https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

Multi-Source Remote Sensing for large-scale biomass estimation in mediterranean olive orchards using GEDI LiDAR and Machine Learning

Francisco Contreras[1], María L. Cayuela[1], Miguel A. Sánchez-Monedero[1], Pedro Pérez-Cutillas[2]

[1] Department of Soil and organic waste management for greenhouse gas mitigation in agriculture, CEBAS-CSIC, Campus Universitario de Espinardo, 30100, Murcia, Spain
[2] Department of Geography, University of Murcia, C. Santo Cristo, 1, 30001, Murcia, Spain

Correspondence to: Francisco Contreras (fcontreras@cebas.csic.es)

Abstract

Accurate estimation of Above-Ground Biomass Density (AGBD) is essential for assessing carbon stocks and promoting sustainable agricultural practices. This study integrates multi-source remote sensing data, including GEDI LiDAR, optical, SAR, and topographic variables, to predict AGBD in Mediterranean olive orchards using a Random Forest regression model implemented on Google Earth Engine (GEE). The volumetric approach, based on GEDI L2A canopy height and dendrometric parameters, provided more accurate predictions than the GEDI L4A product, which is limited by its global stratification methodology. The model's predictive performance varied depending on data combinations, with the fully multi-source configuration achieving the highest accuracy (R^2 = 0.62, RMSE = 5.95 Mg·ha^{-1}). NDWI, slope, and NDVI were identified as the most influential predictors. The spatial analysis revealed that Spain exhibited the highest total AGBD among the studied countries, followed by Italy and Greece, reflecting their dominance in olive production. The model effectively captured biomass variability across different regions, demonstrating its suitability for large-scale applications. This study highlights the potential of integrating LiDAR, optical, and SAR data for biomass estimation, offering a scalable and cost-effective approach for monitoring carbon stocks and optimizing agricultural resource management. By providing accurate AGBD predictions, this methodology supports climate-smart agriculture and facilitates data-driven decision-making for both farmers and policymakers, contributing to the advancement of sustainable agricultural systems in Mediterranean olive orchards.

Keywords: Biomass Modeling, Carbon Sequestration, Vegetation Structure, Crop Monitoring, Ecosystem Services, Precision Agriculture, Remote Sensing, Geospatial Data Integration

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

1 Introduction

Olive tree groves (*Olea europaea* L.) are among the most significant crops globally, serving as a mainstay of agricultural and economic systems, particularly in Mediterranean-climate regions where conditions are highly favorable for their cultivation (Proietti et al., 2014). Beyond their economic relevance, olive groves play a crucial role in global climate dynamics due to their ecological and agronomic characteristics. As perennial plants with extensive root systems, olive trees contribute to soil carbon sequestration, effectively storing carbon in both their biomass and the soil over long periods. This natural process helps mitigate climate change by reducing greenhouse gas emissions (Pardo et al., 2017). Therefore, integrating innovative management strategies in olive cultivation presents a promising approach to enhancing climate change resilience and promoting sustainable agricultural practices.

A fundamental parameter for assessing biomass in these ecosystems is Above-Ground Biomass Density (AGBD). The Land Product Validation Subgroup of the Committee on Earth Observation Satellites (CEOS) defines AGBD as the standing dry mass of live or dead woody matter derived from trees or shrubs, typically expressed in megagrams per hectare (Mg/ha). This definition is particularly relevant for Earth Observation (EO) applications, as it encompasses both living and dead biomass. While some scientific disciplines differentiate between live wood, leaf mass, or organic debris, EO methodologies primarily focus on standing biomass. For instance, LiDAR sensors can detect both live and dead biomass, whereas radar systems mainly measure live woody biomass due to their sensitivity to moisture content and scattering object size. However, converting EO-derived measurements into biomass estimates often involves allometric models, which may inconsistently account for the proportions of live and dead wood in calibration datasets (Duncanson et al., 2021).

Despite significant advances in remote sensing, accurately estimating biomass in olive groves presents unique challenges. Unlike forested ecosystems, where tree structure and biomass distribution follow well-established patterns, olive orchards exhibit high structural variability due to differences in planting density, pruning practices, and irrigation regimes (Brunori et al., 2017). Additionally, olive trees have a distinct architecture, with a substantial portion of their biomass concentrated in the crown, which complicates the application of generic biomass estimation models (Velàzquez-Martì et al., 2014). The complex relationship between canopy cover and total biomass further increases uncertainty in satellite-derived biomass estimates (Rodríguez-Lizana et al., 2023). These factors highlight the necessity of developing specific methodologies tailored to olive groves.

To improve biomass estimation accuracy, alternative approaches have been proposed. Velázquez-Martí et al., (2011) developed a methodology that incorporates multiple allometric characteristics, including olive tree variety, tree height, and irrigation conditions, allowing for highly accurate predictions in specific crop types. Conventionally, forest inventories have been the primary source of biomass data for national assessments (Hunka et al., 2025; Nesha et al., 2022). These methods are widely recognized for their accuracy and are extensively applied in forestry sector inventories. However, they require substantial financial resources and manual labor, making them less feasible for large-scale agricultural applications.

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

An alternative method estimates total biomass using geometric parameters to derive the volume of woody material in different tree structures (Kebede et al., 2018; Velàzquez-Martì et al., 2014). This approach suggests that olive trees typically allocate approximately 60% of their biomass to the crown and 40% to the stem or trunk (Brunori et al., 2017; Velàzquez-Martì et al., 2014). Allometric modeling techniques, which derive tree volume for biomass estimation, have shown promising results in capturing structural variations in olive groves. While allometric models provide high accuracy in localized studies, their large-scale applicability is often constrained by the need for extensive field data collection (Brede et al., 2022). Conversely, volumetric models offer a scalable alternative but may introduce greater uncertainty when applied to heterogeneous landscapes. Although remote sensing has significantly improved the ability to quantify biomass at large scales, estimating biomass in olive orchards remains challenging. Airborne LiDAR data can provide highly accurate three-dimensional structural information but is costly and limited in temporal coverage (Dubayah et al., 2020). Meanwhile, optical and radar satellite sensors offer frequent observations over large areas but are constrained by spatial resolution and sensitivity to canopy characteristics. High-resolution imagery captured by Remotely Piloted Aircraft Systems (RPAS) enables detailed local measurements, often outperforming conventional satellite-based approaches in precision (Perna et al., 2024). However, RPAS-based methods lack scalability and are labor-intensive, restricting their application to regional studies (Estornell et al., 2015). The integration of multiple data sources, including LiDAR, optical, radar, and topographic datasets, has emerged as a potential solution to mitigate these limitations (Shendryk, 2022).

Facing such a new challenge, Google Earth Engine (GEE) has emerged as a powerful cloud-based platform for geospatial analysis, enabling efficient processing of large-scale Earth observation datasets while supporting multitemporal and multiscale investigations (Gorelick et al., 2017). Its extensive repository of remote sensing imagery, coupled with high-performance computing capabilities, facilitates the harmonization of disparate datasets. This capability is particularly relevant for biomass estimation, where the temporal variability of vegetation dynamics and the spatial heterogeneity of landscapes necessitate adaptive and scalable analytical frameworks. By leveraging GEE's capacity to process vast amounts of remote sensing data in near real-time, it is possible to track biomass fluctuations across different temporal resolutions and apply models across diverse ecological and climatic regions (Pérez-Cutillas et al., 2023).

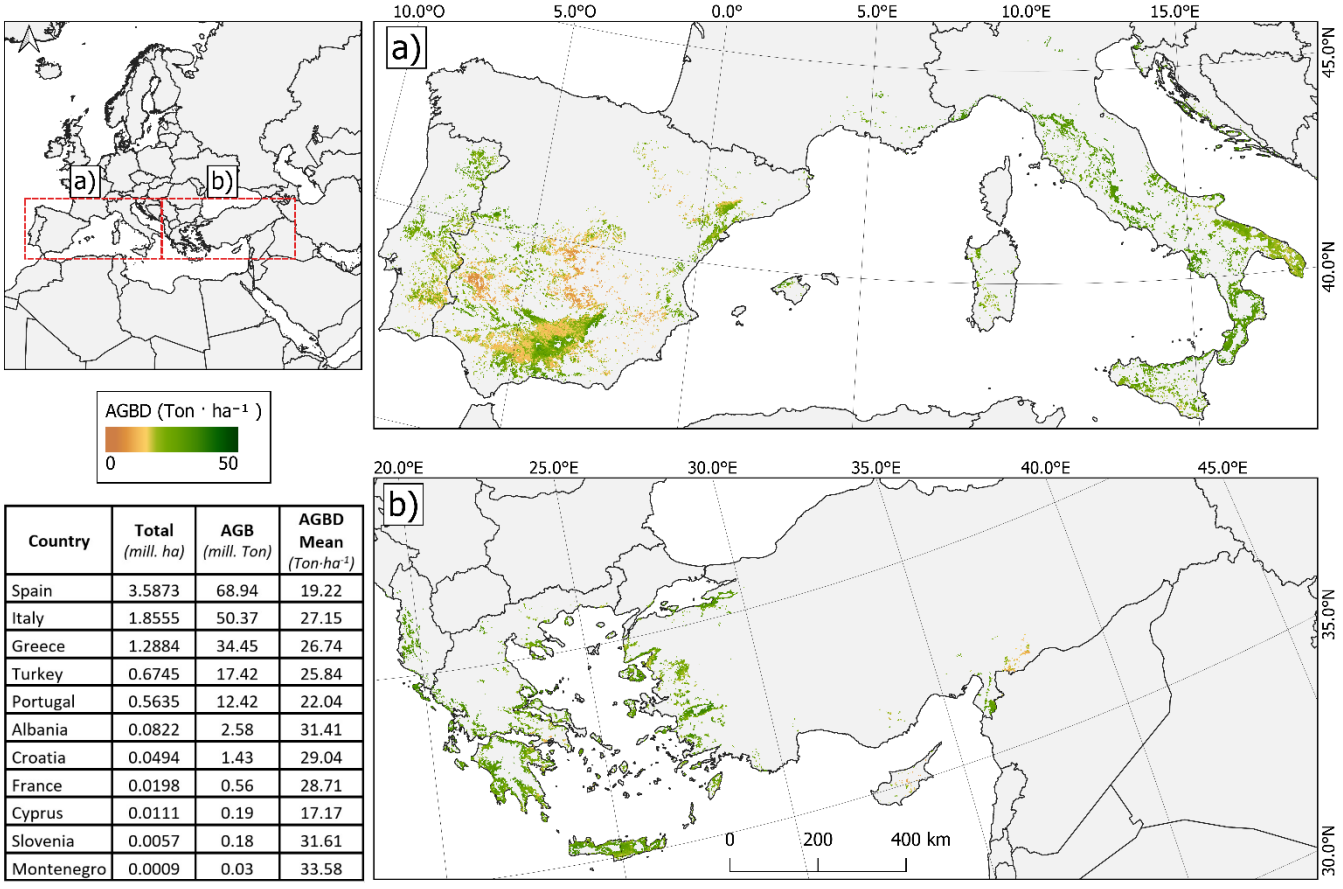
Given these challenges, this study aims to develop a methodology for estimating AGBD in olive groves through the integration of multi-source remote sensing data, including the Global Ecosystem Dynamics Investigation data (GEDI LiDAR), optical and infrared satellite imagery, synthetic aperture radar (SAR), and topographic variables. By evaluating the relative contribution of each data type, the study seeks to determine the most effective combination of remote sensing inputs for biomass estimation. Additionally, the proposed methodology will be applied at a European scale, leveraging datasets such as Corine Land Cover (CLC) as other datasets at greater precision to ensure spatial consistency. Ultimately, this research aims to provide a tool for monitoring carbon sequestration and biomass resources in olive orchards, supporting sustainable agricultural management and climate change mitigation strategies.

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

2 Material and methods

2.1. Study area

The study focuses on developing a methodology for biomass estimation in olive groves at a European scale. To identify olive cultivation areas, the CLC database is used, which includes a specific "Olive Groves" category **(Figure 1)**. However, CLC has limitations in accurately delineating parcel boundaries due to its generalized spatial resolution, making it challenging to create a reliable dataset for model training. To address this limitation, the calibration analysis is conducted in Spain, utilizing the 'Sistema de Información Geográfica de Parcelas Agrícolas' (MAPA, 2024). SIGPAC is a governmental agronomic management tool that provides highly detailed cadastral information on agricultural parcel boundaries. With 2,695,055 hectares of olive plantations registered in the database, SIGPAC enables a more precise selection of olive groves for model calibration.



Country	Total *(mill. ha)*	AGB *(mill. Ton)*	AGBD Mean *(Ton·ha⁻¹)*
Spain	3.5873	68.94	19.22
Italy	1.8555	50.37	27.15
Greece	1.2884	34.45	26.74
Turkey	0.6745	17.42	25.84
Portugal	0.5635	12.42	22.04
Albania	0.0822	2.58	31.41
Croatia	0.0494	1.43	29.04
France	0.0198	0.56	28.71
Cyprus	0.0111	0.19	17.17
Slovenia	0.0057	0.18	31.61
Montenegro	0.0009	0.03	33.58

Figure 1. Study Area of Above-Ground Biomass Density (AGBD) mapping in Europe. The map shows the distribution of AGBD in olive orchards across Mediterranean countries applying the "Fully-Multisource" model. Panel 'a' shows the AGBD distribution in the Iberian Peninsula and Italy, while Panel 'b' displays the AGBD distribution in Greece, Turkey, and surrounding regions. The color gradient represents AGBD values, ranging from 0 (brown) to 50 Ton/ha (dark green). The inset table summarizes the total cultivated area (mill. ha), total AGB (mill. Ton), and mean AGBD (Ton/ha) for each country.

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

Furthermore, Spain is a key reference for olive cultivation, representing approximately 41% of Europe's total olive grove area. This extensive coverage provides a robust dataset for model training and validation. Within Spain, the Mediterranean agro-climatic region is particularly relevant for olive production, characterized by warm and dry summers and mild winters (Pardo et al., 2017). In these environments, annual precipitation varies between 200 and 600 mm, with a markedly seasonal distribution. Such conditions favor drought-resistant crops such as olives, which thrive in shallow, calcareous soils under high temperatures and intense solar radiation (Deitch et al., 2017; Urdiales-Flores et al., 2024).

The extensive cultivation of olive trees in Spain has significant economic and environmental implications. Beyond being a major contributor to the regional economy, olive groves offer opportunities for biomass valorization and carbon sequestration, supporting both agricultural sustainability and bioenergy development (Rosúa, 2012). A key example is Jaén province, the world's leading olive oil-producing region, with 550,000 hectares of olive groves. This accounts for over 25% of Spain's total olive-growing area and 42% of Andalusia's cultivated land (Fernández-Lobato et al., 2024), making it an essential site for biomass assessment and model validation.

2.2. Methodological framework approach

This study utilizes GEDI data to estimate Above-Ground Biomass Density (AGBD) through two modeling approaches for comparative analysis. The first approach derives AGBD estimates from the GEDI L2A product, which provides canopy height metrics transformed into biomass values using a volumetric model tailored to olive tree morphology. The second approach uses the GEDI L4A product, which offers stratified AGBD predictions based on plant functional types. Both datasets are integrated with remote sensing variables and analyzed to assess the relative importance of each data type in biomass estimation. GEE was employed to acquire, pre-process, filter, and transform the satellite images. All spectral bands and derived variables, including vegetation indices, slope, aspect, and SAR textures, were used in the training process. **Figure 2** illustrates the workflow for data harmonization, model training, and mapping.

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

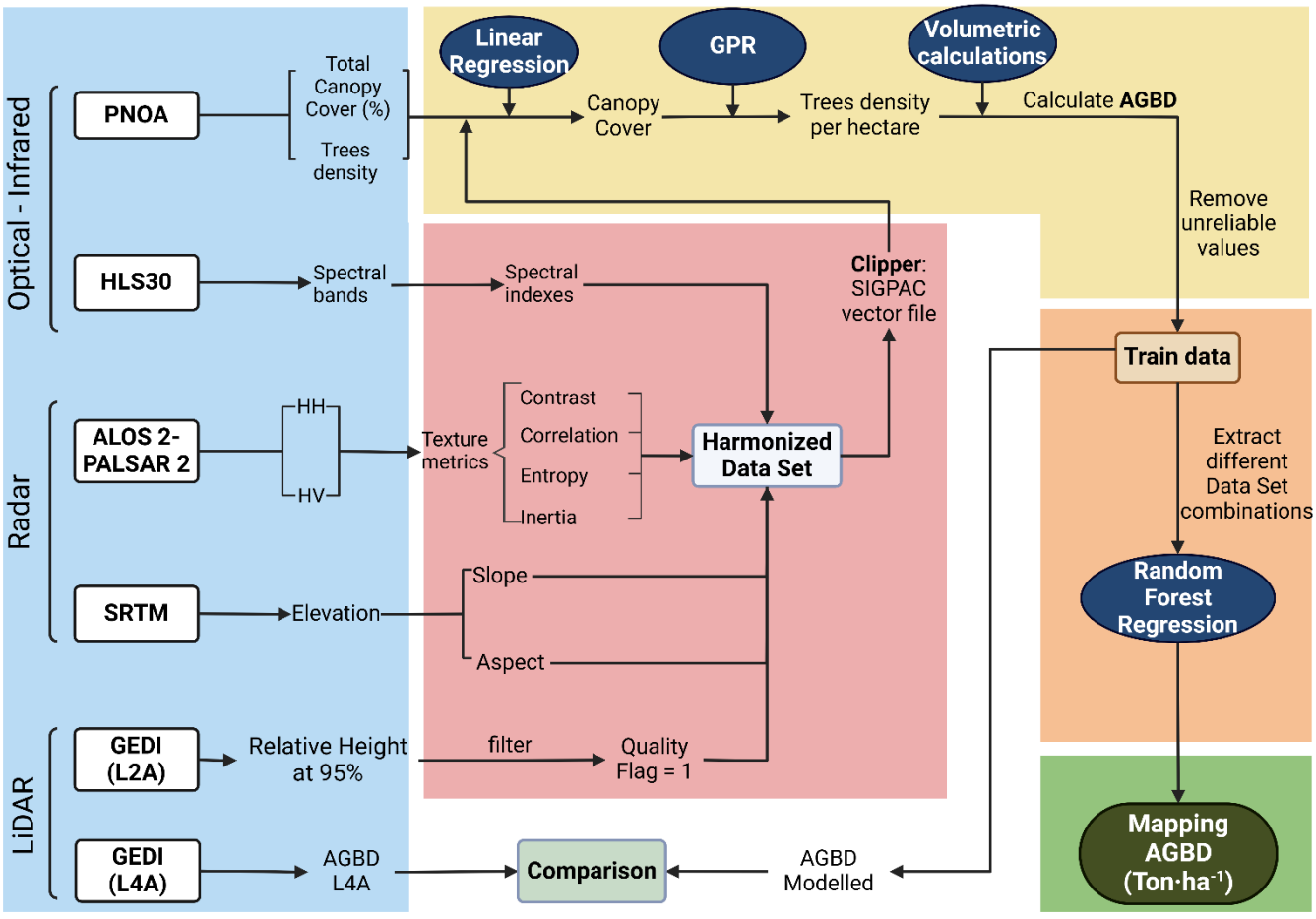


Figure 2. Methodological Flowchart for Above Ground Biomass Density (AGBD) mapping in Olive Orchards. The workflow comprises five sequential stages, each highlighted by a distinct color frame: data acquisition (blue frame), preprocessing (red frame), training dataset generation (yellow frame), model training (orange frame), and biomass mapping (green frame). Remote sensing data sources include GEDI (Global Ecosystem Dynamics Investigation) LiDAR onboard the ISS (International Space Station), optical imagery (Sentinel-2, Landsat 8/9), SAR (Synthetic Aperture Radar) data (ALOS-2 PALSAR-2), and topographic information (SRTM – Shuttle Radar Topography Mission). To improve plot delineation, SIGPAC (Sistema de Información Geográfica de Parcelas Agrícolas), a Spanish Geographic Information System for Agricultural Plots, is used instead of the CLC (Corine Land Cover) database, which lacks cadastral precision. GEDI-derived biomass values (L2A canopy height metrics and L4A AGBD estimates) are correlated with spectral indices (NDVI – Normalized Difference Vegetation Index, NDWI – Normalized Difference Water Index) and radar-texture variables. A Random Forest (RF) model is trained and applied at a large scale to generate a biomass density map, supporting biomass monitoring, carbon sequestration assessment, and sustainable agricultural management.

The methodology begins with the acquisition of multi-source remote sensing data, including GEDI LiDAR, optical imagery (Sentinel-2, Landsat 8/9), synthetic aperture radar (SAR) from ALOS-2 PALSAR-2, and topographic data from SRTM. These datasets are then preprocessed to ensure consistency, including atmospheric correction, geometric normalization, and the derivation of key indices (NDVI, NDWI) and radar texture metrics. To improve spatial accuracy in olive grove delineation,

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

SIGPAC cadastral data is used to refine the initial classification from the Corine Land Cover (CLC) database. Training Dataset Construction and Biomass Estimation has been carried out using the filtered GEDI footprints are linked to spectral, radar, and topographic variables, creating a comprehensive training dataset. L2A canopy height metrics and L4A biomass estimates serve as reference values for AGBD estimation. The Random Forest (RF) regression model is employed to analyze the contribution of each predictor variable, optimizing biomass prediction accuracy. Lastly, Model Application and Biomass Mapping, once trained, is applied on a large spatial scale to generate a high-resolution biomass density map for olive groves. This final output supports biomass monitoring, carbon sequestration analysis, and sustainable agricultural management across Mediterranean regions.

2.3 Data Sources and Preprocessing

2.3.1 GEDI LiDAR Data

GEDI is a spaceborne LiDAR instrument onboard the International Space Station (ISS) used to capture detailed vertical vegetation structure data. It operates as an active remote sensing system emitting laser pulses and analyzing the returned waveform, which encodes the vertical distribution of vegetation and ground surfaces within each footprint (Dubayah et al., 2020). This waveform enables the extraction of key structural metrics, such as canopy height, relative height percentiles (RH metrics), and ground elevation, even under dense vegetation cover(Asner et al., 2012).

GEDI provides multiple data products, classified by processing levels. The L2A product offers relative height metrics (RH metrics), which describe the vertical structure of vegetation, while the L4A product estimates Above-Ground Biomass Density (AGBD) using allometric models. These models are stratified based on Plant Functional Types (PFTs) and calibrated with field measurements (Kellner et al., 2023). For this study, both L2A and L4A products were used. L2A data were filtered based on a quality control flag, with only footprints labeled as reliable being retained. The L4A product, which provides biomass estimates at 25 m spatial resolution, was used to compare and validate biomass predictions. AGBD values from L4A were derived using parametric models that establish linear relationships between L2A canopy height metrics and field-measured biomass (Indirabai et al., 2024).

2.3.2 Optical and Infrared Data

The Harmonized Landsat-Sentinel (HLS) dataset integrates Landsat 8-9 and Sentinel-2A/B imagery to provide consistent, high-resolution optical and infrared data at a 30 m spatial resolution (Masek et al., 2021). By merging these sensors, HLS enables global observations with a revisit frequency of 2–3 days, offering the highest temporal resolution among freely available high-resolution imagery. Sentinel-2's MultiSpectral Instrument (MSI) operates in the visible (VIS), near-infrared (NIR), and shortwave infrared (SWIR) ranges, whereas Landsat's Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) provide coverage in the visible, infrared, and thermal regions (Masek et al., 2018). To ensure consistency, the HLS product undergoes rigorous preprocessing, including atmospheric correction, cloud-shadow masking, illumination and view angle normalization, spatial co-registration, and spectral band harmonization.

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

Spectral indices derived from HLS bands were computed within the Google Earth Engine (GEE) platform. The indices, detailed in **Table 1**, were extracted from the Index Database (IDB) (Henrich, 2012) and included vegetation metrics such as the Soil-Adjusted Vegetation Index (SAVI) and the Bare Soil Index (BSI), which were optimized for semi-arid Mediterranean environments. Moreover, high-resolution aerial orthoimages from the National Aerial Orthophotography Plan (PNOA) were used as ground truth data, supporting the development of a linear model for precise tree cover estimation within Sentinel-2 grid cells.

2.3.3 Synthetic Aperture Radar (SAR) and Topographic Data

For this study, PALSAR-2 data were processed to generate an annual global mosaic composed of multiple SAR strips. ALOS-2 PALSAR-2 is a Synthetic Aperture Radar (SAR) system that operates as an active remote sensing method. It emits microwave pulses and measures their backscatter to detect surface features. SAR wavelengths are long enough to penetrate cloud cover, ensuring reliable data acquisition regardless of atmospheric conditions (Shimada et al., 2014). The dataset underwent radiometric and geometric corrections, including elevation-based terrain correction and slope normalization to reduce distortions related to radar viewing angles. An angular correction algorithm was applied to standardize the incidence angle, improving consistency across different acquisition times. A median filter was also implemented to minimize speckle noise, enhancing image clarity.

Regarding topographical data, this study employed of the Shuttle Radar Topography Mission (SRTM) Global 1 Arc-Second v003 Digital Elevation Model (DEM), which provides global elevation data at a 30 m spatial resolution (NASA JPL, 2013). SRTM-derived terrain variables, including slope and aspect, were integrated to improve biomass estimation models, accounting for the influence of topography on vegetation growth patterns.

2.4 AGBD Estimation Methodology

2.4.1 Theoretical Framework for Biomass Modeling

The estimation of Above-Ground Biomass Density (AGBD) was based on a volumetric modeling approach that combined GEDI L2A LiDAR metrics with canopy cover predictions derived from optical imagery. The Relative Height at 95% (RH95) from GEDI L2A was used as a proxy for canopy top height, representing tree height within each footprint. Canopy cover was calculated to indicate the proportion of the GEDI footprint occupied by the tree canopy. To estimate crown volume, the model proposed by Velàzquez-Martì et al., (2014) for olive trees was applied, using a hemispherical volumetric model. This model converts crown volume to biomass using a wood density (WD) of 0.76 $g \cdot cm^{-3}$, which is consistent with values reported for olive trees (Brunori et al., 2017; Kebede et al., 2018; Velàzquez-Martì et al., 2014).

Crown volume (Cvol) was estimated using the crown radius (Cdiam/2) and crown height, both derived from GEDI and canopy cover metrics. The crown height was calculated from RH95, assuming that approximately 65% of the total tree height corresponds to the crown, reflecting typical olive tree morphology (Velàzquez-Martì et al., 2014). The crown diameter (Cdiam)

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

was obtained from the total canopy cover (Cover), footprint area (Farea), and the number of trees per footprint (Mfoot), following Equation (1):

$$Cdiam = 2\sqrt{\frac{(\text{cover} \cdot \text{Farea})/\text{Mfoot}}{\pi}} \quad (1)$$

The hemispherical crown volume (Hvol) was calculated using the formula for a hemisphere (Equation 2):

$$Hvol = \frac{2}{3}\pi r^2 h \quad (2)$$

Where, $r = \frac{C_{diam}}{2}$ is the crown radius, and h is the crown height (65% of RH95).

To obtain the actual crown volume (Cvol), the hemispherical volume was adjusted by an Occupational Factor (OF) specific to olive tree crown morphology (Velàzquez-Martì et al., 2014) as shown in Equation 3:

$$Cvol = \text{Hvol} \cdot OF \quad (3)$$

The stem volume was approximated based on its relationship with crown volume, following the proportional model described by Velàzquez-Martì et al., (2014). Crown and stem volumes were converted into biomass using the wood density (WD) value of 0.76 g·cm^{-3}. The total above-ground biomass density (AGBD) was then computed by multiplying the total tree volume (Tvol, sum of crown and stem volumes) by wood density and the number of trees per hectare (Mi), as shown in Equation 4:

$$AGBD = \text{Tvol} \cdot \text{WD} \cdot \text{M}i \quad (4)$$

The total number of trees (Mfoot) within each GEDI footprint was estimated using Gaussian Process Regression (GPR), trained with supervised learning on field data. The model predicted tree size per hectare, which was then converted to the number of trees per footprint by scaling to the GEDI footprint area. Lastly, the computed AGBD values per footprint were subsequently used as training data in a RF model, which was applied across the study area to generate a spatially continuous map of AGBD for olive orchards. The RF model was trained with a combination of LiDAR-derived metrics, canopy cover from optical indices, and topographic variables, enabling large-scale biomass estimation.

2.4.2 Dataset Construction and Footprint Filtering

The training phase was conducted using two distinct Above-Ground Biomass Density (AGBD) estimation approaches: (i) the direct retrieval of biomass estimates from the GEDI L4A product and (ii) a modeled AGBD derived from volumetric and dendrometric parameters. The construction of the AGBD dataset involved integrating GEDI LiDAR metrics, canopy cover predictions, and remote sensing variables from optical, SAR, and topographic sources. The process began with filtering GEDI footprints obtained from the L2A and L4A products, covering acquisitions from 2020 to 2022. To ensure high data quality, only footprints with a quality flag of zero, indicating reliable signal returns, were retained (Dubayah et al., 2020). Footprints were then spatially filtered using olive grove parcels identified through the Sistema de Información Geográfica de Parcelas Agrícolas (SIGPAC), a cadastral geographic information system that provides high-precision agricultural boundaries(MAPA, 2024).

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

This spatial intersection ensured that only footprints fully within olive groves were included, minimizing contamination from other land uses. Additionally, temporal consistency was maintained by synchronizing GEDI acquisitions with HLS (Harmonized Landsat and Sentinel-2) imagery dates to reduce discrepancies between LiDAR and optical observations (Masek et al., 2021). Canopy cover for each GEDI footprint was estimated using spectral indices derived from HLS imagery, including NDVI and SAVI. A linear regression model, trained on 250 ground truth points from high-resolution orthophotos provided by the Plan Nacional de Ortofotografía Aérea (PNOA), was applied to improve canopy cover estimates. The estimated canopy cover was then used to calculate crown diameter based on the relationship defined in the previous section, linking tree cover to crown geometry.

To avoid edge effects, particularly those caused by mixed land cover classes, olive grove polygons from the SIGPAC database were rasterized, and an "erode" filter was applied to reduce parcel boundaries. This process helped eliminate footprints partially overlapping non-olive areas, ensuring data consistency (Indirabai et al., 2024). The dataset was subsequently partitioned into training and testing subsets, with 75% of the footprints assigned to the training set and the remaining 25% to the test set. The partitioning was stratified to maintain a balanced distribution of canopy heights and cover values, essential for robust model generalization (Ma et al., 2019).

Predictor variables were extracted for each GEDI footprint using the Google Earth Engine (GEE) platform, incorporating a comprehensive set of remote sensing features. LiDAR-derived metrics included RH95, canopy cover, and L4A-derived AGBD estimates (Duncanson et al., 2022). Optical predictors, including NDVI, SAVI, and BSI, were computed from HLS imagery (Masek et al., 2021). To capture additional structural information, SAR texture metrics were derived from ALOS-2 PALSAR-2 imagery using the Gray Level Co-occurrence Matrix (GLCM) method, which provides valuable insights into vegetation patterns through measures of entropy and contrast (Haralick et al., 1973; Shimada et al., 2014). Additionally, topographic variables, including slope, aspect, and elevation, were derived from the Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) with a spatial resolution of 30 meters (NASA JPL, 2013). These diverse predictors ensured that the dataset captured the multi-dimensional characteristics necessary for accurate biomass modeling.

2.5 Machine Learning Approach for AGBD Prediction

The selected model for this study was RF, a machine learning algorithm widely recognized for its resilience to noise and effectiveness in addressing overfitting issues. Although RF was initially developed for classification tasks, its use in regression problems, such as biomass estimation, has significantly increased (Pérez-Cutillas et al., 2023). The model was trained using the comprehensive feature dataset detailed in **Table 1**, which integrates LiDAR metrics (GEDI L2A and L4A), optical indices (HLS NDVI, SAVI, BSI), SAR texture features (PALSAR-2 GLCM metrics), and topographic attributes (SRTM slope, aspect, and elevation). The RF algorithm employs a bootstrap aggregation (bagging) approach, wherein each decision tree is trained on a random subset of the data with replacement. This ensemble method aggregates predictions from multiple trees to reduce variance and improve accuracy (Breiman, 2001).

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

Table 1. Summary of Remote Sensing Variables and Features Used for AGBD Estimation in Olive Orchards

Data type	Image Data	Band	Description
LiDAR	GEDI	RH95	L2A: Canopy Top Height at Relative Height 95% (m)
		Quality Flag	L2A: Flag indicating likely invalid waveform
		AGBD	L4A: AGBD prediction (Ton ha-1)
		AGBD std	L4A: AGBD prediction standard error (Ton ha-1)
Optical-Infrared	HLS: Landsat 8/9 and Sentinel 2	Band 1	Coastal Aerosol
		Band 2	Blue
		Band 3	Green
		Band 4	Red
		Band 5	NIR
		Band 6	SWIR 1
		Band 7	SWIR 2
		Band 9	Cirrus
		NDVI	(NIR-RED) / (NIR+RED)
		GI	GREEN / RED
		NDWI	(NIR-SWIR) / (NIR+SWIR)
		GNDBI	(NIR-GREEN) / (NIR+GREEN)
		MCARI1	1.2 * ((2.5*(NIR-RED)) - (1.3*(NIR-GREEN)))
		GLI	2*(GREEN + RED + BLUE) / 2*(GREEN - RED - BLUE)
		EVI	2.5 * ((NIR - RED) / (NIR + 6 * RED - 7.5 * BLUE + 1))
		SAVI	((NIR - RED) / (NIR + RED + 0.5)) * 1.5
	PNOA	Cover	Total Canopy Cover (%)
SAR	ALOS2 PALSAR2	HH	Polarization backscattering coefficient L-Band
		HV	Polarization backscattering coefficient L-Band
		Contrast	GLCM Texture: Contrast
		Correlation	GLCM Texture: Correlation
		Entropy	GLCM Texture: Entropy
		Inertia	GLCM Texture: Inertia
		RVI	(4 * HV) / (HH + HV)
Topographic	SRTM	Elevation	SRTM V3 product (SRTM Plus)
		Slope	Derived by “Elevation”
		Aspect	Derived by “Elevation”

Note: Acronyms used in this table are as follows: GEDI (Global Ecosystem Dynamics Investigation), AGBD (Above-Ground Biomass Density), L2A (GEDI product providing Canopy Height Metrics), L4A (GEDI product providing Above-Ground Biomass Density estimates), RH95 (Relative Height at 95%), HLS (Harmonized Landsat-Sentinel), NIR (Near-Infrared), SWIR (Shortwave Infrared), NDVI (Normalized Difference Vegetation Index), NDWI (Normalized Difference Water Index),

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

GI (Green Index), GNDBI (Green Normalized Difference Built-up Index), MCARI1 (Modified Chlorophyll Absorption Ratio Index 1), GLI (Green Leaf Index), EVI (Enhanced Vegetation Index), SAVI (Soil-Adjusted Vegetation Index), PNOA (National Aerial Orthophotography Plan), SAR (Synthetic Aperture Radar), ALOS-2 (Advanced Land Observing Satellite-2), PALSAR-2 (Phased Array type L-band Synthetic Aperture Radar-2), HH (Horizontal-Horizontal polarization), HV (Horizontal-Vertical polarization), GLCM (Gray Level Co-occurrence Matrix), RVI (Radar Vegetation Index), SRTM (Shuttle Radar Topography Mission), and DEM (Digital Elevation Model).

Model hyperparameters were configured to balance computational efficiency and predictive performance. The maximum tree depth was set to 10, with a minimum of 1 sample per leaf node and 2 samples per split, ensuring that trees remained efficient and avoided overfitting. The number of trees in the forest was fixed at 100, a value determined to provide stable predictions without excessive computational costs. These parameters were selected to optimize model performance within the GEE environment, where complex models with excessive depth or tree counts may become inoperable due to memory constraints. To ensure robust performance across different data configurations, the model was trained under various input combinations of remote sensing variables, as outlined in **Table 2**.

2.6 Model Validation and Large-Scale Application

The model validation was conducted through a cross-validation process using the test dataset, which represented 25% of the total data. Standard performance metrics were calculated, including the coefficient of determination (R^2), and root mean square error (RMSE).To enhance the robustness of biomass estimation, different configurations of remote sensing datasets were systematically combined and analyzed, allowing for a comprehensive assessment of feature importance and the contribution of each geospatial dataset.

The validation process involved multiple combinations of input variables derived from optical, radar, and topographic datasets (**Table 2**). Initially, optical bands from the HLS dataset, including bands B1 to B9, were tested independently to establish a baseline accuracy. Subsequently, spectral index derivatives were incorporated to assess their impact on biomass prediction accuracy.

Table 2.Analyses performed on the training set depending on the typology of data sets combination

Model	Features	Geospatial dataset
Optical bands	-B1, B2, B3, B4, B5, B6, B7, B9	HLS L2A
Spectral derivatives	-EVI, GI, GLI, GNDBI, MCARI1, NDWI, NDVI, SAVI	HLS L2A
SAR polarization, textures, and Radar Vegetation Index (RVI)	-HH, HV -HH Contrast, HH Correlation, HH Entropy, HH Inertia -HV Contrast, HV Correlation, HV Entropy, HV Inertia -RVI	ALOS2-PALSAR2

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

Optical bands, SAR textures, and RVI	-B1, B2, B3, B4, B5, B6, B7, B9 -HH Contrast, HH Correlation, HH Entropy, HH Inertia -HV Contrast, HV Correlation, HV Entropy, HV Inertia -RVI	HLS L2A| ALOS2-PALSAR2
Optical bands and SAR polarization	-B1, B2, B3, B4, B5, B6, B7 -HH, HV	HLS L2A |ALOS2-PALSAR2
SAR polarization, textures, RVI, and Topographic data	-HH, HV -HH Contrast, HH Correlation, HH Entropy, HH Inertia -HV Contrast, HV Correlation, HV Entropy, HV Inertia -RVI -Aspect, Elevation, Slope	ALOS2-PALSAR2 | SRTM v003
Optical bands and Topographic data	-B1, B2, B3, B4, B5, B6, B7, B9 -Aspect, Elevation, Slope	HLS L2A| SRTM v003
Fully Multi-Source	-B1, B2, B3, B4, B5, B6, B7, B9 -EVI, GI, GLI, GNDBI, MCARI1, NDBI, NDVI, SAVI, FVC -HH, HV -HH Contrast, HH Correlation, HH Entropy, HH Inertia -HV Contrast, HV Correlation, HV Entropy, HV Inertia -RVI -Aspect, Elevation, Slope	HLS L2A| ALOS2-PALSAR2 | SRTM v003

Note: Acronyms used in this table are as follows: B1-B9 (Spectral bands from Harmonized Landsat-Sentinel), HLS L2A (Harmonized Landsat-Sentinel Level 2A product), EVI (Enhanced Vegetation Index), GI (Green Index), GLI (Green Leaf Index), GNDBI (Green Normalized Difference Built-up Index), MCARI1 (Modified Chlorophyll Absorption Ratio Index 1), NDWI (Normalized Difference Water Index), NDVI (Normalized Difference Vegetation Index), SAVI (Soil-Adjusted Vegetation Index), SAR (Synthetic Aperture Radar), HH (Horizontal-Horizontal polarization), HV (Horizontal-Vertical polarization), Contrast, Correlation, Entropy, and Inertia (GLCM Texture metrics derived from SAR data), RVI (Radar Vegetation Index), ALOS2-PALSAR2 (Advanced Land Observing Satellite-2 Phased Array type L-band Synthetic Aperture Radar-2), SRTM v003 (Shuttle Radar Topography Mission version 3), Aspect, Elevation, and Slope (Topographic variables derived from SRTM).

Further evaluations integrated SAR features derived from ALOS-2 PALSAR-2, considering polarization channels (HH, HV), texture metrics (contrast, correlation, entropy, and inertia), and the RVI. These variables were tested individually and in combination with optical bands to examine their synergy in biomass estimation. To explore topographic influences, SRTM data were added, including terrain-related predictors such as slope, aspect, and elevation. Finally, the most comprehensive model, termed Fully Multi-Source, incorporated all available data sources, including optical bands, spectral indices, SAR-derived features, and topographic variables.

For large-scale application, the trained model was deployed on the GEE platform, enabling the generation of a high-resolution AGBD map of olive groves across Europe. The spatially explicit biomass estimates support biomass monitoring, carbon sequestration assessments, and sustainable agricultural management in Mediterranean landscapes (Shendryk, 2022).

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

Additionally, the model was tested under different regional satellite image coverage scenarios, validating its generalization capability and robustness across diverse agro-climatic conditions. The integration of multi-source remote sensing data within a cloud-based platform ensures scalability and operational feasibility, making it a valuable tool for large-scale biomass assessments.

3 Results

3.1. Training Phase and AGBD Estimation Inputs

The training dataset was compiled using filtered GEDI footprints from the study area, primarily covering the extensive olive orchards of Andalusia, Spain. **Figure 3** illustrates the spatial distribution of the training samples along with the histograms of AGBD values obtained from both GEDI L4A and the volumetric model (GEDI L2A).

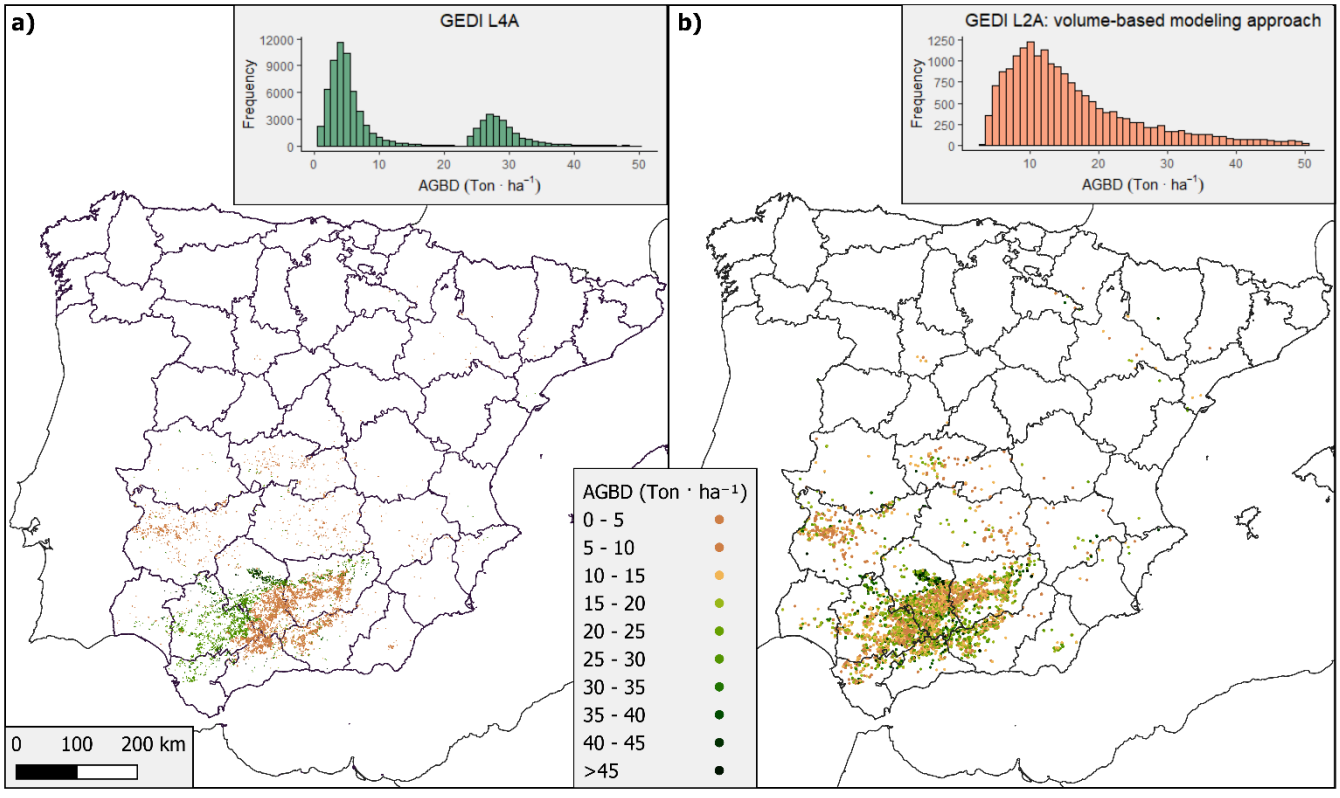


Figure 3. Spatial distribution of training points (GEDI footprints) filtered for olive orchard plots in Spain from 2020 to 2022. The figure compares AGBD estimates from GEDI L4A and GEDI L2A-derived models. (a) GEDI L4A provides direct aboveground biomass density (AGBD) predictions, whereas (b) AGBD estimates derived from GEDI L2A data are obtained through a theoretical volume-based modeling approach

The GEDI L4A product exhibited a bimodal distribution of AGBD values, which is attributable to the classification scheme employed in its estimation algorithm. The first and most prominent cluster of biomass estimates fell within the range of 0–15

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

Ton·ha^{-1}, primarily corresponding to areas classified as Grasslands, Shrubs, and Woodlands, which are prevalent within olive-growing regions. A secondary peak in the data was observed between 25 and 35 Ton·ha^{-1}, aligning with areas classified as Deciduous Broadleaf Forests. The stratification applied by GEDI's retrieval algorithm, which differentiates biomass estimation models based on Plant Functional Types (PFTs), contributes to this distinct grouping pattern (Kellner et al., 2023).

Conversely, the modelled AGBD dataset derived from GEDI L2A canopy height and volumetric calculations exhibited a more continuous and uniform frequency distribution. Biomass values displayed broader variability, with the highest occurrence near Ton·ha^{-1} and a distribution tail extending towards 50 Ton·ha^{-1}. The spatial distribution of the modelled values did not exhibit distinct clusters, suggesting a more consistent representation of olive grove biomass. The volumetric approach, based on canopy structure and tree morphology, provided a more direct estimation of olive tree biomass, independent of global stratified models.

3.2. Model Performance and AGBD Prediction Accuracy

The predictive performance of the model varied depending on the dataset combinations used during the training phase. The "Fully Multi-Source" model consistently achieved the highest accuracy, with an R^2 of 0.56 and an RMSE of 9.09 Mg·ha^{-1} for GEDI L4A, and 0.62 and 5.95 Mg·ha^{-1} for GEDI L2A, demonstrating the advantage of integrating multiple remote sensing sources, including optical, SAR, LiDAR, and topographic variables. Among the individual data sources, topographic and optical datasets significantly improved model accuracy (**Figure 4**).

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

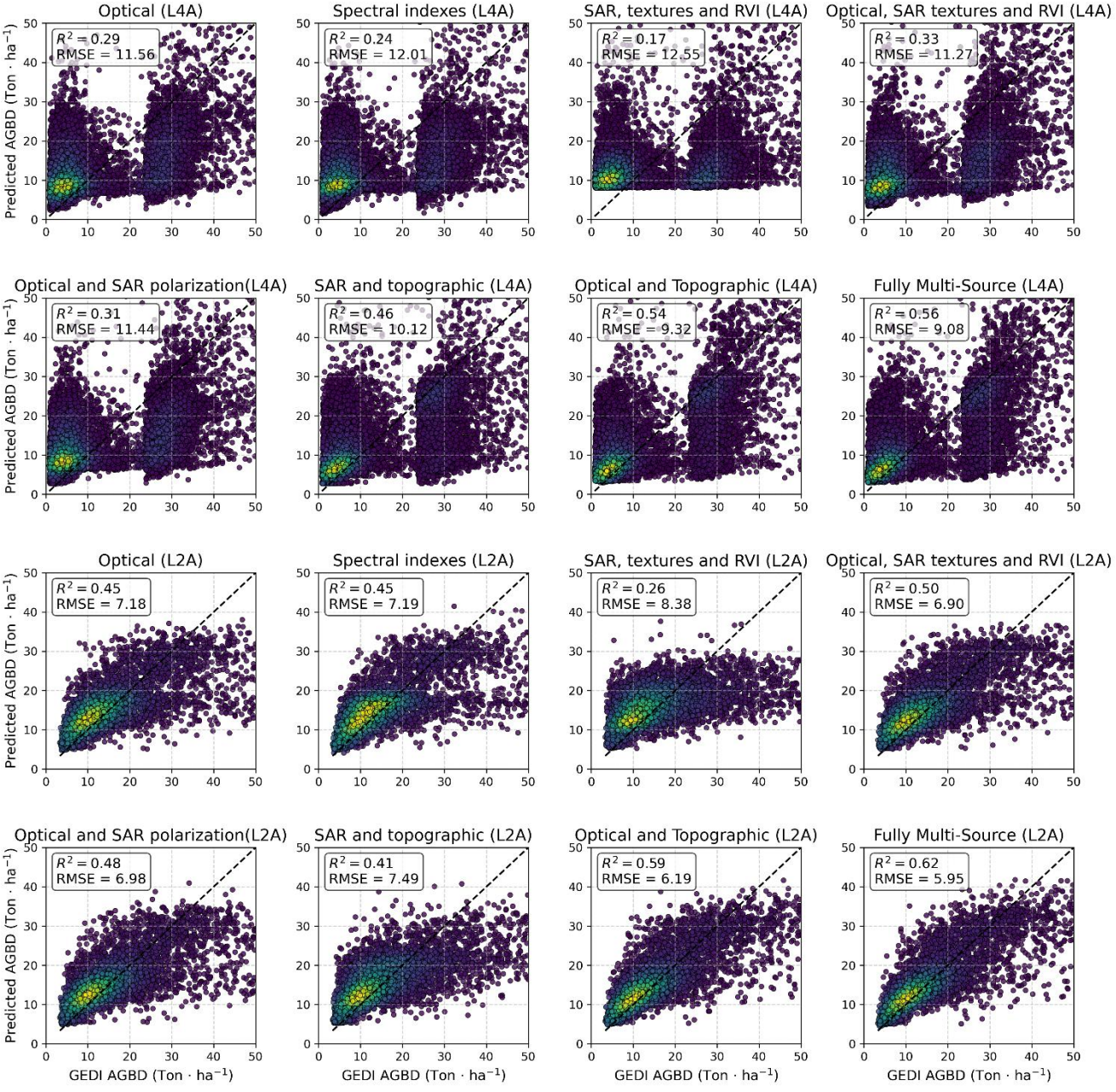


Figure 4. Density plots for the regression model AGBD estimation. All tested models are represented, L4A is referred to models using GEDI L4A data (AGBD predictions depending on PFTs) and L2A is referred to models using GEDIA L2A data (AGBD volume-based method). The dashed black line represents the 1:1 line. For interpreting the colour of the dots it is necessary to know that the yellow colour represents high density of dots while the dark blue colour represents low density of dots.

The "Optical bands and Topographic data" model achieved the second-best performance, with an R^2 of 0.54 and an RMSE of 9.33 Ton·ha^{-1} for GEDI L4A, and 0.59 and 6.19 Ton·ha^{-1} for GEDI L2A. The inclusion of slope and elevation variables proved to be particularly beneficial in cultivated areas with significant terrain variability. Models relying exclusively on optical bands

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

or spectral indices yielded lower predictive power. The "Optical bands" model produced R^2 values of 0.30 (GEDI L4A) and 0.45 (GEDI L2A), while the "Spectral Indexes" model performed similarly, with R^2 values of 0.24 (GEDI L4A) and 0.45 (GEDI L2A). These results suggest that, while vegetation indices contribute relevant spectral information, they alone are insufficient for accurate biomass estimation. Radar-derived models showed the lowest predictive accuracy.

The "SAR polarization, textures, and RVI" model exhibited the weakest performance, with an R^2 of 0.17 and RMSE of 12.54 Ton·ha^{-1} for GEDI L4A, and 0.25 and 8.38 TonMg·ha^{-1} for GEDI L2A. The incorporation of topographic data alongside SAR variables improved the predictions slightly (R^2 of 0.46 for GEDI L4A and 0.41 for GEDI L2A), but RMSE values remained significantly higher compared to models integrating optical data. The combination of optical and SAR data showed moderate improvements over individual sources. The "Optical bands and SAR textures and RVI" model achieved an R^2 of 0.33 and RMSE of 11.27 Ton·ha^{-1} for GEDI L4A, and 0.50 and 6.89 Ton·ha^{-1} for GEDI L2A. Similarly, the "Optical bands and SAR polarization" model reached an R^2 of 0.31 (GEDI L4A) and 0.48 (GEDI L2A), with corresponding RMSE values of 11.39 Ton·ha^{-1} and 6.97 Ton·ha^{-1}.

In general, all model configurations showed higher accuracy when predicting GEDI L2A-derived AGBD compared to GEDI L4A. The RMSE values were consistently lower for GEDI L2A, and the R^2 values were systematically higher across all dataset combinations, suggesting a stronger relationship between predicted and observed biomass values for the GEDI L2A model. These findings highlight that the most accurate models were those that integrated topographic and optical data, with the fully integrated multi-source model producing the best overall performance. In contrast, models relying exclusively on spectral indices or radar data exhibited lower predictive capacity, reinforcing the importance of integrating multiple remote sensing sources for improved AGBD estimation.

3.3. Spatial Distribution and Large-Scale Mapping of AGBD

The spatial distribution of Above-Ground Biomass Density (AGBD) across Mediterranean olive-growing regions exhibits significant variations at both national and sub-national levels. Spain presents the highest total AGB among the analyzed countries, with extensive areas of high-density biomass predominantly located in Andalusia. The highest biomass values, exceeding 50 Ton·ha^{-1}, are concentrated in southern Spain, particularly in regions characterized by intensive olive cultivation. Other areas in Spain, such as Castilla-La Mancha and Catalonia, display moderate AGBD values, ranging between 20 and 40 Ton·ha^{-1}, while lower biomass densities, below 15 Ton·ha^{-1}, are observed in more marginal olive-growing areas (**Figure 1**). Italy follows Spain in total AGB, but the latter is more heterogeneous distribution. The highest biomass densities, reaching 40–50 TonMg·ha^{-1}, are mainly observed in central and southern Italy, particularly in Tuscany, Apulia, and Calabria. Northern and insular regions such as Sardinia and Sicily exhibit more variable biomass values, with patches of high-density AGBD interspersed with areas below 20 Ton·ha^{-1}. Greece ranks third in total AGBD but displays a distinctive biomass distribution pattern. The highest densities, ranging from 30 to 45 Ton·ha^{-1}, are concentrated in Crete and the Peloponnese, which are key regions for olive production. Other olive-growing areas in mainland Greece show lower biomass densities, generally below 25 Ton·ha^{-1}, with more fragmented and dispersed high-density patches.

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

The large-scale AGBD mapping provides a detailed representation of biomass gradients across olive orchards, enabling a refined analysis at regional and local scales. In all three countries, intensive production zones consistently exhibit higher biomass values, while traditional and less-managed orchards tend to show lower AGBD densities, often below 15 Ton·ha^{-1}. These spatial variations emphasize the model's ability to accurately capture biomass heterogeneity across diverse agro-ecological contexts, offering a valuable tool for biomass monitoring and sustainable resource management in Mediterranean olive-growing regions. $AGBD_{mean}$

3.4. Impact of Remote Sensing Data Integration on Model Performance

Overall, the results confirm that integrating diverse geospatial datasets enhances the robustness and precision of AGBD predictions. The combination of LiDAR, optical, SAR, and topographic data allowed for a more comprehensive characterization of olive biomass distribution, reducing uncertainty in predictions and improving spatial resolution across Mediterranean agricultural landscapes.

Feature importance analysis provided insights into the relative contributions of different predictors to the model. Each model identifies key features relevant to estimation, aiding in the interpretation of their characteristics and influence on predictions. The integration of spectral indices derived from HLS further enhanced the model, with NDVI and SAVI emerging as key variables due to their strong correlation with vegetation vigor and canopy structure provided by GEDI sensor.

Among the spectral indices, Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI) emerged as the most influential predictors of AGBD. Optical bands consistently outperformed SAR-derived features, demonstrating greater sensitivity to vegetation structure and biomass variations. Within the SAR datasets, HV polarization contributed the most to the predictions, with simple HV polarization proving more effective than its associated texture metrics. However, models relying solely on SAR features tended to underestimate biomass values, leading to reduced accuracy compared to those integrating optical data.

The results indicate that models integrating spectral data, HV polarization, and topographic variables achieved the highest predictive performance, highlighting the strong influence of topography—particularly slope and elevation—on biomass estimations. In the optical domain, bands B7, B5, and B4 provided the highest predictive value, reinforcing the importance of vegetation-sensitive spectral regions in biomass modeling. Overall, the fully integrated multi-source model provided the most reliable AGBD estimates, reducing RMSE by approximately 20% compared to LiDAR-only models. These findings underscore the necessity of incorporating diverse remote sensing inputs to enhance predictive accuracy and ensure robust biomass assessments in Mediterranean olive orchards.

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

4 Discussion

4.1. Model Performance and Data Integration

The modeled AGBD using volumetric information derived from dendrometric data provided more accurate estimates for olive orchards compared to GEDI L4A measurements. This result aligns with the findings of (Velàzquez-Martì et al., (2014) who demonstrated that crown diameter and volume are strong predictors of biomass in olive trees. However, the limited availability of large-scale field data remains a significant challenge for validating biomass predictions (Fernández-Sarría et al., 2019). Although volumetric calculations offer a direct estimation approach, uncertainties inherent to dendrometric measurements can propagate into the model, affecting prediction accuracy.

The model's predictive performance varied depending on the combination of remote sensing datasets used during training. Optical and topographic variables played a crucial role in improving model accuracy, with slope and elevation contributing significantly, especially in regions with complex terrain. This is consistent with previous studies highlighting the influence of topography on vegetation growth and biomass distribution (Indirabai et al., 2024; Rodríguez-Lizana et al., 2023). Among optical indices, NDWI and NDVI exhibited the highest importance, reflecting their sensitivity to vegetation structure and water content (Kebede et al., 2018). Notably, spectral bands B7, B5, and B4 contributed valuable information, supporting findings by Estornell et al., (2015) regarding the relationship between spectral reflectance and biomass estimation.

The comparative analysis of different dataset combinations showed that integrating optical, SAR, and topographic variables produced the most accurate predictions, with the fully multi-source model achieving the highest R^2 and lowest RMSE. This result underscores the benefits of combining complementary data sources to capture the diverse factors influencing biomass. While SAR data, particularly HV polarization, added structural information, its predictive capacity was lower than that of optical and topographic variables. This observation aligns with Shendryk, (2022), who noted that SAR-derived metrics are more effective in forest environments due to their sensitivity to canopy structure.

On the other hand, it should be added that the model's scalability and computational efficiency are notable advantages, facilitated by Google Earth Engine's (GEE) cloud-based infrastructure. GEE's capacity to process large geospatial datasets enables the application of the model over extensive areas without significant computational constraints. This scalability is essential for generating high-resolution biomass maps that support carbon stock assessments and sustainable land management. Furthermore, the use of Random Forest regression within GEE enhances model robustness (Pérez-Cutillas et al., 2023), due to its resilience to noise and non-linear relationships. Overall, the integration of diverse remote sensing datasets, combined with the computational capabilities of GEE, enables accurate and scalable AGBD estimation across Mediterranean olive orchards. This methodological approach provides a practical solution for large-scale biomass monitoring, aligning with current efforts to improve carbon accounting and resource management in agricultural landscapes.

Several studies have employed GEDI datasets or similar remote sensing technologies, such as ICESAT, to map biomass, primarily focusing on extensive forested regions (Indirabai and Nilsson, 2024; Jiang et al., 2022; Li et al., 2024; López-Serrano et al., 2019; May et al., 2024; Shendryk, 2022). However, the application of remote sensing techniques to perennial crops,

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

particularly olive orchards, remains relatively unexplored. While a few studies have quantified biomass in olive trees and assessed pruning residues (Estornell et al., 2015; Rodríguez-Lizana et al., 2023; Velàzquez-Martì et al., 2014), these efforts have generally been limited to localized contexts, highlighting the need for scalable approaches capable of providing large-scale estimates.

4.2. Implications for Biomass Estimation in Mediterranean Olive Orchards

The spatial mapping of biomass in olive groves provides practical benefits for farmers, environmental managers, and policymakers. Accurate biomass quantification supports carbon accounting, enabling more sustainable agricultural practices through optimized residue management and reduced carbon emissions. Identifying areas with high biomass density facilitates decisions on biomass plant locations, promoting efficient resource utilization and enhancing the economic viability of energy production from agricultural residues (Rodríguez-Lizana et al., 2023; Velàzquez-Martì et al., 2014).

By quantifying biomass accumulation, the model enables the estimation of carbon stocks and tracks carbon sequestration, aligning with global climate mitigation goals (Rodríguez-Lizana et al., 2023). High-density olive orchards, with AGBD values reaching 50 $Ton \cdot ha^{-1}$, act as key carbon sinks, capable of offsetting atmospheric CO_2 (Velàzquez-Martì et al., 2014). Identifying these areas promotes sustainable land management practices, including the use of pruning residues as organic soil amendments, which further enhances soil carbon retention while reducing emissions from conventional waste disposal (Kebede et al., 2018).

Moreover, AGBD predictions help improve crop efficiency by identifying orchards with lower biomass productivity, enabling targeted interventions such as optimized irrigation, nutrient management, and canopy pruning to enhance growth and yield. At the regional level, the model supports biomass inventories, providing essential data for both policymakers and farmers to assess the feasibility of biomass-based energy production. For example, the geostatistical approach of (Rodríguez-Lizana et al., 2023) demonstrates the potential to identify optimal biomass plant locations, reducing transportation costs and emissions while supporting renewable energy within local economies. Additionally, monitoring biomass dynamics over time allows for evaluating the long-term sustainability of agricultural practices, ensuring that productivity improvements do not compromise soil health or carbon sequestration capacity. By integrating remote sensing technologies like GEDI and multispectral imagery, the model offers a scalable and cost-effective solution to support climate-smart agriculture and sustainable land use in Mediterranean olive orchards.

The biomass values obtained in this study are consistent with prior research estimating biomass in both forested and agricultural landscapes. Unlike dense and unordered forest vegetation, olive groves exhibit structured planting frameworks, resulting in lower biomass values per unit area (Fernández-Sarría et al., 2019). Regarding model performance, previous studies have reported similar predictive accuracy, despite variations in vegetation types and geographic settings. For instance, Li et al. (2024) achieved an R^2 of 0.59, aligning closely with the R^2 of 0.56 obtained in this study. Conversely, Shendryk, (2022)

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

reported higher prediction accuracy (R^2 = 0.72) when applying GEDI data to forested regions, suggesting that vegetation structure and canopy density influence the predictive capacity of remote sensing models.

Despite its advantages, improving the precision of biomass estimation in olive groves remains a key challenge. Remote sensing models calibrated with satellite data may struggle to capture fine-scale variability within orchards, limiting their accuracy at localized levels. While UAV-based approaches have demonstrated high precision in small-scale biomass assessments (Estornell et al., 2015; Fernández-Sarría et al., 2019), their limited spatial coverage restricts their applicability for regional or continental-scale assessments. Consequently, integrating high-resolution UAV data with large-scale satellite observations could enhance model performance, enabling more accurate and scalable biomass mapping in Mediterranean olive orchards (Roma et al., 2024).

4.3. Limitations of AGBD model and Future Research

The AGBD model developed in this study presents several limitations related to input data variability, environmental factors, computational constraints, and generalization potential. A key challenge is the variability and accuracy of input data, particularly the dependence on volumetric estimates derived from dendrometric measurements. These measurements, while essential, introduce uncertainties due to variations in tree structure and orchard management practices (Rodríguez-Lizana et al., 2023; Velàzquez-Martì et al., 2014). The use of GEDI L4A data, although globally available, is not specifically calibrated for olive crops, leading to potential discrepancies in biomass predictions (Kellner et al., 2023). Additionally, the spatial filtering of GEDI footprints to isolate olive groves may exclude mixed land covers, further affecting data accuracy (Indirabai et al., 2024).

One potential limitation of the model is that the estimated AGBD mean (**Figure 1**) values for Spain are relatively low compared to other countries within the study area, despite Spain having the largest olive-growing area in the European Union. This discrepancy raises the question of whether olive trees in Spain inherently have lower biomass density or if the model overestimates biomass density in regions with lower olive tree density. A key factor influencing this outcome is the environmental conditions prevalent in Spain, particularly the arid and semi-arid climates that characterize extensive areas of the country. The primary predictors used in the model (NDWI, NDVI, and SAVI) are sensitive to vegetation cover and moisture content (Fern et al., 2018), which may result in lower predicted AGBD values in regions with sparse vegetation and greater exposure of bare soil. Conversely, areas with more abundant vegetation cover, such as Slovenia, Croatia, and Italy, exhibit higher average AGBD values, highlighting the influence of spectral indices in capturing vegetation density and health.

This effect underscores the importance of considering environmental factors, such as precipitation and water availability, directly impacting biomass accumulation in olive orchards, but these factors were not directly integrated into the model. The observed influence of slope on model performance in Spain suggests that including climatic and soil properties could enhance prediction accuracy, as reported in previous studies on olive groves and other crops (Kebede et al., 2018; Rodríguez-Lizana et al., 2023). Differences in irrigation practices and soil composition, particularly soil organic carbon (SOC) and clay content,

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

can also impact biomass estimates, highlighting the need for more comprehensive datasets in future research (Fernández-Sarría et al., 2019). Consequently, while Spain's extensive olive cultivation contributes significantly to total biomass production, the average biomass density is comparatively lower possibly due to the region's climatic constraints.

From a computational perspective, the cloud-based Google Earth Engine (GEE) platform facilitates large-scale biomass mapping, yet its reliance on pre-processed datasets can limit model customization and input diversity. The RF algorithm, while robust and efficient, is sensitive to data quality and may exhibit reduced performance when input data are sparse or noisy (Shendryk, 2022). Moreover, the computational complexity of integrating multisource data, including optical, SAR, and LiDAR metrics, requires optimizing model parameters to balance accuracy and processing time, as demonstrated in large-scale biomass studies using GEDI and Sentinel data (Indirabai et al., 2024; Shendryk, 2022).

Remote sensing limitations also affect biomass estimation, particularly in regions with heterogeneous canopy structures. SAR data, despite its ability to penetrate cloud cover, exhibited limited predictive capacity compared to optical data, consistent with findings in previous studies (Kellner et al., 2023; Rodríguez-Lizana et al., 2023). The saturation effect observed in high-biomass areas remains a challenge, although combining SAR and optical data partially mitigated this issue (Shendryk, 2022). Additionally, the spatial resolution of GEDI footprints (25 m) may not capture fine-scale variations within olive orchards, especially in areas with irregular planting patterns.

Uncertainty in biomass estimations arises from both input data variability and model assumptions. The volumetric approach, while scalable, inherently carries higher uncertainty than allometric methods, which rely on direct measurements of tree mass (Rodríguez-Lizana et al., 2023; Velàzquez-Martì et al., 2014). The exclusion of certain environmental variables, such as soil moisture and nutrient availability, further contributes to prediction uncertainty. Nevertheless, the model's performance, with R^2 values up to 0.62 and RMSE as low as 5.95 $Ton \cdot ha^{-1}$, is comparable to previous studies in forestry and agricultural contexts, indicating its suitability for large-scale applications (Indirabai et al., 2024; Shendryk, 2022).

Model generalization and transferability represent additional challenges. While the model demonstrated reasonable accuracy in Spanish olive groves, its applicability to other regions may be limited by differences in climate, soil properties, and orchard management practices. However, the relatively consistent environmental conditions across major olive-growing regions increase the likelihood of successful model transfer, provided that region-specific calibration is performed (Rodríguez-Lizana et al., 2023). Future research should focus on expanding the training dataset to include diverse environmental conditions and incorporating additional variables such as soil texture, SOC, and irrigation regimes to enhance model robustness and generalizability.

5 Conclusion

This study demonstrates the potential of integrating multi-source remote sensing data to estimate Above-Ground Biomass Density (AGBD) in Mediterranean olive orchards. By combining GEDI LiDAR-derived metrics with optical, SAR, and topographic variables, the model achieved a robust predictive performance, with the fully multi-source model obtaining the

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

highest accuracy (R^2 = 0.62, RMSE = 5.95 Mg·ha^{-1}). Among the remote sensing variables, NDWI, slope, and NDVI emerged as the most influential predictors, highlighting the relevance of both vegetation structure and environmental conditions for biomass estimation.

The volumetric approach using GEDI L2A data proved more accurate than the GEDI L4A product, primarily due to its direct relationship with tree morphology. However, uncertainties remain due to variations in tree structure and orchard management practices, particularly in estimating crown and stem volumes. Despite these limitations, the model successfully captured the spatial heterogeneity of biomass across Mediterranean olive-growing regions, with Spain exhibiting the highest total AGBD, followed by Italy and Greece.

The model's large-scale applicability, facilitated by Google Earth Engine, enables efficient biomass mapping over extensive areas, supporting carbon stock assessments and resource management. Its scalability and computational efficiency make it a practical tool for governmental organizations to monitor carbon dynamics, evaluate annual emissions, and assess the agricultural sector's carbon sequestration potential. In addition, the model could provide farmers with a digital support tool for crop monitoring, facilitating optimized irrigation, nutrient management, and biomass utilization.

Future research should focus on integrating additional environmental variables such as climate, soil properties, and irrigation regimes to further improve prediction accuracy. Expanding the training dataset to include diverse olive-growing regions will enhance model generalization and transferability. Moreover, combining high-resolution UAV data with large-scale satellite observations could help capture fine-scale variability within orchards, enhancing both accuracy and scalability.

In conclusion, this study presents a scalable and cost-effective methodology for estimating AGBD in olive orchards, contributing to carbon accounting, sustainable agriculture, and climate change mitigation in Mediterranean ecosystems. The integration of remote sensing technologies and machine learning techniques offers a valuable tool for supporting data-driven decision-making in both agricultural management and environmental policy.

Acknowledgements

This study is part of the project OLIVCHAR Ref: TED2021-131907B-I00, financed by the Spanish MCIN/AEI /10.13039/501100011033 and the European NextGenerationEU/PRTR funds

References

Asner, G. P., Mascaro, J., Muller-Landau, H. C., Vieilledent, G., Vaudry, R., Rasamoelina, M., Hall, J. S., and van Breugel, M.: A universal airborne LiDAR approach for tropical forest carbon mapping, Oecologia, 168, 1147–1160, https://doi.org/10.1007/s00442-011-2165-z, 2012.

Brede, B., Terryn, L., Barbier, N., Bartholomeus, H. M., Bartolo, R., Calders, K., Derroire, G., Krishna Moorthy, S. M., Lau, A., Levick, S. R., Raumonen, P., Verbeeck, H., Wang, D., Whiteside, T., van der Zee, J., and Herold, M.: Non-destructive

estimation of individual tree biomass: Allometric models, terrestrial and UAV laser scanning, Remote Sens Environ, 280, 113180, https://doi.org/10.1016/j.rse.2022.113180, 2022.

Breiman, L.: Random forests, Mach Learn, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Brunori, A., Dini, F., Cantini, C., Sala, G., La Mantia, T., Caruso, T., Marra, F. P., Trotta, C., Nasini, L., Regni, L., and Proietti, P.: Biomass and volume modeling in Olea europaea L. cv "Leccino," Trees - Structure and Function, 31, 1859–1874, https://doi.org/10.1007/s00468-017-1592-9, 2017.

Deitch, M., Sapundjieff, M., and Feirer, S.: Characterizing Precipitation Variability and Trends in the World's Mediterranean-Climate Areas, Water (Basel), 9, 259, https://doi.org/10.3390/w9040259, 2017.

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P. L., Qi, W., and Silva, C.: The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography, Science of Remote Sensing, 1, https://doi.org/10.1016/j.srs.2020.100002, 2020.

Duncanson, L., Disney, M., Armston, J., Nickeson, J., Minor, D., and Camacho, F.: Committee on Earth Observation Satellites Working Group on Calibration and Validation Land Product Validation Subgroup Aboveground Woody Biomass Product Validation Good Practices Protocol Version 1.0, Good Practices for Satellite Derived Land Product Validation, https://doi.org/10.5067/doc/ceoswgcv/lpv/agb.001, 2021.

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., Alonso, A., Andersen, H.-E., Aplin, P., Baker, T. R., Barbier, N., Bastin, J. F., Biber, P., Boeckx, P., Bogaert, J., Boschetti, L., Boucher, P. B., Boyd, D. S., Burslem, D. F. R. P., Calvo-Rodriguez, S., Chave, J., Chazdon, R. L., Clark, D. B., Clark, D. A., Cohen, W. B., Coomes, D. A., Corona, P., Cushman, K. C., Cutler, M. E. J., Dalling, J. W., Dalponte, M., Dash, J., de-Miguel, S., Deng, S., Ellis, P. W., Erasmus, B., Fekety, P. A., Fernandez-Landa, A., Ferraz, A., Fischer, R., Fisher, A. G., García-Abril, A., Gobakken, T., Hacker, J. M., Heurich, M., Hill, R. A., Hopkinson, C., Huang, H., Hubbell, S. P., Hudak, A. T., Huth, A., Imbach, B., Jeffery, K. J., Katoh, M., Kearsley, E., Kenfack, D., Kljun, N., Knapp, N., Král, K., Krůček, M., Labrière, N., Lewis, S. L., Longo, M., Lucas, R. M., Main, R., Manzanera, J. A., Martínez, R. V., Mathieu, R., Memiaghe, H., Meyer, V., Mendoza, A. M., Monerris, A., Montesano, P., Morsdorf, F., Næsset, E., Naidoo, L., Nilus, R., O'Brien, M., Orwig, D. A., Papathanassiou, K., Parker, G., Philipson, C., Phillips, O. L., Pisek, J., Poulsen, J. R., Pretzsch, H., et al.: Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission, Remote Sens Environ, 270, 112845, https://doi.org/10.1016/j.rse.2021.112845, 2022.

Estornell, J., Ruiz, L. A., Velázquez-Martí, B., López-Cortés, I., Salazar, D., and Fernández-Sarría, A.: Estimation of pruning biomass of olive trees using airborne discrete-return LiDAR data, Biomass Bioenergy, 81, 315–321, https://doi.org/10.1016/j.biombioe.2015.07.015, 2015.

Fern, R. R., Foxley, E. A., Bruno, A., and Morrison, M. L.: Suitability of NDVI and OSAVI as estimators of green biomass and coverage in a semi-arid rangeland, Ecol Indic, 94, 16–21, https://doi.org/10.1016/j.ecolind.2018.06.029, 2018.

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

Fernández-Lobato, L., Ruiz-Carrasco, B., Tostado-Véliz, M., Jurado, F., and Vera, D.: Environmental impact of the most representative Spanish olive oil farming systems: A life cycle assessment study, J Clean Prod, 442, 141169, https://doi.org/10.1016/j.jclepro.2024.141169, 2024.

Fernández-Sarría, A., López-Cortés, I., Estornell, J., Velázquez-Martí, B., and Salazar, D.: Estimating residual biomass of olive tree crops using terrestrial laser scanning, International Journal of Applied Earth Observation and Geoinformation, 75, 163–170, https://doi.org/10.1016/j.jag.2018.10.019, 2019.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sens Environ, 202, 18–27, https://doi.org/10.1016/j.rse.2017.06.031, 2017.

Haralick, R. M., Shanmugam, K., and Dinstein, I.: Textural Features for Image Classification, IEEE Trans Syst Man Cybern, SMC-3, 610–621, https://doi.org/10.1109/TSMC.1973.4309314, 1973.

Henrich, V.: IDB - Index-Database; Development of a database for remote sensing indices, ZFL-Colloquium, Bonn, 21. 06. 2012., 2012.

Hunka, N., May, P., Babcock, C., de la Rosa, J. A. A., de los Ángeles Soriano-Luna, M., Saucedo, R. M., Armston, J., Santoro, M., Suarez, D. R., Herold, M., Málaga, N., Healey, S. P., Kennedy, R. E., Hudak, A. T., and Duncanson, L.: A geostatistical approach to enhancing national forest biomass assessments with Earth Observation to aid climate policy needs, Remote Sens Environ, 318, 114557, https://doi.org/10.1016/j.rse.2024.114557, 2025.

Indirabai, I. and Nilsson, M.: Estimation of above ground biomass in tropical heterogeneous forests in India using GEDI, Ecol Inform, 82, 102712, https://doi.org/10.1016/j.ecoinf.2024.102712, 2024.

Jiang, F., Deng, M., Tang, J., Fu, L., and Sun, H.: Integrating spaceborne LiDAR and Sentinel-2 images to estimate forest aboveground biomass in Northern China, Carbon Balance Manag, 17, https://doi.org/10.1186/s13021-022-00212-y, 2022.

Kebede, B. and Soromessa, T.: Allometric equations for aboveground biomass estimation of *Olea europaea* L. subsp. *cuspidata* in Mana Angetu Forest, Ecosystem Health and Sustainability, 4, 1–12, https://doi.org/10.1080/20964129.2018.1433951, 2018.

Kellner, J. R., Armston, J., and Duncanson, L.: Algorithm Theoretical Basis Document for GEDI Footprint Aboveground Biomass Density, Earth and Space Science, 10, https://doi.org/10.1029/2022EA002516, 2023.

Li, H., Hiroshima, T., Li, X., Hayashi, M., and Kato, T.: High-resolution mapping of forest structure and carbon stock using multi-source remote sensing data in Japan, Remote Sens Environ, 312, 114322, https://doi.org/10.1016/j.rse.2024.114322, 2024.

López-Serrano, P. M., Cárdenas Domínguez, J. L., Corral-Rivas, J. J., Jiménez, E., López-Sánchez, C. A., and Vega-Nieva, D. J.: Modeling of Aboveground Biomass with Landsat 8 OLI and Machine Learning in Temperate Forests, Forests, 11, 11, https://doi.org/10.3390/f11010011, 2019.

Ma, J., Li, Y., Chen, Y., Du, K., Zheng, F., Zhang, L., and Sun, Z.: Estimating above ground biomass of winter wheat at early growth stages using digital images and deep convolutional neural network, European Journal of Agronomy, 103, 117–129, https://doi.org/10.1016/j.eja.2018.12.004, 2019.

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

MAPA: Sistema de Información Geográfica de Parcelas Agrícolas, SIGPAC. (Accesed November 2024), 2024.

Masek, J., Ju, J., Roger, J.-C., Skakun, S., Claverie, M., and Dungan, J.: Harmonized Landsat/Sentinel-2 Products for Land Monitoring, in: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, 8163–8165, https://doi.org/10.1109/IGARSS.2018.8517760, 2018.

Masek, J., Ju, J., Roger, J., Skakun, S., Vermote, E., Claverie, M., Dungan, J., Yin, Z., Freitag, B., and Justice, C.: HLS Operational Land Imager Surface Reflectance and TOA Brightness Daily Global 30m v2.0 [Data set], 2021.

May, P. B., Schlund, M., Armston, J., Kotowska, M. M., Brambach, F., Wenzel, A., and Erasmi, S.: Mapping aboveground biomass in Indonesian lowland forests using GEDI and hierarchical models, Remote Sens Environ, 313, 114384, https://doi.org/10.1016/j.rse.2024.114384, 2024.

NASA JPL: NASA Shuttle Radar Topography Mission Global 1 arc second [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. Accessed 2024-11-22 from https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1.003, 2013.

Nesha, K., Herold, M., De Sy, V., de Bruin, S., Araza, A., Málaga, N., Gamarra, J. G. P., Hergoualc'h, K., Pekkarinen, A., Ramirez, C., Morales-Hidalgo, D., and Tavani, R.: Exploring characteristics of national forest inventories for integration with global space-based forest biomass data, Science of The Total Environment, 850, 157788, https://doi.org/10.1016/j.scitotenv.2022.157788, 2022.

Pardo, G., del Prado, A., Martínez-Mena, M., Bustamante, M. A., Martín, J. A. R., Álvaro-Fuentes, J., and Moral, R.: Orchard and horticulture systems in Spanish Mediterranean coastal areas: Is there a real possibility to contribute to C sequestration?, Agric Ecosyst Environ, 238, 153–167, https://doi.org/10.1016/j.agee.2016.09.034, 2017.

Pérez-Cutillas, P., Pérez-Navarro, A., Conesa-García, C., Zema, D. A., and Amado-Álvarez, J. P.: What is going on within google earth engine? A systematic review and meta-analysis, https://doi.org/10.1016/j.rsase.2022.100907, 1 January 2023.

Perna, C., Pagliai, A., Sarri, D., Lisci, R., and Vieri, M.: Can a Light Detection and Ranging (LiDAR) and Multispectral Sensor Discriminate Canopy Structure Changes Due to Pruning in Olive Growing? A Field Experimentation, Sensors, 24, 7894, https://doi.org/10.3390/s24247894, 2024.

Proietti, S., Sdringola, P., Desideri, U., Zepparelli, F., Brunori, A., Ilarioni, L., Nasini, L., Regni, L., and Proietti, P.: Carbon footprint of an olive tree grove, Appl Energy, 127, 115–124, https://doi.org/10.1016/j.apenergy.2014.04.019, 2014.

Rodríguez-Lizana, A., Ramos, A., Pereira, M. J., Soares, A., and Ribeiro, M. C.: Assessment of the Spatial Variability and Uncertainty of Shreddable Pruning Biomass in an Olive Grove Based on Canopy Volume and Tree Projected Area, Agronomy, 13, https://doi.org/10.3390/agronomy13071697, 2023.

Roma, E., Catania, P., Vallone, M., and Orlando, S.: Assessing the Effectiveness of Pruning in an Olive Orchard Using a Drone and a Multispectral Camera: A Three-Year Study, Agronomy, 14, 1023, https://doi.org/10.3390/agronomy14051023, 2024.

https://doi.org/10.5194/egusphere-2025-917
Preprint. Discussion started: 2 April 2025

Rosúa, J. M. and Pasadas, M.: Biomass potential in Andalusia, from grapevines, olives, fruit trees and poplar, for providing heating in homes, Renewable and Sustainable Energy Reviews, 16, 4190–4195, https://doi.org/10.1016/j.rser.2012.02.035, 2012.

Shendryk, Y.: Fusing GEDI with earth observation data for large area aboveground biomass mapping, https://doi.org/10.1016/j.jag.2022.103108, 1 December 2022.

Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R., and Lucas, R.: New global forest/non-forest maps from ALOS PALSAR data (2007–2010), Remote Sens Environ, 155, 13–31, https://doi.org/10.1016/j.rse.2014.04.014, 2014.

Urdiales-Flores, D., Zittis, G., Hadjinicolaou, P., Cherchi, A., Alessandri, A., Peleg, N., and Lelieveld, J.: A Global Analysis of Historical and Future Changes in Mediterranean Climate-Type Regions, International Journal of Climatology, 44, 5607–5620, https://doi.org/10.1002/joc.8655, 2024.

Velázquez-Martí, B., Fernández-González, E., López-Cortés, I., and Salazar-Hernández, D. M.: Quantification of the residual biomass obtained from pruning of trees in Mediterranean olive groves, Biomass Bioenergy, 35, 3208–3217, https://doi.org/10.1016/j.biombioe.2011.04.042, 2011.

Velàzquez-Martì, B., Cortès, I. L., and Salazar-Hernàndez, D. M.: Dendrometric analysis of olive trees for wood biomass quantification in Mediterranean orchards, Agroforestry Systems, 88, 755–765, https://doi.org/10.1007/s10457-014-9718-1, 2014.