Universidad de Valladolid

## Campus de Palencia

ESCUELA TÉCNICA SUPERIOR

DE INGENIERÍAS AGRARIAS

## Máster en Ingeniería de Montes

# Análisis del secuestro de carbono orgánico del suelo en masas de *Pinus halepensis* Mill. empleando tecnología LiDAR y métodos de inventario clásicos


Estudiante: David Moreno Pérez

Tutora: Celia Herrero de Aza

Cotutor: Frederico Tupinambá Simões


Junio de 2025

Universidad de Valladolid

Campus de Palencia

ESCUELA TÉCNICA SUPERIOR

DE INGENIERÍAS AGRARIAS

Máster en Ingeniería de Montes

# Analysis of soil organic carbon sequestration in *Pinus halepensis* Mill. stands using LiDAR technology and traditional inventory methods

Student: David Moreno Pérez

Supervisor: Celia Herrero de Aza

Co-supervisor: Frederico Tupinambá Simões

June 2025

# INDEX

# 1. ABSTRACT & RESUMEN

## 1.1. Resumen

La estimación precisa del carbono orgánico del suelo (COS) en ecosistemas forestales es esencial para cuantificar su contribución como sumideros de carbono y mejorar las estrategias de gestión frente al cambio climático. El objetivo de este estudio fue modelizar el COS en masas de *Pinus halepensis* Mill. mediante métricas estructurales derivadas de datos LiDAR procedentes del Plan Nacional de Ortofotografía Aérea (PNOA).

El área de estudio comprendió una superficie de 46,8 hectáreas ubicada en el término municipal de Ampudia, Palencia (España). Para el desarrollo del trabajo se llevó a cabo un muestreo sistemático de suelos y un inventario forestal. Asimismo, se aplicó tecnología LiDAR y se obtuvieron 87 métricas estructurales. Estas métricas se integraron con variables edáficas y datos de biomasa aérea para construir modelos predictivos del stock de carbono mediante técnicas de regresión multivariante.

Entre los modelos evaluados, el algoritmo Random Forest mostró el mejor rendimiento en validación cruzada ($R^2$ = 0,81; RMSE = 7,73 Mg/ha), evidenciando una adecuada capacidad predictiva frente a otros modelos. El enfoque propuesto permitió evaluar el potencial de los datos LiDAR procedentes de escaneo láser aerotransportado (ALS), adquiridos en el marco de programas cartográficos de carácter general, como herramienta eficaz para la estimación espacial del COS. Este procedimiento, validado sobre una base empírica, ofrece una base metodológica útil para avanzar en la estimación del COS mediante teledetección, contribuyendo a mejorar la cuantificación de los servicios ecosistémicos relacionados con el suelo.

**Palabras clave: COS, Teledetección, Gestión forestal sostenible, Modelos predictivos, Biomasa forestal.**

## 1.2. Abstract

Accurate estimation of soil organic carbon (SOC) in forest ecosystems is essential for quantifying their contribution as carbon sinks and improving management strategies in the face of climate change. The objective of this study was to model SOC in *Pinus halepensis* Mill. stands using structural metrics derived from LiDAR data from the National Aerial Orthophotography Plan (PNOA).

The study area covered 46.8 hectares located in the municipality of Ampudia, Palencia (Spain). To carry out the work, systematic soil sampling and a forest inventory were conducted. LiDAR technology was also applied and 87 structural metrics were obtained. These metrics were integrated with edaphic variables and above-ground biomass data to build predictive models of carbon stock using multivariate regression techniques.

Among the models evaluated, the Random Forest algorithm showed the best performance in cross-validation ($R^2 = 0.81$; RMSE = 7.73 Mg/ha), demonstrating adequate predictive capacity compared to other models. The proposed approach made it possible to evaluate the potential of LiDAR data from airborne laser scanning (ALS), acquired within the framework of general mapping programmes, as an effective tool for the spatial estimation of SOC. This procedure, validated on an empirical basis, provides a useful methodological basis for advancing in the estimation of SOC through remote sensing, contributing to improve the quantification of soil-related ecosystem services.

**Keywords: SOC, Remote sensing, Sustainable forest management, Predictive modelling, Forest biomass.**

# 2. INTRODUCTION

## 2.1. Soil monitoring act

Soil monitoring has become a central focus of the European Union environmental strategies, with the aim of ensuring the sustainability of terrestrial ecosystems and mitigating the effects of climate change (Panagos et al., 2020). Quantifying soil carbon is a basic prerequisite for assessing the role of forest ecosystems in mitigating climate change, particularly in the context of climate policies such as the EU Soil Protection Strategy for 2030 (European Commission, 2021). The recent legislative proposal for the Soil Monitoring Act, presented by the European Commission on 5 July 2023, establishes a comprehensive framework for the assessment, conservation and sustainable management of soils in all Member States, with the aim of achieving the soil status by 2050 (European Comission, 2023). The carbon market is an important tool for meeting global climate goals in the short and medium term by promoting emissions reductions, carbon offsets, and investment in technologies to reduce emissions (Jiao et al., 2023). However, accurate estimation of soil carbon remains a challenge due to the high spatial and temporal heterogeneity, as well as the variability introduced by different measurement methods (Will, 2017).

In order to ensure transparency and scientific credibility in these mechanisms, especially in land-based carbon initiatives, it is necessary to establish harmonized monitoring systems. In this line, current European legislation proposes the implementation of a harmonised soil monitoring system, with standardised methodologies for assessing soil quality and ecological functionality (European Comission, 2023). In this policy and scientific framework the present study gets relevance, contributing with empirical evidence on soil carbon estimation under Mediterranean forest conditions. This law represents a socioeconomic opportunity for the agroforestry sector as a whole.

## 2.2. Importance of soil organic carbon (SOC) fixation in forest ecosystems

Soil is an essential component in the provision of ecosystem services, playing an important role in soil-atmosphere interactions, facilitating basic processes such as photosynthesis, improving the water cycle and contributing carbon fixation (Silva & Lambers, 2020; Oishy et al., 2025). SOC fixation and accumulation result from the balance between organic matter input (mainly from litter, fine roots and root exudates) and microbial decomposition levels, which are conditioned by soil, climate and forest management factors (Lal, 2005; Schmidt et al., 2011).

The capture of $CO_2$ by forest ecosystems through photosynthesis is essential both for net primary production and for reducing the effects of climate change (Navarro Cerrillo et al., 2018). Numerous authors have studied global forest carbon stocks. So, Pan et al. (2011) estimated that global forest carbon stocks were 861 Pg C, with soil up to 1 m deep being the main reserve (44%), followed by biomass (42%), dead wood (8%) and litter (5%). Carbon fixation is distributed throughout the different ecosystems (Mayer et al., 2020) (Figure 1). There is marked variability in the vertical distribution of organic carbon, with greater accumulation in soils of humid boreal ones in contrast to tropical biotypes,

where the dominant fraction is found in above-ground biomass. This heterogeneity highlights the importance of considering the edaphic component in the global carbon balance, particularly in regions where soil is the main reservoir of organic carbon.

The magnitude of carbon stocks in forest soil is affected by the interaction of factors that regulate site fertility and productivity, such as climate, vegetation, topography, chemical, physical and biological properties of the soil, and parent material, in addition to land use and management practices (Mayer et al., 2020). Many researchers have highlighted that soil carbon stocks can be influenced by factors such as low water retention capacity, the presence of coarse materials and limited effective soil depth, reinforcing the need to implement forest management strategies adapted to these conditions (Ruiz-Peinado et al., 2017).

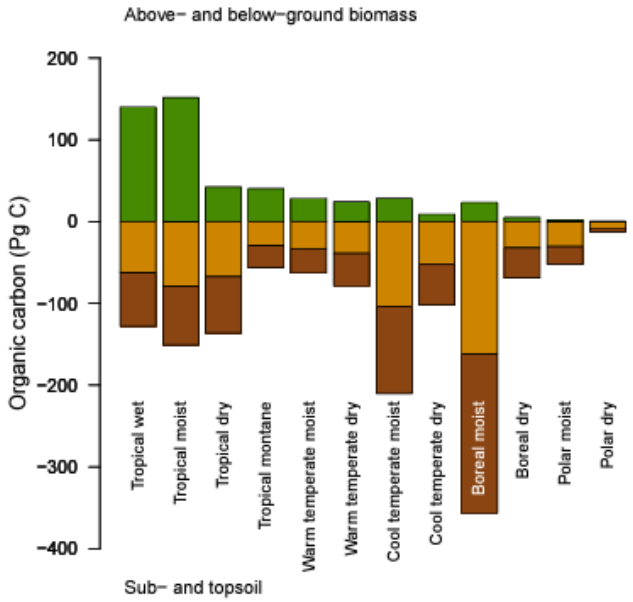


*Figure 1. Organic carbon stocks in the subsoil (brown) and topsoil (orange), and in above-ground and below-ground biomass (green). Figure from Mayer et al. (2020) modified from Scharlemann et al. (2014).*

Therefore, it is increasingly important to establish assessment parameters to estimate soil health, both at the local level, enabling farmers and forest owners to analyse the impact of their management practices, and at the national level, so that countries can provide more information on the state of their soil resources (Rabot et al., 2024). In this context, soil organic carbon (SOC) is one of the most important soil elements, as it directly influences plant growth and essential nutrient cycles (Navarrete-Poyatos et al., 2019).

In the Mediterranean context, where soil resilience to degradation processes is particularly limited, the role of SOC becomes even more important. Its conservation contributes to regulating the carbon balance and maintaining ecosystem functions such as fertility, water retention and soil structural stability (Muñoz-Rojas et al., 2015). In this type of ecosystem, the capacity of forest soils to store carbon is particularly relevant,

given that these are highly vulnerable to aridity, climate variability and historical land use degradation (Chevallier et al., 2016). Consequently, an accurate and particularly explicit estimate of SOC content is important for understanding the underlying ecological processes and designing effective forest policies aimed at sustainable management and adaptation to climate change.

### 2.3. Relevance of *Pinus halepensis* Mill. in carbon storage in the Mediterranean region

*Pinus halepensis*, commonly known as Aleppo pine, is a species native to the Mediterranean region, extending from the Iberian Peninsula to the eastern Mediterranean, including countries such as Spain, France, Italy, Greece and North Africa. Its distribution covers approximately 3.5 million hectares, concentrated mainly in the western Mediterranean basin, especially in Spain and southern France (Figure 2).

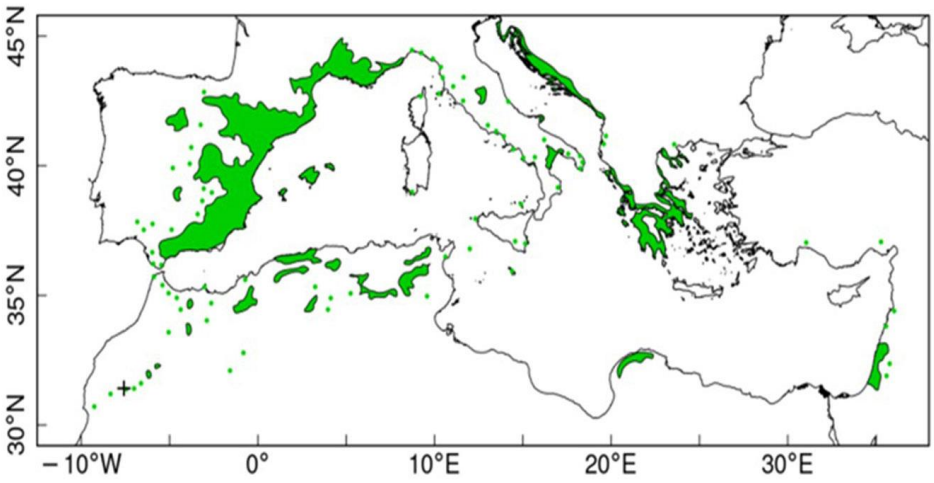


*Figure 2. Native range of Pinus halepensis Mill. in the Mediterranean region. Adapted from Vieira et al. (2022), as cited in Alsanousi et al. (2025).*

Climate change is altering the dynamics of Mediterranean forests, affecting their carbon storage capacity and increasing the risk of desertification (del Río et al., 2008).

Its adaptability to semi-arid conditions and poor soils allows it to colonise degraded areas, contributing to the restoration of disturbed environments and the provision of ecosystem services (Derak & Cortina, 2014; Alsanousi et al., 2025). Thanks to its high ecological plasticity, this species acts as an early coloniser in disturbed ecosystems, being able to regenerate efficiently on bare soils with relatively poor hydrological conditions (Serrada et al., 2008). In turn, *Pinus halepensis* has proven to be an efficient carbon sink (Ruíz-Navarro et al., 2009; López-Senespleda et al., 2021; Santonja et al., 2022). The conversion of agricultural land to forest through reforestation with *Pinus halepensis* can generate significant carbon sequestration over several decades (Charro et al., 2008). The implementation of silvicultural practices, such as systematic clear-cutting and selective

thinning, have been shown to influence growth, carbon storage and sequestration in Aleppo pine reforestations in central Spain (De las Heras et al., 2012; Lull et al., 2024).

### 2.4. **LiDAR as a tool for predicting SOC in forest ecosystems**

LiDAR (Light Detection and Ranging) is an active laser technology widely used in forest studies (Dassot et al., 2011; Borsah et al., 2023) due to its ability to accurately characterise canopy structure and estimate variables such as biomass or tree height (Oehmcke et al., 2021). LiDAR technology allows us to characterize the three-dimensional characterisation of forest canopy by emitting and recording laser pulses, generating metrics that describe the vertical distribution of vegetation, canopy density, average or maximum tree height, terrain roughness and other attributes associated with biomass (Lefsky et al., 2002; Tupinambá-Simões et al., 2025). Currently, advances in remote sensing techniques have allowed field inventory to be replaced and/or complemented by airborne laser scanning (ALS) (Navarro Cerrillo et al., 2018). Although significant progress has been made in estimating global soil carbon stocks, the high spatial and temporal variability of carbon reserves hinders uniform and accurate assessment (Wang et al., 2021; Usman & Begum., 2023; Aroca-Fernandez et al., 2025). In this context, ALS datasets from initiatives such as the National Orthophotography Plan (PNOA) offers an innovative opportunity to improve traditional estimates of soil carbon storage. LiDAR technology allows us to obtain forest structure, reducing uncertainty in the quantification of SOC.

Numerous authors have worked on modelling forest structural variables using ALS LiDAR data and algorithms such as k-nearest neighbours (KNN) or Random Forest (RF) (Yavari & Sohrabi, 2019; Adhikari et al., 2023; Pereira et al., 2023; Strunk & McGaughey, 2023). However, there are not many studies about modelling SOC using LiDAR metrics. Rasel et al. (2017) estimated variables obtained using ALS LiDAR (elevation, forest type and above-ground biomass) to carry out modelling that predicted SOC reserves using the Random Forest (RF) algorithm. Navarro Cerrillo et al. (2018) designed a methodology to facilitate silvicultural decision-making in forest management by estimating SOC and other variables such as above-ground biomass in stands of *Pinus halepensis* in southern Spain using the KNN algorithm. Moreno Muñoz et al. (2024) developed a model using the RF algorithm to estimate organic carbon storage of mangrove ecosystems in the Colombia southern Pacific coast, demonstrating the potential of machine learning techniques to predict edaphic variables in tropical coastal environments.

Moreover, several studies in soil have demonstrated high accuracy in estimating soil properties, such as SOC and total nitrogen (TN), using high-resolution digital elevation models (DEMs) specifically generated by LiDAR technology (Zhou et al., 2020; Zou et al., 2024). In addition, environmental variables derived from site variables have been used to improve predictions (Mendes & Sommer, 2023). Spatial and temporal dynamics of carbon have been analyzed at regional level, such as Farina et al. (2017), who proposed studies based on biophysical models (RothC10N) linked to GIS combined with a database of soil, land use and climate, or Reddy et al. (2015), who carried out an estimate of the

extent of soil carbon loss after a forest fire and uncertainty. On the other hand, high-resolution multispectral sensors have already been put into orbit, such as Sentinel-2 (S2), whose ability to quantify SOC content is comparable to that of future hyperspectral space sensors (Castaldi et al., 2019). Topography-based Relief Prediction and Classification Systems (RPCS) offer advantages when simulating soil redistribution and associated SOC dynamics. Topographic information can be easily obtained from DEMs. The recent increase in the accessibility of high-spatial-resolution LiDAR data can help improve the accuracy of landscape topography derived from DEMs and benefit research in regions with data use constraints. Topography-based models can effectively quantify soil redistribution and SOC distribution patterns (Li & McCarty, 2018). Therefore, one of the main challenges in determining SOC distribution is significant spatial variability, along with the uncertainty it generates (Will, 2017).

This context highlights the importance of continuing to optimise methodologies for quantifying and monitoring SOC, which underlines the relevance of this study in the use of advanced technologies such as LiDAR to improve the accuracy of carbon estimation in forest soils. Integrating LiDAR-derived structural data with soil and biomass variables requires addressing issues such as scale mismatches, spatial resolution, and model generalisability, particularly in heterogeneous Mediterranean landscapes. The findings of this study may inform forest managers and policymakers in designing evidence-based strategies for soil carbon conservation and climate mitigation under Mediterranean conditions.

### 2.5. **Objectives of the study**

The main objective of this study is to model the SOC content in *Pinus halepensis* forests by integrating structural variables derived from airborne LiDAR data with soil and forest inventory data. The aim is to generate SOC estimates applicable to forest management in Mediterranean environments, contributing to the development of more accurate and efficient methodologies for soil carbon monitoring. This objective aligns with the EU goalof achieving reliable SOC stimates, a key indicator for assessing forest health.

# 3. MATERIALS AND METHODS

## 3.1. Description of the study area

The study area is located in Ampudia, Palencia (Spain) in the north of the Iberian Peninsula (41º 51'48'' N; 4º 46'13'' W) (Figure 3). The area, situated at 860 meters above sea level (masl), shows a sub-Mediterranean continental climate (Dsb) (AEMET, 2024). The average temperature in the study area is 11.3 ºC and the average annual rainfall is 393 mm for the time series between 1995 and 2024 (Figure 4). The present study is sited in *Pinus halepensis* afforested stand (60 years old), with sporadic presence of *Quercus faginea* Lam., which is sited in the lower stratum and clearly subordinate in terms of coverage and dominance. Table 1 shows the main dendrometric and forest structural variables of the study area (Bueis Mellado, 2017).

*Table 1. Main dendrometric and dasometric variables obtained through the forest inventory*

| N (trees/ha) | Ht (m) | Ho (m) | Dbh (cm) | G ($m^2$/ha) |
|---|---|---|---|---|
| 845 | 8.5 | 10.12 | 17.73 | 23.20 |

*Note: N = number of trees per hectare; Ht = average total height; Ho = dominant height; Dbh= diameter at breast height; G = basal area

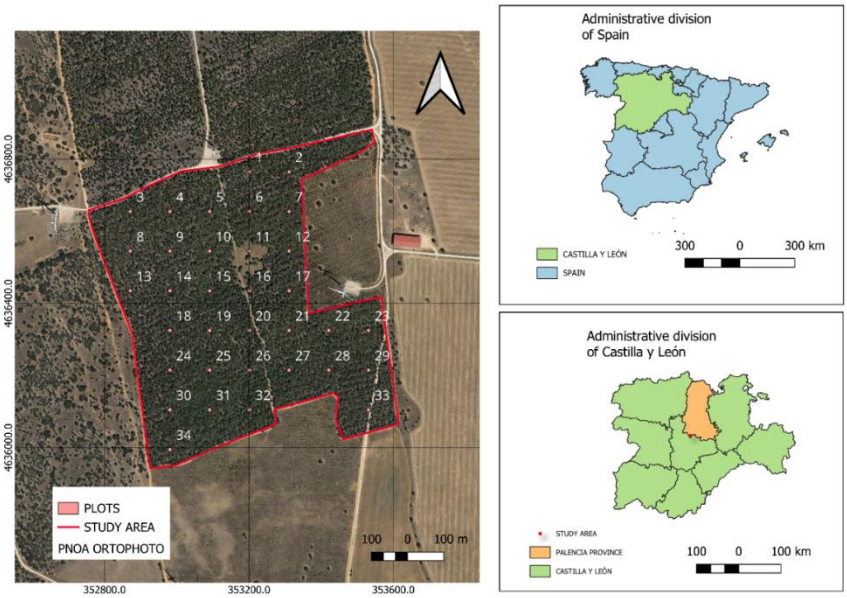


*Figure 3. Location and spatial distribution of the study plots within the experimental study area. Note: the plots are overlaid on a PNOA orthophoto using UTM Zone 30N projection and ETRS89 coordinate system.*

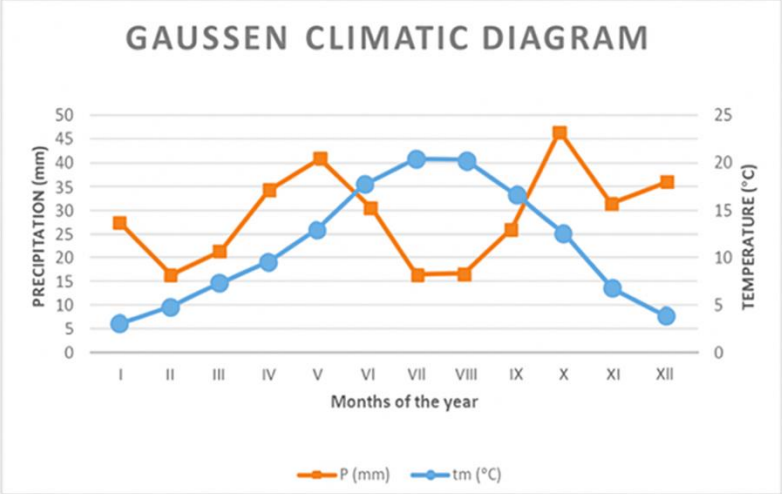


*Figure 4. Gaussen climate diagram of the study area by time series 1995-2024*

### 3.2. **General methodology**

Different data sources (soil data, forest inventory data and LiDAR metrics) were used to model SOC (Figure 5). Soil sampling data parameters were determinated to estimate soil carbon stock. The forest inventory provided dendrometric variables to estimate tree biomass and C using allometric equations. Finally, processed LiDAR dataset from the PNOA were used to determine structural stand variables.

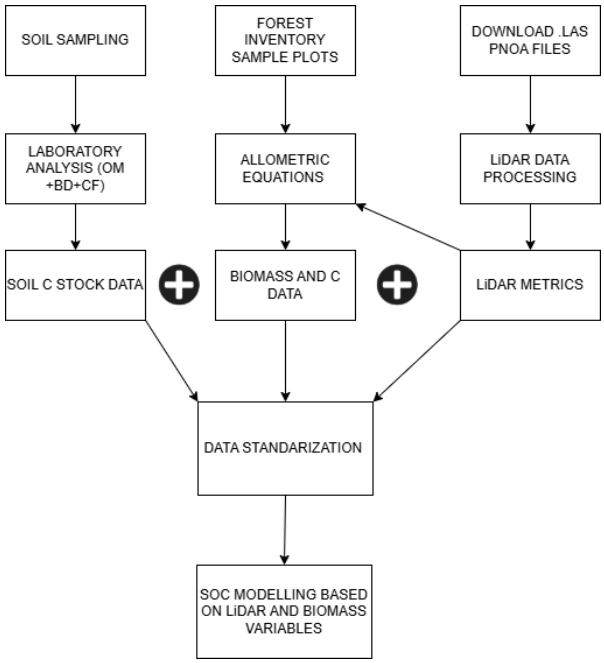


*Figure 5. Road map of the methodology used in this study*

### 3.3. **Soil sampling and laboratory analyses**

34 plots were established to carry out the soil sampling. A systematic sampling design was carried out in the whole study area (46.8 ha). All plots were georeferenced using high-precision sub-meter GPS, ensuring the replicability of the sampling design and enabling its accurate integration with the other dataset (forest and LiDAR metrics).

Mineral soil samples were taken in each plot at depth of 0-10 cm (Figure 6). Each plot was divided into four parts, from 6 to 8 individual randomized samples were taken and composited into one homogenized sample per plot. Samples were transported to the laboratory. Mineral soil samples were air-dried, sieved and the percentage of coarse soil materials ([Ø > 2 mm]) was calculated. In each sample, bulk density was measured through the core method (Blake & Hartge, 1986) in the field with volumetric steel rings and soil dry weight.

Soil organic matter was determinated by Walkley and Black method (de Vos et al., 2007), based on the partial oxidation of organic carbon with potassium dichromate in an acidic medium. After the reaction, the excess dichromate not consumed was titrated with ferrous ammonium sulphate (Mohr's salt) in the presence of diphenylamine as an indicator. The percentage of easily oxidizable organic carbon was calculated from the volume of this compound used and the dry weight of the sample. To determine the total carbon in the soil, the total organic matter was divided by 1.724 (MAPA, 1994) [Equation 1].

$$SOC = \frac{SOM}{1.724} \qquad \text{Equation 1}$$

Accumulated $SOC_i$ (Mg C $ha^{-1}$) [Equation 2] was calculated considering C concentration, bulk density, thickness and the percentage of gravels (MAPA, 1994) using the following equation (Lee et al., 2009):

$$SOCi = OCi * BD * (1 - CFi) * ti * 100 \qquad \text{Equation 2}$$

*Note: $SOC_i$ is the stock of C (Mg C $ha^{-1}$) at depth i; $OC_i$ is the organic C content of the fine soil fraction at depth I; $BD_i$ is the bulk density at depth i ($Mg/m^3$); $CF_i$ is the volumetric content of the coarse fraction at depth i (%) and t is the horizon thickness (m).

All the soil analysis was carried out at the soil laboratory of the ETSIIAA (University of Valladolid). For more information on laboratory analyses, see Appendix 2.

*Figure 6. Example of sampling soil samples*

### 3.4. **Forest inventory and generation of equations for quantifying biomass and Carbon**

Diameter at breast height (Dbh, with CODIMEX L manual caliper) and total height (Ht, with high-precision Laser GEO Vertex) were recorded in 250 selected trees (Dissanayake, 2024) to develop a site-specific height – diameter equation. Multiple functional forms were tested (linear, exponential, logarithmic, polynomial and potential models) using SAS software with the aim of identifying the best-fit relationship between Dbh and Ht for the study area (SAS Institute Inc., 2023). The selected equation was subsequently used to estimate tree biomass using species-specific allometric models for *Pinus halepensis*, established by the Spanish National Forest Inventory (IFN4). So, the following tree biomass components were obtained: Ws: stem with bark (commercial volume, up to a top diameter of 7 cm), Wmb: medium branches (diameter between 2 and 7 cm), Wthinb: thin branches (diameter smaller than 2 cm) and Wr: coarse roots (Table 2). Aboveground biomass (Wa) was defined as the sum of the aboveground biomass fractions of all live trees (Ws, Wmb, Wthinb). Tree carbon (C) fixation was determined by multiplying each biomass value by a generic C concentration of 50.0 %, according to Kollmann (1959) and the Intergovernmental Panel on Climate Change (IPCC) recommendations (Penman et al. 2003).

*Table 2. Equations for estimating biomass for Pinus halepensis. Source: IFN4*

| Variable | Equation |
|---|---|
| **Ws** | $Ws = 0.0139 * d^2 * h$ |
| **Wr** | $Wr = (0.0785 * d^2)$ |
| **Wb2-7** | $Wb2-7 = 4.257 + 0.00506 * d^2 * h - 0.0722 * d * h$ |
| **Wb0,5-2** | $Wb0,5-2 = 6.197 + 0.00932 * h^2 * 1 - 0.0686 * h$ |

*Note: Ws = stem biomass; Wr = root biomass; Wb 2-7 = biomass of branches from 2 cm to 7 cm; Wb 0.5-2 = biomass of branches from 0.5 cm to 2 cm

### 3.5. **ALS data from the PNOA and data processing**

For the ALS data, Leica ALS80 was used as the main sensor. Technical specifications of the data, obtained in 2019, are presented in Table 3. The flight over the study area took place in 2019 (Figure 7).

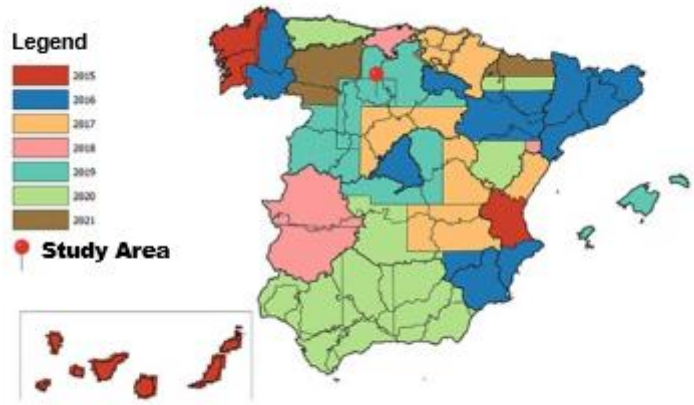


*Figure 7. Year of commencement of the second cycle of the PNOA-LiDAR project by zone. Source: PNOA*

*Table 3. Technical specifications of the second coverage of the PNOA – LiDAR project. Source: PNOA*

| | |
|---|---|
| **Minimum point density** | 0.5 – 2 points/$m^2$ |
| **Year of flight** | 2019 |
| **Geodetic reference system** | ETRS89 huso 30 N |
| **RMSE Z** | ≤ 20 cm |
| **Estimated planimetric accuracy** | ≤ 30 cm |
| **Simultaneous image** | Si |
| **File size** | 2 x 2 km |
| **File format** | LAS 1.2 format 3 |
| **MDE grid spacing** | 2 m x 2 m |
| **RMSE Z (MDE)** | ≤ 25 cm |
| **Estimated planimetric accuracy (MDE)** | ≤ 50 cm |

*Note: RMSE Z = Root mean square error in the Z component; MDE = Digital Elevation Model

LiDAR data was downloaded from the National Geographic Information Centre portal in .LAS format, selecting the files corresponding to the map sheets that completely covered the study area. The .LAS files were then spatially cropped using a previously digitized vector layer that precisely delimited the perimeter of the analyzed area. The LiDAR data was processed at the plot scale, addressing each sampling unit individually. To do this, the R programming environment (version 4.4.1) was used (R Core Team, 2024), employing a combination of packages specialised in geospatial analysis and point cloud processing, such as lidR, terra and sf (Pebesma, 2016; Hijmans, 2025). The lidR package provides a wide range of functions for manipulating and visualising airborne LiDAR data, including reading and writing .LAS and .LAZ files, point classification, digital terrain and canopy model generation, height normalisation and individual tree segmentation (Roussel et al., 2020; Peter et al., 2021).

First, the files were read to extract a point cloud that included both the file header information (coordinate reference system and format type) and the position data and attributes associated with each point in the cloud. Specific attributes were then selected, such as X, Y, and Z coordinates and intensity, in order to optimize memory management and speed up processing. In addition, specific attributes and filters were determined to eliminate non-essential data, such as points that did not correspond to the first return, which increased the efficiency of the analysis by reducing the amount of data to be processed. Next, a filter was applied to extract vegetation points, considering that points corresponding to the terrain had heights of less than 2 m. The Digital Terrain Model (DTM) and the Canopy Height Model (CHM) were generated for the entire study area. Subsequently, the structural metrics of the vegetation were calculated, such as the average height, maximum height and minimum height in meters. Based on the CHM, the height peaks representing tree crowns were identified, and the CHM was segmented to delimit or identify individual trees. To generate the DTM, only points classified as terrain were used, applying the Inverse Distance Weighting (IDW) method. This approach assigns values to unsampled locations using a weighted average of neighboring points, with weights inversely proportional to the distance raised to a specified power, commonly used in geospatial studies to interpolate continuous surfaces (Shepard, 1968; Moussa & Abboud, 2024). In this study, interpolation parameters were defined with k=10 nearest neighbours and a power p=2. Subsequently, the LiDAR point cloud was normalized by adjusting the heights of non-terrestrial points in relation to the DTM, allowing for a

representation of the vertical structure of the forest canopy. To generate the CHM, the 'pit-free' algorithm was applied, designed to avoid artificial depressions in canopy models by combining multiple surfaces generated at different height thresholds (Khosravipour et al., 2014). This CHM was used to identify and characterize individual tree crowns, extracting metrics such as the maximum height of each tree.

A conceptual map of LiDAR data processing is presented (Figure 8).

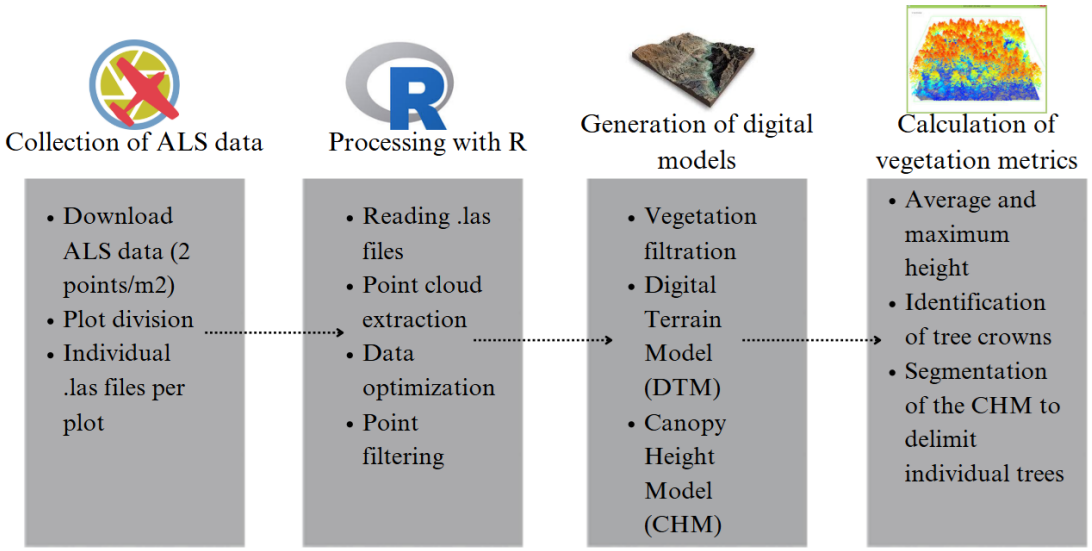


*Figure 8. Concept map of LiDAR data processing*

Once the point cloud was processed, multiple structural metrics derived from LiDAR returns at the plot level were calculated using the lidar_metrics function from the lidR package (Roussel & Auty, 2022). The variables extracted included descriptive height statistics such as total number of returns, maximum height, minimum height, mean height, standard deviation, coefficient of variation, skewness, and kurtosis (n, zmax, zmin, zmean, zsd, zcv, zskew, zkurt). Height percentiles were also calculated in increments from 1 to 99% (zq1 to zq99), proportions of returns above certain thresholds (e.g., 2 m, 5 m, zmean), as well as cumulative metrics and proportions within specific height ranges. Likewise, indices derived from L moments (L1 to L4, Lskew, Lkurt), metrics based on the leaf area density profile (lad_min, lad_max, lad_mean, lad_sum, lad_cv) and statistics associated with the position of the return within the pulse (n_first, n_last, n_single, n_multiple, p_first, p_last, etc.). All metrics were calculated individually for each plot and normalised for subsequent use as predictors in the SOC estimation models. For a detailed description of each variable, see Appendix 4.

### 3.6. **SOC modelling and generation of the predicted SOC map**

The modeling approach consisted of applying different models and machine learning algoritms, with the aim of evaluating the predictive capacity of SOC through soil data, tree biomass variables and LiDAR metrics. So, firstly, Lidar data was processed. Predictor variables with a variance of less than 0.01 were removed as they were considered non-informative. The collinearity between the variables was examined by calculating the Pearson correlation matrix, removing those with correlation coefficients greater than 0.9 in order to reduce redundancy and prevent multicollinearity problems. The remaining variables were normalized through standardization (mean zero and standard deviation one).

Secondly, the processed dataset was randomly divided into two subsets: a training subset comprising 75% of the observations and a test subset comprising the remaining 25%. This division was performed in a stratified and reproducible manner, using a fixed random seed. Four regression models with different levels of complexity were defined and implemented: a simple linear model, a second-degree polynomial model, a logarithmic model (log(x+1)) and a Random Forest model. All models were validated using a cross-validation scheme with 10 partitions (k-fold cross validation, v=10) on the training set. In the case of the Random Forest model, hyperparameter optimization was carried out using a random search with 20 different combinations of the parameters mtry (number of predictors considered in each tree division) and min_n (minimum number of observations in a terminal leaf). The performance of each model was evaluated using the statistical parameters root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination ($R^2$).

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \qquad \text{Equation 3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - x_i)^2}{n}} \qquad \text{Equation 4}$$

$$R^2 = 1 - \frac{\sum (y_i - x_i)^2}{\sum (y_i - \mu_y)} \qquad \text{Equation 5}$$

The statistical parameters obtained for each model allowed us to identify the best alternative in terms of predictive accuracy.

To generate a spatially continuous map of predicted SOC, a wall-to-wall rasterization approach was implemented using R (version 4.4.1) and the terra package (Hijmans, 2025). First, the complete dataset of predictor variables was structured as a spatial point

object using the corresponding UTM coordinates (EPSG:25830). The area of interest was rasterized at a spatial resolution of 20 × 20 meters, consistent with the spatial density of the LiDAR-derived metrics (Trouvé et al., 2023). A multi-band SpatRaster template was constructed, covering the full extent of the study area. The predictor variables used in the selected model were then rasterized layer by layer over this template. Once rasterized, the trained selected model was applied to the full stack of predictor layers using the predict() function, yielding a continuous surface of SOC values in Mg C/ha. The resulting raster was clipped using a shapefile defining the limits of the forest stand. The file containing the SOC information for each pixel was exported in GeoTIFF format for visualisation and subsequent spatial analysis. Following the modelling and prediction procedure, the resulting SOC map was laid out using QGIS (version 3.40). For a detailed description of the SOC modelling script and the resulting raster data, is shown in Appendix 5.

# 4. RESULTS

## 4.1. Results of soil data

Carbon concentration ranged from 0.63% to 5.08%, while TOM ranged between 1.08% and 8.74%. Higher SOC values (Mg C/ ha) were associated with higher concentrations of organic carbon (OC%) and total organic matter (TOM%).

Bulk density (BD) ranged between 1.16 and 1.85 g / cm³, while coarse fraction percentage (CF%) showed a broader dispersion, from low values (<1%) in several plots to higher proportions (e.g., 30% in plot 22). The results of the edaphic properties analysed (Table 4) showed that SOC (Mg C / ha) varied among plots, ranging from 9.49 Mg C $ha^{-1}$ (plot 28) to 53.16 Mg C $ha^{-1}$ (plot 2).

*Table 4. Laboratory results of soil properties per plot: soil organic carbon (SOC0–10), total organic carbon (TOC), total organic matter (TOM), bulk density (BD) and percentage of coarse elements.*

| Plot | $SOC_{0-10}$ (Mg C / ha) | OC (%) | TOM (%) | BD (g/cm3) | CF (%) |
|---|---|---|---|---|---|
| **1** | 14.60 | 1.36 | 2.34 | 1.24 | 13.64 |
| **2** | 53.16 | 5.08 | 8.74 | 1.21 | 13.67 |
| **3** | 16.75 | 1.31 | 2.25 | 1.56 | 17.72 |
| **4** | 38.48 | 1.97 | 3.39 | 1.99 | 1.58 |
| **5** | 19.20 | 2.31 | 3.97 | 1.05 | 20.68 |
| **6** | 19.83 | 1.36 | 2.34 | 1.48 | 1.52 |
| **7** | 18.83 | 1.51 | 2.60 | 1.57 | 20.50 |
| **8** | 17.21 | 1.02 | 1.76 | 1.69 | 0.51 |
| **9** | 16.04 | 1.01 | 1.74 | 1.59 | 0.60 |
| **10** | 20.71 | 1.25 | 2.15 | 1.66 | 0.29 |
| **11** | 18.13 | 1.46 | 2.51 | 1.32 | 5.63 |
| **12** | 30.82 | 1.69 | 2.91 | 1.85 | 1.59 |
| **13** | 18.59 | 1.21 | 2.08 | 1.54 | 0.42 |
| **14** | 18.70 | 1.37 | 2.36 | 1.36 | 0.22 |
| **15** | 14.50 | 0.80 | 1.38 | 1.82 | 0.16 |
| **16** | 47.25 | 3.51 | 6.03 | 1.52 | 11.41 |
| **17** | 15.96 | 1.02 | 1.76 | 1.56 | 0.22 |
| **18** | 12.43 | 0.91 | 1.57 | 1.57 | 13.26 |
| **19** | 22.15 | 1.68 | 2.90 | 1.45 | 9.15 |
| **20** | 20.15 | 1.44 | 2.48 | 1.45 | 3.39 |
| **21** | 22.62 | 1.45 | 2.50 | 1.58 | 1.64 |
| **22** | 13.72 | 1.49 | 2.56 | 1.39 | 33.68 |
| **23** | 25.67 | 1.56 | 2.69 | 1.67 | 1.33 |

*Table 4. Cont. Laboratory results of soil properties per plot: soil organic carbon (SOC0–10), total organic carbon (TOC), total organic matter (TOM), bulk density (BD) and percentage of coarse elements.*

| Plot | $SOC_{0-10}$ | OC (%) | TOM (%) | BD (g/cm3) | CF (%) |
|---|---|---|---|---|---|
| **24** | 26.08 | 1.81 | 3.11 | 1.51 | 4.37 |
| **25** | 15.38 | 1.03 | 1.76 | 1.53 | 1.92 |
| **26** | 17.02 | 1.32 | 2.27 | 1.44 | 10.50 |
| **27** | 14.53 | 0.94 | 1.62 | 1.57 | 1.67 |
| **28** | 9.49 | 0.63 | 1.08 | 1.51 | 0.51 |
| **29** | 19.11 | 1.69 | 2.90 | 1.29 | 12.07 |
| **30** | 11.20 | 0.71 | 1.22 | 1.60 | 1.60 |
| **31** | 17.48 | 1.23 | 2.12 | 1.52 | 6.55 |
| **32** | 15.41 | 1.64 | 2.81 | 1.16 | 18.83 |
| **33** | 18.44 | 1.45 | 2.49 | 1.45 | 11.96 |
| **34** | 11.58 | 0.78 | 1.35 | 1.52 | 3.12 |

*Note: $SOC_{0-10}$ = Carbon Stock from 0-10 cm; OC = organic carbon content of the fine soil fraction (< 2 mm); TOM = Total Organic Matter; BD = bulk density; CF = volumetric content of the coarse fraction.

*Table 5. Descriptive statistics of soil variables analyzed in the laboratory (n = 34)*

| | Mean | Standard deviation | Minimum | Maximum | Coefficient of variation (%) |
|---|---|---|---|---|---|
| **$SOC_{0-10}$** | 20.33 | 9.44 | 9.49 | 53.16 | 46.46 |
| **OC (%)** | 1.50 | 0.82 | 0.63 | 5.08 | 54.78 |
| **TOM (%)** | 2.58 | 1.41 | 1.08 | 8.74 | 54.74 |
| **BD (g/cm3)** | 1.51 | 0.19 | 1.05 | 1.99 | 12.63 |
| **CF (%)** | 7.23 | 8.11 | 0.16 | 33.68 | 112.14 |

The average SOC was 20.33 Mg/ha (Table 5), ranging from 9.49 to 53.16 Mg/ha. A similar trend was found in total organic matter values, with an average value of 2.58% and a coefficient of variation of 54.78%. The bulk density ranged from 1.05 to 1.99 g/cm³. In contrast, the percentage of gravels showed the greatest dispersion of all the variables analyzed (CV = 112.10%).

### 4.2. **Results of tree biomass estimation**

For the h-d model, the linear equation was selected as the best (Table 6) showing a higher $R^2$ (0.546) (Figure 9).

*Table 6. Height diameter models fitted with forest inventory data.*

| Model | Equation | $R^2$ |
|---|---|---|
| Linear | $y = 0.2731x + 4.9865$ | 0.546 |
| Exponential | $y = 6.4408\ e^{0.0233x}$ | 0.532 |
| Logarithmic | $y = 6.6774\ln(x) - 9.4012$ | 0.519 |
| Polynomial | $y = -0.001x^2 + 0.3299x + 4.2507$ | 0.535 |
| Potential | $y = 1.837x^{0.5778}$ | 0.542 |

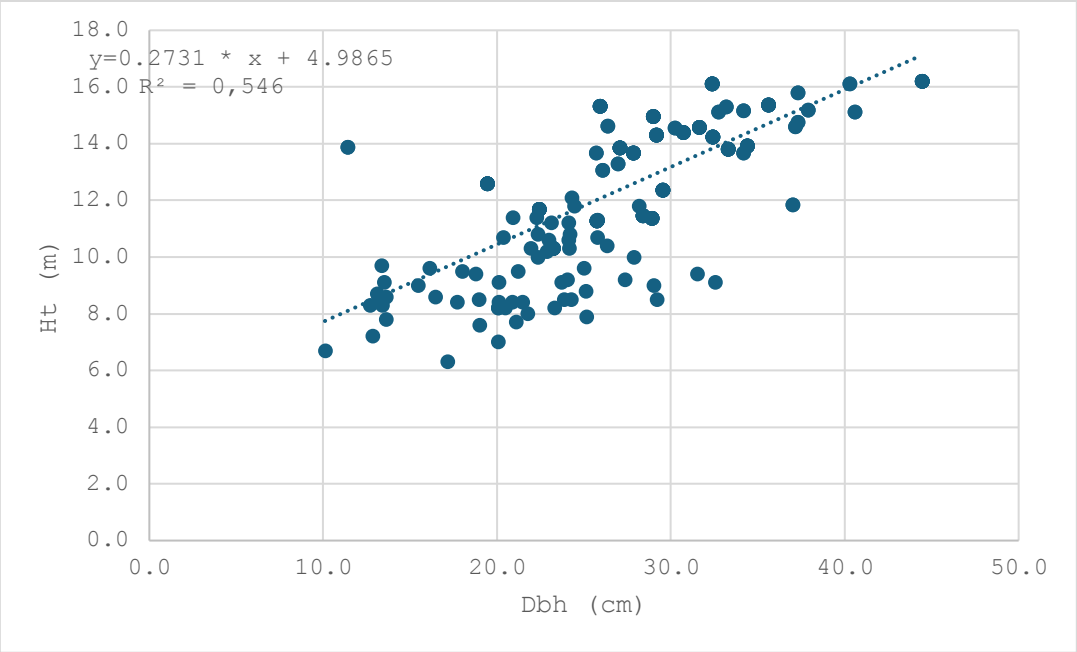


*Figure 9. Fit between height and diameter of measured trees using the linear model.*

$$Ht = 0.2731 * Dbh + 4.9865 \qquad \text{Equation 6}$$

The results of biomass estimation in all plots are shown in Appendix 3 (Table 8).

Stem biomass (Ws) represented the dominant fraction of aboveground biomass across the plots, accounting for an average of 57.23% of the total. It was followed by thick branches (Wb2–7), contributing 27.42%, while small branches (Wb0.5–2) accounted for 15.35% of the aboveground biomass. Root biomass (Wr), as the belowground compartment, accounted for the total Carbon tree biomass.

### 4.3. SOC modelling based on LiDAR and biomass metrics

Ten-fold cross-validation on the training set showed differential performance between models, with higher performance in the case of Random Forest model (Table 7), far exceeding the results of the parametric models. Given the performance achieved during cross-validation, the Random Forest model was selected as the optimal candidate for estimating the SOC. Table 7 summarizes the mean values of each metric for the four models evaluated during cross-validation. Figure 10 shows the relationship between observed and predicted SOC values, showing that predictions ranged between 15 and 30 Mg C/ha.

*Table 7. Average performance metrics obtained through cross-validation (10-fold) for soil organic carbon (SOC) estimation models*

| Model | RMSE | MAE | $R^2$ |
|---|---|---|---|
| **Random Forest** | 7.73 | 6.13 | 0.811 |
| **Logarithmic** | 331 | 264 | 0.871 |
| **Polynomial** | 745 | 455 | 0.794 |
| **Linear** | 661 | 498 | 0.721 |

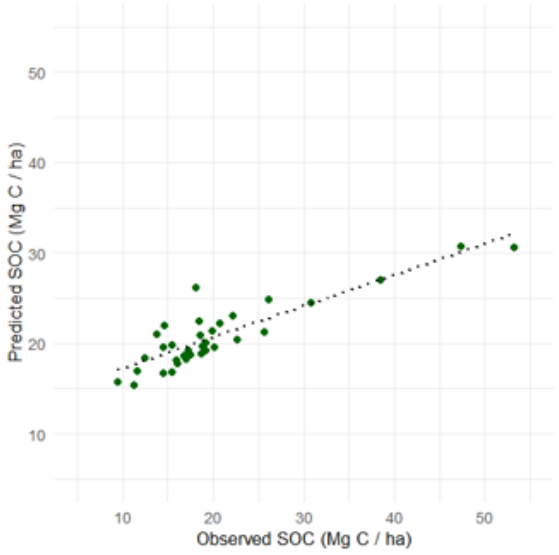


*Figure 10. Random Forest model performance: Predicted vs Observed Soil Organic Carbon (SOC) after cross-validation*

### 4.4. **Mapping of the SOC**

SOC content was spatially distributed (Figure 11) and most values ranged between 16.2 and 23.5 Mg C/ha, with a remarkable spatial variability throughout the forest stand.

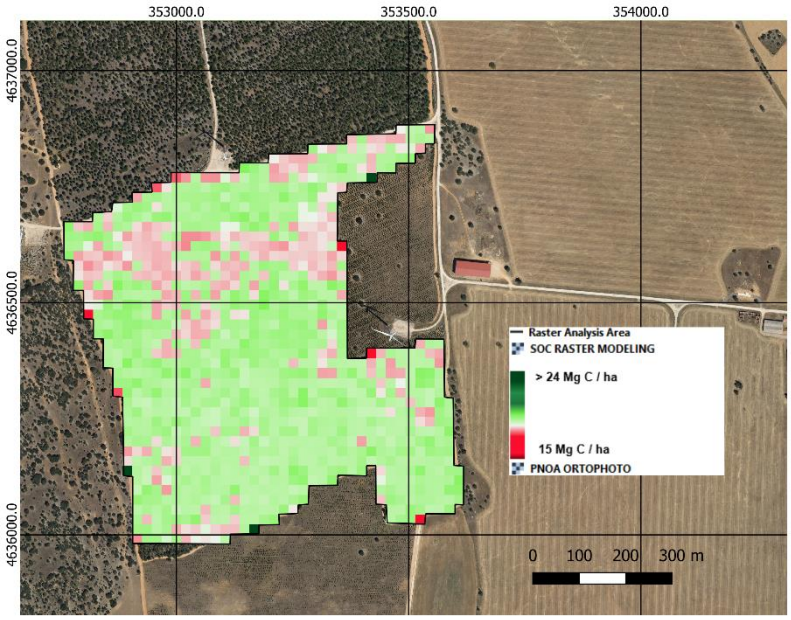


*Figure 11. Predicted spatial distribution of soil organic carbon (Mg C/ha) using the Random Forest model.*

# 5. DISCUSSION

This study integrated structural data obtained from soil, tree biomass and airborne laser scanning (ALS) to estimate SOC content in a Mediterranean *Pinus halepensis* ecosystem. This approach was based on (a) soil carbon stock data, (b) auxiliary biomass variables calculated from forest inventory (c) structural metrics derived from LiDAR point clouds processed at the plot level, and (d) modelling and machine learning techniques to identify relationships between forest structure and soil carbon contents. The results allowed us to generate relevant predictors of SOC, even under conditions of high vegetation cover, demonstrating the robustness of the modelling framework to operate under structurally heterogeneous forest environments. Our findings emphasize the potential of our approach to be scaled to other semi-arid or Mediterranean ecosystems where field sampling is limited, but airborne LiDAR data is available. Our approach could be applied across broader spatial domains, aligning with the current efforts to develop harmonized soil carbon monitoring strategies at international level.

## 5.1. LiDAR ALS for estimating the SOC

In this study, the application of airborne LiDAR (ALS) data has proven to be effective in capturing structural metrics of forest canopy, providing valuable proxies for estimating SOC in Mediterranean ecosystems dominated by *Pinus halepensis*. Studies using ALS sensors to estimate forest biomass are common (Breidenbach et al., 2021; Perea-Ardila et al., 2021; Li et al., 2024) and different authors carried out studies of SOC modeling based on satellite images such as Sentinel (Zhou et al., 2020; Amarnath et al., 2024), however there is a lack of studies using ALS sensors to estimate SOC. Our findings showed that metrics derived from ALS can correlate significantly with soil carbon stocks, especially when they were integrated with forest inventory data.

The upper soil layer (0-10 cm) usually concentrates most of the active organic matter and is particularly influenced by processes of carbon accumulation (Hoyle et al., 2011). Previous studies have shown estimations similar to our findings. So, Rasel et al. (2017) fitted a model that estimated 69% of SOC based on biomass variables obtained through ALS sensors and Stumpf et al. (2024), were able to estimate up to 64% of the carbon in the top 1.2 m of soil. Moreover, Navarro-Cerrillo et al. (2018) found that combining low-density ALS data with nearest neighbor (kNN) models showed 82% of accuracy in carbon stocks in the top 10 cm of soil in *Pinus halepensis* plantations in southeastern Spain. Even Navarrete-Poyatos et al. (2019) identified a significant improvement in the prediction of SOC using ALS data in combination with machine learning models such as Random Forest. Finally, Pascual et al. (2023) addressed the quantification of stored carbon by integrating LiDAR metrics with estimatations of above-ground biomass and litter, highlighting the relevance of incorporating multiple ecosystem pools, looking for a more complete assessment of carbon content.

Other studies highlighted the usefulness of LiDAR ALS in predicting soil properties. For example, Li et al. (2016) evaluated the effectiveness of LiDAR-derived variables for estimating surface soil horizon properties in a *Pinus koraiensis* forest and showed that LiDAR-derived variables could be predictors of soil properties, with coefficients of

determination ($R^2$) ranging from 0.46 to 0.66. In addition, Hounkpatin et al. (2021) compared global and local models for predicting SOC stocks in Swedish forests, using national forest inventory data and digital soil mapping approaches. Their findings suggested that local calibration has the potential to obtain higher accuracy. Other studies carried out a complex methodology that combined field data with metrics derived from LiDAR ALS to estimate changes in SOC stocks at the stand level, using the Yasso15 model to simulate edaphic carbon dynamics (Strîmbu et al., 2023). In the present study, the Random Forest algorithm was applied to model SOC from LiDAR metrics obtained from the PNOA. Unlike previous studies that rely on dense sampling networks, detailed spectral data, or mechanistic models like Yasso15, our approach showed that it is possible to achieve accurate SOC predictions using solely airborne LiDAR metrics derived from public datasets. This adds value by confirming the operational potential of low-cost, replicable methodologies based on open-access remote sensing data and machine learning techniques, especially in Mediterranean forest systems where field data are often highly heterogeneous.

### 5.2. Evaluation of the predictive model performance

Despite not having an independent test set evaluation due to sample effort limitations, the robustness of the model in cross-validation allowed these results to be considered as representative of the expected behavior of the model under similar conditions. The Random Forest (RF) model applied in this study showed the best performance among the evaluated approaches for SOC estimation, achieving a coefficient of determination of 0.811 during cross-validation. Beyond its predictive accuracy, this result reinforces the capacity of non-parametric methods to capture nonlinear relationships and interactions between structural LiDAR metrics and soil carbon content. The robustness of RF in handling high-dimensional datasets with potential multicollinearity makes it particularly suitable for heterogeneous Mediterranean forest systems. The associated errors were relatively low, with a mean RMSE of 7.73 Mg/ha and a MAE of 6.13 Mg/ha. In comparison, parametric models (linear, polynomial and logarithmic) had higher errors and lower explanatory power, which limited their predictive utility. These results are consistent with other studies that showed more accurate estimates from machine learning techniques for estimating soil properties by LiDAR data (Navarro Cerrillo et al., 2018; Alonso-Sarria et al., 2025). Rasel et al. (2017) highlighted that RF is particularly effective in capturing nonlinear relationships and managing highly correlated variables, thus outperforming linear regression-based approaches when working with structural proxies derived from remote sensing. Similarly, Misebo et al. (2024) used generalized additive models to estimate SOC in areas restored after mining and obtained significantly lower $R^2$ values, demonstrating the advantages of the approach used in this study. However, the absence of certain key edaphic variables should be considered when extrapolating the results to more complex or heterogeneous forest scenarios.

The robustness of the RF model agrees with the findings of Hu et al. (2023), who applied remote sensing-based models and ecological simulations in fire-affected boreal forests, finding that SOC depended on both metrics quality and the suitability of predictive methods. In their study, the combination of field data with remote information allowed

models to be calibrated with coefficients of determination greater than 0.80, even under severe disturbance conditions. In addition, Hengl et al. (2017) confirmed the widespread use of Random Forest as a generic framework for modeling complex spatial variables, highlighting its usefulness in predicting edaphic carbon with multisensor data. The performance observed in the present study is within the range of accuracy reported in these studies, reinforcing the applicability of the model in Mediterranean forest contexts with moderate-resolution ALS data coverage.

A tendency of the Random Forest model to underestimate SOC values was observed, particularly in plots with higher stocks values. This systematic underestimation is consistent with findings from previous research. For instance, Agaba (2024) reported that Random Forest models underestimated SOC in heterogeneous mountainous landscapes, especially at high altitudes and in areas with steep slopes. They attributed this pattern to insufficient sampling density in complex terrain and the limited representation of extreme SOC values in the training data. Similarly, Ou et al. (2024) demonstrated that the inclusion of site attributes and climatic variables significantly improved SOC prediction in cropland soils, highlighting that the absence of such covariates can reduce model performance. Future research could explore whether the inclusion of site or climatic covariates—absent in the present model—might help reduce the underestimation observed in certain SOC predictions. Similarly, increasing the number of field plots could contribute to improving model robustness and reducing potential bias, particularly in heterogeneous forest environments.

### 5.3. Relationship between forest structure and SOC

Structural parameters such as tree density, tree size, species diversity and aboveground biomass can estimate both atmospheric carbon fluxes and carbon incorporation and stabilization in the soil through the contribution of litter, roots and exudates (Muñoz-Rojas et al., 2016). In *Pinus halepensis* forests, the relationship between forest structure and SOC is particularly relevant, due to their distribution in many areas of semi-arid Mediterranean ecosystems. However, several studies have shown that *P. halepensis*, a species widely used in reforestation in the Mediterranean area plantations do not always fully restore original SOC levels after disturbances such as fires or previous degradation (Goberna et al., 2007). These results suggest that, although the above-ground biomass of *P. halepensis* may be considerable, its effect on soil carbon reservoirs could be limited by factors such as the poor quality of the organic matter contributed or accelerated decomposition rates under arid conditions (Lull et al., 2024). This pattern is consistent with the results obtained in the present study, where despite observing plots with high values of biomass and aboveground carbon, no direct and consistent relationship was identified between these values and SOC content. This suggests that carbon input from vegetation does not necessarily translate into greater edaphic carbon fixation, reinforcing the idea that other factors such as microbial dynamics, textural characteristics or coarse element content may modulate SOC retention and stabilization in the surface soil horizon (Doetterl et al., 2025).

The marked heterogeneity in CF values across plots suggested significant differences in the content of rock fragments or lithology within the surface soil profile, which may have important implications for water storage capacity, nutrient availability and root development of vegetation. The results obtained showed variability in biomass values between plots associated with differences in stand density, stand structure and forest management intensity. These results highlighted that the structural development plays a key role in aboveground biomass accumulation. Plots with greater structural complexity—such as plots 12, and 9—showed the highest aboveground biomass values, exceeding 80 Mg/ha in the most developed cases, reflecting advanced stand maturity and significant carbon storage capacity. In contrast, plot 11 with no tree vegetation, pointing to early successional conditions or perhaps disturbance history, showed limited structural development (Hernández-Alonso et al., 2023).

### 5.4. **Projections and scalability of the model for monitoring SOC**

The integration of remote sensing techniques such as ALS with machine learning-based prediction models has proven to be an effective strategy for estimating SOC, not only at the local scale, but also with potential for scalability to regional or international levels. In this study, a set of structural metrics derived from low-density ALS data and forest inventory was used to model the SOC content in a Mediterranean forest ecosystem with a predominance of *Pinus halepensis*. The results obtained, especially with the Random Forest model ($R^2 = 0.811$; RMSE = 7.73 Mg/ha), demonstrated the ability of this type of variable to capture soil patterns with a reasonable level of error.

This approach agrees with studies such as that by Stevens et al. (2013), who integrated site variables derived from digital terrain models with field data to generate SOC predictions in agricultural soils. Moreover, Brogniez et al. (2015), also used generalized additive models based on the Land Use/Cover Area frame statistical Survey (LUCAS) database to map SOC in the surface horizon at a continental scale. Both studies highlighted the usefulness of multivariate approaches that combine environmental variables derived from remote sensors with limited but representative edaphic information. Our work contributes to this line of research, demonstrating that, even in scenarios with a small number of plots, the structural information obtained by ALS can be sufficient to generate robust SOC maps at an operational scale.

In addition, Panagos et al. (2022) highlighted the importance of infrastructures such as the European Soil Data Centre (ESDAC) for harmonizing data and facilitating its use in soil management policies and climate strategies. Similarly, Ballabio et al. (2016) demonstrated how harmonized databases such as LUCAS can be used to map high-resolution physical soil properties using non-linear modelling techniques such as adaptive regression splines (MARS). In the present study, although with a more restricted scope, similar methodological procedures based on multivariate regression and integration of topographic and structural variables derived from LiDAR were used to estimate SOC content on a smaller scale. In this context, the results of our study could generate future carbon mapping schemes on a larger scale, contributing replicable methodologies based on open data and interpretable models. This possibility is particularly relevant in the

context of European climate neutrality objectives, where the accurate quantification of carbon in forest soils is a priority (Panagos et al., 2020).

### 5.5. Study limitations and methodological recommendations

Estimating SOC using metrics derived from LiDAR and forest inventory data offers an innovative and non-destructive approach to assessing carbon reservoirs in forest ecosystems. However, this study has certain limitations that should be considered when interpreting the results and planning future research. One limitation is related to the resolution and coverage of the LiDAR data used, which were obtained from the PNOA. Although these data are publicly available and cover large areas, their point density and temporal resolution may be limited, which affects the accuracy of the detailed characterization of forest structure and, consequently, the estimation of SOC (Johnson et al., 2022). This restriction has direct implications for the generalizability of the predictive models developed. Furthermore, although relevant variables such as bulk density and percentage of coarse elements were incorporated, other important edaphic factors such as soil texture, pH or indicators of biological activity were not included, variables that in other studies have been shown to be determinants for the accuracy of SOC estimation (Muñoz-Rojas et al., 2015).

In order to improve future methodological approaches, the integration of multiple data sources is recommended. Combining LiDAR metrics with information obtained from hyperspectral images or radar data would allow for the capture of a wider range of variables related to both vegetation and soil properties, thus strengthening the robustness of the models (Tafur et al., 2022). It is also advisable to increase the size of the field sample, increasing the number of plots and their spatial distribution, to better represent the environmental heterogeneity of the study area and reduce biases in modelling. The incorporation of additional soil variables, such as soil texture, nutrient content or microbial biomass, could also enrich the understanding of the processes that control carbon storage in the soil. Another important limitation is the spatial and temporal scale of the study. Although the approach applied has shown good results in a relatively homogeneous area, its applicability to more heterogeneous landscapes with marked variations in topography, land use or climatic conditions may be compromised. Furthermore, the use of point data limits the possibility of analyzing the evolution of SOC over time. The integration of multi-temporal or repeated data would not only improve the accuracy of the models but also allow the stability of edaphic carbon in the face of disturbances to be assessed (Guillaume et al., 2021). Finally, it is important to implement rigorous cross-validation procedures and use independent test sets to assess the stability and generalizability of predictive models under different forest scenarios.

## 6. CONCLUSIONS

This study confirms the potential of airborne LiDAR data, combined with estimates of aboveground and belowground biomass and soil data, for predicting SOC content in *Pinus halepensis* forest ecosystems. The Random Forest model showed the best predictive performance, achieving a coefficient of determination ($R^2$) of 0.811 and moderate mean errors (RMSE = 7.73 Mg/ha; MAE = 6.13 Mg/ha) during cross-validation, clearly outperforming the parametric models used. The integration of structural metrics derived from publicly available LiDAR data allowed the spatial variability of SOC to be considered efficiently, without the need for intensive field sampling. These results are consistent with previous studies that support the use of machine learning models in the prediction of complex soil properties, especially in systems with high structural heterogeneity. However, limitations related to LiDAR data resolution and the absence of certain key soil variables that could improve our results.

Our results reinforce the usefulness of remote sensing and machine learning tools for estimating carbon stocks in forest soils, providing a methodological approach to support sustainable forest management and environmental monitoring strategies in Mediterranean contexts.

## 7. ACKNOWLEDGEMENTS

This study was made possible thanks to funding from the IMFLEX project (PID2021-1262750B-C22) funded by Agencia estatal de investigación, fondos FEDER, Micinn. Ministerio de Ciencia e innovación. Plan de recuperación, transformación y resiliencia, Unión Europea.

I would like to thank my tutors, Celia and Frederico, for their invaluable help and advice, as well as my colleagues, especially Rubén de Prado, for his help in collecting field samples, and Marina Ortiz and Elisa Pérez for their assistance in the laboratory.

Of course, I would also like to thank my family, especially my girlfriend, for their constant support over the last few months.

## 8. REFERENCES

Adhikari, A., Montes, C. R., & Peduzzi, A. (2023). A Comparison of Modeling Methods for Predicting Forest Attributes Using Lidar Metrics. *Remote Sensing 2023, Vol. 15, Page 1284*, *15*(5), 1284. https://doi.org/10.3390/RS15051284

Agaba, S. (2024). *Mapping Soil Organic Carbon using different machine learning models as an application of Digital Soil Mapping*. https://boa.unimib.it/handle/10281/512359

Alonso-Sarria, F., Blanco-Bernardeau, A., Gomariz-Castillo, F., Jiménez-Bastida, H., & Romero-Diaz, A. (2025). Estimation of soil properties using machine learning techniques to improve hydrological modeling in a semiarid environment: Campo de Cartagena (Spain). *Earth Science Informatics 2025 18:3*, *18*(3), 1–24. https://doi.org/10.1007/S12145-025-01833-W

Alsanousi, A. A., Abdul-Hamid, H., Mohamed, J., & Masoud, M. (2025). *Pinus halepensis* Mill. in the Mediterranean region: a review of ecological significance, growth patterns, and soil interactions. *IForest - Biogeosciences and Forestry*, *18*(1), 30. https://doi.org/10.3832/IFOR4566-017

Amarnath, A., Dasar, G. V., & Gowda, G. B. (2024). Geospatial mapping of soil organic carbon in Sirsi Forst division using remote sensing techniques. *International Journal of Research in Agronomy*, *7*(9S), 878–884. https://doi.org/10.33545/2618060X.2024.V7.I9SL.1625

Aroca-Fernandez, J. M., Diez-Pastor, J. F., Latorre-Carmona, P., Elvira, V., Camps-Valls, G., Pascual, R., & Garcia-Osorio, C. (2025). *A Collaborative Platform for Soil Organic Carbon Inference Based on Spatiotemporal Remote Sensing Data*. https://arxiv.org/pdf/2504.13962

Ballabio, C., Panagos, P., & Monatanarella, L. (2016). Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma*, *261*, 110–123. https://doi.org/10.1016/J.GEODERMA.2015.07.006

Blake, G. R., & Hartge, K. H. (1986). Bulk density. En A. Klute (Ed.), *Methods of soil analysis: Part 1. Physical and mineralogical methods* (2nd ed., pp. 363–375). American Society of Agronomy and Soil Science Society of America. https://doi.org/10.2136/sssabookser5.1.2ed.c13

Borsah, A. A., Nazeer, M., & Wong, M. S. (2023). LIDAR-Based Forest Biomass Remote Sensing: A Review of Metrics, Methods, and Assessment Criteria for the Selection of Allometric Equations. *Forests 2023, Vol. 14, Page 2095*, *14*(10), 2095. https://doi.org/10.3390/F14102095

Breidenbach, J., Ivanovs, J., Kangas, A., Nord-Larsen, T., Nilsson, M., & Astrup, R. (2021). Improving living biomass c-stock loss estimates by combining optical satellite,

airborne laser scanning, and nfi data. *Canadian Journal of Forest Research*, *51*(999), 1–14. https://doi.org/10.1139/cjfr-2020-0518

Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., & van Wesemael, B. (2019). Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. *ISPRS Journal of Photogrammetry and Remote Sensing*, *147*, 267–282. https://doi.org/10.1016/j.isprsjprs.2018.11.026Get rights and content

Centro de Descargas del CNIG (IGN). (n.d.). Retrieved March 12, 2025, from https://centrodedescargas.cnig.es/CentroDescargas/lidar-segunda-cobertura

Centro Nacional de Información Geográfica (CNIG). (2020). Datos LiDAR del Plan Nacional de Ortofotografía Aérea (PNOA). *Ministerio de Transportes, Movilidad y Agenda Urbana*. https://centrodedescargas.cnig.es

Chevallier, T., Hamdi, S., Gallali, T., Brahim, N., Cardinael, R., Bounouara, Z., Cournac, L., Chenu, C., & Bernoux, M. (2016). Soil carbon as an indicator of Mediterranean soil quality. En *Soils and climate change: Challenges and perspectives* (pp. 627–636). IRD Éditions. https://books.openedition.org/irdeditions/24018?lang=en

Charro, E., & Gardini, A. M. (2008). Estimation of carbon sequestration by *Pinus halepensis* forest soils in Castilla y León (Spain).

Comisión Europea. (2021). *Estrategia de la UE para la protección del suelo para 2030: Aprovechar los beneficios de unos suelos sanos para las personas. los alimentos. la naturaleza y el clima* (COM (2021) 699 final). https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52021DC0699

Dassot, M., Constant, T., & Fournier, M. (2011). The use of terrestrial LiDAR technology in forest science: Application fields, benefits and challenges. *Annals of Forest Science*, *68*(5), 959–974. https://doi.org/10.1007/s13595-011-0102-2

de las Heras, J., Moya, D., López-Serrano, F. R., & Rubio, E. (2013). Carbon sequestration of naturally regenerated Aleppo pine stands in response to early thinning. *New Forests*, *44*(3), 457–470. https://doi.org/10.1007/s11056-012-9356-2

de los Bueis Mellado, T. (2017). *Relationships between the dynamics of Pinus halepensis Mill. and Pinus sylvestris L. plantations and environmental parameters: A basis for sustainable management of stands / Relaciones entre las dinámicas de las plantaciones de Pinus halepensis Mill. y Pinus sylvestris L. y factores del medio: Base para la gestión sostenible de las masas* [Tesis doctoral, Universidad de Valladolid].

del Río, M., Barbeito, I., Bravo-Oviedo, A., Calama, R., Cañellas, I., Herrero, C., & Bravo, F. (2008). *Carbon Sequestration in Mediterranean Pine Forests*. 221–245. https://doi.org/10.1007/978-1-4020-8343-3_13

Derak, M., & Cortina, J. (2014). Multi-criteria participative evaluation of *Pinus halepensis* plantations in a semiarid area of southeast Spain. *Ecological Indicators*, *43*, 56–68. https://doi.org/10.1016/J.ECOLIND.2014.02.017

de Vos, B., Lettens, S., Muys, B., & Deckers, J. A. (2007). Walkley-Black analysis of forest soil organic carbon: Recovery, limitations and uncertainty. *Soil Use and Management*, *23*(3), 221–229. https://doi.org/10.1111/j.1475-2743.2007.00084.x

Dissanayake, C. T. M. (2024). *LiDAR and photogrammetry for automated stem volume estimation at the single-tree level* [Master's thesis, Erasmus Mundus Master in Mediterranean Forestry and Natural Resources – MEDFOR]. Universidad de Valladolid & Università degli Studi di Padova.

Doetterl, S., Berhe, A. A., Heckman, K., Lawrence, C., Schnecker, J., Vargas, R., Vogel, C., & Wagai, R. (2025). A landscape-scale view of soil organic matter dynamics. *Nature Reviews Earth & Environment 2025 6:1*, *6*(1), 67–81. https://doi.org/10.1038/s43017-024-00621-2

European Commission. (2021). *EU Soil Strategy for 2030: Reaping the benefits of healthy soils for people, food, nature and climate* (COM(2021) 699 final). Brussels: European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021DC0699

European Commission. (2023). *Proposal for a Directive of the European Parliament and of the Council on Soil Monitoring and Resilience (Soil Monitoring Law)*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52023PC0416

Farina, R., Marchetti, A., Francaviglia, R., Napoli, R., & di Bene, C. (2017). Modeling regional soil C stocks and CO2 emissions under Mediterranean cropping systems and soil types. *Agriculture, Ecosystems & Environment*, *238*, 128–141. https://doi.org/10.1016/J.AGEE.2016.08.015

Goberna, M., Sánchez, J., Pascual, J. A., & García, C. (2007). *Pinus halepensis* Mill. plantations did not restore organic carbon, microbial biomass and activity levels in a semi-arid Mediterranean soil. *Applied Soil Ecology*, *36*(2–3), 107–115. https://doi.org/10.1016/J.APSOIL.2006.12.003

Guillaume, T., Bragazza, L., Levasseur, C., Libohova, Z., & Sinaj, S. (2021). Long-term soil organic carbon dynamics in temperate cropland-grassland systems. *Agriculture, Ecosystems & Environment*, *305*, 107184. https://doi.org/10.1016/J.AGEE.2020.107184

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, *2018*(8), e5518. https://doi.org/10.7717/PEERJ.5518/SUPP-1

Hernández-Alonso, H., Madrigal-González, J., Tornos-Estupiña, L., Santiago-Rodríguez, A., Alonso-Rojo, P., Morera-Beita, A., & Silla, F. (2023). Tree-size heterogeneity modulates the forest age-dependent carbon density in biomass and top soil stocks on Mediterranean woodlands. *Plant and Soil*, *486*(1–2), 361–373. https://doi.org/10.1007/s11104-023-05874-2

Hijmans, R. J. (2025). Spatial Data Analysis [R package terra version 1.8-50]. *CRAN: Contributed Packages*. https://doi.org/10.32614/CRAN.PACKAGE.TERRA

Hoyle, F. C., Baldock, J. A., Murphy, D. v, Hoyle, F. C., Baldock, J. A., & Murphy, D. v. (2011). Soil Organic Carbon – Role in Rainfed Farming Systems. *Rainfed Farming Systems*, 339–361. https://doi.org/10.1007/978-1-4020-9132-2_14

Hounkpatin, K. O. L., Stendahl, J., Lundblad, M., & Karltun, E. (2021). Predicting the spatial distribution of soil organic carbon stock in Swedish forests using a group of covariates and site-specific data. *SOIL*, *7*(2), 377–398. https://doi.org/10.5194/SOIL-7-377-2021,

Hu, T., Yu, C., Dou, X., Zhang, Y., Li, G., & Sun, L. (2023). Simulation of Soil Organic Carbon Dynamics in Postfire Boreal Forests of China by Incorporating High-Resolution Remote Sensing Data and Field Measurement. *Fire 2023, Vol. 6, Page 414*, *6*(11), 414. https://doi.org/10.3390/FIRE6110414

Jiao, Y., Wang, D., Yao, X., Wang, S., Chi, T., & Meng, Y. (2023). Forest Emissions Reduction Assessment Using Optical Satellite Imagery and Space LiDAR Fusion for Carbon Stock Estimation. *Remote Sensing 2023, Vol. 15, Page 1410*, *15*(5), 1410. https://doi.org/10.3390/RS15051410

Johnson, L. K., Mahoney, M. J., Bevilacqua, E., Stehman, S. v., Domke, G. M., & Beier, C. M. (2022). Fine-resolution landscape-scale biomass mapping using a spatiotemporal patchwork of LiDAR coverages. *International Journal of Applied Earth Observation and Geoinformation*, *114*, 103059. https://doi.org/10.1016/J.JAG.2022.103059

Khosravipour, A., Skidmore, A. K., Isenburg, M., Wang, T., & Hussin, Y. A. (2014). Generating Pit-free Canopy Height Models from Airborne Lidar. *Photogrammetric Engineering and Remote Sensing*, *80*(9), 863–872. https://doi.org/10.14358/PERS.80.9.863

Kollmann F (1959) Tecnología de la madera y sus aplicaciones. IFIE, Madrid.

Lal, R. (2005). Forest soils and carbon sequestration. *Forest Ecology and Management*, *220*(1–3), 242–258. https://doi.org/10.1016/J.FORECO.2005.08.015

Lee, J., Hopmans, J. W., Rolston, D. E., Baer, S. G., & Six, J. (2009). Determining soil carbon stock changes: Simple bulk density corrections fail. *Agriculture, Ecosystems & Environment*, *134*(3–4), 251–256. https://doi.org/10.1016/J.AGEE.2009.07.006

Lefsky, M. A., Cohen, W. B., Parker, G. G., & Harding, D. J. (2002). Lidar Remote Sensing for Ecosystem Studies: Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *BioScience*, *52*(1), 19–30. https://doi.org/10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO;2

Li, W., Hu, X., Su, Y., Tao, S., Ma, Q., & Guo, Q. (2024). A new method for voxel-based modelling of three-dimensional forest scenes with integration of terrestrial and airborne LiDAR data. *Methods in Ecology and Evolution*, *15*(3), 569–582. https://doi.org/10.1111/2041-210X.14290

Li, C., Xu, Y., Liu, Z., Tao, S., Li, F., & Fang, J. (2016). Estimation of Forest Topsoil Properties Using Airborne LiDAR-Derived Intensity and Topographic Factors. *Remote Sensing 2016, Vol. 8, Page 561*, 8(7), 561. https://doi.org/10.3390/RS8070561

López-Senespleda, E., Calama, R., & Ruiz-Peinado, R. (2021). Estimating forest floor carbon stocks in woodland formations in Spain. *Science of The Total Environment*, *788*, 147734. https://doi.org/10.1016/J.SCITOTENV.2021.147734

Lull, C., Gil-Ortiz, R., Bautista, I., del Campo, A., & Lidón, A. (2024). The Short-Term Effects of Heavy Thinning on Selected Soil Carbon Pools and Microbial Activity in a Young Aleppo Pine Forest. *Forests*, *15*(4), 658. https://doi.org/10.3390/f15040658

MAPA; 1994. Métodos oficiales de análisis. Ministerio de Agricultura, Pesca y Alimentación. 662 pp. Madrid.

Mapas climáticos de España (1981-2010) y ETo (1996-2016) - Agencia Estatal de Meteorología - AEMET. Gobierno de España. (2024).

Mayer, M., Prescott, C. E., Abaker, W. E. A., Augusto, L., Cécillon, L., Ferreira, G. W. D., James, J., Jandl, R., Katzensteiner, K., Laclau, J. P., Laganière, J., Nouvellon, Y., Paré, D., Stanturf, J. A., Vanguelova, E. I., & Vesterdal, L. (2020). Tamm Review: Influence of forest management activities on soil organic carbon stocks: A knowledge synthesis. *Forest Ecology and Management*, *466*, 118127. https://doi.org/10.1016/J.FORECO.2020.118127

Misebo, A. M., Hawryło, P., Szostak, M., & Pietrzykowski, M. (2024). Spatial estimation of soil organic carbon, total nitrogen, and soil water storage in reclaimed post-mining site based on remote sensing data. *Ecological Indicators*, *166*, 112228. https://doi.org/10.1016/J.ECOLIND.2024.112228

Moreno Muñoz, A. S., Guzmán Alvis, Á. I., & Benavides Martínez, I. F. (2024). A random forest model to predict soil organic carbon storage in mangroves from Southern *Colombian Pacific coast. Estuarine, Coastal and Shelf Science, 299*, 108674. https://doi.org/10.1016/J.ECSS.2024.108674

Moussa, H., & Abboud, M. (2024). The Methodology of Applying Inverse Distance Weighting Interpolation Method in Determining Normal Heights. *Resourceedings, 4*(1), 01–06. https://doi.org/10.21625/resourceedings.v4i1.1068

Muñoz-Rojas, M., Erickson, T. E., Martini, D. C., Dixon, K. W., & Merritt, D. J. (2016). Climate and soil factors influencing seedling recruitment of plant species used for dryland restoration. *SOIL*, *2*(2), 287–298. https://doi.org/10.5194/SOIL-2-287-2016

Muñoz-Rojas, M., Jordán, A., Zavala, L. M., de la Rosa, D., Abd-Elmabod, S. K., & Anaya-Romero, M. (2015). Impact of Land Use and Land Cover Changes on Organic Carbon Stocks in Mediterranean Soils (1956–2007). *Land Degradation & Development*, *26*(2), 168–179. https://doi.org/10.1002/LDR.2194

Navarrete-Poyatos, M. A., Navarro-Cerrillo, R. M., Lara-Gómez, M. A., Duque-Lazo, J., Varo, M. de los A., & Rodriguez, G. P. (2019). Assessment of the Carbon Stock in Pine Plantations in Southern Spain through ALS Data and K-Nearest Neighbor Algorithm Based Models. *Geosciences 2019, Vol. 9, Page 442*, *9*(10), 442. https://doi.org/10.3390/GEOSCIENCES9100442

Navarro-Cerrillo, R. M., Duque-Lazo, J., Rodríguez-Vallejo, C., Varo-Martínez, M. Á., & Palacios-Rodríguez, G. (2018). Airborne Laser Scanning Cartography of On-Site Carbon Stocks as a Basis for the Silviculture of *Pinus Halepensis* Plantations. *Remote Sensing 2018, Vol. 10, Page 1660*, *10*(10), 1660. https://doi.org/10.3390/RS10101660

Oehmcke, S., Li, L., Revenga, J. C., Nord-Larsen, T., Trepekli, K., Gieseke, F., & Igel, C. (2021). Deep Learning Based 3D Point Cloud Regression for Estimating Forest Biomass. *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*. https://doi.org/10.1145/3557915.3561471

Oishy, M. N., Shemonty, N. A., Fatema, S. I., Mahbub, S., Mim, E. L., Hasan Raisa, M. B., & Anik, A. H. (2025). Unravelling the effects of climate change on the soil-plant-atmosphere interactions: A critical review. *Soil & Environmental Health*, *3*(1), 100130. https://doi.org/10.1016/J.SEH.2025.100130

Ou, J., Wu, Z., Yan, Q., Feng, X., & Zhao, Z. (2024). Improving soil organic carbon mapping in farmlands using machine learning models and complex cropping system information. *Environmental Sciences Europe, 36(1)*, 1–15. https://doi.org/10.1186/s12302-024-00912-x

Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., Phillips, O. L., Shvidenko, A., Lewis, S. L., Canadell, J. G., Ciais, P., Jackson, R. B., Pacala, S. W.,

McGuire, A. D., Piao, S., Rautiainen, A., Sitch, S., & Hayes, D. (2011). A large and persistent carbon sink in the world's forests. *Science*, *333*(6045), 988–993. https://doi.org/10.1126/science.1201609

Panagos, Panos., Ballabio, Cristiano., Scarpa, Simone., Borrelli, Pasquale., Lugato, Emanuele., & Montanarella, Luca. (2020). *Soil related indicators to support agro-environmental policies*. https://doi.org/10.2760/011194

Panagos, P., van Liedekerke, M., Borrelli, P., Köninger, J., Ballabio, C., Orgiazzi, A., Lugato, E., Liakos, L., Hervas, J., Jones, A., & Montanarella, L. (2022). European Soil Data Centre 2.0: Soil data and knowledge in support of the EU policies. *European Journal of Soil Science, 73(6)*, e13315. https://doi.org/10.1111/EJSS.13315

Pascual, A., Godinho, S., & Guerra-Hernández, J. (2023). Integrated LiDAR-supported valuation of biomass and litter in forest ecosystems. A showcase in Spain. *Science of The Total Environment*, *897*, 165364. https://doi.org/10.1016/J.SCITOTENV.2023.165364

Pebesma, E. (2016). sf: Simple Features for R. CRAN: *Contributed Packages*. https://doi.org/10.32614/CRAN.PACKAGE.SF

Penman, J., Gytarsky, M., Hiraishi, T., Krug, T., Kruger, D., Pipatti, R., Buendia, L., Miwa, K., Ngara, T., Tanabe, K., & Wagner, F. (2003). *Good practice guidance for land use, land-use change and forestry*. IPCC/IGES. http://www.ipcc-nggip.iges.or.jp/public/gpglulucf/gpglulucf_contents

Perea-Ardila, M. A., Andrade-Castañeda, H. J., & Segura-Madrigal, M. A. (2021). Estimación de biomasa aérea y carbono con teledetección en bosques alto-andinos de Boyacá, Colombia. *Revista Cartográfica*, *102*, 99–123. https://doi.org/10.35424/rcarto.i102.821

Pereira, F. F., Sussel, T., Mendes, G., Silvio, ·, Coelho, J., Márcio, S. ·, Magalhães De Andrade, R., Luiz, M., Reiss, L., Fortes, J., Renk, C., Correia, T., Santos, S., Jorge, S., & Simões, C. (2023). Comparison of LiDAR- and UAV-derived data for landslide susceptibility mapping using Random Forest algorithm. *Landslides 2023 20:3*, *20*(3), 579–600. https://doi.org/10.1007/S10346-022-02001-7

Peter, J. S., Drake, J., Medley, P., & Ibeanusi, V. (2021). Forest Structural Estimates Derived Using a Practical, Open-Source Lidar-Processing Workflow. *Remote Sensing 2021, Vol. 13, Page 4763*, *13*(23), 4763. https://doi.org/10.3390/RS13234763

Publications - IPCC-TFI. (n.d.). Retrieved October 25, 2024, from https://www.ipcc-nggip.iges.or.jp/public/2006gl/spanish/index.html

QGIS Development Team. (2024). *QGIS Geographic Information System* (Version 3.40) [Software]. Open Source Geospatial Foundation. https://qgis.org

R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.4.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rabot, E., Saby, N. P. A., Martin, M. P., Barré, P., Chenu, C., Cousin, I., Arrouays, D., Angers, D., & Bispo, A. (2024). Relevance of the organic carbon to clay ratio as a national soil health indicator. *Geoderma*, *443*, 116829. https://doi.org/10.1016/J.GEODERMA.2024.116829

Rasel, S. M. M., Groen, T. A., Hussin, Y. A., & Diti, I. J. (2017). Proxies for soil organic carbon derived from remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, *59*, 157–166. https://doi.org/10.1016/J.JAG.2017.03.004

Reddy, A. D., Hawbaker, T. J., Wurster, F., Zhu, Z., Ward, S., Newcomb, D., & Murray, R. (2015). Quantifying soil carbon loss and uncertainty from a peatland wildfire using multi-temporal LiDAR. *Remote Sensing of Environment*, *170*, 306–316. https://doi.org/10.1016/J.RSE.2015.09.017

Roussel, J.-R., & Auty, D. (2022). *lidR: Airborne LiDAR data manipulation and visualization for forestry applications* (Version 4.0.1) [R package]. https://cran.r-project.org/package=lidR

Roussel, J. R., Auty, D., Coops, N. C., Tompalski, P., Goodbody, T. R. H., Meador, A. S., Bourdon, J. F., de Boissieu, F., & Achim, A. (2020). lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sensing of Environment*, *251*, 112061. https://doi.org/10.1016/J.RSE.2020.112061

Ruiz-Navarro, A., Barberá, G. G., Navarro-Cano, J. A., Albaladejo, J., & Castillo, V. M. (2009). Soil dynamics in *Pinus halepensis* reforestation: Effect of microenvironments and previous land use. *Geoderma*, *153*(3–4), 353–361. https://doi.org/10.1016/J.GEODERMA.2009.08.024

Ruiz-Peinado, R., Bravo-Oviedo, A., López-Senespleda, E., Bravo, F., & del Río, M. (2017). Forest management and carbon sequestration in the Mediterranean region: A review. *Forest Systems*, *26*(2), eR04S-eR04S. https://doi.org/10.5424/FS/2017262-11205

Santonja, M., Pereira, S., Gauquelin, T., Quer, E., Simioni, G., Limousin, J. M., Ourcival, J. M., Reiter, I. M., Fernandez, C., & Baldy, V. (2022). Experimental Precipitation Reduction Slows Down Litter Decomposition but Exhibits Weak to No Effect on Soil Organic Carbon and Nitrogen Stocks in Three Mediterranean Forests of Southern France. *Forests*, *13*(9), 1485. https://doi.org/10.3390/f13091485

SAS Institute Inc. (2023). *SAS software (Version 9.4M8)* [Computer software]. Cary, NC: SAS Institute Inc.

Scharlemann, J. P. W., Tanner, E. V. J., Hiederer, R., & Kapos, V. (2014). Global soil carbon: Understanding and managing the largest terrestrial carbon pool. *Carbon Management*, 5(1), 81–91. https://doi.org/10.4155/cmt.13.77

Schmidt, M. W. I., Torn, M. S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I. A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D. A. C., Nannipieri, P., Rasse, D. P., Weiner, S., & Trumbore, S. E. (2011). Persistence of soil organic matter as an ecosystem property. *Nature 2011 478:7367*, *478*(7367), 49–56. https://doi.org/10.1038/nature10386

Serrada Hierro, R., Montero González, G., & Reque Kilchenmann, J. A. (Eds.). (2008). *Compendio de selvicultura aplicada en España* (2 vols.). Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA).

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 23rd ACM National Conference*, *ACM 1968*, 517–524. https://doi.org/10.1145/800186.810616

Silva, L. C. R., & Lambers, H. (2020). Soil-plant-atmosphere interactions: structure, function, and predictive scaling for climate change mitigation. *Plant and Soil 2020 461:1*, *461*(1), 5–27. https://doi.org/10.1007/S11104-020-04427-1

Stevens, A., Nocita, M., Tóth, G., Montanarella, L., & van Wesemael, B. (2013). Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. *PLOS ONE*, 8(6), e66409. https://doi.org/10.1371/JOURNAL.PONE.0066409

Strîmbu, V. F., Næsset, E., Ørka, H. O., Liski, J., Petersson, H., & Gobakken, T. (2023). Estimating biomass and soil carbon change at the level of forest stands using repeated forest surveys assisted by airborne laser scanner data. *Carbon Balance and Management*, *18*(1), 1–20. https://doi.org/10.1186/s13021-023-00222-4

Strunk, J. L., & McGaughey, R. J. (2023). Stand validation of lidar forest inventory modeling for a managed southern pine forest. *Canadian Journal of Forest Research*, *53*(2), 71–89. https://doi.org/10.1139/cjfr-2022-0032

Stumpf, F., Behrens, T., Schmidt, K., & Keller, A. (2024). Exploiting Soil and Remote Sensing Data Archives for 3D Mapping of Multiple Soil Properties at the Swiss National Scale. *Remote Sensing*, *16*(15), 2712. https://doi.org/10.3390/rs16152712

Tafur, E., Veneros, J., García, L., Gamarra, Ó., Farje, J., Santistevan, M., Tafur, E., Veneros, J., García, L., Gamarra, Ó., Farje, J., & Santistevan, M. (2022). Técnicas no destructivas para la estimación de la biomasa forestal aérea. *Idesia (Arica)*, *40*(3), 7–17. https://doi.org/10.4067/S0718-34292022000300007

Trouvé, R., Jiang, R., Fedrigo, M., White, M. D., Kasel, S., Baker, P. J., & Nitschke, C. R. (2023). Combining Environmental, Multispectral, and LiDAR Data Improves Forest Type Classification: A Case Study on Mapping Cool Temperate Rainforests and Mixed Forests. *Remote Sensing*, *15*(1), 60. https://doi.org/10.3390/rs15010060

Tupinambá-Simões, F., Pascual, A., Guerra-Hernández, J., Ordóñez, C., Barreiro, S., & Bravo, F. (2025). Combining hand-held and drone-based lidar for forest carbon monitoring: insights from a Mediterranean mixed forest in central Portugal. *European Journal of Forest Research*, 1–16. https://doi.org/10.1007/s10342-025-01772-7

Usman, J., & Begum, S. (2023). SOC stocks prediction on the basis of spatial and temporal variation in soil properties by using partial least square regression. *Scientific Reports*, *13*(1), 1–12. https://doi.org/10.1038/s41598-023-34607-9

Vieira, J., Nabais, C., & Campelo, F. (2022). Dry and hot years drive growth decline of *Pinus halepensis* at its southern range limit in the Moroccan High Atlas Mountains. *Trees - Structure and Function*, *36*(5), 1585–1595. https://doi.org/10.1007/s00468-022-02314-z

Wang, S., Xu, L., Zhuang, Q., & He, N. (2021). Investigating the spatio-temporal variability of soil organic carbon stocks in different ecosystems of China. *Science of The Total Environment*, *758*, 143644. https://doi.org/10.1016/J.SCITOTENV.2020.143644

Will, R. M. (2017). Mapping Soil Organic Carbon (SOC) in a Semi-Arid Mountainous Watershed Using Variables From Hyperspectral, Lidar and Traditional Datasets. *Boise State University Theses and Dissertations*. https://doi.org/10.18122/B2Q598

Yavari, F., & Sohrabi, H. (2019). Estimation of Available Canopy Fuel of Coppice Oak Stands Using Low-Density Airborne Laser Scanning (LiDAR) Data. *Advances in Science, Technology and Innovation*, 171–173. https://doi.org/10.1007/978-3-030-01440-7_40

Zhou, T., Geng, Y., Chen, J., Pan, J., Haase, D., & Lausch, A. (2020). High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. *Science of The Total Environment*, *729*, 138244. https://doi.org/10.1016/J.SCITOTENV.2020.138244

Zou, J., Wei, Y., Zhang, Y., Liu, Z., Gai, Y., Chen, H., Liu, P., & Song, Q. (2024). Remote sensing inversion of soil organic matter in cropland combining topographic factors with spectral parameters. *Frontiers in Environmental Science*, *12*, 1420557. https://doi.org/10.3389/fenvs.2024.1420557

# APPENDIX 1. PHOTOGRAPHIC DOSSIER

*Figure 12. General view of plot 6 from its centre, with several specimens of Pinus halepensis. Image taken during the soil sampling day.*

*Figure 13. General view of Plot 7 showing low natural regeneration density under a pine-dominated canopy.*

*Figure 14. Central area of Plot 16 showing higher seedling density and more abundant natural regeneration under similar canopy conditions.*

*Figure 15. Soil sampling on plot 13 during the sampling day on 1 April 2025.*

*Figure 16. Soil sampling at 0–10 cm depth for subsequent laboratory analysis.*

*Figure 17. Soil samples drying prior to sieving for granulometric and chemical analysis*

## APPENDIX 2. DETAIL OF THE EDAPHOLOGICAL ANALYSIS.

This appendix describes in detail the procedures followed for the determination of bulk density and soil organic matter (SOM) in the 34 composite soil samples taken in the study area. All analyses were conducted in the soil laboratory of ETSIIAA (Escuela Técnica Superior de Ingenierías Agrarias) in Palencia.

- **Bulk density (BD) determination**

The bulk density of the mineral soil (0–10 cm) was determined in situ using the core method, following the protocol by Blake & Hartge (1986). The procedure was as follows:

A. Sampling:
   In each plot, undisturbed soil cores were extracted using stainless steel cylinders (100 $cm^3$ volume), inserted vertically into the ground to avoid compaction or loss of material.

B. Sample handling:
   The cores were carefully extracted, trimmed flush with the edges of the cylinder using a spatula or knife, and sealed for transport (Figure 18).

*Figure 18. Preparation of soil core samples by trimming excess material prior to sealing.*

C. Drying:
   In the laboratory, each fresh (field-moist) sample was weighed to determine its

initial mass before drying. Then, samples were oven-dried at 105 °C for 48 hours (Figure 19).

*Figure 19. Soil samples in beakers prepared for drying in an oven to estimate bulk density*

D. Weight measurement:
Once dried, the samples were cooled in a desiccator and weighed using an analytical balance (±0.01 g) (Figure 20).


*Figure 20. Determination of soil sample weight after 48 h oven-drying at 105 °C using a precision balance.*

E. Calculation:
Bulk density was calculated as the ratio between dry mass and core volume (in g/cm³), discounting the weight of the container:

$$BD = \frac{Dry\ soil\ mass\ (g)}{Core\ volume\ (cm^3)}$$

Equation A.1

- **Organic matter determination – Walkley and Black method**

The soil organic matter content was quantified by the Walkley and Black method, which estimates the fraction of easily oxidizable organic carbon via wet chemical oxidation. The procedure was adapted from de Vos et al. (2014) and executed as follows:

A. Sample preparation:
A representative aliquot (0.2–1 g) of air-dried, sieved soil (<2 mm) was placed in a 500 mL Erlenmeyer flask (Figure 21).

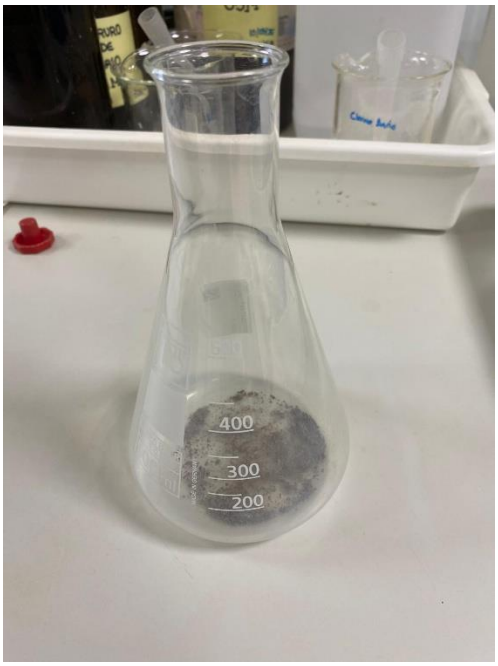


*Figure 21. Aliquot of air-dried, sieved soil placed in an Erlenmeyer flask for chemical analysis.*

B. Oxidation:

- 10 mL of 0.5 N potassium dichromate ($K_2Cr_2O_7$) were added to the flask (Figure 22).

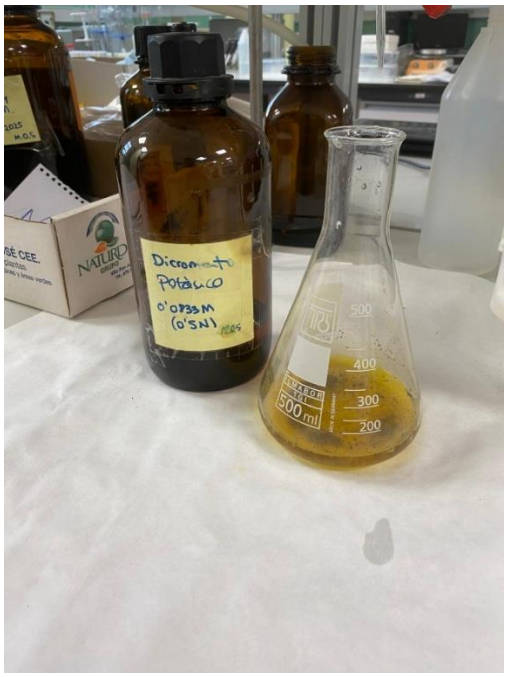


*Figure 22. Addition of 10 mL potassium dichromate solution (0.5 N) to soil samples during oxidation step.*

- Then, 20 mL of concentrated sulfuric acid ($H_2SO_4$) were carefully poured down the side of the flask to initiate the oxidation of organic carbon (Figure 23).

*Figure 23. Dispensing of concentrated sulfuric acid ($H_2SO_4$) for the oxidation of organic carbon in soil samples.*

- The mixture was gently swirled and allowed to react for 30 minutes at room temperature.

C. Dilution and titration:

- After oxidation, 150 mL of distilled water, 10 mL of 50% phosphoric acid ($H_3PO_4$) and 1 mL of diphenylamine indicator were added.
- The excess dichromate (not consumed in the reaction) was titrated with 0.25 N ferrous ammonium sulphate (Mohr's salt), until the solution shifted from blue-violet to bright green, indicating the endpoint (Figure 24).

*Figure 24. Assessment of dichromate in the determination of organic carbon using the Walkley-Black method*

D. Blank and sample correction:
A blank (no soil) was run in parallel to correct for baseline dichromate levels (Figure 25).

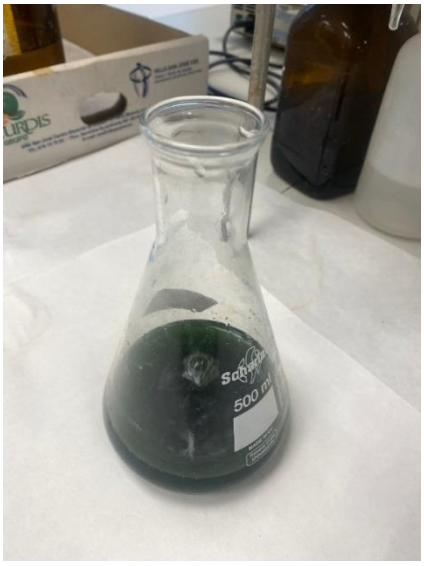

*Figure 25. Visual result of the equivalence point in the determination of organic carbon*

E. Calculations:
The difference in volume between the blank and the sample was used to calculate the amount of oxidized carbon, and the result was expressed as percentage of oxidizable organic carbon (%Cox). Then:

$$SOM\% = \frac{Cox\%}{0.58}$$ Equation A.2

The 0.58 factor assumes that organic matter contains 58% carbon (MAPA, 1994).

## APPENDIX 3. RESULTS OF ABOVEGROUND AND ROOT BIOMASS ESTIMATION

*Table 8. Estimates of above-ground biomass. root biomass and carbon stored by biomass fraction in each plot (Mg/ha)*

| Plot | Ws (Mg/ha) | Wr (Mg/ha) | Wb2-7 (Mg/ha) | Wb 0.5-2 (Mg/ha) | Above-ground biomass (Mg/ha) | Root biomass (Mg/ha) | Above-ground C (Mg/ha) | Root C (Mg/ha) |
|---|---|---|---|---|---|---|---|---|
| 1 | 9.20 | 4.75 | 1.83 | 1.15 | 12.16 | 4.75 | 6.08 | 2.37 |
| 2 | 17.35 | 9.13 | 3.40 | 2.28 | 23.03 | 9.13 | 11.52 | 4.55 |
| 3 | 31.43 | 15.38 | 6.28 | 2.90 | 40.61 | 15.38 | 20.29 | 7.69 |
| 4 | 12.67 | 7.14 | 2.28 | 1.99 | 16.93 | 7.14 | 8.47 | 3.58 |
| 5 | 37.60 | 18.28 | 7.36 | 2.79 | 47.75 | 18.28 | 23.87 | 9.13 |
| 6 | 20.60 | 10.32 | 4.00 | 2.01 | 26.64 | 10.32 | 13.31 | 5.17 |
| 7 | 44.87 | 19.81 | 9.53 | 2.25 | 56.65 | 19.81 | 28.32 | 9.90 |
| 8 | 24.20 | 11.25 | 5.44 | 2.01 | 31.65 | 11.25 | 15.83 | 5.61 |
| 9 | 54.80 | 23.21 | 12.00 | 2.30 | 69.08 | 23.21 | 34.55 | 11.61 |
| 10 | 44.41 | 19.03 | 9.64 | 1.97 | 56.01 | 19.03 | 28.01 | 9.53 |
| 11 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 12 | 67.11 | 30.20 | 14.06 | 3.74 | 84.88 | 30.20 | 42.44 | 15.10 |
| 13 | 16.25 | 8.73 | 3.14 | 2.41 | 21.77 | 8.73 | 10.90 | 4.35 |
| 14 | 19.19 | 9.55 | 3.80 | 1.72 | 24.74 | 9.55 | 12.36 | 4.77 |
| 15 | 12.62 | 6.48 | 2.72 | 1.97 | 17.33 | 6.48 | 8.67 | 3.25 |
| 16 | 3.14 | 2.01 | 0.69 | 1.11 | 4.93 | 2.01 | 2.45 | 1.02 |
| 17 | 23.98 | 11.83 | 4.80 | 2.30 | 31.10 | 11.83 | 15.54 | 5.92 |
| 18 | 4.97 | 3.23 | 1.48 | 2.48 | 8.93 | 3.23 | 4.47 | 1.61 |
| 19 | 12.58 | 6.39 | 2.43 | 1.28 | 16.31 | 6.39 | 8.16 | 3.18 |
| 20 | 23.19 | 11.63 | 4.49 | 2.17 | 29.84 | 11.63 | 14.92 | 5.81 |

*Table 8. Cont. Estimates of above-ground biomass. root biomass and carbon stored by biomass fraction in each plot (Mg/ha)*

| Plot | Ws (Mg/ha) | Wr (Mg/ha) | Wb2-7 (Mg/ha) | Wb 0.5-2 (Mg/ha) | Above-ground biomass (Mg/ha) | Root biomass (Mg/ha) | Above-ground C (Mg/ha) | Root C (Mg/ha) |
|---|---|---|---|---|---|---|---|---|
| **21** | 41.93 | 19.85 | 8.47 | 2.94 | 53.34 | 19.85 | 26.66 | 9.93 |
| **22** | 18.02 | 8.05 | 3.87 | 1.17 | 23.06 | 8.05 | 11.52 | 4.02 |
| **23** | 17.90 | 9.51 | 3.51 | 2.54 | 23.98 | 9.51 | 12.00 | 4.75 |
| **24** | 7.76 | 4.40 | 1.53 | 1.55 | 10.83 | 4.40 | 5.42 | 2.19 |
| **25** | 16.98 | 9.39 | 3.03 | 2.28 | 22.28 | 9.39 | 11.14 | 4.69 |
| **26** | 17.60 | 8.49 | 3.71 | 1.72 | 23.03 | 8.49 | 11.52 | 4.24 |
| **27** | 9.13 | 5.57 | 1.88 | 2.37 | 13.37 | 5.57 | 6.68 | 2.79 |
| **28** | 12.91 | 7.38 | 2.28 | 2.12 | 17.31 | 7.38 | 8.67 | 3.69 |
| **29** | 33.64 | 16.60 | 6.68 | 3.03 | 43.37 | 16.60 | 21.68 | 8.29 |
| **30** | 24.09 | 13.11 | 4.33 | 3.01 | 31.41 | 13.11 | 15.69 | 6.57 |
| **31** | 21.82 | 12.05 | 4.00 | 3.25 | 29.09 | 12.05 | 14.54 | 6.03 |
| **32** | 16.62 | 8.93 | 3.21 | 2.41 | 22.24 | 8.93 | 11.12 | 4.47 |
| **33** | 22.94 | 12.25 | 4.33 | 2.98 | 30.28 | 12.25 | 15.14 | 6.12 |
| **34** | 19.76 | 8.55 | 4.24 | 0.91 | 24.91 | 8.55 | 12.45 | 4.27 |

## APPENDIX 4. DESCRIPTION OF LIDAR VARIABLES USED IN SOC MODELING

*Table 9. Height and forest structure metrics derived from LiDAR point clouds (a).*

| PLOT | n | zmax | zmin | zmean | zvar | zsd | zcv | zskew | zkurt |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 620.00 | 12.59 | -0.05 | 2.14 | 14.71 | 3.83 | 179.13 | 1.36 | 3.07 |
| 2 | 739.00 | 12.32 | -0.07 | 3.75 | 18.16 | 4.26 | 113.65 | 0.43 | 1.45 |
| 3 | 720.00 | 14.59 | -0.07 | 4.45 | 24.15 | 4.91 | 110.44 | 0.36 | 1.43 |
| 4 | 681.00 | 11.26 | -0.06 | 3.21 | 16.58 | 4.07 | 126.74 | 0.60 | 1.55 |
| 5 | 782.00 | 12.94 | -0.07 | 4.74 | 25.14 | 5.01 | 105.89 | 0.20 | 1.17 |
| 6 | 709.00 | 13.15 | -0.05 | 4.34 | 22.65 | 4.76 | 109.66 | 0.29 | 1.28 |
| 7 | 808.00 | 14.92 | -0.07 | 5.82 | 30.23 | 5.50 | 94.40 | -0.01 | 1.17 |
| 8 | 709.00 | 13.90 | -0.10 | 4.12 | 22.65 | 4.76 | 115.52 | 0.60 | 1.71 |
| 9 | 725.00 | 14.80 | -0.09 | 5.31 | 33.09 | 5.75 | 108.23 | 0.24 | 1.18 |
| 10 | 694.00 | 14.43 | -0.07 | 4.61 | 31.14 | 5.58 | 121.13 | 0.45 | 1.32 |
| 11 | 526.00 | 9.28 | -0.07 | 0.16 | 0.77 | 0.88 | 535.21 | 6.90 | 55.09 |
| 12 | 775.00 | 14.32 | -0.07 | 5.81 | 29.66 | 5.45 | 93.73 | 0.00 | 1.20 |
| 13 | 694.00 | 12.13 | -0.33 | 3.64 | 16.98 | 4.12 | 113.26 | 0.45 | 1.52 |
| 14 | 699.00 | 12.83 | -0.07 | 3.87 | 22.52 | 4.75 | 122.48 | 0.54 | 1.47 |
| 15 | 691.00 | 12.84 | -0.12 | 3.28 | 15.78 | 3.97 | 121.21 | 0.70 | 1.96 |
| 16 | 604.00 | 9.61 | -0.06 | 2.28 | 9.59 | 3.10 | 135.76 | 0.87 | 2.11 |
| 17 | 764.00 | 12.90 | -0.07 | 4.39 | 21.06 | 4.59 | 104.43 | 0.29 | 1.41 |
| 18 | 727.00 | 10.00 | -0.05 | 2.79 | 10.22 | 3.20 | 114.66 | 0.55 | 1.70 |
| 19 | 678.00 | 12.52 | -0.07 | 3.50 | 17.82 | 4.22 | 120.63 | 0.55 | 1.61 |
| 20 | 783.00 | 12.81 | -0.07 | 4.29 | 19.69 | 4.44 | 103.51 | 0.25 | 1.36 |
| 21 | 834.00 | 13.98 | -0.06 | 4.92 | 25.09 | 5.01 | 101.90 | 0.23 | 1.30 |

*Table 9. Cont. Height and forest structure metrics derived from LiDAR point clouds (a).*

| PLOT | n | zmax | zmin | zmean | zvar | zsd | zcv | zskew | zkurt |
|---|---|---|---|---|---|---|---|---|---|
| 22 | 677.00 | 14.06 | -0.05 | 3.53 | 20.74 | 4.55 | 128.96 | 0.76 | 2.01 |
| 23 | 793.00 | 12.66 | -0.05 | 3.99 | 18.67 | 4.32 | 108.36 | 0.39 | 1.46 |
| 24 | 656.00 | 11.60 | -0.08 | 2.98 | 14.20 | 3.77 | 126.40 | 0.60 | 1.58 |
| 25 | 689.00 | 11.54 | -0.06 | 3.55 | 17.82 | 4.22 | 119.01 | 0.47 | 1.39 |
| 26 | 730.00 | 13.69 | -0.27 | 3.94 | 19.59 | 4.43 | 112.30 | 0.51 | 1.64 |
| 27 | 758.00 | 11.24 | -0.09 | 3.53 | 14.48 | 3.81 | 107.82 | 0.41 | 1.47 |
| 28 | 787.00 | 12.06 | -0.06 | 4.04 | 17.78 | 4.22 | 104.41 | 0.26 | 1.29 |
| 29 | 793.00 | 13.24 | -0.06 | 4.79 | 21.36 | 4.62 | 96.41 | 0.15 | 1.31 |
| 30 | 748.00 | 11.82 | -0.09 | 4.09 | 19.47 | 4.41 | 107.78 | 0.27 | 1.22 |
| 31 | 801.00 | 12.58 | -0.07 | 4.35 | 16.37 | 4.05 | 93.02 | 0.11 | 1.41 |
| 32 | 740.00 | 12.11 | -0.04 | 3.80 | 16.12 | 4.02 | 105.54 | 0.38 | 1.57 |
| 33 | 767.00 | 12.22 | -0.19 | 3.73 | 18.22 | 4.27 | 114.32 | 0.47 | 1.49 |
| 34 | 182.00 | 15.08 | -0.07 | 6.45 | 32.56 | 5.71 | 88.49 | -0.12 | 1.23 |

*Note: **n**, total number of LiDAR returns within the plot; **zmax**, maximum Z value; **zmin**, minimum Z value; **zmean**, mean of Z values; **zvar**, variance of Z values; **zsd**, standard deviation of Z values; **zcv**, coefficient of variation (zsd/zmean); **zskew**, skewness of Z values; **zkurt**, kurtosis of Z values.

*Table 10. Height and forest structure metrics derived from LiDAR point clouds (b).*

| PLOT | zq1 | zq5 | zq10 | zq15 | zq20 | zq25 | zq30 | zq35 | zq40 | zq45 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -0.04 | -0.02 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| 2 | -0.04 | -0.02 | -0.01 | -0.01 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 | 0.03 |
| 3 | -0.04 | -0.02 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 | 0.04 |
| 4 | -0.04 | -0.02 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 |
| 5 | -0.04 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.04 |
| 6 | -0.03 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 |
| 7 | -0.04 | -0.02 | -0.01 | -0.01 | 0.00 | 0.01 | 0.01 | 0.02 | 0.04 | 0.09 |
| 8 | -0.05 | -0.02 | -0.01 | -0.01 | 0.00 | 0.00 | 0.01 | 0.01 | 0.03 | 0.07 |
| 9 | -0.04 | -0.02 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 | 0.04 |
| 10 | -0.04 | -0.02 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.02 |
| 11 | -0.05 | -0.03 | -0.02 | -0.02 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| 12 | -0.03 | -0.02 | -0.01 | 0.00 | 0.00 | 0.01 | 0.02 | 0.03 | 0.06 | 2.86 |
| 13 | -0.04 | -0.02 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.03 |
| 14 | -0.04 | -0.02 | -0.01 | -0.01 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 | 0.03 |
| 15 | -0.04 | -0.02 | -0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.03 |
| 16 | -0.04 | -0.02 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 |
| 17 | -0.05 | -0.02 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.03 | 0.07 |
| 18 | -0.04 | -0.02 | -0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.04 |
| 19 | -0.04 | -0.02 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 |
| 20 | -0.04 | -0.02 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 | 0.05 |
| 21 | -0.03 | -0.02 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.05 | 0.09 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 22 | -0.03 | -0.02 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 |

*Table 10. Cont Height and forest structure metrics derived from LiDAR point clouds (b).*

| PLOT | zq1 | zq5 | zq10 | zq15 | zq20 | zq25 | zq30 | zq35 | zq40 | zq45 |
|---|---|---|---|---|---|---|---|---|---|---|
| 23 | -0.04 | -0.02 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.03 | 0.06 |
| 24 | -0.04 | -0.02 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 |
| 25 | -0.05 | -0.02 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 | 0.03 |
| 26 | -0.03 | -0.02 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.03 | 0.06 |
| 27 | -0.04 | -0.02 | -0.01 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 | 0.05 | 0.14 |
| 28 | -0.04 | -0.02 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.04 | 0.07 |
| 29 | -0.03 | -0.01 | -0.01 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 | 0.05 | 2.36 |
| 30 | -0.04 | -0.02 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 | 0.06 |
| 31 | -0.04 | -0.02 | -0.01 | 0.00 | 0.00 | 0.01 | 0.01 | 0.04 | 0.16 | 3.18 |
| 32 | -0.04 | -0.02 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 | 0.06 |
| 33 | -0.04 | -0.02 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.03 |
| 34 | -0.04 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.06 | 6.85 |

*Note: variables **zq1** to **zq45** represent the 1st, 5th, 10th, 15th, 20th, 25th, 30th, 35th, 40th and 45th percentiles of height values (Z).

*Table 11. Height and forest structure metrics derived from LiDAR point clouds (c).*

| PLOT | zq50 | zq55 | zq60 | zq65 | zq70 | zq75 | zq80 | zq85 | zq90 | zq95 | zq99 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.01 | 0.01 | 0.02 | 0.03 | 0.06 | 0.91 | 7.45 | 8.34 | 9.02 | 10.18 | 11.63 |
| 2 | 0.07 | 2.74 | 5.63 | 6.66 | 7.71 | 8.22 | 8.68 | 9.10 | 9.60 | 10.60 | 11.30 |
| 3 | 0.07 | 6.04 | 7.35 | 8.13 | 8.73 | 9.11 | 9.67 | 10.15 | 11.07 | 12.45 | 13.60 |
| 4 | 0.03 | 0.06 | 2.29 | 5.94 | 7.04 | 7.73 | 8.13 | 8.76 | 9.39 | 9.87 | 10.62 |
| 5 | 0.07 | 7.36 | 8.34 | 9.03 | 9.55 | 9.90 | 10.21 | 10.72 | 11.11 | 11.64 | 12.26 |
| 6 | 0.05 | 6.00 | 7.51 | 8.32 | 8.79 | 9.10 | 9.55 | 10.00 | 10.52 | 11.37 | 12.79 |
| 7 | 8.07 | 9.17 | 9.81 | 10.23 | 10.52 | 10.97 | 11.40 | 11.79 | 12.29 | 13.00 | 13.95 |
| 8 | 1.65 | 2.56 | 4.33 | 6.62 | 7.81 | 8.90 | 9.77 | 10.61 | 11.67 | 12.28 | 13.29 |
| 9 | 0.12 | 8.49 | 9.60 | 10.15 | 10.80 | 11.28 | 11.77 | 12.25 | 12.63 | 13.29 | 14.23 |
| 10 | 0.04 | 0.06 | 7.16 | 9.51 | 10.22 | 10.84 | 11.34 | 11.82 | 12.38 | 13.02 | 13.78 |
| 11 | 0.00 | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 | 0.03 | 0.05 | 0.07 | 0.14 | 5.24 |
| 12 | 7.85 | 8.83 | 9.44 | 10.17 | 10.64 | 11.04 | 11.37 | 11.82 | 12.47 | 13.01 | 13.80 |
| 13 | 0.08 | 2.92 | 5.69 | 6.52 | 7.30 | 7.81 | 8.25 | 8.67 | 9.39 | 10.15 | 11.74 |
| 14 | 0.06 | 0.13 | 4.70 | 7.07 | 8.63 | 9.18 | 9.58 | 10.15 | 10.84 | 11.38 | 12.29 |
| 15 | 0.06 | 1.87 | 3.83 | 5.07 | 5.85 | 6.84 | 7.69 | 8.39 | 9.51 | 10.58 | 11.82 |
| 16 | 0.03 | 0.05 | 0.74 | 2.84 | 3.88 | 5.51 | 6.07 | 6.56 | 7.32 | 7.99 | 9.04 |
| 17 | 4.01 | 5.75 | 7.00 | 7.45 | 8.07 | 8.68 | 9.36 | 10.01 | 10.67 | 11.53 | 12.67 |
| 18 | 0.12 | 2.48 | 3.77 | 4.63 | 5.25 | 5.76 | 6.31 | 6.82 | 7.70 | 8.28 | 9.19 |
| 19 | 0.03 | 0.08 | 5.39 | 6.45 | 7.14 | 7.76 | 8.12 | 8.66 | 9.41 | 10.75 | 12.01 |
| 20 | 3.52 | 5.93 | 7.08 | 7.64 | 8.09 | 8.55 | 8.90 | 9.33 | 10.11 | 11.14 | 12.18 |
| 21 | 3.73 | 6.09 | 7.89 | 8.84 | 9.57 | 10.00 | 10.63 | 11.03 | 11.48 | 12.07 | 13.04 |
| 22 | 0.03 | 0.05 | 3.75 | 5.80 | 7.05 | 7.88 | 8.56 | 9.01 | 10.53 | 12.32 | 13.46 |
| 23 | 1.67 | 4.07 | 5.78 | 7.09 | 7.95 | 8.45 | 8.96 | 9.39 | 9.95 | 10.78 | 11.99 |
| 24 | 0.02 | 0.04 | 1.38 | 5.64 | 6.50 | 7.04 | 7.63 | 8.07 | 8.41 | 8.91 | 10.70 |

*Table 11. Cont. Height and forest structure metrics derived from LiDAR point clouds (c).*

| PLOT | zq50 | zq55 | zq60 | zq65 | zq70 | zq75 | zq80 | zq85 | zq90 | zq95 | zq99 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 25 | 0.07 | 0.19 | 5.21 | 6.94 | 7.77 | 8.17 | 8.60 | 9.03 | 9.51 | 10.17 | 10.79 |
| 26 | 0.90 | 3.42 | 4.95 | 6.65 | 7.54 | 8.43 | 9.05 | 9.50 | 10.36 | 11.39 | 12.68 |
| 27 | 1.60 | 2.94 | 5.26 | 6.04 | 6.92 | 7.44 | 8.15 | 8.53 | 8.82 | 9.32 | 10.89 |
| 28 | 2.11 | 4.50 | 6.73 | 7.61 | 8.15 | 8.43 | 8.75 | 9.06 | 9.42 | 9.99 | 11.47 |
| 29 | 4.45 | 6.58 | 7.59 | 8.29 | 8.81 | 9.36 | 9.73 | 10.29 | 10.91 | 11.47 | 12.19 |
| 30 | 0.84 | 5.37 | 7.26 | 7.91 | 8.38 | 8.65 | 9.05 | 9.47 | 9.91 | 10.32 | 11.10 |
| 31 | 5.16 | 6.08 | 6.71 | 7.23 | 7.65 | 8.11 | 8.44 | 8.79 | 9.31 | 10.34 | 11.70 |
| 32 | 2.69 | 4.67 | 5.59 | 6.40 | 7.16 | 7.59 | 8.07 | 8.48 | 9.22 | 10.65 | 11.60 |
| 33 | 0.07 | 3.02 | 5.35 | 6.52 | 7.29 | 8.04 | 8.79 | 9.38 | 10.02 | 10.61 | 11.27 |
| 34 | 9.34 | 9.89 | 10.60 | 10.97 | 11.16 | 11.51 | 11.78 | 12.30 | 12.63 | 13.84 | 14.84 |

*Note: Variables **zq50** to **zq99** represent the 50th, 55th, 60th, 65th, 70th, 75th, 80th, 85th, 90th, 95th and 99th percentiles of height values (Z).

*Table 12. Height and forest structure metrics derived from LiDAR point clouds (d).*

| PLOT | pzabovemean | pzabove2 | pzabove5 | ziqr | zMADmean | zMADmedian | CRR | zentropy | zpcum1 | zpcum2 | zpcum3 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 24.03 | 24.03 | 23.39 | 0.91 | 3.20 | 0.02 | 0.17 | 0.52 | 75.69 | 76.01 | 76.01 |
| 2 | 43.84 | 46.14 | 42.49 | 8.22 | 4.04 | 0.10 | 0.31 | 0.78 | 52.78 | 54.41 | 56.17 |
| 3 | 46.94 | 47.08 | 46.81 | 9.11 | 4.70 | 0.10 | 0.31 | 0.72 | 52.72 | 53.00 | 53.00 |
| 4 | 39.65 | 40.09 | 36.71 | 7.73 | 3.84 | 0.05 | 0.29 | 0.70 | 59.94 | 59.94 | 60.53 |
| 5 | 47.95 | 48.34 | 47.83 | 9.90 | 4.89 | 0.10 | 0.37 | 0.70 | 51.67 | 51.67 | 51.92 |
| 6 | 46.83 | 46.83 | 46.40 | 9.10 | 4.61 | 0.07 | 0.33 | 0.71 | 53.18 | 53.18 | 53.18 |
| 7 | 53.47 | 54.33 | 53.71 | 10.96 | 5.33 | 5.08 | 0.39 | 0.74 | 45.53 | 45.78 | 46.03 |
| 8 | 40.62 | 48.10 | 37.66 | 8.90 | 4.34 | 1.67 | 0.30 | 0.87 | 49.22 | 55.45 | 59.26 |
| 9 | 46.48 | 46.62 | 46.48 | 11.28 | 5.63 | 0.14 | 0.36 | 0.68 | 53.11 | 53.53 | 53.53 |
| 10 | 41.07 | 41.35 | 41.07 | 10.84 | 5.39 | 0.05 | 0.32 | 0.64 | 58.67 | 58.67 | 58.67 |
| 11 | 4.94 | 2.85 | 1.14 | 0.03 | 0.30 | 0.01 | 0.03 | 0.16 | 95.99 | 97.33 | 97.52 |
| 12 | 52.77 | 55.23 | 53.68 | 11.04 | 5.24 | 5.15 | 0.41 | 0.76 | 44.50 | 44.89 | 45.41 |
| 13 | 43.52 | 46.11 | 42.51 | 7.81 | 3.89 | 0.12 | 0.32 | 0.78 | 52.02 | 53.90 | 56.50 |
| 14 | 40.63 | 41.77 | 39.91 | 9.18 | 4.50 | 0.08 | 0.31 | 0.72 | 57.82 | 58.68 | 59.40 |
| 15 | 41.53 | 44.28 | 35.31 | 6.84 | 3.60 | 0.08 | 0.26 | 0.84 | 53.85 | 57.18 | 59.80 |
| 16 | 35.93 | 36.75 | 27.15 | 5.51 | 2.78 | 0.04 | 0.24 | 0.79 | 60.07 | 62.73 | 64.89 |
| 17 | 49.61 | 50.79 | 47.51 | 8.68 | 4.32 | 4.01 | 0.34 | 0.80 | 48.75 | 49.41 | 49.67 |
| 18 | 44.29 | 45.25 | 32.32 | 5.76 | 2.95 | 0.16 | 0.28 | 0.85 | 51.86 | 54.62 | 55.86 |
| 19 | 42.63 | 42.92 | 41.00 | 7.76 | 3.99 | 0.05 | 0.28 | 0.74 | 56.95 | 57.40 | 57.40 |
| 20 | 48.91 | 50.70 | 47.64 | 8.55 | 4.23 | 3.52 | 0.34 | 0.80 | 48.53 | 49.81 | 50.45 |
| 21 | 47.72 | 52.40 | 47.48 | 9.99 | 4.75 | 3.73 | 0.35 | 0.81 | 47.12 | 47.84 | 51.08 |
| 22 | 40.62 | 40.92 | 37.37 | 7.88 | 4.16 | 0.04 | 0.25 | 0.71 | 59.11 | 59.11 | 61.63 |

*Table 12. Cont. Height and forest structure metrics derived from LiDAR point clouds (d).*

| PLOT | pzabovemean | pzabove2 | pzabove5 | ziqr | zMADmean | zMADmedian | CRR | zentropy | zpcum1 | zpcum2 | zpcum3 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 23 | 45.27 | 49.31 | 41.99 | 8.45 | 4.04 | 1.70 | 0.32 | 0.82 | 49.75 | 51.65 | 53.92 |
| 24 | 39.79 | 39.79 | 36.74 | 7.04 | 3.57 | 0.03 | 0.26 | 0.70 | 59.63 | 60.24 | 60.24 |
| 25 | 41.94 | 42.38 | 40.35 | 8.17 | 4.03 | 0.09 | 0.31 | 0.72 | 56.48 | 57.64 | 57.93 |
| 26 | 42.88 | 48.63 | 40.00 | 8.43 | 4.08 | 0.93 | 0.30 | 0.82 | 50.14 | 52.06 | 57.01 |
| 27 | 43.27 | 48.42 | 40.77 | 7.44 | 3.56 | 1.62 | 0.32 | 0.82 | 47.09 | 52.12 | 56.35 |
| 28 | 46.51 | 50.19 | 44.22 | 8.42 | 4.03 | 2.14 | 0.34 | 0.77 | 48.15 | 50.57 | 52.99 |
| 29 | 49.56 | 55.99 | 49.05 | 9.36 | 4.37 | 4.45 | 0.36 | 0.83 | 43.36 | 45.26 | 48.80 |
| 30 | 45.32 | 46.79 | 45.32 | 8.65 | 4.29 | 0.87 | 0.35 | 0.74 | 50.54 | 53.62 | 54.56 |
| 31 | 52.18 | 56.80 | 50.81 | 8.10 | 3.79 | 4.40 | 0.35 | 0.84 | 42.30 | 43.93 | 46.31 |
| 32 | 47.57 | 51.35 | 42.97 | 7.59 | 3.73 | 2.70 | 0.32 | 0.83 | 47.90 | 48.98 | 51.70 |
| 33 | 43.55 | 46.28 | 41.33 | 8.04 | 4.01 | 0.10 | 0.32 | 0.80 | 52.81 | 53.99 | 56.34 |
| 34 | 56.04 | 58.24 | 57.14 | 11.50 | 5.44 | 4.04 | 0.43 | 0.75 | 41.67 | 41.67 | 42.22 |

*Note: **pzabovemean**, proportion of returns above the mean height; **pzabove2**, proportion of returns above 2 meters; **pzabove5**, proportion of returns above 5 meters; **ziqr**, interquartile range of height values (Q3 − Q1); **zMADmean**, mean absolute deviation from the mean; **zMADmedian**, mean absolute deviation from the median; **CRR**, canopy return ratio; **zentropy**, Shannon entropy of the height distribution; **zpcum1**, **zpcum2**, **zpcum3**, cumulative percentage of returns above 1 %, 2 %, and 3 % of maximum height, respectively.

*Table 13. Height and forest structure metrics derived from LiDAR point clouds (e).*

| PLOT | zpcum4 | zpcum5 | zpcum6 | zpcum7 | zpcum8 | zpcum9 | L1 | L2 | L3 | L4 | Lskew |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 76.66 | 77.96 | 80.55 | 88.98 | 94.65 | 98.38 | 2.14 | 1.68 | 0.95 | 0.29 | 0.57 |
| 2 | 57.26 | 63.36 | 66.89 | 78.83 | 91.45 | 97.96 | 3.75 | 2.26 | 0.53 | -0.27 | 0.23 |
| 3 | 54.39 | 59.55 | 70.01 | 85.08 | 92.19 | 98.19 | 4.45 | 2.61 | 0.55 | -0.32 | 0.21 |
| 4 | 62.30 | 64.06 | 68.48 | 76.29 | 86.75 | 96.76 | 3.21 | 2.08 | 0.65 | -0.19 | 0.31 |
| 5 | 52.31 | 52.95 | 56.41 | 65.13 | 81.15 | 94.87 | 4.74 | 2.65 | 0.38 | -0.46 | 0.14 |
| 6 | 54.17 | 55.87 | 61.39 | 76.52 | 89.82 | 97.03 | 4.34 | 2.51 | 0.47 | -0.37 | 0.19 |
| 7 | 46.65 | 48.26 | 53.72 | 67.87 | 85.98 | 97.15 | 5.82 | 2.96 | 0.15 | -0.51 | 0.05 |
| 8 | 62.80 | 66.05 | 72.84 | 79.63 | 87.55 | 96.04 | 4.12 | 2.55 | 0.70 | -0.21 | 0.27 |
| 9 | 53.53 | 53.67 | 56.02 | 66.11 | 80.36 | 95.16 | 5.31 | 3.03 | 0.49 | -0.51 | 0.16 |
| 10 | 58.96 | 59.97 | 61.27 | 68.35 | 82.51 | 94.65 | 4.61 | 2.86 | 0.73 | -0.38 | 0.26 |
| 11 | 98.66 | 99.05 | 99.05 | 99.24 | 100.00 | 100.00 | 0.16 | 0.17 | 0.14 | 0.13 | 0.87 |
| 12 | 47.09 | 48.51 | 53.69 | 63.13 | 81.24 | 93.53 | 5.81 | 2.95 | 0.16 | -0.50 | 0.06 |
| 13 | 57.08 | 60.98 | 69.22 | 81.79 | 92.34 | 96.53 | 3.64 | 2.19 | 0.52 | -0.24 | 0.24 |
| 14 | 60.69 | 62.41 | 66.86 | 72.60 | 85.51 | 96.13 | 3.87 | 2.45 | 0.70 | -0.27 | 0.28 |
| 15 | 65.02 | 71.99 | 79.54 | 87.52 | 93.90 | 98.40 | 3.28 | 2.10 | 0.66 | -0.11 | 0.31 |
| 16 | 69.55 | 72.05 | 76.71 | 86.19 | 92.35 | 97.50 | 2.28 | 1.57 | 0.61 | -0.02 | 0.39 |
| 17 | 53.22 | 57.42 | 68.33 | 77.00 | 86.86 | 95.27 | 4.39 | 2.49 | 0.44 | -0.32 | 0.18 |
| 18 | 61.10 | 67.59 | 77.79 | 85.93 | 92.41 | 98.90 | 2.79 | 1.71 | 0.45 | -0.14 | 0.26 |
| 19 | 59.02 | 63.17 | 72.04 | 85.50 | 92.31 | 97.04 | 3.50 | 2.20 | 0.62 | -0.19 | 0.28 |
| 20 | 52.37 | 57.11 | 65.04 | 80.54 | 91.04 | 96.80 | 4.29 | 2.40 | 0.38 | -0.33 | 0.16 |
| 21 | 53.97 | 57.21 | 62.26 | 72.00 | 86.78 | 97.36 | 4.92 | 2.72 | 0.39 | -0.42 | 0.14 |
| 22 | 64.00 | 69.93 | 78.22 | 88.89 | 92.15 | 96.44 | 3.53 | 2.34 | 0.82 | -0.08 | 0.35 |

*Table 13. Height and forest structure metrics derived from LiDAR point clouds (e).*

| PLOT | zpcum4 | zpcum5 | zpcum6 | zpcum7 | zpcum8 | zpcum9 | L1 | L2 | L3 | L4 | Lskew |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 23 | 58.23 | 63.04 | 67.22 | 78.35 | 91.52 | 97.97 | 3.99 | 2.33 | 0.48 | -0.29 | 0.21 |
| 24 | 61.62 | 65.60 | 74.16 | 85.47 | 97.55 | 98.78 | 2.98 | 1.93 | 0.59 | -0.17 | 0.31 |
| 25 | 59.10 | 61.43 | 64.63 | 73.22 | 87.92 | 96.94 | 3.55 | 2.20 | 0.57 | -0.27 | 0.26 |
| 26 | 60.85 | 65.38 | 73.08 | 85.16 | 92.86 | 98.21 | 3.94 | 2.38 | 0.59 | -0.23 | 0.25 |
| 27 | 58.33 | 62.30 | 68.65 | 77.51 | 91.80 | 98.54 | 3.53 | 2.06 | 0.43 | -0.25 | 0.21 |
| 28 | 55.29 | 57.58 | 61.78 | 74.90 | 92.10 | 97.96 | 4.04 | 2.27 | 0.35 | -0.36 | 0.15 |
| 29 | 51.45 | 54.99 | 62.33 | 74.21 | 86.98 | 98.36 | 4.79 | 2.54 | 0.28 | -0.38 | 0.11 |
| 30 | 54.69 | 55.76 | 58.45 | 67.83 | 84.58 | 97.45 | 4.09 | 2.34 | 0.39 | -0.38 | 0.17 |
| 31 | 49.19 | 56.70 | 68.09 | 84.86 | 94.49 | 97.75 | 4.35 | 2.24 | 0.20 | -0.28 | 0.09 |
| 32 | 56.17 | 62.96 | 70.83 | 84.80 | 92.40 | 95.52 | 3.80 | 2.18 | 0.43 | -0.23 | 0.20 |
| 33 | 57.91 | 62.61 | 69.93 | 78.17 | 88.63 | 97.78 | 3.73 | 2.27 | 0.56 | -0.26 | 0.25 |
| 34 | 43.33 | 46.67 | 48.33 | 60.00 | 83.33 | 95.00 | 6.45 | 3.10 | 0.01 | -0.50 | 0.00 |

*Note: **zpcum4** to **zpcum9**, cumulative percentage of returns above 4 % to 9 % of the maximum height; **L1** to **L4**, first to fourth L-moments of the height distribution (Z); **Lskew**, L-moment-based skewness (L-skewness).

*Table 14. Height and forest structure metrics derived from LiDAR point clouds (f).*

| PLOT | Lkurt | Lcoefvar | lad_min | lad_max | lad_mean | lad_cv | lad_sum | pz_below_0 | pz_0.0.15 | pz_0.15.2 | pz_2.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.17 | 0.78 | 0.00 | 0.19 | 0.05 | 1.11 | 0.55 | 31.13 | 41.45 | 3.39 | 0.65 |
| 2 | -0.12 | 0.60 | 0.00 | 0.30 | 0.11 | 0.78 | 1.24 | 24.63 | 27.06 | 2.17 | 3.65 |
| 3 | -0.12 | 0.59 | 0.00 | 0.26 | 0.10 | 0.96 | 1.27 | 20.83 | 31.81 | 0.28 | 0.28 |
| 4 | -0.09 | 0.65 | 0.01 | 0.23 | 0.10 | 0.83 | 1.02 | 25.40 | 33.63 | 0.88 | 3.38 |
| 5 | -0.17 | 0.56 | 0.00 | 0.35 | 0.12 | 1.03 | 1.32 | 22.12 | 29.54 | 0.00 | 0.51 |
| 6 | -0.15 | 0.58 | 0.00 | 0.35 | 0.11 | 1.10 | 1.26 | 25.81 | 27.08 | 0.28 | 0.42 |
| 7 | -0.17 | 0.51 | 0.00 | 0.40 | 0.12 | 1.07 | 1.57 | 21.04 | 24.38 | 0.25 | 0.62 |
| 8 | -0.08 | 0.62 | 0.03 | 0.17 | 0.11 | 0.39 | 1.31 | 24.12 | 23.55 | 4.23 | 10.44 |
| 9 | -0.17 | 0.57 | 0.00 | 0.28 | 0.10 | 1.17 | 1.26 | 21.66 | 29.38 | 2.34 | 0.14 |
| 10 | -0.13 | 0.62 | 0.00 | 0.27 | 0.08 | 1.16 | 1.07 | 25.50 | 32.85 | 0.29 | 0.29 |
| 11 | 0.80 | 1.01 | 0.00 | 0.03 | 0.01 | 1.23 | 0.06 | 44.11 | 50.95 | 2.09 | 1.71 |
| 12 | -0.17 | 0.51 | 0.01 | 0.33 | 0.12 | 0.92 | 1.61 | 18.84 | 25.42 | 0.52 | 1.55 |
| 13 | -0.11 | 0.60 | 0.01 | 0.28 | 0.11 | 0.82 | 1.24 | 23.20 | 27.95 | 2.74 | 3.60 |
| 14 | -0.11 | 0.63 | 0.01 | 0.28 | 0.10 | 0.81 | 1.08 | 24.03 | 31.62 | 2.58 | 1.86 |
| 15 | -0.05 | 0.64 | 0.01 | 0.18 | 0.11 | 0.47 | 1.17 | 26.34 | 26.48 | 2.89 | 8.97 |
| 16 | -0.01 | 0.69 | 0.02 | 0.21 | 0.11 | 0.55 | 0.92 | 27.48 | 31.95 | 3.81 | 9.60 |
| 17 | -0.13 | 0.57 | 0.01 | 0.30 | 0.13 | 0.65 | 1.42 | 20.42 | 27.49 | 1.31 | 3.27 |
| 18 | -0.08 | 0.61 | 0.02 | 0.28 | 0.15 | 0.56 | 1.20 | 24.90 | 25.86 | 3.99 | 12.93 |
| 19 | -0.09 | 0.63 | 0.01 | 0.23 | 0.10 | 0.86 | 1.12 | 26.55 | 30.24 | 0.29 | 1.92 |
| 20 | -0.14 | 0.56 | 0.02 | 0.32 | 0.13 | 0.81 | 1.41 | 21.97 | 26.44 | 0.89 | 3.07 |
| 21 | -0.15 | 0.55 | 0.02 | 0.26 | 0.12 | 0.66 | 1.48 | 20.50 | 25.54 | 1.56 | 4.92 |
| 22 | -0.03 | 0.66 | 0.00 | 0.24 | 0.08 | 0.78 | 1.05 | 24.37 | 34.56 | 0.15 | 3.55 |
| 23 | -0.12 | 0.59 | 0.02 | 0.27 | 0.12 | 0.61 | 1.36 | 20.93 | 26.86 | 2.90 | 7.31 |

*Table 14. Cont. Height and forest structure metrics derived from LiDAR point clouds (f).*

| PLOT | Lkurt | Lcoefvar | lad_min | lad_max | lad_mean | lad_cv | lad_sum | pz_below_0 | pz_0.0.15 | pz_0.15.2 | pz_2.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 24 | -0.09 | 0.65 | 0.00 | 0.26 | 0.10 | 1.00 | 1.01 | 29.42 | 30.18 | 0.61 | 3.05 |
| 25 | -0.12 | 0.62 | 0.01 | 0.31 | 0.11 | 0.92 | 1.10 | 22.79 | 31.49 | 3.34 | 2.03 |
| 26 | -0.10 | 0.60 | 0.01 | 0.21 | 0.11 | 0.52 | 1.33 | 21.78 | 27.12 | 2.47 | 8.63 |
| 27 | -0.12 | 0.58 | 0.01 | 0.32 | 0.13 | 0.72 | 1.32 | 18.87 | 26.39 | 6.33 | 7.65 |
| 28 | -0.16 | 0.56 | 0.00 | 0.43 | 0.13 | 0.98 | 1.39 | 20.46 | 26.94 | 2.41 | 5.97 |
| 29 | -0.15 | 0.53 | 0.00 | 0.28 | 0.14 | 0.64 | 1.64 | 18.41 | 23.96 | 1.64 | 6.94 |
| 30 | -0.16 | 0.57 | 0.00 | 0.39 | 0.13 | 1.06 | 1.26 | 21.79 | 27.01 | 4.41 | 1.47 |
| 31 | -0.13 | 0.51 | 0.02 | 0.35 | 0.15 | 0.78 | 1.68 | 18.85 | 21.10 | 3.25 | 5.99 |
| 32 | -0.11 | 0.57 | 0.01 | 0.26 | 0.13 | 0.62 | 1.44 | 22.84 | 24.73 | 1.08 | 8.38 |
| 33 | -0.11 | 0.61 | 0.00 | 0.19 | 0.11 | 0.62 | 1.24 | 25.42 | 26.60 | 1.69 | 4.95 |
| 34 | -0.16 | 0.48 | 0.00 | 0.45 | 0.12 | 1.11 | 1.75 | 20.88 | 20.88 | 0.00 | 1.10 |

*Note: **Lkurt**, L-moment-based kurtosis (L-kurtosis); **Lcoefvar**, L-moment-based coefficient of variation; **lad_min**, minimum value of the leaf area density (LAD) profile; **lad_max**, maximum value of the LAD profile; **lad_mean**, mean value of the LAD profile; **lad_cv**, coefficient of variation of the LAD profile; **lad_sum**, total sum of LAD values; **pz_below_0**, proportion of returns with negative height (below ground level); **pz_0.0.15**, proportion of returns between 0 and 0.15 meters; **pz_0.15.2**, proportion of returns between 0.15 and 2 meters; **pz_2.5**, proportion of returns above 2.5 meters.

*Table 15. Height and forest structure metrics derived from LiDAR point clouds (g).*

| PLOT | pz_5.10 | pz_10.20 | pz_20.30 | pz_above_30 | n_first | n_intermediate | n_last | n_single | n_multiple | p_first | p_intermidiate |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 17.58 | 5.81 | 0.00 | 0.00 | 511.00 | 5.00 | 104.00 | 406.00 | 214.00 | 82.42 | 0.81 |
| 2 | 34.10 | 8.39 | 0.00 | 0.00 | 519.00 | 10.00 | 210.00 | 305.00 | 434.00 | 70.23 | 1.35 |
| 3 | 30.42 | 16.39 | 0.00 | 0.00 | 479.00 | 13.00 | 228.00 | 261.00 | 459.00 | 66.53 | 1.81 |
| 4 | 32.75 | 3.96 | 0.00 | 0.00 | 507.00 | 10.00 | 164.00 | 343.00 | 338.00 | 74.45 | 1.47 |
| 5 | 24.94 | 22.89 | 0.00 | 0.00 | 511.00 | 15.00 | 256.00 | 259.00 | 523.00 | 65.35 | 1.92 |
| 6 | 31.31 | 15.09 | 0.00 | 0.00 | 510.00 | 6.00 | 193.00 | 319.00 | 390.00 | 71.93 | 0.85 |
| 7 | 15.47 | 38.24 | 0.00 | 0.00 | 535.00 | 16.00 | 257.00 | 288.00 | 520.00 | 66.21 | 1.98 |
| 8 | 19.32 | 18.34 | 0.00 | 0.00 | 482.00 | 16.00 | 211.00 | 277.00 | 432.00 | 67.98 | 2.26 |
| 9 | 10.76 | 35.72 | 0.00 | 0.00 | 503.00 | 9.00 | 213.00 | 284.00 | 441.00 | 69.38 | 1.24 |
| 10 | 8.65 | 32.42 | 0.00 | 0.00 | 520.00 | 5.00 | 169.00 | 346.00 | 348.00 | 74.93 | 0.72 |
| 11 | 1.14 | 0.00 | 0.00 | 0.00 | 511.00 | 3.00 | 12.00 | 503.00 | 23.00 | 97.15 | 0.57 |
| 12 | 16.77 | 36.90 | 0.00 | 0.00 | 530.00 | 16.00 | 229.00 | 301.00 | 474.00 | 68.39 | 2.06 |
| 13 | 37.03 | 5.48 | 0.00 | 0.00 | 494.00 | 11.00 | 189.00 | 306.00 | 388.00 | 71.18 | 1.59 |
| 14 | 23.46 | 16.45 | 0.00 | 0.00 | 498.00 | 11.00 | 190.00 | 310.00 | 389.00 | 71.24 | 1.57 |
| 15 | 27.50 | 7.81 | 0.00 | 0.00 | 517.00 | 16.00 | 158.00 | 358.00 | 333.00 | 74.82 | 2.32 |
| 16 | 27.15 | 0.00 | 0.00 | 0.00 | 512.00 | 7.00 | 85.00 | 425.00 | 179.00 | 84.77 | 1.16 |
| 17 | 32.07 | 15.45 | 0.00 | 0.00 | 517.00 | 19.00 | 228.00 | 290.00 | 474.00 | 67.67 | 2.49 |
| 18 | 32.32 | 0.00 | 0.00 | 0.00 | 504.00 | 8.00 | 215.00 | 288.00 | 439.00 | 69.33 | 1.10 |
| 19 | 33.19 | 7.82 | 0.00 | 0.00 | 506.00 | 7.00 | 165.00 | 343.00 | 335.00 | 74.63 | 1.03 |
| 20 | 37.55 | 10.09 | 0.00 | 0.00 | 508.00 | 20.00 | 255.00 | 255.00 | 528.00 | 64.88 | 2.55 |
| 21 | 22.54 | 24.94 | 0.00 | 0.00 | 521.00 | 26.00 | 287.00 | 229.00 | 605.00 | 62.47 | 3.12 |
| 22 | 26.29 | 11.08 | 0.00 | 0.00 | 534.00 | 11.00 | 132.00 | 398.00 | 279.00 | 78.88 | 1.62 |
| 23 | 32.79 | 9.21 | 0.00 | 0.00 | 527.00 | 10.00 | 256.00 | 270.00 | 523.00 | 66.46 | 1.26 |

*Table 16. Cont. Height and forest structure metrics derived from LiDAR point clouds (h).*

| PLOT | pz_5.10 | pz_10.20 | pz_20.30 | pz_above_30 | n_first | n_intermediate | n_last | n_single | n_multiple | p_first | p_intermidiate |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 24 | 35.06 | 1.68 | 0.00 | 0.00 | 498.00 | 4.00 | 154.00 | 343.00 | 313.00 | 75.91 | 0.61 |
| 25 | 33.82 | 6.53 | 0.00 | 0.00 | 511.00 | 3.00 | 175.00 | 337.00 | 352.00 | 74.17 | 0.44 |
| 26 | 28.22 | 11.78 | 0.00 | 0.00 | 505.00 | 17.00 | 208.00 | 293.00 | 437.00 | 69.18 | 2.33 |
| 27 | 39.18 | 1.58 | 0.00 | 0.00 | 517.00 | 8.00 | 233.00 | 281.00 | 477.00 | 68.21 | 1.06 |
| 28 | 39.26 | 4.96 | 0.00 | 0.00 | 525.00 | 9.00 | 253.00 | 273.00 | 514.00 | 66.71 | 1.14 |
| 29 | 32.16 | 16.90 | 0.00 | 0.00 | 524.00 | 22.00 | 247.00 | 276.00 | 517.00 | 66.08 | 2.77 |
| 30 | 36.90 | 8.42 | 0.00 | 0.00 | 499.00 | 9.00 | 240.00 | 262.00 | 486.00 | 66.71 | 1.20 |
| 31 | 44.94 | 5.87 | 0.00 | 0.00 | 512.00 | 21.00 | 268.00 | 241.00 | 560.00 | 63.92 | 2.62 |
| 32 | 36.35 | 6.62 | 0.00 | 0.00 | 498.00 | 26.00 | 216.00 | 279.00 | 461.00 | 67.30 | 3.51 |
| 33 | 31.03 | 10.30 | 0.00 | 0.00 | 537.00 | 11.00 | 219.00 | 318.00 | 449.00 | 70.01 | 1.43 |
| 34 | 13.19 | 43.96 | 0.00 | 0.00 | 123.00 | 4.00 | 55.00 | 56.00 | 126.00 | 67.58 | 2.20 |

*Note: **pz_5.10**, proportion of returns between 5 and 10 meters in height;**pz_10.20**, proportion of returns between 10 and 20 meters; **pz_20.30**, proportion of returns between 20 and 30 meters; **pz_above_30**, proportion of returns above 30 meters; **n_first**, number of returns classified as first of each pulse; **n_intermediate**, number of intermediate returns; **n_last**, number of returns classified as last of each pulse; **n_single**, number of single returns per pulse (only one return recorded); **n_multiple**, number of pulses with multiple returns; **p_first**, proportion of returns that were first; **p_intermidiate**, proportion of returns that were intermediate.

*Table 17. Height and forest structure metrics derived from LiDAR point clouds (i).*

| PLOT | p_last | p_single | p_multiple | ratio_last_first | ratio_intermediate_first | ratio_multiple_single | n_return_1 | n_return_2 | n_return_3 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 16.77 | 65.48 | 34.52 | 0.20 | 0.01 | 0.53 | 511.00 | 104.00 | 5.00 |
| 2 | 28.42 | 41.27 | 58.73 | 0.40 | 0.02 | 1.42 | 519.00 | 210.00 | 10.00 |
| 3 | 31.67 | 36.25 | 63.75 | 0.48 | 0.03 | 1.76 | 479.00 | 228.00 | 13.00 |
| 4 | 24.08 | 50.37 | 49.63 | 0.32 | 0.02 | 0.99 | 507.00 | 165.00 | 9.00 |
| 5 | 32.74 | 33.12 | 66.88 | 0.50 | 0.03 | 2.02 | 511.00 | 256.00 | 15.00 |
| 6 | 27.22 | 44.99 | 55.01 | 0.38 | 0.01 | 1.22 | 510.00 | 193.00 | 6.00 |
| 7 | 31.81 | 35.64 | 64.36 | 0.48 | 0.03 | 1.81 | 535.00 | 259.00 | 14.00 |
| 8 | 29.76 | 39.07 | 60.93 | 0.44 | 0.03 | 1.56 | 482.00 | 209.00 | 18.00 |
| 9 | 29.38 | 39.17 | 60.83 | 0.42 | 0.02 | 1.55 | 503.00 | 213.00 | 9.00 |
| 10 | 24.35 | 49.86 | 50.14 | 0.33 | 0.01 | 1.01 | 520.00 | 169.00 | 5.00 |
| 11 | 2.28 | 95.63 | 4.37 | 0.02 | 0.01 | 0.05 | 511.00 | 12.00 | 3.00 |
| 12 | 29.55 | 38.84 | 61.16 | 0.43 | 0.03 | 1.57 | 530.00 | 227.00 | 18.00 |
| 13 | 27.23 | 44.09 | 55.91 | 0.38 | 0.02 | 1.27 | 494.00 | 189.00 | 11.00 |
| 14 | 27.18 | 44.35 | 55.65 | 0.38 | 0.02 | 1.25 | 498.00 | 190.00 | 11.00 |
| 15 | 22.87 | 51.81 | 48.19 | 0.31 | 0.03 | 0.93 | 517.00 | 157.00 | 17.00 |
| 16 | 14.07 | 70.36 | 29.64 | 0.17 | 0.01 | 0.42 | 512.00 | 85.00 | 7.00 |
| 17 | 29.84 | 37.96 | 62.04 | 0.44 | 0.04 | 1.63 | 517.00 | 229.00 | 18.00 |
| 18 | 29.57 | 39.61 | 60.39 | 0.43 | 0.02 | 1.52 | 504.00 | 215.00 | 8.00 |
| 19 | 24.34 | 50.59 | 49.41 | 0.33 | 0.01 | 0.98 | 506.00 | 165.00 | 7.00 |
| 20 | 32.57 | 32.57 | 67.43 | 0.50 | 0.04 | 2.07 | 508.00 | 255.00 | 20.00 |
| 21 | 34.41 | 27.46 | 72.54 | 0.55 | 0.05 | 2.64 | 521.00 | 287.00 | 26.00 |
| 22 | 19.50 | 58.79 | 41.21 | 0.25 | 0.02 | 0.70 | 534.00 | 132.00 | 11.00 |
| 23 | 32.28 | 34.05 | 65.95 | 0.49 | 0.02 | 1.94 | 527.00 | 255.00 | 11.00 |
| 24 | 23.48 | 52.29 | 47.71 | 0.31 | 0.01 | 0.91 | 498.00 | 154.00 | 4.00 |

*Table 17. Cont. Height and forest structure metrics derived from LiDAR point clouds (i).*

| PLOT | p_last | p_single | p_multiple | ratio_last_first | ratio_intermediate_first | ratio_multiple_single | n_return_1 | n_return_2 | n_return_3 |
|---|---|---|---|---|---|---|---|---|---|
| 25 | 25.40 | 48.91 | 51.09 | 0.34 | 0.01 | 1.04 | 511.00 | 175.00 | 3.00 |
| 26 | 28.49 | 40.14 | 59.86 | 0.41 | 0.03 | 1.49 | 505.00 | 208.00 | 16.00 |
| 27 | 30.74 | 37.07 | 62.93 | 0.45 | 0.02 | 1.70 | 517.00 | 233.00 | 8.00 |
| 28 | 32.15 | 34.69 | 65.31 | 0.48 | 0.02 | 1.88 | 525.00 | 253.00 | 9.00 |
| 29 | 31.15 | 34.80 | 65.20 | 0.47 | 0.04 | 1.87 | 524.00 | 247.00 | 22.00 |
| 30 | 32.09 | 35.03 | 64.97 | 0.48 | 0.02 | 1.85 | 499.00 | 239.00 | 10.00 |
| 31 | 33.46 | 30.09 | 69.91 | 0.52 | 0.04 | 2.32 | 512.00 | 268.00 | 21.00 |
| 32 | 29.19 | 37.70 | 62.30 | 0.43 | 0.05 | 1.65 | 498.00 | 214.00 | 28.00 |
| 33 | 28.55 | 41.46 | 58.54 | 0.41 | 0.02 | 1.41 | 537.00 | 218.00 | 12.00 |
| 34 | 30.22 | 30.77 | 69.23 | 0.45 | 0.03 | 2.25 | 123.00 | 57.00 | 2.00 |

*Note: **p_last**, **p_single**, **p_multiple**: proportions of last, single, and multiple returns; **ratio_last_first**, **ratio_intermediate_first**, **ratio_multiple_single**: ratios between return types; **n_return_1**, **n_return_2**, **n_return_3**: count of first-, second-, and third-order returns per pulse.

## APPENDIX 5. SCRIPT USED IN SOC MODELING

```
# ----------------------------------------------------
# SOIL ORGANIC CARBON (SOC) MODELING
# ----------------------------------------------------

# 1. Load libraries and data

# Install and load required packages
# install.packages("pacman")
library(pacman)
p_load(caret, tidyverse, janitor, readxl, tidymodels, dplyr, rsample, parsnip, ranger)

# Load the data
data <- read.csv("E:/DATA.csv") %>%
  select(-id) %>%
  remove_empty("cols") %>%
  drop_na()

# 2. Advanced preprocessing of variables

# Remove the response variable to filter predictors
data_split <- data %>%
  select(-SOC)

# Remove low variance variables (< 0.01)
data_filtered <- data_split %>%
  select(where(~ var(.) > 0.01))

# Remove highly correlated variables (> 0.9)
data_cor <- cor(data_filtered)
high_corr <- findCorrelation(data_cor, cutoff = 0.9)
data_uncorr <- data_filtered %>%
  dplyr::select(-high_corr)

# Normalize variables and reattach SOC
data_final <- data_uncorr %>%
  mutate(across(everything(), scale)) %>%
  as.data.frame() %>%
  bind_cols(COS = data$COS)

# 3. Split data into training and testing sets
set.seed(123)

split <- initial_split(data_final, prop = 0.75)
```

```
train <- training(split)
test <- testing(split)

# 4. Model definitions

# Simple linear model
linear_model <- linear_reg() %>%
  set_engine("lm")

# Polynomial model (degree 2)
poly_recipe <- recipe(COS ~ ., data = train) %>%
  step_normalize(all_predictors()) %>%
  step_poly(all_predictors(), degree = 2, options = list(raw = TRUE))

poly_model <- linear_reg() %>%
  set_engine("lm")

# Logarithmic model
log_recipe <- recipe(COS ~ ., data = train) %>%
  step_log(all_predictors(), offset = 1)

log_model <- linear_reg() %>%
  set_engine("lm")

# Random Forest model
rf_model <- rand_forest(mtry = tune(), trees = 500, min_n = tune()) %>%
  set_engine("ranger") %>%
  set_mode("regression")

# 5. Cross-validation
set.seed(123)
cv <- vfold_cv(train, v = 10)

# 6. Workflows and model evaluation

# Linear
linear_wf <- workflow() %>%
  add_model(linear_model) %>%
  add_formula(COS ~ .)

linear_res <- fit_resamples(linear_wf, cv, metrics = metric_set(rmse, rsq, mae))

# Polynomial
poly_wf <- workflow() %>%
  add_model(poly_model) %>%
  add_recipe(poly_recipe)
```

```
poly_res <- fit_resamples(poly_wf, cv, metrics = metric_set(rmse, rsq, mae))

# Logarithmic
log_wf <- workflow() %>%
  add_model(log_model) %>%
  add_recipe(log_recipe)

log_res <- fit_resamples(log_wf, cv, metrics = metric_set(rmse, rsq, mae))

# Random Forest with tuning
rf_wf <- workflow() %>%
  add_model(rf_model) %>%
  add_formula(COS ~ .)

rf_grid <- grid_random(mtry(range = c(2, ncol(train) - 1)),
                       min_n(range = c(2, 10)),
                       size = 20)

rf_res <- tune_grid(rf_wf, resamples = cv, grid = rf_grid, metrics = metric_set(rmse,
rsq, mae))

# Select best RF model
rf_best <- select_best(rf_res, metric = "rmse")

# 7. Model comparison
linear_metrics <- collect_metrics(linear_res) %>% mutate(Model = "Linear")
poly_metrics <- collect_metrics(poly_res) %>% mutate(Model = "Polynomial")
log_metrics <- collect_metrics(log_res) %>% mutate(Model = "Logarithmic")
rf_metrics <- collect_metrics(rf_res) %>%
  filter(mtry == rf_best$mtry & min_n == rf_best$min_n) %>%
  mutate(Model = "Random Forest")

final_metrics <- bind_rows(linear_metrics, poly_metrics, log_metrics, rf_metrics) %>%
  select(Model, .metric, mean, std_err) %>%
  arrange(.metric, mean)

print(final_metrics)

# 8. Final evaluation of the best model on test set

# Train final model
final_rf <- rf_wf %>%
  finalize_workflow(rf_best) %>%
  fit(data = train)
```

```
# Predict on test set
pred_rf <- predict(final_rf, test) %>%
  bind_cols(test)

# Final evaluation
final_metrics <- pred_rf %>%
  metrics(truth = SOC, estimate = .pred)

print(final_metrics)

# ------------------------------------------------
# 9. Visualization of results
# ------------------------------------------------

library(ggplot2)

# ------------------------------------------------
# Plot 1: Observed vs Predicted SOC - All 34 points (scale from 5 to 55, labels 10 to 50)
# ------------------------------------------------

# Fit the final model using all available data
final_rf_34 <- rf_wf %>%
  finalize_workflow(rf_best) %>%
  fit(data = dados_final)

# Predict SOC for the 34 points
pred_rf_34 <- predict(final_rf_34, new_data = dados_final) %>%
  bind_cols(COS = dados_final$COS)

# Plot with trend line and adjusted axes
ggplot(pred_rf_34, aes(x = SOC, y = .pred)) +
  geom_point(color = "darkgreen", size = 2) +
  geom_smooth(method = "lm", se = FALSE, color = "black", linetype = "dotted",
linewidth = 1) +
  scale_x_continuous(limits = c(5, 55), breaks = seq(10, 50, 10)) +
  scale_y_continuous(limits = c(5, 55), breaks = seq(10, 50, 10)) +
  coord_fixed() +
  labs(title = "Graph 3: Observed vs Predicted SOC - Random Forest (34 points)",
       x = "Observed SOC (Mg C / ha)",
       y = "Predicted SOC (Mg C / ha)") +
  theme_minimal(base_size = 13)

# ------------------------------------------------
# Table: Predicted SOC Values
# ------------------------------------------------
```

```
# Add point identifier
pred_rf_34 <- pred_rf_34 %>%
  mutate(Point = row_number())

# Create table with only predicted values
soc_table <- pred_rf_34 %>%
  select(Point, Predicted_SOC = .pred) %>%
  arrange(Point)

# Display in console
print(soc_table)


# OBTAINING THE RASTER WITH THE SOC PREDICTION

library(terra)

# Read the Excel file
M_w2w <- readxl::read_excel("C:/data_pixels.xlsx")
M_w2w <- M_w2w %>%
  dplyr::select(-PLOT, -Z) %>%  # Remove the columns PLOT and Z
  dplyr::mutate(across(everything(), as.numeric))  # Convert all columns to numeric

# Convert to point object with all variables
pts <- vect(M_w2w, geom = c("X", "Y"), crs = "EPSG:25830")  # Use the correct
EPSG for your system
plot(pts)

# Define the extent of the area and resolution (based on the data)
resolution <- min(diff(sort(unique(M_w2w$X))))  # assuming regular spacing

# Convert to SpatRaster with multiple layers
# Create SpatRaster object from points
r_template <- rast(
  extent = ext(min(M_w2w$X), max(M_w2w$X), min(M_w2w$Y), max(M_w2w$Y)),
  resolution = resolution,
  crs = "EPSG:25830"
)
# Select only the variable names (excluding X and Y)
vars <- setdiff(names(M_w2w), c("X", "Y"))

# Rasterize all variables
r_stack <- rast(lapply(vars, function(v) {
  rasterize(pts, r_template, field = v)
}))
```

```
# Assign names to the raster bands
names(r_stack) <- vars

# Plot the values
Gpred <- raster::predict(r_stack, final_rf)

# Load the Region of Interest (ROI) shapefile
ROI <- vect("C:/ 02-Vectors/SHP/FOREST_AREA.shp")

# Mask the prediction using the ROI
Gpre <- raster::mask(Gpred, ROI)

# Write the predicted values to a raster file
writeRaster(Gpre, file.path("Predict_FULL.tif"), overwrite = TRUE)
```