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

5   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)

10   **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, offered a promising framework for large-scale biomass estimation using lidar technologies on satellite platforms. The model's predictive performance varied depending on data combinations, with the fully multi-source configuration achieving the highest accuracy. NDBI, slope, HV polarization, and MCARI1 were identified as the most influential predictors. The spatial analysis revealed that Spain exhibited the highest total AGB stock 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, although it exhibited certain limitations in terms of precision at finer spatial scales. This study presents an exploratory analysis that 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

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 the concentration of greenhouse gases in the atmosphere (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^{-1}). 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 components due to their structural sensitivity, whereas radar systems are typically more responsive to moisture content and object size, which often results in a stronger signal from live woody biomass. Nonetheless, SAR can also detect dead biomass, especially when structural elements persist. 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. Forest ecosystems typically exhibit complex structures shaped by natural processes (Sinha et al., 2015), while olive orchards are generally more structurally ordered due to intentional planting arrangements. However, this apparent regularity can be misleading, as olive plantations still display substantial structural variability across sites. Differences in planting frameworks, irrigation regimes, pruning practices, and management intensities introduce considerable heterogeneity, even within this controlled agricultural context. 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 (Velázquez-Martí 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), where the availability of allometric databases from diverse locations supports their implementation (Jucker 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.

65 An alternative method estimates total biomass using geometric parameters to derive the volume of woody material in different tree structures (Kebede and Soromessa, 2018; Velázquez-Martí 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). 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

70 is often constrained by the need for extensive field data collection (Brede et al., 2022).

Conversely, volumetric models represent a scalable and potentially accurate alternative for biomass estimation (Li et al., 2024). However, when applied to heterogeneous landscapes, their reliability strongly depends on the resolution and accuracy of the input data. In the case of GEDI, while it provides valuable structural information, limitations in both vertical and horizontal precision introduce considerable uncertainty. Airborne LiDAR data can provide highly accurate three-dimensional structural

75 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).

80 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

85 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).

90 Given the limited availability of remote sensing-based biomass estimates in olive groves, 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

95 applied at a European scale, leveraging datasets such as Corine Land Cover (CLC) as other datasets at greater precision to ensure to ensure the application of the model on the correct land use. Ultimately, this research aims to provide a tool for monitoring carbon stocks in form of biomass in olive orchards,. In the case of olive orchards, this is particularly relevant given the possibility of transforming residual biomass into stable carbon forms, enabling the development of carbon sequestration protocols, supporting sustainable agricultural management and climate change mitigation strategies.

100 **2 Material and methods**

2.1. Study area

To identify olive cultivation areas, the CLC database was used. This is a European land use dataset that includes a specific category for "Olive Groves". 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
105 analysis is conducted in Spain, utilizing the 'Sistema de Información Geográfica de Parcelas Agrícolas'. 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. In comparison, CLC database reports 3,587,300 hectares, clearly overestimating the actual olive grove area in Spain.

110 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
115 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 and Pasadas, 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%
120 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

AGBD prediction in olive orchards focuses on quantifying the biomass associated with olive trees within a gridded spatial framework. This approach involves estimating the total number of trees per pixel, the average crown diameter per pixel, and
125 applying a volumetric modelling method to derive biomass estimates.

Currently, there are products such as GEDI L4A product that provides AGBD estimates based on plant functional types (PFTs). This product employs a zonal stratification approach, applying specific models depending on the geographic region (Kellner et al., 2023). These models are developed based on the dominant vegetation types within each area. While this method is appropriate for large-scale mapping and generalization, it becomes insufficient when focusing on specific tree species. In such cases, more detailed approaches are required to enable more accurate monitoring of biomass using remote sensing techniques. This study utilizes GEDI L2A data to estimate AGBD through a volumetric approach. This approach derives AGBD estimates from the GEDI L2A product (Dubayah et al., 2021), which provides canopy height metrics transformed into biomass values using a volumetric model tailored to olive tree morphology, supported by the estimations of tree density and canopy cover. This dataset is integrated with remote sensing variables and analyzed to assess the relative importance of each data type in biomass estimation. GEE was employed to download, pre-process, filter, and transform the satellite images.

The workflow was structured into six main phases: (1) acquisition of remote sensing data; (2) processing of all datasets to generate derived subproducts from the original inputs; (3) modeling of AGBD using a volumetric approach; (4) application of machine learning techniques to estimate both intermediate subproducts and AGBD; (5) large-scale application of the model; and (6) model validation using 25% of the training dataset, along with an additional validation using ALS-derived referenced data (Figure 1).
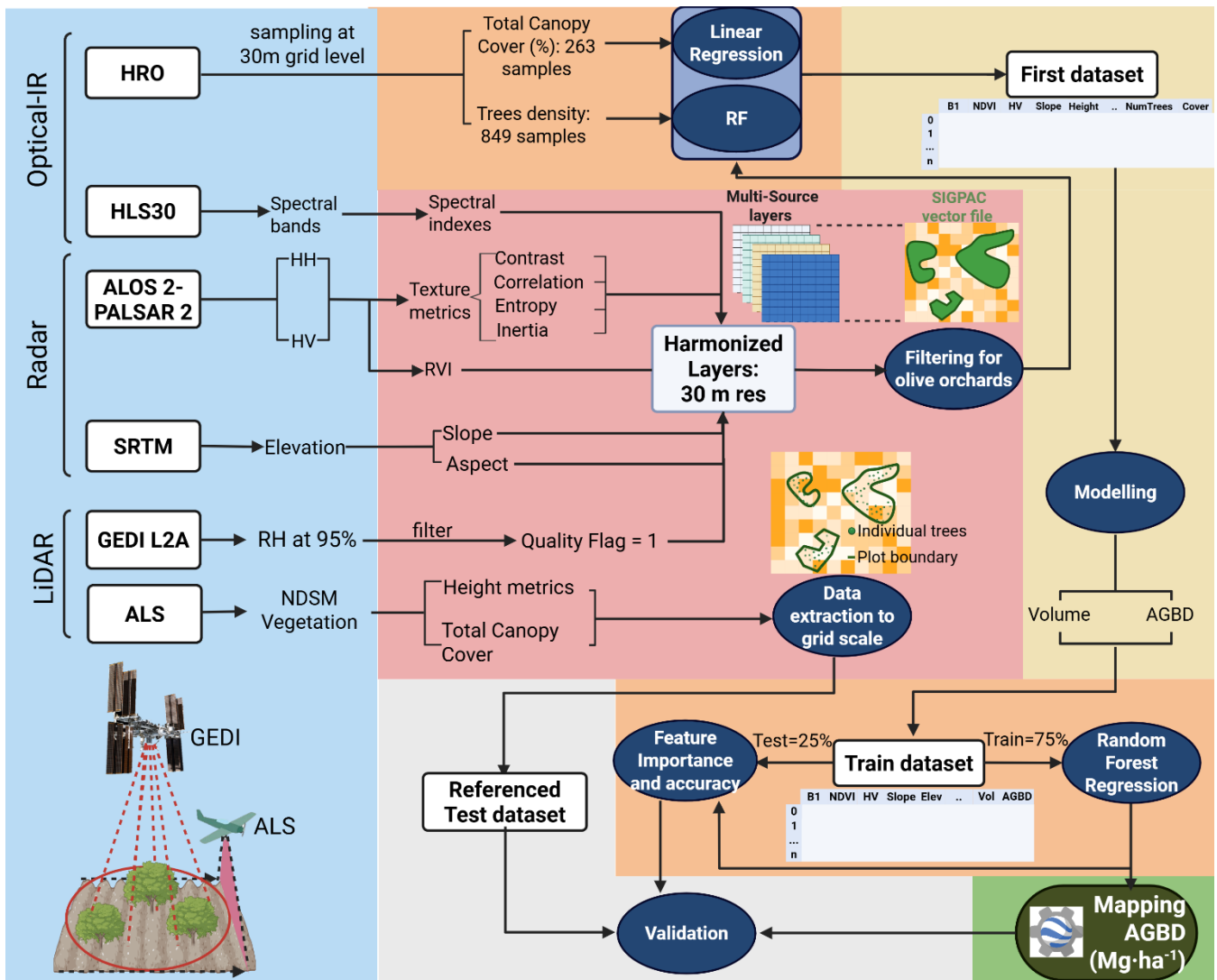
Figure 1. Methodological Flowchart for AGBD mapping in Olive Orchards. The workflow comprises six sequential stages, each highlighted by a distinct color frame: data acquisition (blue frame), preprocessing (red frame), training dataset generation (yellow frame), model training (orange frame), biomass mapping (green frame), and testing (grey frame). Remote sensing data sources include GEDI (Global Ecosystem Dynamics Investigation) LiDAR onboard the ISS (International Space Station), Airbone Laser Scanning (ALS), High Resolution Orthoimagery (HRO), optical imagery (HLS: 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. GEDI-derived biomass values (L2A canopy height metrics) are correlated with spectral indices and radar-derivatives. A Random Forest (RF) model is trained and applied at a large scale to generate a biomass density map.

The methodology begins with the processing 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. The

spatial distribution of AGBD across Mediterranean accuracy in olive grove delineation, SIGPAC cadastral data is used to refine the initial classification from the CLC database, that is, SIGPAC was used for the filtering of selected sites for training and testing. Training dataset construction and biomass estimation have been carried out using the filtered GEDI data that are linked to spectral, radar, and topographic variables, creating a comprehensive training dataset. L2A canopy height metrics serves as canopy height values for AGBD estimation after its processing using a volumetric approach adapted to olive cultivation.

Remote sensing data provides a continuous spatial grid, while GEDI data provides data only at certain points at the footprint level. It is for this reason that we are assessing the ability of common remotely sensed data to reliably estimate biophysical variables by integrating both types of data. The Random Forest (RF) regression model is employed to large-scale predictions in the GEE platform and 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 (30 m) biomass density map for olive groves only in the areas considered "Olive Orchards" by the CLC database.

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).

Each GEDI footprint has a diameter of approximately 25 meters and is separated 60 meters apart along track, allowing for detailed sampling of vertical structure at fine spatial resolution (Duncanson et al., 2022). Also, the L2A product has a relative uncertainty associated that interferes in the sampling vertical and positional accuracy.

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 (Kellner et al., 2023). The GEDI L2A product delivers these RH metrics, which represent the height above ground below which a given percentage of the total returned energy has been accumulated, which is useful as biomass model inputs.

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). This resolution is used as a key layer for harmonizing all multi-source layers. 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 (Masek et al., 2021). For the HLS data processing, all available images from the annual period between 2020 and 2022 were selected. A filtering step was then applied to exclude pixels with more than 10% cloud cover. Subsequently, a temporal median filter was applied to the remaining data. This process enables the generalization of the spectral response over the entire annual period, reducing the influence of outliers and transient atmospheric effects.

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).

High-Resolution Orthophotography (HRO) from the Spanish National Aerial Orthophotography Plan- PNOA (https://pnoa.ign.es/web/portal/pnoa-lidar/presentacion) was also incorporated as a key optical data source in this study, which served as the basis for the sampling process to estimate the crops' planting framework and the canopy cover. These orthophotos have a spatial resolution of 0.25 meters. This dataset was used exclusively during the sampling phase, due to its high spatial resolution and its suitability for accurately guiding the sampling process.

2.3.3 Synthetic Aperture Radar (SAR) and Topographic Data

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). PALSAR-2 is a highly valuable data source, as its operating frequency in the L-band allows for a more effective characterization of overall vegetation structure compared to other SAR systems operating in different frequency bands. For instance, C-band sensors offer higher spatial resolution but are more sensitive to seasonal variations (Qin et al., 2025). The objective of this study was to work with annual SAR data, making ALOS-2 PALSAR-2 the most suitable option. Moreover, the Google Earth Engine (GEE) catalog provides a ready-to-use annual mosaic generated from multiple image strips, which facilitates consistent temporal analysis and simplifies data preprocessing. ALOS-2 PALSAR-2 data were processed directly within GEE to enhance radiometric quality and correct for topographic effects. First, radiometric calibration was applied by converting the digital number (DN) values into backscatter coefficients expressed in decibels (dB) expressed in Equation 1:

$$\gamma_0 = 10 \cdot log_{10}(DN^2) - 83 \tag{1}$$

Subsequently, topographic correction was performed using complementary SRTM data to estimate slope and aspect, allowing for the calculation of the local incidence angle. This angle was then used to adjust the backscatter values, compensating for the influence of terrain relief on SAR signal intensity. Finally, a spatial median filter with a 50-meter radius was applied to reduce speckle noise. The HH and HV polarizations were processed separately and later combined to derive subsequent SAR-based products.

Regarding topographical data, this study employed 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.

Table 1. Summary of Remote Sensing Variables and Features Employed

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
	ALS	NDMS-Veg	Normalized Digital Model Surface of the vegetation class
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
		NDBI	(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
	HR Orthoimagery	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"

225 Note: Acronyms used in this table are as follows: GEDI (Global Ecosystem Dynamics Investigation), , L2A (GEDI product providing Canopy Height Metrics), RH95 (Relative Height at 95%), NDMS (Normalized Digital Model Surface), HLS

(Harmonized Landsat-Sentinel), NIR (Near-Infrared), SWIR (Shortwave Infrared), NDVI (Normalized Difference Vegetation Index), NDWI (Normalized Difference Water Index), 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), HRO-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).

2.4 AGBD Estimation Methodology

2.4.1. Tree Density and Canopy Cover Modeling

Tree density per hectare and tree crown diameter are essential variables for applying the proposed approach at pixel level. A sampling strategy was implemented to obtain these variables using data collected from various points. This sampling allowed producing two derivate products that subsequently interacted with GEDI canopy top metrics to produce the volumetric estimations at 30m grid level.

For the estimation of tree crown size, a method was sought that could incorporate the presence of multiple trees within a single pixel. This enabled the calculation of the average crown diameter across all trees within a pixel, thereby allowing for the generalization of this attribute at the 30-meter-pixel level. Specifically, knowing the proportion of the pixel covered by canopy and the total number of trees within it makes it possible to reliably estimate the average crown diameter.

To this end, 263 'Total Canopy Cover' sampling points were collected across Spain olive orchards (Figure 2). To simplify the canopy cover estimation process, a simple linear regression analysis was conducted, using SAVI as the predictor variable and canopy cover as the response variable. This approach enabled the use of the resulting regression equation to estimate the approximate canopy cover for each pixel using spectral bands.

Additionally, before calculating SAVI, the Bare Soil Index (BSI) was used as an indicator to assess the type of ground cover present. This step improved the characterization of soil conditions, enhancing the reliability of SAVI-based canopy cover estimations. In this case, a threshold-based approach was applied to define the SAVI L-factor: pixels with BSI values less than 0 (indicating non-bare soil) were assigned an 'L' value of 0, while pixels with BSI values greater than 0 (indicating a significant influence of bare soil) were assigned an 'L' value of 0.5. This adjustment is particularly beneficial in Mediterranean environments, where high variability in soil cover is common. For instance, in olive groves with partial vegetative ground cover, distinguishing between canopy and bare soil can be challenging. SAVI helps mitigate these limitations by reducing the soil background effect in such heterogeneous landscapes.

Regarding the number of trees per hectare, the sampling method involved manually counting individual trees within a 30-meter grid (Figure 2). This approach enabled the estimation of tree density at the hectare scale by applying a spatial conversion from the 30-meter grid (900 m²) to a 100-meter grid (10000 m²). The sample number of tree density was 849 across Spain, using sampling in multiple kinds of olive tree cultivation frames. To perform the estimation, Random Forest Regression was used.
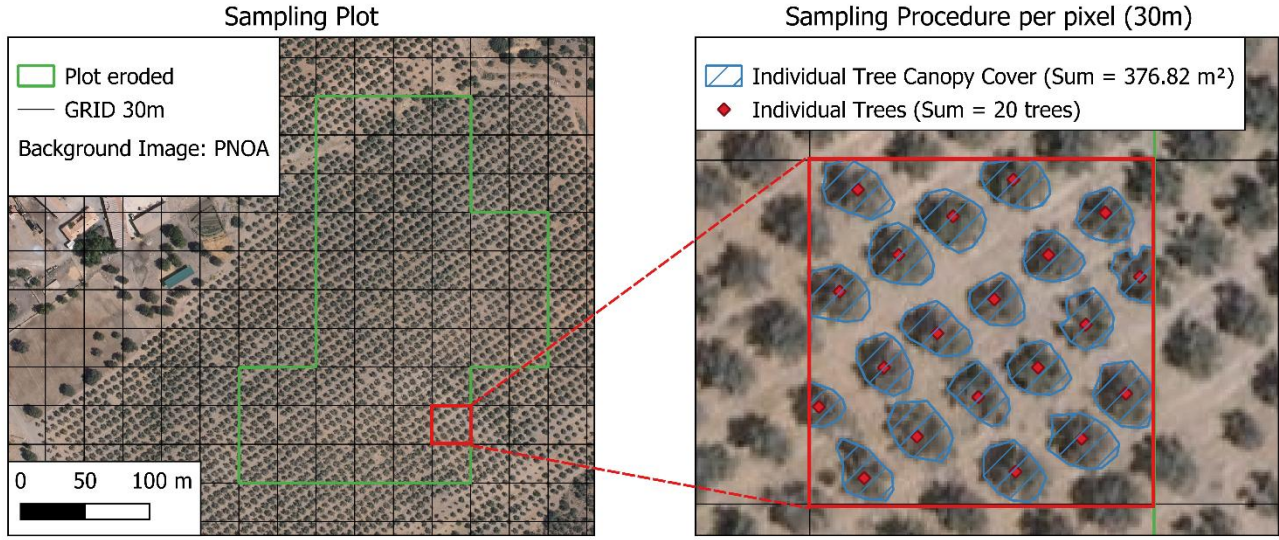
Figure 2. Illustration of the sampling procedure implemented to estimate tree density (trees·ha⁻¹) and canopy cover. The left panel displays an example of single eroded olive grove parcel, extracted from the SIGPAC database, with a 30-meter pixel selected for sampling highlighted in red. The right panel depicts the methodological workflow applied during the sampling phase, which was used to derive the training dataset for model development.

2.4.2 Theoretical Framework for Biomass Modeling

The estimation of AGBD was based on a volumetric modeling approach that combined GEDI L2A LiDAR metrics with canopy cover and tree density predictions derived from multi-source 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 Velázquez-Martí et al. (2014) for olive trees, using a hemispherical volumetric model. This model converts crown volume to biomass using a wood density (WD) of 0.76 g·cm⁻³, which is consistent with values reported for olive trees (Brunori et al., 2017; Kebede and Soromessa, 2018).

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 (Velázquez-Martí et al., 2014). The crown diameter (Cdiam) is the average crown diameter per tree within the 30m pixel, that was obtained from the total canopy cover (Cover), pixel area (Parea), and the number of trees per pixel (Mpix), following Equation (2):

$$Cdiam = 2\sqrt{\frac{(cover \cdot Parea)/Mpix}{\pi}} \tag{2}$$

The hemispherical crown volume (Hvol) was calculated using the formula for a hemisphere (Equation 3):

$$Hvol = \frac{2}{3}\pi r^2 h \tag{3}$$

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

285   To obtain the actual crown volume (Cvol), the hemispherical volume was adjusted by an Occupational Factor (OF) specific to olive tree crown morphology (Velázquez-Martí et al., 2014) as shown in Equation 4:

$$Cvol = \text{Hvol} \cdot OF \tag{4}$$

The stem volume was approximated based on its relationship with crown volume, following the proportional model described by Velázquez-Martí et al. (2014). Crown and stem volumes were converted into biomass using the wood density (WD) value

290   of 0.76 g·cm⁻³. The total 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:

$$AGBD = \text{Tvol} \cdot \text{WD} \cdot \text{M}i \tag{5}$$

Lastly, the computed AGBD values at 30m pixel level 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

295   with a combination of LiDAR-derived metrics, canopy cover from spectral indices, and topographic variables, enabling large-scale biomass estimation.

2.4.3 Dataset Construction and Footprint Filtering

The construction of the AGBD dataset involved integrating GEDI L2A LiDAR metrics, canopy cover predictions, tree density

300   per hectare predictions, and remote sensing variables from optical-infrared, SAR, and topographic sources. The process began with filtering GEDI footprints obtained from the L2A product, covering acquisitions from 2020 to 2022. To ensure high data quality, only footprints with a quality flag of one, 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

305   (MAPA, 2024).

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 imagery dates to reduce discrepancies between GEDI-LiDAR and Optical-Infrared observations (Masek et al., 2021). To avoid edge effects, particularly those caused by mixed land cover classes, olive grove polygons from the SIGPAC database were

310   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 and Nilsson, 2024).

Predictor variables were extracted for each 30m pixel using the Google Earth Engine (GEE) platform, incorporating a comprehensive set of remote sensing features, the resampling method to align all multisource data was Nearest Neighbor. Fully optical-infrared predictors were computed from HLS imagery. To capture additional structural information, SAR texture

315   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 SRTM -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.

320 For the validation dataset, the PNOA-Normalized Digital Surface Model for vegetation (NDSM-veg) was used to ensure consistency in both positional and vertical accuracy. This dataset is a 2.5-meter resolution product derived from LiDAR point clouds collected through Airborne Laser Scanning (ALS) missions. This product is already processed through interpolation of height relative to the ground, based on vegetation classes (low, medium, and tall vegetation derived from LiDAR flights conducted during the first coverage of PNOA. This information was used to generate a robust validation dataset at the 30-

325 meter grid scale, allowing for the estimation of both average crown diameter and the number of trees per pixel. The validation dataset was processed using the same methodology applied during the biomass modeling phase (Figure 1), enabling a volumetric-based and aboveground biomass density (AGBD) validation approach.

2.5 Machine Learning Approach for AGBD Prediction

330 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 2**, which integrates optical indices, SAR polarizations and SAR derivatives (PALSAR-2 GLCM metrics), and topographic attributes (SRTM slope, aspect, and elevation). The RF algorithm employs a

335 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).

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

340 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**.

345

2.6 Model Validation and Large-Scale Application

The model validation was conducted through a cross-validation process using the test dataset, which represented the totally PNOA-ALS points sampled (4792 grid samples). This process is detailed in Figure 1, within the task titled "Data extraction to grid.". Standard performance metrics, including the coefficient of determination (R^2) and root mean square error (RMSE),

350   were calculated. To enhance the robustness of biomass estimation, various input variables derived from optical, radar, and topographic datasets (Table 2) were systematically evaluated. This approach allowed for a comprehensive assessment of feature importance and the individual contribution of each geospatial data source.

The model's accuracy was also evaluated against the training dataset derived from the GEDI L2A-based modeling process. For this purpose, the training dataset was randomly split into 75% for model training and 25% for testing (Figure 1). The
355   total number of samples is 44,393, following a filtering process aimed at ensuring that all AGBD ranges are approximately equally represented within the training dataset. This approach was applied to create a balanced training set across the entire AGBD spectrum, as illustrated in the histogram in Figure 3. The performance of the Random Forest (RF) model was assessed using the test subset, and the evaluation metrics included the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), as well as the relative importance of each predictor variable.

360   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
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	HLS L2A| ALOS2-
	-EVI, GI, GLI, GNDBI, MCARI1, NDBI, NDVI, SAVI, FVC	PALSAR2 | SRTM v003
	-HH, HV	
	-HH Contrast, HH Correlation, HH Entropy, HH Inertia	
	-HV Contrast, HV Correlation, HV Entropy, HV Inertia	
	-RVI	
	-Aspect, Elevation, Slope	

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. 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: biomass modelling framework and AGBD inputs

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

390
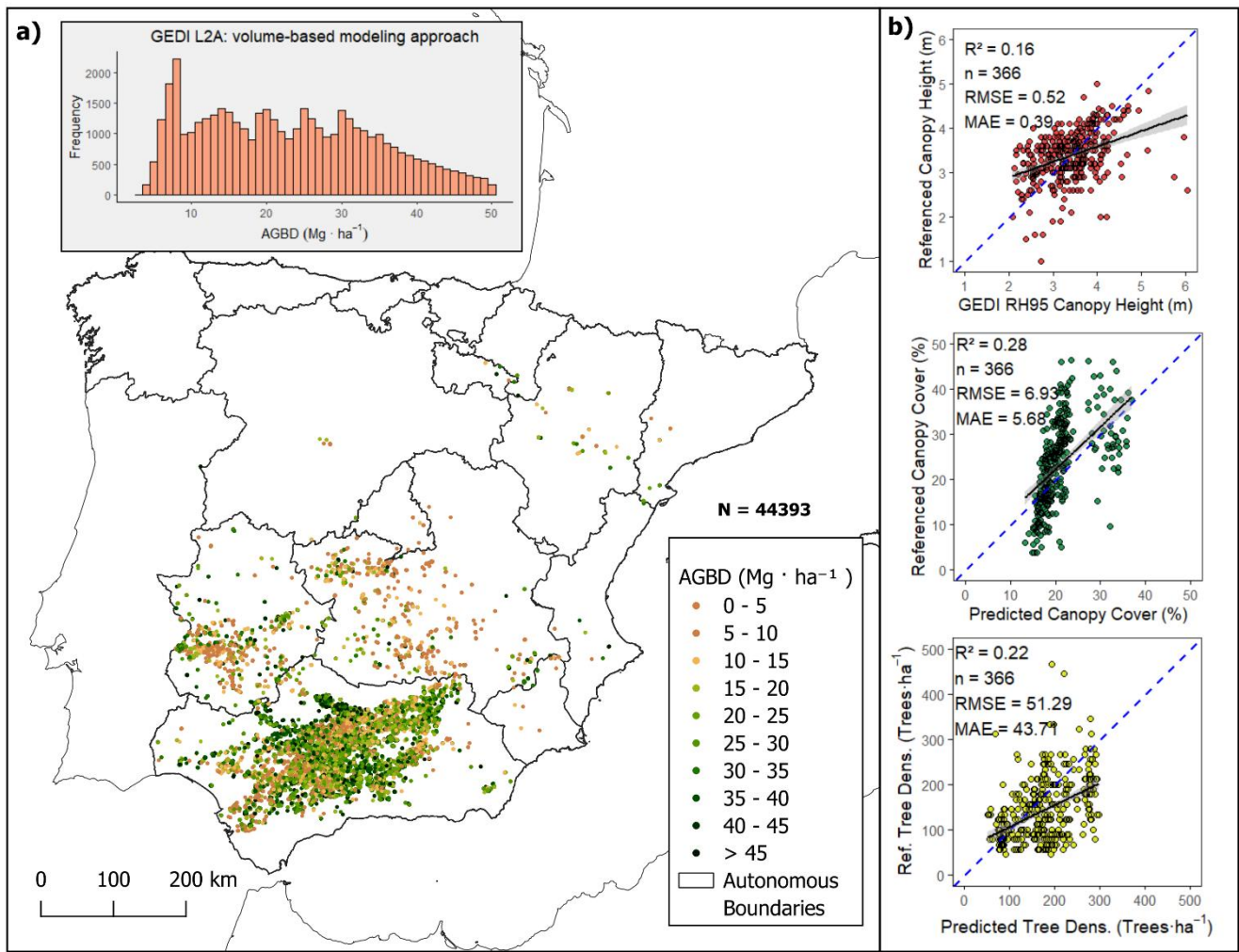
Figure 3. Spatial distribution of training points (GEDI L2A footprints) filtered for olive orchard plots in Spain from 2020 to 2022. (Figure 3 shows in panel "a" the spatial distribution of the points to which the volumetric model was applied, along with the corresponding histogram of the AGBD results. Panel "b" shows the GEDI vegetation height validation, Total Canopy Cover predictions, and Tree Density predictions)

395 The model derived from GEDI L2A enabled the construction of a dataset that spans the entire range of AGBD values for olive groves, while also maintaining balance across biomass classes. This dataset was used for subsequent training with RF for large-scale biomass mapping. Figure 3a shows that the frequency distribution is representative across all biomass intervals, which is critical for the robustness of RF training. The training points are primarily concentrated in southern Spain due to the high density of olive plantations in the region, which allowed us to capture the variability of cultivation practices across this area.

400 However, the dataset includes three key variables with a certain level of uncertainty that are essential for estimating biomass using the proposed approach: (i) vegetation relative height (from GEDI L2A), (ii) canopy cover, and (iii) tree density (trees per hectare). To validate these variables, a total of 366 manually interpreted reference points were used (Figure 3). For the

GEDI L2A height metric (RH95), validation was performed by selecting pixels with an overlapping GEDI footprint. The same set of reference samples was used for validating canopy cover and tree density estimates.

405   The canopy height estimates from GEDI L2A were compared to the referenced data using regression analysis, yielding an R^2 of 0.16, a mean absolute error (MAE) of 0.39 m, and a root mean square error (RMSE) of 0.52 m. Canopy cover estimates, derived from linear regression trained on PNOA sampling data, were validated against reference values, resulting in an R^2 of 0.28, MAE of 5.68%, and RMSE of 6.93%. Finally, tree density predictions obtained through the Random Forest model were also compared to the reference data, with a performance of $R^2 = 0.22$, MAE = 43.71 trees·ha^{-1}, and RMSE = 51.29 trees·ha^{-1}.

410   **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 models were evaluated by comparing the Random Forest predictions against the reference dataset, which consisted of 4,782 independent test samples. The "Fully Multi-Source" model consistently achieved one of the highest accuracies, with an R^2 of 0.14 and an RMSE of 6.68 Mg·ha^{-1} when compared to referenced biomass dataset. When compared to modeled biomass

415   datasets, the model yielded an R^2 of 0.56 and RMSE of 7.73 Mg·ha^{-1} for GEDI L2A volume-based approach, 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 1**).
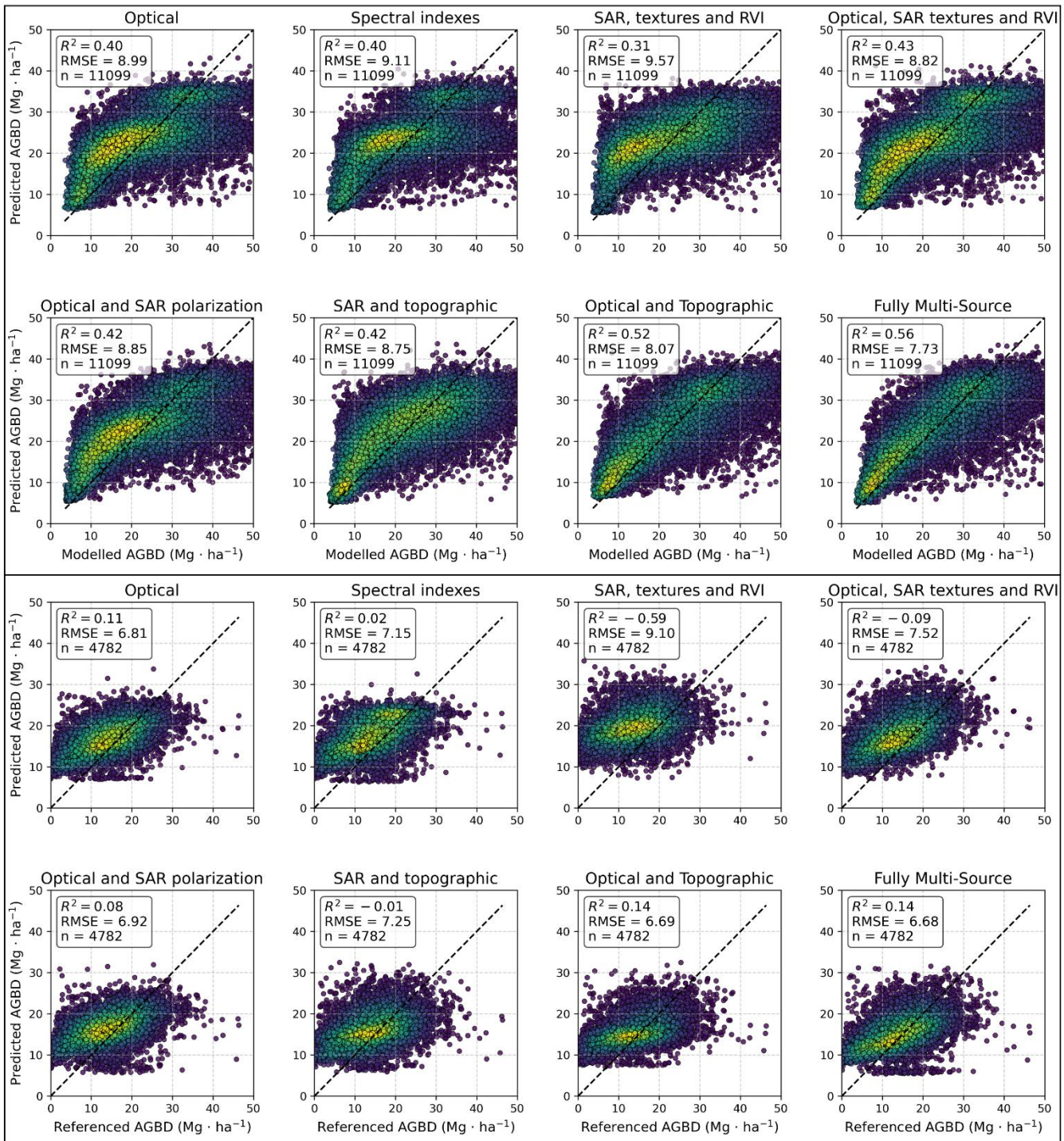
Figure 1. Density plots for the regression model AGBD estimation. The upper panel represents the test dataset derived from the modeled biomass using the GEDI L2A volume-based approach, compared against the model predictions. The lower panel shows the comparison between the reference dataset (PNOA-ALS-based) and the model predictions. 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 0.14 and an RMSE of 6.69 Mg·ha^{-1} compared with the referenced test dataset and compared with modelled dataset with R2 of 0.52 and an RMSE of 8.07. 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 or spectral indices yielded lower predictive power. The "Optical bands" model produced R^2=0.11 , while the "Spectral Indexes" model underperformed, with R^2=0.02. These results indicated 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=-0.59 and RMSE= 9.10 Mg·ha^{-1} compared with the referenced dataset (Figure 4). The incorporation of topographic data alongside SAR variables improved the predictions slightly (R^2 = -0.01), 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= -0.09, representing a slight improvement over models using only SAR data. Similarly, the "Optical bands and SAR polarization" model reached an R^2= 0.08, with corresponding RMSE values of 6.92 Mg·ha^{-1}. When comparing the predicted values with the modeled biomass datasets (as shown in the first eight scatterplots of Figure 4), the R^2 values improve substantially. However, the RMSE values are generally higher, reflecting increased dispersion despite stronger correlation.

Overall, all model configurations demonstrated relatively low accuracy when compared to the referenced dataset. However, certain combinations proved more effective than others, yielding more reliable results. In particular, the two best-performing configurations were the "Optical and Topographic" and "Fully Multi-Source" models, both achieving similar R^2 values and comparable RMSE. 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 due to the lowest RMSE. 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 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, a region in southern Spain. The highest biomass values, exceeding 50 Mg·ha^{-1}, are concentrated in this southern part of the country, 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 Mg·ha^{-1}, while lower biomass densities, below 15 Mg·ha^{-1}, are observed in more marginal olive-growing areas (**Figure 5**).

455 Italy follows Spain in total AGB_stock, but the latter is more heterogeneous distribution. For Italy, the highest biomass densities, reaching 40–50 Mg·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 Mg·ha⁻¹. Greece ranks third in total AGBD but displays a distinctive biomass distribution pattern. The highest densities, ranging from 30 to 45 Mg·ha⁻¹, are concentrated in Crete and the Peloponnese,

460 which are key regions for olive production. Other olive-growing areas in mainland Greece show lower biomass densities, generally below 25 Mg·ha⁻¹, with more fragmented and dispersed high-density patches.

Country	Total	AGB	AGBD mean
Unit	mill. ha	mill. Mg	Mg · ha⁻¹
Spain	3.5873	86.25	24.04
Italy	1.8555	60.92	32.82
Greece	1.2884	41.58	32.28
Turkey	0.6745	21.21	31.44
Portugal	0.5635	14.76	26.20
Albania	0.0822	3.05	37.17
Croatia	0.0494	1.69	34.28
France	0.0198	0.67	34.07
Cyprus	0.0111	0.25	22.97
Slovenia	0.0057	0.21	37.53
Montenegro	0.0009	0.04	38.65
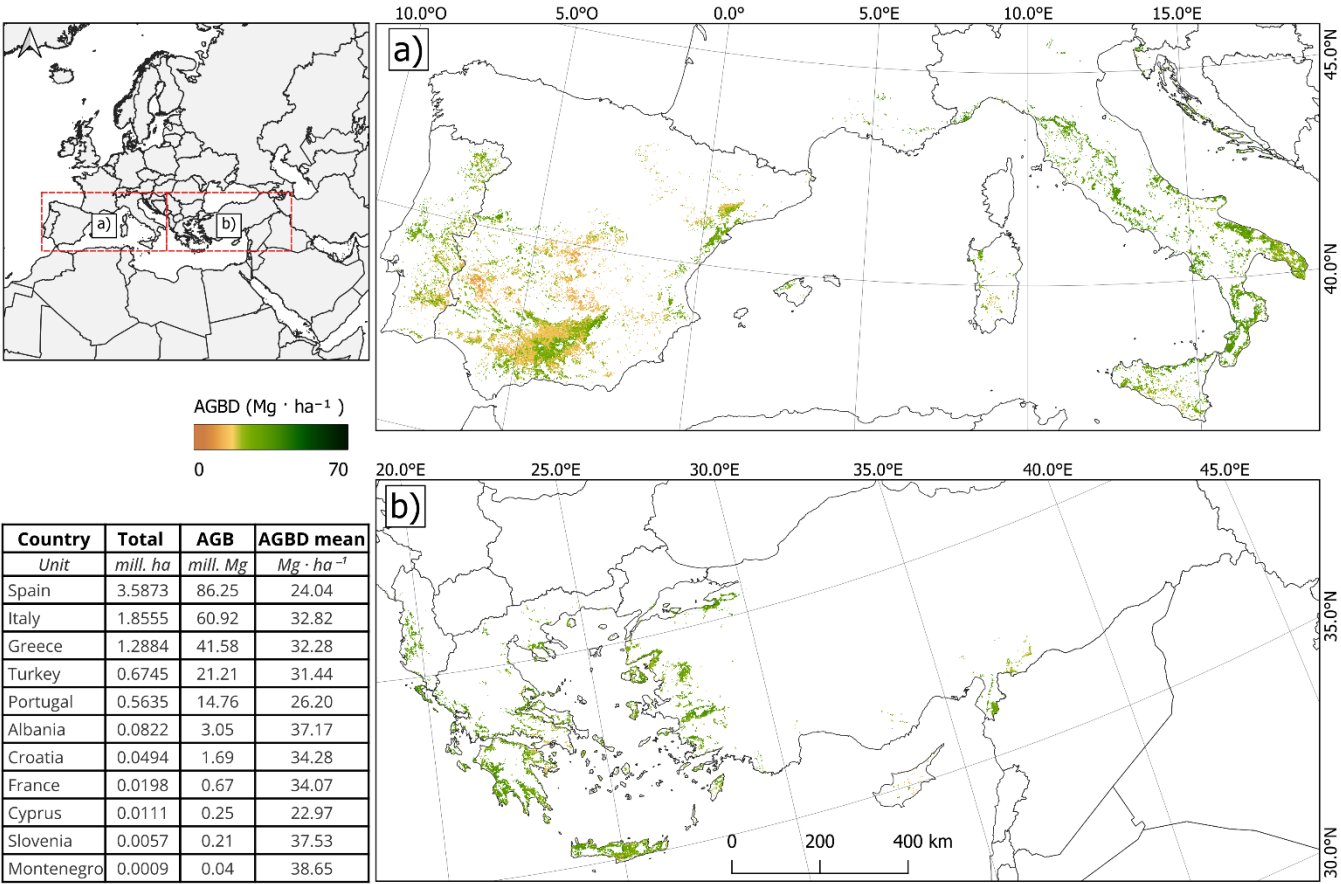
Figure 5. The map shows the distribution of AGBD in olive orchards across Mediterranean countries applying the "Fully Multi-Source" model. Panel "a" shows the AGBD distribution in the Iberian Peninsula and Italy, while Panel 'b' displays the AGBD distribution in Greece,

465 Turkey, and surrounding regions. The color gradient represents AGBD values, ranging from 0 (brown) to 70 Mg·ha⁻¹ (dark green). The inset table summarizes the total cultivated area (mill. ha), total AGB stock (mill. Mg), and mean AGBD (Mg·ha⁻¹) for each country.

The large-scale AGBD mapping provides a detailed representation of biomass gradients across CLC-based 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 20

470 Mg·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.

3.4. Impact of Remote Sensing Data Integration on Model Performance

475 Each model was analyzed based on the performance of RF in predicting AGBD, using the training dataset, as indicated by the orange block in Figure 1. This analysis reflects the learning capacity of RF depending on the input data used during training, with 75% of the data allocated for training and the remaining 25% for testing. The dataset was built from previously generated products, all of which carry an associated level of uncertainty that was previously evaluated (Figure 3). 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 480 biomass distribution, reducing uncertainty in predictions and improving spatial resolution across Mediterranean agricultural landscapes. The Random Forest "Fully Multi-Source" model achieved the highest performance, with an R² of 0.56 and an RMSE of 7.73 Mg·ha⁻¹ when evaluated against the test dataset. It was closely followed by the "Optical Bands and Topographic" model, which reached an R² of 0.52 and an RMSE of 8.07 Mg·ha⁻¹ (Table S1).

Feature importance analysis provided insights into the relative contributions of different predictors to the model. Each model 485 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 vegetation index emerging as key variables due to their strong correlation with vegetation vigor and canopy structure provided by GEDI sensor.

Among the spectral indices, MCARI1, Normalized Difference Vegetation Index (NDVI) and Normalized Difference Built-up Index (NDBI) emerged as the most influential predictors of AGBD (Table S1). Optical bands consistently outperformed SAR-490 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 495 predictive performance, highlighting the strong influence of topography—particularly slope and elevation—on biomass estimations. Notably, 'Slope' was the second most important predictor in the "Fully Multi-Source" model, accounting for 27% of the total variable importance, second only to NDBI, which contributed 42% (Table S1). 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 500 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.

4 Discussion

4.1. Model Performance and Data Integration

505 The volumetric modeling approach for AGBD has proven to be a potentially accurate method. However, as implemented in this study, the modeling phase incorporates multiple sources of error that are subsequently propagated into the final model—representing one of the major weaknesses of this methodology. While the volumetric approach yields realistic biomass estimates at the individual tree level, its application at the geospatial (per-pixel) scale—requiring the derivation of intermediate predictive sub-products—introduces additional uncertainty into the training dataset, which in turn reduces the overall accuracy

510 of the final AGBD estimates.

The validation results of the sub-products, including GEDI L2A canopy height, total canopy cover, and tree density, revealed relatively low predictive performance, which negatively impacted the biomass estimation. The model's heavy dependence on these intermediate variables undermines its reliability, as generating sub-products with limited precision compromises the predictive accuracy of the overall model. Nevertheless, evaluating the performance of each of these components is essential

515 for identifying the model's weaknesses and understanding the key challenges that need to be addressed in future research.

This result aligns with the findings (Velázquez-Martí et al., 2014) of 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

520 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 and Nilsson, 2024; Rodríguez-Lizana et al., 2023).

525 Among optical indices, NDBI and NDVI exhibited the highest importance, reflecting their sensitivity to vegetation structure and water content (Kebede and Soromessa, 2018). Notably, spectral bands B7, B5, and B4 contributed valuable information, supporting findings by (Estornell et al., 2012) 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 R² and lowest RMSE. This

530 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,

535 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 (Figure S1). 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, 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; Velázquez-Martí 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; Velázquez-Martí 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. High-density olive orchards, with AGBD values reaching 50 $Mg \cdot ha^{-1}$, act as key carbon sinks, capable of offsetting atmospheric CO_2. 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 and Soromessa, 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 and lower canopy heights, 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 R² of 0.59, aligning closely with the R² of 0.56 obtained in this study. Conversely, reported higher prediction accuracy (R² = 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 high resolution LiDAR approaches have demonstrated high precision in small-scale biomass assessments (Estornell et al., 2012; 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 .

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; Velázquez-Martí et al., 2014). Additionally, the spatial filtering of GEDI footprints to isolate olive groves may exclude mixed land covers, further affecting data accuracy; however, spatial misalignments between the various remote sensing layers and the GEDI footprints prevent the model from capturing finer-scale details within each pixel.

One potential limitation of the model is that the estimated AGBD mean (**Figure 5**) 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 (NDBI, NDVI, and MCARI1) 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 and Soromessa, 2018; Rodríguez-Lizana et al., 2023). At finer spatial scales, biomass and residue production have been shown to vary significantly depending on canopy cover (Rodríguez-Lizana et al., 2023). This variability may also be associated with other factors influencing biomass, such as differences in irrigation practices and soil composition, particularly soil organic carbon (SOC) and clay content, can also impact biomass estimates, highlighting the need for more comprehensive datasets in future research. 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 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 and Nilsson, 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 . 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. Additionally, ALOS PALSAR has proven to be a reliable predictor of biomass in complex forest structures. However, in highly structured plantation frameworks with relatively low canopies, its performance is suboptimal. Therefore, the use of Sentinel-1 (C-band) data may offer a promising alternative for future approaches, given its higher spatial resolution and sensitivity to finer structural variations (Adrah et al., 2025).

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 height and diameter at tree breast. The exclusion of certain environmental variables, such as soil moisture and nutrient availability, further contributes to prediction uncertainty. Nevertheless, the most critical factor for achieving accurate results is the availability of a robust and reliable training dataset. In this regard, GEDI has not proven to be a dependable source for estimating tree height in olive groves. Furthermore, the associated subproducts "canopy cover" and "tree density" also exhibited limited reliability, which ultimately constrained the model's performance and increased prediction uncertainty. The

model's performance, with R² values up to 0.56 and RMSE as low as 7.73 Mg·ha⁻¹, is comparable to previous studies in forestry and agricultural contexts, indicating its suitability for large-scale applications (Indirabai and Nilsson, 2024; Shendryk, 2022). However, it exhibits substantial limitations at fine spatial scales, primarily due to the generalized nature of the training dataset and the low reliability of the derived subproducts. These weaknesses are clearly reflected in the poor performance of the model when compared to the PNOA-referenced validation dataset.

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.

Moreover, to improve the accuracy of biomass estimation models in olive groves, it is essential to avoid using subproducts derived from intermediate estimations, as they introduce considerable uncertainty that ultimately degrades model performance. This study demonstrates that training models with the most robust and direct input data is a critical factor, particularly given the poor reliability observed in canopy height, canopy cover, and tree density subproducts derived from GEDI volume-based biomass modelling approach. For future research, the use of artificial intelligence techniques applied to high-resolution optical imagery, such as that provided by the Spanish PNOA program, may offer an effective solution for tree segmentation, thereby eliminating the need for GEDI-based height estimates and avoiding the introduction of uncertainty from derived products. This would result in a more reliable training dataset and significantly improve model accuracy. In addition, integrating alternative remote sensing sources such as Sentinel-1 and Sentinel-2 (Adrah et al., 2025) could enhance the spatial resolution of the final product and better capture fine-scale structural variations for agricultural monitoring.

5 Conclusion

This study demonstrates the potential of integrating multi-source remote sensing data to estimate AGBD in Mediterranean olive orchards. By combining GEDI LiDAR-derived metrics with optical, SAR, and topographic variables, the model achieved a robust predictive performance when evaluated against the internal validation dataset derived from the training process. However, its performance was considerably weaker when validated against the independent reference dataset, indicating limitations in the generalizability of the model at fine spatial scales, with the fully multi-source model obtaining the highest accuracy. Among the remote sensing variables, NDBI, slope, and MCARI1 emerged as the most influential predictors, highlighting the relevance of both vegetation structure and environmental conditions for biomass estimation.

The volumetric approach based on GEDI L2A height metrics, along with the estimated subproducts 'Total Canopy Cover' and 'Tree Density,' revealed certain limitations due to the low accuracy of these variables for subsequent training with Random

Forest. While the volumetric method has shown reasonable precision for individual tree estimations, its application at the grid level significantly amplifies uncertainty, highlighting a major weakness of the GEDI-based approach in this context. Besides, the model presents notable limitations when applied to small-scale crop monitoring, which must be addressed before using it for purposes such as multi-temporal evaluations or assessing crop productivity at finer spatial resolutions. However, despite these shortcomings, the RMSE obtained by the fully multi-source model remains relatively acceptable for large-scale applications, allowing for approximate assessment of biomass stocks at national and regional levels.

The model's large-scale applicability, facilitated by GEE, 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 implementing advanced artificial intelligence techniques to enhance the quality and robustness of training datasets. Additionally, the inclusion of diverse variables such as climatic conditions, soil properties, and irrigation regimes could significantly improve model accuracy. Expanding the training dataset to encompass a broader range of olive-growing regions would further strengthen the model's generalization and transferability.

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.

Code availability

Code is available upon request from the corresponding author.

Author contributions.

FC: methodology, data processing, visualization, results analysis, and writing. MC: conceptualization, work coordination, paper review, and editing. MASM: conceptualization, work coordination, paper review, and editing. PPC: conceptualization, methodology, results analysis, writing.

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

Adrah, E., Wong, J. P., and Yin, H.: Integrating GEDI, Sentinel-2, and Sentinel-1 imagery for tree crops mapping, Remote Sens Environ, 319, 114644, https://doi.org/10.1016/j.rse.2025.114644, 2025.

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.

Dubayah, R., Hofton, M., Blair, J., Armston, J., Tang, H., and Luthcke, S.: GEDI L2A Elevation and Height Metrics Data Global Footprint Level V002 [Data set], 2021.

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., and Hermosilla, T.: Estimation of biomass and volume of shrub vegetation using LiDAR and spectral data in a Mediterranean environment, Biomass Bioenergy, 46, 710–721, https://doi.org/10.1016/j.biombioe.2012.06.023, 2012.

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.

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

760   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
765   aboveground biomass in Northern China, Carbon Balance Manag, 17, https://doi.org/10.1186/s13021-022-00212-y, 2022.

Jucker, T., Fischer, F. J., Chave, J., Coomes, D. A., Caspersen, J., Ali, A., Loubota Panzou, G. J., Feldpausch, T. R., Falster, D., Usoltsev, V. A., Adu-Bredu, S., Alves, L. F., Aminpour, M., Angoboy, I. B., Anten, N. P. R., Antin, C., Askari, Y., Muñoz, R., Ayyappan, N., Balvanera, P., Banin, L., Barbier, N., Battles, J. J., Beeckman, H., Bocko, Y. E., Bond-
770   Lamberty, B., Bongers, F., Bowers, S., Brade, T., van Breugel, M., Chantrain, A., Chaudhary, R., Dai, J., Dalponte, M., Dimobe, K., Domec, J., Doucet, J., Duursma, R. A., Enríquez, M., van Ewijk, K. Y., Farfán-Rios, W., Fayolle, A., Forni, E., Forrester, D. I., Gilani, H., Godlee, J. L., Gourlet-Fleury, S., Haeni, M., Hall, J. S., He, J., Hemp, A., Hernández-Stefanoni, J. L., Higgins, S. I., Holdaway, R. J., Hussain, K., Hutley, L. B., Ichie, T., Iida, Y., Jiang, H., Joshi, P. R., Kaboli, H., Larsary, M. K., Kenzo, T., Kloeppel, B. D., Kohyama, T., Kunwar, S., Kuyah, S., Kvasnica,
775   J., Lin, S., Lines, E. R., Liu, H., Lorimer, C., Loumeto, J., Malhi, Y., Marshall, P. L., Mattsson, E., Matula, R., Meave, J. A., Mensah, S., Mi, X., Momo, S., Moncrieff, G. R., Mora, F., Nissanka, S. P., O'Hara, K. L., Pearce, S., Pelissier, R., Peri, P. L., Ploton, P., Poorter, L., Pour, M. J., Pourbabaei, H., Dupuy-Rada, J. M., Ribeiro, S. C., Ryan, C., Sanaei, A., Sanger, J., Schlund, M., Sellan, G., et al.: Tallo: A global tree allometry and crown architecture database, Glob Chang Biol, 28, 5254–5268, https://doi.org/10.1111/gcb.16302, 2022.

780   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.

785   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,
790   11, 11, https://doi.org/10.3390/f11010011, 2019.

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.

795 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.

800 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
805 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.

810 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,
815 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.

Qin, S., Wang, H., Rogers, C., Bermúdez, J., Lourenço, R. B., Zhang, J., Li, X., Chau, J., Tompalski, P., and Gonsamo, A.: Aboveground biomass mapping of Canada with SAR and optical satellite observations aided by active learning,
820 ISPRS Journal of Photogrammetry and Remote Sensing, 226, 204–220, https://doi.org/10.1016/j.isprsjprs.2025.05.022, 2025.

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.

825 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.

830 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.

Sinha, S., Jeganathan, C., Sharma, L. K., and Nathawat, M. S.: A review of radar remote sensing for biomass estimation, International Journal of Environmental Science and Technology, 12, 1779–1792, https://doi.org/10.1007/s13762-
835 015-0750-0, 2015.

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
840 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.

Velázquez-Martí, B., Cortés, I. L., and Salazar-Hernández, 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.

845