NDVI	$\sqrt{\frac{B08 - B04}{B08 + B04}} + 0.5$	-
Pan NDVI	$\frac{B08 - (B02 + B03 + B04)}{B08 + (B02 + B03 + B04)}$	-
SLAVI	$\frac{B08}{B08 + B11}$	-
MSAVI	$\frac{2 \cdot B08 + 1 - \sqrt{(2 \cdot B08 + 1)^2 - 8(B08 - B04)}}{2}$	-
GSAVI	$0.5 \frac{B08 - B03}{B08 + B03 + 0.5}$	-
GARI	$\frac{B08 - (B03 - (B02 - B04))}{B08 - (B03 + (B02 - B04))}$	-
GCVI	$\frac{B08}{B03} - 1$	-
TCT Brightness	$0.30370B02 + 0.27030B03 + 0.47430B04 + 0.55850B08 + 0.50821B11 + 0.18630B12$	-
TCT Greenness	$-0.2848B02 - 0.2435B03 - 0.5436B04 + 0.7243B08 + 0.0840B11 - 0.1800B12$	-
TCT Wetness	$0.1509B02 + 0.1973B03 + 0.3279B04 + 0.3406B08 + 0.7112B11 + 0.4572B12$	-
TCT Distance	$\sqrt{TCT\ Greenness^2 + TCT\ Brightness^2}$	-
TCT Angle	$\arctan\left(\frac{TCT\ Greenness}{TCT\ Brightness}\right)$	-

- iii) Elimination of the least important according to extreme gradient boosting feature gain
- iv) Repeat from i) if there is more than one feature available.

After running all iterations, the error metrics response to the decreasing number of features was analysed to identify the best features for each biomass component.

4.4 Model Validation

The model validation was ensured by a five-fold cross-validation being computed the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Bias, Pearson correlation (R), and Coefficient of Variation (CV), see equations (5) to (8). Concerning the RMSE, MAE, and Bias, the respective normalised versions of the were also computed, by the ratio of their absolute value with the dataset biomass mean value. This approach was employed during the model building, and after defining the best hyperparameters and predictors a last round of cross-validation was performed to compute the final estimated error.

Table 4: Tested hyperparameters during the Grid Search.

Hyperparameter	Tested Values
Number of Estimators	100, 200, 300, 400, 500, 600, 700, 800, 900, 1000
Maximum Depth	2, 3, 4, 5, 6, 7, 8
Learning Rate	0.01, 0.05, 0.1, 0.3

Table 5: Biomass component regressors final hyperparameter and number of features.

Biomass Pool	Hyperparameter			Number of Features
	# Estimators	Maximum Depth	Learning Rate	
AGBD	900	5	0.05	148
BGBD	700	5	0.05	110
LBD	900	5	0.05	36
SBD	800	4	0.05	51

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{N}} \quad (5)$$

$$MAE = \frac{\sum_{i=1}^N |x_i - \hat{x}_i|}{N} \quad (6)$$

$$Bias = \frac{\sum_{i=1}^N (x_i - \hat{x}_i)}{N} \quad (7)$$

$$R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} \quad (8)$$

$$CV = \frac{\sigma}{\mu} \quad (9)$$

5. Results and Discussion

Beginning with the dynamic factor a for the quartile filter, an analysis of different possible values led to the decision of using a factor of 2.6 (filtering values above 310 Mg/ha) and 2.2 (filtering values above 100 Mg/ha) for the AGBD and BGBD, respectively. Despite the significance of applying this filtering stage to ensure that a meaningful range in which the regression produces reliable predictions, the IQR approach was only capable to successfully applying an upper limit (since the lower limit was always a negative value). Therefore, a non-symmetric filtering should be investigated, to avoid establishing an operational range for the regressors where there is still a strong overestimation of biomass at the lower values.

This was followed by the hyperparameters' adjustment and the feature selection (see Table 5 for the final values). An analysis to these results shows that regressors with a high number of estimators but with a relatively small depth were preferred. This was expected, since large maximum depths usually lead to overfitting, being more beneficial the use of a wider set of apparently weaker regressors. Another interesting detail is that AGBD and BGBD required a substantial larger number of features than LBD and SBD. This reflects the higher complexity the former biomass components. This is confirmed by the results presented in Table 6, which shows that AGBD and fBGBD achieved %RMSEs of 63.29% and 62.90%, respectively, and for the LBD and the SBD this result was 41.07% and 28.34%, respectively. Despite the IQR filtering already applied to AGBD and BGBD, the regressors of the biomass components are significantly worse than the ones for the LBD and SBD.

It is also worth noting that despite Sentinel-2 derived predictors were the most significant for the regressors, predictors from all feature sets were selected for all biomass components regressors. Namely, S1 SLC (which is not commonly considered in biomass regression with machine learning) had more features in the final sets than S1 GRD and ALOS2, for both AGBD and BGBD. This shows the applicability of the proposed models to produce biomass maps and the importance of including interferometric features in the prediction models.

The ability of the proposed models to provide reliable biomass predictions across a wide range of values is evidenced in the scatter plots presented in Figure 2. This shows the possibility of using the developed models for large scale mapping in the mediterranean regions of Portugal and Spain. Notwithstanding their potential, these models still face difficulties when mapping extremely low and high biomass values. This is particularly evident for AGBD and BGBD, where values above 150 and 50 Mg/ha are strongly underestimated; and values close to 0 Mg/ha strongly overestimated. This becomes more evident when analysing the percentile plots presented in Figure 3. With exception, of the LBD that seems to accurately predict values on the lower biomass range, all of them exhibit the previously mentioned tendency of over and underestimating the lower and higher biomass values.

6. Conclusion and Future Work

The results achieved in this article demonstrate the possibility of predicting additional biomass components from space, namely, the BGBD. This opens the door for the product of more accurate and informative biomass maps, describing the AGBD not just as the aerial components found in the trees, but also litter components found in the forest floor and shrubs. It was also demonstrated the benefits of the integration of SAR and multispectral datasets, especially the advantages of including Sentinel-1 SLC predictors, that are usually disregarded when using machine learning for biomass prediction.

In the continuation of this work, several steps should be considered. With the increasing availability of embeddings such as Tessera and the AlphaEarth, the use of these products should be evaluated for biomass mapping. In addition to this, the launch of the NASA-ISRO Synthetic Aperture Radar, will offer new opportunities when it comes to the SAR L-band, such as computation of interferometric predictors as it was done for Sentinel-1. Also, more extensive research should be done concerning the temporal composite extension. Evaluate if longer or shorter periods lead to even better biomass predictions. And finally, the large-scale application of the proposed methodology should be tested, for began producing maps for Portugal and Spain.

Acknowledgements

This research was supported by the Fundação para a Ciência e Tecnologia under Grant UID/00066/2025 (awarded to Centre of Technology and Systems – Uninova) and the Associated Lab of Intelligent Systems (LASI), UID/00239/2025 (awarded to Forest Research Centre), LA/P/0092/2020 (awarded to Associate Laboratory TERRA), and 2023.15225.TENURE.008 (awarded to Dr. Guerra- Hernández for supporting is 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).

Table 6: Result metrics per biomass component achieved by the final regressors.

Biomass Pool	Metrics							
	RMSE (Mg/ha)	%RMSE	MAE (Mg/ha)	%MAE	Bias (Mg/ha)	%Bias	R	CV
AGBD	44.72	63.29	32.37	45.82	-0.62	-0.90	0.68	0.62
BGBD	16.78	62.90	12.41	46.52	-0.34	-1.28	0.62	0.53
LBD	7.97	41.07	5.41	27.92	-0.13	-0.68	0.83	0.60
SBD	6.52	22.12	5.09	22.12	0.00	-0.01	0.71	0.29

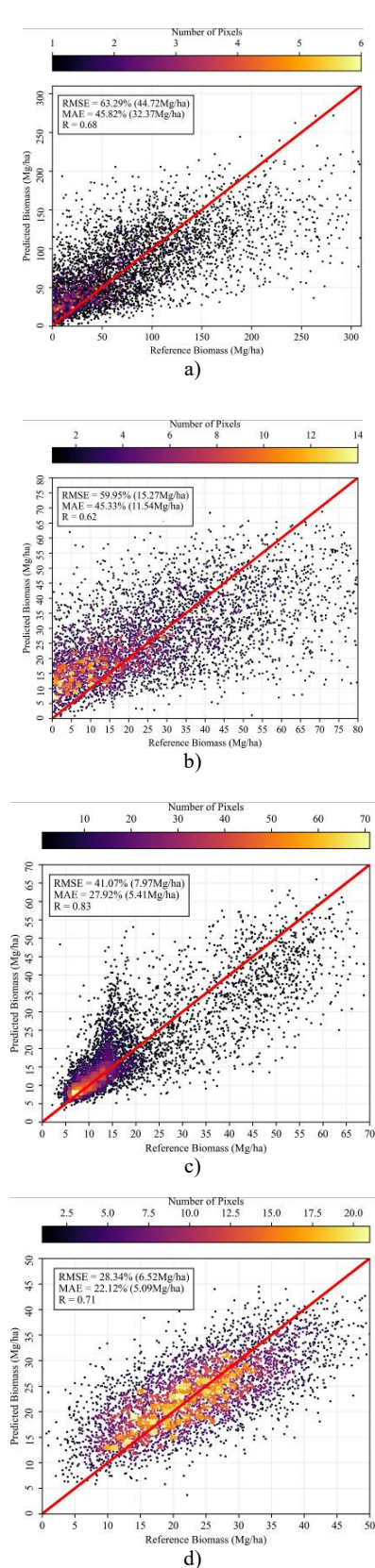


Figure 2: Biomass regressors scatter plots for the: a) AGBD, BGBD, LBD, and SBD.

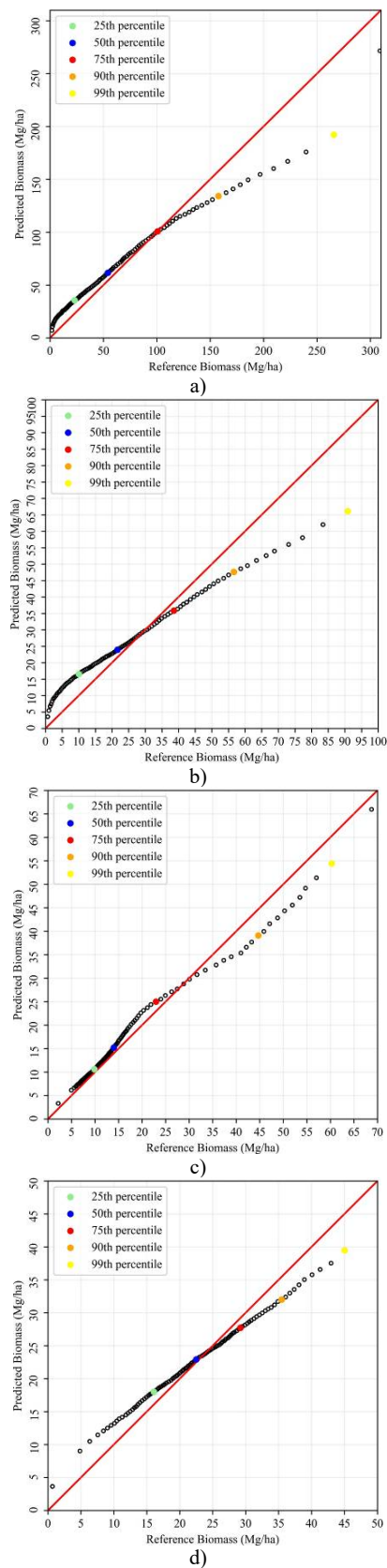


Figure 3: Biomass regressors percentile plots for the: a) AGBD, BGBD, LBD, and SBD.

References

- Chen, A., Xu, C., Zhang, M., Guo, J., Xing, X., Yang, D., Xu, B., & Yang, X. (2024). Cross-scale mapping of above-ground biomass and shrub dominance by integrating UAV and satellite data in temperate grassland. *Remote Sensing of Environment*, 304. <https://doi.org/10.1016/j.rse.2024.114024>
- Chen, Y., Feng, X., Fu, B., Ma, H., Zohner, C. M., Crowther, T. W., Huang, Y., Wu, X., & Wei, F. (2023). Maps with 1 km resolution reveal increases in above- and belowground forest biomass carbon pools in China over the past 20 years. *Earth System Science Data*, 15(2), 897–910. <https://doi.org/10.5194/essd-15-897-2023>
- Dlamini, C. M., Odindi, J., Matongera, T. N., & Mutanga, O. (2025). Exploring the utility of remote sensing technology in vegetation below ground biomass (BGB) estimation: a critical review of methods and challenges. In *Frontiers in Remote Sensing* (Vol. 6). Frontiers Media SA. <https://doi.org/10.3389/frsen.2025.1668676>
- Duncanson, L., Kellner, J. R., Armston, J., Dubayah, R., Minor, D. M., Hancock, S., Healey, S. P., Patterson, P. L., Saarela, S., Marselis, S., Silva, C. E., Bruening, J., Goetz, S. J., Tang, H., Hofton, M., Blair, B., Luthcke, S., Fatoyinbo, L., Abernethy, K., ... Zraggen, C. (2022). Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission. *Remote Sensing of Environment*, 270. <https://doi.org/10.1016/j.rse.2021.112845>
- Enes, T., Lousada, J., Fonseca, T., Viana, H., Calvão, A., & Aranha, J. (2020). Large scale shrub biomass estimates for multiple purposes. *Life*, 10(4). <https://doi.org/10.3390/life10040033>
- Greaves, H. E., Vierling, L. A., Eitel, J. U. H., Boelman, N. T., Magney, T. S., Prager, C. M., & Griffin, K. L. (2016). High-resolution mapping of aboveground shrub biomass in Arctic tundra using airborne lidar and imagery. *Remote Sensing of Environment*, 184, 361–373. <https://doi.org/10.1016/j.rse.2016.07.026>
- Guerra-Hernández, J., Narine, L. L., Pascual, A., Gonzalez-Ferreiro, E., Botequim, B., Malambo, L., Neuenschwander, A., Popescu, S. C., & Godinho, S. (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>
- Hardiman, B., Gough, C., Butnor, J., Bohrer, G., Detto, M., & Curtis, P. (2017). Coupling Fine-Scale Root and Canopy Structure Using Ground-Based Remote Sensing. *Remote Sensing*, 9(2), 182. <https://doi.org/10.3390/rs9020182>
- ICNF. (2022). Projecto áGIL - Mafra, Monsanto, Pombal, Serra da Lousã, Sintra-Cascais, Vila Pouca de Aguiar. <https://geocatalogo.icnf.pt/geovisualizador/agil.html>
- Lee, J. S., Jurkevich, L., Dewaele, P., Wambacq, P., & Oosterlinck, A. (1994). Speckle filtering of synthetic aperture radar images: A review. *Remote Sensing Reviews*, 8(4), 313–340. <https://doi.org/10.1080/02757259409532206>
- Liu, W., Wang, J., Hu, Y., Ma, T., Otgonbayar, M., Li, C., Li, Y., & Yang, J. (2024). Mapping Shrub Biomass at 10 m Resolution by Integrating Field Measurements, Unmanned Aerial Vehicles, and Multi-Source Satellite Observations. *Remote Sensing*, 16(16). <https://doi.org/10.3390/rs16163095>
- Montero, G., López-Leiva, C., Ruiz-Peinado, R., López-Senespleda, E., Onrubia, R., & Pasalodos, M. (2020). Producción de Biomasa y Fijación de Carbono Por Los Matorrales Españoles y Por El Horizonte Orgánico Superficial de Los Suelos Forestales. In *Ministerio de Agricultura, Pesca y Alimentación: Madrid, Spain*.
- Pascual, A., Godinho, S., & Guerra-Hernández, J. (2023). Integrated LiDAR-supported valuation of biomass and litter in forest ecosystems. A showcase in Spain. *Science of The Total Environment*, 897, 165364. <https://doi.org/10.1016/j.scitotenv.2023.165364>
- Pascual, A., & Guerra-Hernández, J. (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). <https://doi.org/10.1016/j.jenvman.2023.118709>
- Pascual, A., & Guerra-Hernández, J. (2026). Integration of Google's Alpha Earth Foundations into biomass estimation combined with GEDI spaceborne lidar and field inventory data. *Forest Ecology and Management*, 606. <https://doi.org/10.1016/j.foreco.2026.123550>
- Pereira-Pires, J. E., Guerra-Hernández, J., Pascual, A., Silva, J. M. N., Fonseca, J. M., & Mora, A. (2025). Belowground, Litter and Shrub Biomass Mapping using Sentinel-1/2 and ALOS-2 Products. 2025 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS), 1–4. <https://doi.org/10.1109/MIGARS67156.2025.11232160>
- PNOA. (2020). LIDAR 2a Cobertura (2015-Actualidad). <http://centrodedescargas.cnig.es/>
- Pereira-Pires, J. E., Guerra-Hernández, J., Silva, J. M. N., Fonseca, J. M., Guida, R., & Mora, A. (2025). Forest Height Mapping with Multifrequency SAR in Mediterranean Forests. *IEEE Transactions on Geoscience and Remote Sensing*, 1–1. <https://doi.org/10.1109/TGRS.2025.3538216>
- Ren, H., Zhang, B., & Guo, X. (2018). Estimation of litter mass in nongrowing seasons in arid grasslands using MODIS satellite data. *European Journal of Remote Sensing*, 51(1), 222–230. <https://doi.org/10.1080/22797254.2017.1418186>
- Ruiz-Peinado Gertrudix, R., Montero, G., & Del Rio, M. (2012). Biomass models to estimate carbon stocks for hardwood tree species. *Forest Systems*, 21(1), 42–52. <https://doi.org/10.5424/fs/2112211-02193>
- Ruiz-Peinado, R., Del Rio, M., & Montero, G. (2011). New models for estimating the carbon sink capacity of Spanish softwood species. *Forest Systems*, 20(1), 176–188. <https://doi.org/10.5424/fs/2011201-11643>
- Silva, D. M., Pereira-Pires, J. E., Guerra-Hernández, J., Pascual, A., Silva, J. M. N., Mora, A., & Fonseca, J. M. (2025). Biomass Estimation at Footprint Level Using GEDI Data and Machine Learning. 2025 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS), 1–4. <https://doi.org/10.1109/MIGARS67156.2025.11232292>
- Tomé, M., Barreiro, S., Paulo, J. A., & Faias, S. P. (2007). Seleção de equações para estimação de variáveis da árvore em inventários florestais a realizar em Portugal.