remote sensing

MDPI

Article

Multisensor Modeling of Old-Growthness in Mediterranean Pine Forests: New Insights from High-Density Airborne Laser Scanning and Satellite Integration

Carlos Guerrero-Cubero, Anabel Calzado Carretero, Cristina Pérez-Carral, Reyes Alejano-Monge and Javier Vázquez-Piqué *

Department of Agroforestry Sciences, University of Huelva, 21007 Huelva, Spain; carlos.guerrero@dcaf.uhu.es (C.G.-C.); carrete@dcaf.uhu.es (A.C.C.); cpcarral@dcaf.uhu.es (C.P.-C.); ralejan@dcaf.uhu.es (R.A.-M.)
* Correspondence: jpique@uhu.es

Highlights

What are the main findings?

- High-density Airborne Laser Scanning (ALS) accounted for most of the predictive accuracy in modeling the Old-Growth Index (OGI) in Mediterranean *Pinus nigra* forests.
- Support Vector Regression (SVR) achieved the highest overall predictive accuracy when combined with the predictor set ALS + Canopy Cover Fraction (CCF) + WorldView-3 (nested $R^2 = 0.79$), outperforming both Random Forest (RF) and Multiple Linear Regression (MLR).
- Models relying exclusively on ALS metrics ($R^2 = 0.71$) or ALS with CCF ($R^2 = 0.77$) yielded statistically comparable performance to models incorporating commercial satellite imagery.

What are the implications of the main findings?

- Neither commercial ultra-high-resolution imagery (WorldView-3) nor free high-resolution optical data (Sentinel-2) provided a clear gain over national ALS and orthophoto data.
- Open-access national LiDAR programs (such as Spanish PNOA) provide a cost-effective, transparent, and scalable framework for mapping forest structural maturity.
- Continuous spatial predictions of OGI offer actionable baseline maps to identify, protect, and monitor candidate old-growth forests under the EU Biodiversity Strategy for 2030.

Abstract

Old-growth forests (OGFs) are conservation priorities under the EU Biodiversity Strategy for 2030, but identifying relict stands in Mediterranean *Pinus nigra* (black pine) forests is challenging due to structural complexity. We evaluated a multisensor remote sensing framework to estimate an Old-Growth Index (OGI) in Cazorla Natural Park (Spain) using high-density Airborne Laser Scanning (ALS) data from the Spanish National Plan for Aerial Orthophotography (PNOA-LiDAR, 2024), RGB–NIR orthophotographs, and multispectral imagery (Sentinel-2 and WorldView-3). Four predictor sets (ALS metrics only; ALS plus canopy cover fraction (CCF); ALS + CCF + Sentinel-2; and ALS + CCF + WorldView-3) were evaluated across Multiple Linear Regression, Random Forest, and Support Vector Regression (SVR) using repeated nested cross-validation. SVR combined with the predictor set ALS + CCF + WorldView-3 achieved the highest nested R^2 (0.79), outperforming Random Forest and Multiple Linear Regression. However, models relying solely on ALS

Check for updates

Academic Editor: Sandra Eckert

Received: 27 August 2026
Revised: 25 October 2026
Accepted: 26 October 2026
Published: 30 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

(R^2 = 0.71) or ALS with CCF (R^2 = 0.77) yielded statistically comparable performance. Furthermore, integrating Sentinel-2 (R^2 = 0.61–0.68) provided no clear gain over ALS-based models. These findings demonstrate that open-access national LiDAR programs provide sufficient 3D structural information to map stand maturity without requiring commercial satellite imagery. Continuous spatial predictions of OGI offer valuable baseline maps to support conservation planning and OGF monitoring in Mediterranean pine ecosystems.

Keywords: Old-Growth Index; Sentinel-2; WorldView-3; airborne laser scanning; Mediterranean forest; Support Vector Regression; Random Forest

1. Introduction

Old-growth forests (OGFs) have attracted significant scientific and policy attention due to their exceptionally high biodiversity and provision of unique habitats and landscapes [1,2]. This interest is further heightened by their extreme scarcity, as OGFs occupy less than 1–3% of Europe's forest area in a highly fragmented state [3]. Consequently, their strict protection has become a cornerstone of the European policy agenda, explicitly mandated by the European Union (EU) Biodiversity Strategy for 2030, which calls for their definition, mapping, and long-term monitoring [4]. Beyond their biodiversity value, a key characteristic of OGFs is their role as long-lasting carbon sinks; as they age, they store significant and stable carbon pools in both aboveground biomass and in soils, making their conservation a more reliable long-term climate mitigation strategy than short-rotation forestry [5,6]. Nevertheless, climate-driven tree mortality can potentially shift these forests into net carbon sources, as recently documented in moist tropical forests [7].

OGFs develop under rare environmental conditions requiring long periods without major disturbances, leading to structural complexity and ecological functions often absent in managed stands [8]. Nowadays, these conditions are becoming even more scarce due to climate change, which is altering disturbance regimes and increasing the frequency of extreme events [9]. One of the central challenges in the management and conservation of OGFs is establishing a definition that is both scientifically robust and operationally applicable. OGFs are defined by a distinct set of structural attributes, including the presence of large and old trees, high variability in tree diameters, a significant density of very large trees, a higher incidence of tree decadence (e.g., top breakage), large accumulations of logs in advanced decay stages, and an understory with greater herbaceous cover and shade-tolerant regeneration [2,10]. A widely adopted approach for quantifying this old-growth condition relies on structural indices or metrics that integrate these key attributes [11]. Among these metrics, the development and application of an Old-Growth Index (OGI) (a dimensionless metric derived from structural characteristics such as the density of large trees and size diversity) offers a practical tool for assessing old-growth status [10,12]. Hevia et al. [12] highlighted that selecting the most appropriate OGI for each forest type and predicting its spatial distribution can substantially improve the identification of OGFs across landscapes; furthermore, they developed an OGI for *Pinus nigra* subsp. *salzmannii* (black pine) forests in Southeastern Spain based on mean tree diameter, diameter variability, the density of large trees, and the total stand density. This structural approach is commonly employed in remote sensing studies, where LiDAR-derived structural metrics and spectral indices are used to characterize and map old-growth forests [2,13,14].

Despite their ecological importance, studies on OGFs in the Mediterranean region remain scarce, as most research has focused on temperate and boreal forests [12,15]. This knowledge gap is particularly critical for Spanish forests of black pine, which have undergone a long history of anthropogenic intervention [16]; today, OGFs persist solely as relict stands

in remote, inaccessible mountainous areas [11]. These forests constitute an extremely rare and highly fragile resource whose conservation urgently requires enhanced monitoring strategies [11,17,18].

Remote sensing-based inventory techniques offer an effective approach to map OGFs. This is particularly relevant because traditional field inventories often fail to capture the structural complexity of late-successional stages due to their limited spatial coverage, the high cost and time required for extensive sampling, and difficulty in measuring fine-scale structural attributes such as canopy gaps and large trees. Furthermore, field sampling is challenging to implement in the remote, inaccessible areas where OGFs typically persist [12,15]. Remote sensing therefore provides unique opportunities to improve OGF detection and characterization. Technologies such as Airborne Laser Scanning (ALS), orthophotos, and satellite imagery have demonstrated strong potential for mapping key forest attributes [2,12,13].

Each remote sensing platform offers complementary strengths and limitations for characterizing OGFs. High-resolution satellite imagery (e.g., Sentinel-2) offers free, frequent, and systematic global coverage, enabling consistent multi-temporal analysis across large regions. However, its spatial resolution (10–20 m) is insufficient to capture the fine-scale structural attributes that characterize old-growth forests [13]. In contrast, ultra-high-resolution satellite data (e.g., WorldView-3) delivers the detailed spectral information and spatial precision needed to analyze canopy heterogeneity and detect subtle structural signatures; nevertheless, its coverage is often limited, and the high acquisition costs constrain its operational use over large areas [15]. Ultra-high-resolution aerial orthophotographs (e.g., RGB-NIR imagery from the Spanish National Aerial Orthophotography Plan, PNOA; 0.25 m/pixel) represent an intermediate alternative, combining extensive regional coverage with sufficient spatial detail to characterize canopy cover and delineate individual tree crowns through spectral indices or object-based image analysis [19]. Finally, Airborne Laser Scanning (ALS) excels at quantifying three-dimensional structural metrics such as canopy height, vertical complexity and surface roughness, which serve as direct proxies for old-growth structural attributes [20]. Nevertheless, ALS acquisition campaigns have traditionally been project-based with low temporal frequency and substantial logistical and financial costs [21]; however, national programs in countries like Spain now provide systematic and periodic ALS coverage.

A critical advancement for regional-scale analyses is the availability of systematic national ALS programs. Spain's PNOA LiDAR 2024 provides ALS data as a periodic, publicly accessible resource with a relatively high point density ($\geq$5 pts/m^2). Initiated in 2009, the PNOA LiDAR program has completed two full acquisition cycles (approximately 2009–2015 and 2015–2021), with a third cycle currently in progress and its most recent update in 2024. This periodic acquisition scheme overcomes the traditional limitations of project-based, spatially fragmented ALS campaigns.

The recent literature consistently shows that overcoming the limitations of individual sensors requires the integration of multiple data sources. Single-sensor approaches (such as relying exclusively on optical imagery) are insufficient to capture the structural complexity that characterizes OGFs [22]. In contrast, the combination of ALS with spectral data has emerged as the predominant paradigm. This combined approach has been shown to enhance not only the identification and mapping of OGFs [23–25] but also the modeling of key attributes relevant for their conservation, including aboveground biomass and carbon stocks [26], tree species diversity [27,28], and successional status [29].

Moreover, the only study mapping old-growthness in Mediterranean pine forests used low-density ALS data (~0.5 pts/m^2) and no satellite imagery [12]. Although studies combining ALS with satellite imagery have been conducted in temperate, boreal and trop-

ical forests [21,23–25], which differ markedly in structure and species composition from Mediterranean pine forests, recent reviews still point to the limited number of remote sensing studies on OGFs and emphasize the need to combine data sources [15]. Consequently, it remains unknown whether integrating high-density ALS with satellite imagery of different spatial resolutions improves the prediction of old-growthness in these ecosystems.

Here, we integrate spectral indices derived from ultra-high-resolution and high-resolution satellite imagery, together with forest structural indices obtained from ALS and orthophotographs, to model the OGI in *Pinus nigra* forests of Cazorla Natural Park (SE Spain). We test three hypotheses: (1) combining information from ultra-high-resolution satellite imagery, ALS, and orthophotos delivers a reliable framework for mapping stand maturity and identifying priority OGFs; (2) freely available high-resolution satellite imagery yields lower OGI predictive accuracy than ultra-high-resolution imagery; and (3) while ALS metrics alone capture most OGI variability by directly measuring canopy structure, adding canopy cover and spectral data further refines predictions by accounting for canopy condition.

2. Materials and Methods

2.1. Study Area

This study was conducted in the southern sector of the Sierra de Cazorla, Segura y Las Villas Natural Park (NP) (Figure 1). The NP covers 209,927.2 ha, and the study area encompasses 21,820.4 ha (10.4% of the NP). The NP is located in northeastern Andalusia, Jaén province (37°51′N, 2°52′W). The climate is Mediterranean, characterized by intense summer droughts and pronounced annual and intra-annual precipitation variability, with an average annual rainfall of 1100 mm and a mean annual temperature of 11.7 °C. Conifers dominate the tree layer, with *Pinus halepensis* (Aleppo pine), *P. pinaster* (Maritime pine), and *P. nigra* subsp. *salzmannii* (black pine) being particularly abundant. Among the broadleaved species, *Quercus ilex* (holm oak) and *Q. faginea* (Portuguese oak) are notable, growing at lower elevations and often co-occurring with other species such as *Acer* spp. (maple), *Populus* spp. (poplar), *Sorbus* spp. (whitebeam), and *Fraxinus angustifolia* (narrow-leafed ash). Black pine is the most widespread tree species in the NP, covering approximately 60,000 ha (~29% of the total NP area) at elevations between 1000 and 2000 m.a.s.l. [30]. It is well adapted to nutrient-poor soils and thrives on steep slopes and rocky outcrops where other more demanding species cannot establish [31]. The study area is dominated by Cambisols and Leptosols, with localized occurrences of Calcisols [32], according to the World Reference Base for Soil Resources (WRB) [33]. These soils are characterized by low water-holding capacity, limited nutrient availability and high stoniness, which constrain tree growth and regeneration.

2.2. Field Data and Calculation of OGI

Field data collection was conducted in the southern sector of the NP. The study focused on areas where *P. nigra* is the dominant tree species (>80%). Thirty square plots (0.25 ha each) were established (Figure 1). Data were collected from 15 young plots and 15 old-growth plots; the coordinates and structural characteristics of each plot are detailed in Table S1 (Supplementary Materials). Young stands were selected based on an average diameter at breast height (DBH) of less than 20 cm, while the selection of old-growth stands was based on forest structure, age and low management intensity. The use of large plots (50 × 50 m) is recommended for structural inventories of mature forests, as it ensures the representation of large trees and canopy gaps.

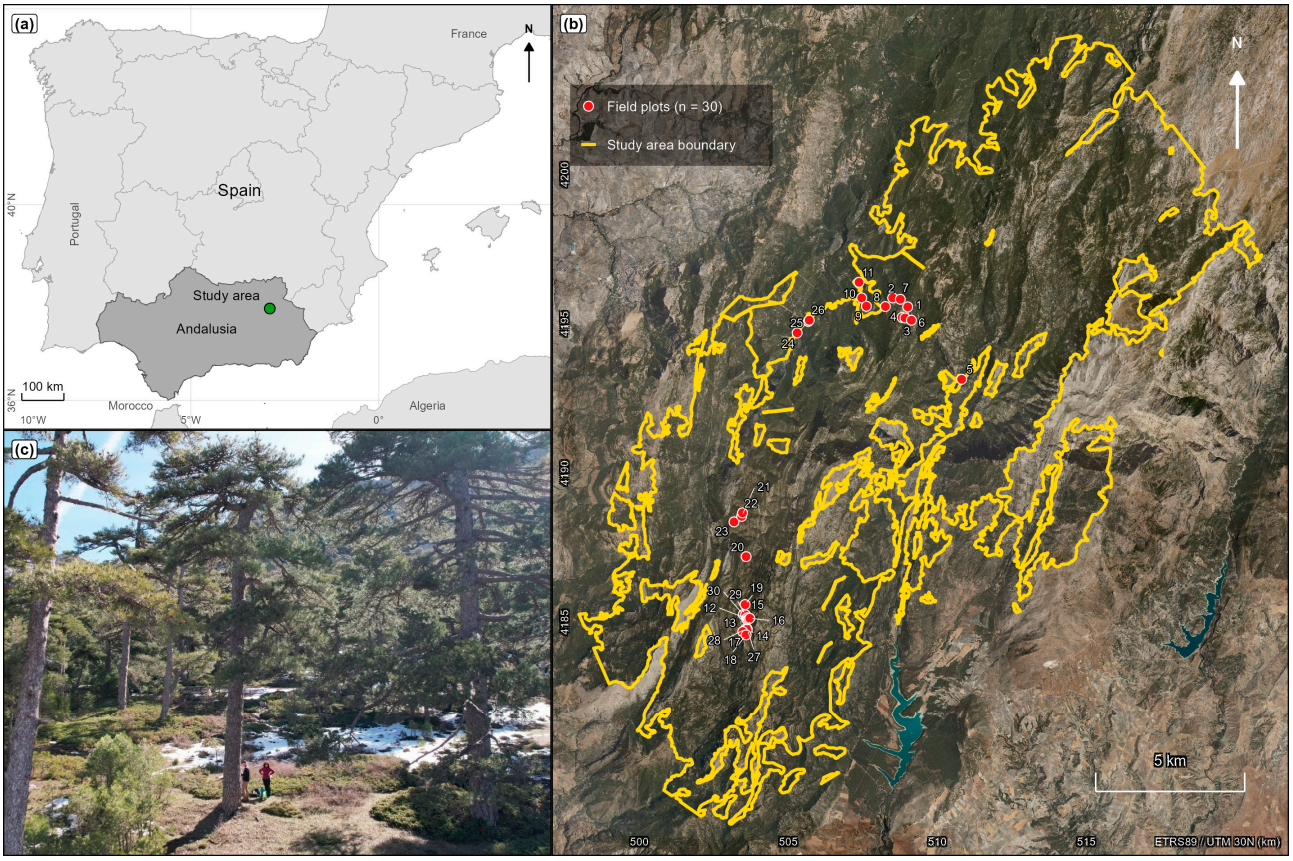


Figure 1. Location of the study area and distribution of field plots. (**a**) Location of the study area (green dot) in Andalusia, Spain; (**b**) boundary of the study area and spatial distribution of the 30 field plots; and (**c**) representative old-growth forest stand within the study area.

Structural data from each plot were acquired using a Terrestrial Laser Scanner (TLS, handheld BLK2GO device, Leica Geosystems AG, Heerbrugg, Switzerland) with an average point density of 420,000 points/m^2 and a range of 25 m (a detailed description of the instrument is available in Dlesk et al. [34]). The scanning of each plot followed the path design proposed by Hunčaga et al. [35], consisting of parallel transects spaced 10 m apart and oriented perpendicular to the maximum slope direction. TLS point cloud processing was conducted using LiDAR360 version 7.2.6.0 (GreenValley International Ltd., Berkeley, CA, USA), and the DBH was generated for each tree within the plot.

The OGI was calculated for each plot using four structural variables [12], based on the index defined by Spies and Franklin [10]: (1) standard deviation of DBH; (2) density of trees with diameter greater than 50 cm (trees/ha); (3) mean DBH; and (4) total tree density (trees/ha).

The threshold of 50 cm DBH was selected because it is widely used in the literature to define large trees in mature forests [36] and because it provided the best discrimination between young and old stands in the study area [12].

These four structural variables were used to calculate the OGI, following Equation (1):

$$\mathrm{OGI} = 25\sum_{i=1}^{i=4}\left|\frac{x_i - x_{i\ young}}{x_{i\ old} - x_{i\ young}}\right| \quad (1)$$

where i corresponds to each of the four structural variables (1–4), x_i is the observed value for the ith structural variable, x_{iyoung} is the characteristic value of the ith structural variable for young stands, and x_{iold} is the characteristic value of the ith structural variable for old stands.

When the observed value of a structural variable in a plot is lower than the characteristic value for young stands, the young-stand value is used. Conversely, when the observed value exceeds that characteristic value for old stands, the old-stand value is applied. As a result, the OGI ranges from 0 (very young stands) to 100 (very old stands).

The characteristic structural values for old stands were obtained from a specific inventory conducted by Hevia et al. [12] in the Cabañas forest, located in the study area. In that inventory, 21 plots of 15 m radius were established in stands where *P. nigra* was the dominant species (>80%) and whose structure, age and low management intensity qualified them as representative old-growth conditions. Reference values for young stands were derived from 18 plots with a mean DBH of less than 20 cm [12].

2.3. Remote Sensing Data

We used four sources of remote sensing data to predict the OGI.

2.3.1. ALS Data (PNOA 2024)

ALS data were obtained from the PNOA LiDAR program (https://centrodedescargas.cnig.es/CentroDescargas/home, accessed on 25 October 2026). We used data from the 2024 flight with a minimum density of 5 points/m^2 and a vertical RMSE lower than 10 cm. Analysis of the 30 sampling plots yielded a mean point density of 19.88 ± 11.15 points/m^2 (range: 7.91–54.37 points/m^2). Point cloud processing was conducted in RStudio (version 2025.09.1+401; Posit Software, Boston, MA, USA) using the lidR package (version 4.2.1) [37]. The point cloud was filtered to retain only returns classified as vegetation with a normalized height between 3 and 32 m, based on field observations. From each plot point cloud, a set of 43 LiDAR metrics (Table 1) was extracted; these metrics have been widely used as predictor variables in forest models based on airborne LiDAR data [38].

Table 1. ALS metrics computed for each plot. h: tree height (m).

ALS Metric	Description
h_{mean}, h_{mode}	mean, mode
h_{min}, h_{max}	minimum, maximum
h_{SD}, h_{CV}	standard deviation, coefficient of variation
h_{SKw}	skewness
h_{Kurt}	kurtosis
h_{IQR}	interquartile range
h_{AAD}	average absolute deviation
$h_{MADmedian}$	median of the absolute deviations from the overall median
$h_{MADmode}$	median of the absolute deviations from the overall mode
h_{L1}, h_{L2}, h_{L3}, h_{L4}	L-moments
h_{Lskw}	L-moments of skewness
h_{Lkur}	L-moments of kurtosis
h_{01}, h_{05}, h_{10}, h_{20}, h_{25}, h_{90}, h_{95}, h_{99}	percentiles
CRR	canopy relief ratio: (mean height–min height)/(max height–min height)
CC	canopy cover: percentage of first returns above 4.5 m/total returns
PARA3	percentage of all returns above 3 m/total all returns

Table 1. *Cont*.

ALS Metric	Description
ARA3.TFR	ratio between all returns above 3 m and total of first returns
PFRAM	percentage of first returns above mean/total all returns
PARAM	percentage of all returns above mean/total all returns
PARAMO	percentage of all returns above mode/total all returns
PERAMO	percentage of first returns above mode/total all returns
ARAM.TFR	ratio between all returns above mean and total of first returns
ARAMO.TFR	ratio between all returns above mode and total of first returns

Note: Skewness measures the asymmetry of the height distribution, whereas kurtosis quantifies the peakedness or flatness of the distribution relative to a normal distribution.

2.3.2. WorldView-3 Imagery

WorldView-3 multispectral images with a spatial resolution of 1.3 m were purchased on 5 February 2025, coinciding with the phenological window when the spectral signature of *Pinus nigra* for NDVI exhibits its highest distinctiveness in Mediterranean environments [39]. The following seven bands were used: B2 (blue, 450–510 nm), B3 (green, 510–580 nm), B4 (yellow, 585–625 nm), B5 (red, 630–690 nm), B6 (red-edge, 705–745 nm), B7 (NIR-1, 770–895 nm), and B8 (NIR-2, 860–1040 nm). From these bands, 12 vegetation indices were calculated, including two exploratory indices (GSI and YNDVI) designed to exploit the yellow band's sensitivity to canopy structure and non-photosynthetic components (Table 2). For each index, the mean, 10th percentile (P10), and 90th percentile (P90) of pixel values were extracted per plot to characterize both the central tendency and within-plot spectral variability.

The formulation of these exploratory indices was guided by the potential of the yellow band (B4: 585–625 nm) to detect biochemical and structural variations associated with forest maturity. This spectral region, absent in most multispectral sensors, has been shown to be sensitive to canopy structure and the spectral influence of shadows and non-photosynthetic elements [40]. Following approaches that demonstrate the value of combining the yellow band with NIR information for estimating traits related to canopy architecture [41,42], we designed specific indices to capture these traits: (1) Gap Soil Index (GSI) normalizes the contrast between the yellow and red bands, where positive values indicate a greater influence of non-photosynthetic components; and (2) Yellow Normalized Vegetation Index (YNDVI) adapts the logic of the classical NDVI by substituting the red band with the yellow band, thereby emphasizing the contrast between structural biomass (NIR) and the spectral influence of shadows and complex crown surfaces characteristic of multi-layered stands. The inclusion of these exploratory indices is supported by the physical plausibility of their underlying spectral mechanisms, as documented in the recent literature [40–42].

2.3.3. Sentinel-2 Imagery

Sentinel-2 Level-2A multispectral images from June 21, 2024 were downloaded from the Copernicus Data Space Ecosystem (https://dataspace.copernicus.eu/ (accessed on 30 June 2026)). Images with cloud cover over the study area were discarded. All spectral bands were resampled to a spatial resolution of 10 m, harmonizing the original 10 m and 20 m resolutions. The following ten bands were used in the analysis: B2 (blue, 490 nm), B3 (green, 560 nm), B4 (red, 665 nm), B5 (red-edge 1, 705 nm), B6 (red-edge 2, 740 nm), B7 (red-edge 3, 783 nm), B8 (NIR, 842 nm), B8A (red-edge 4, 865 nm), B11 (SWIR 1, 1610 nm)

and B12 (SWIR 2, 2190 nm). From these bands, 12 vegetation indices were computed (Table 3). For each index and plot, the mean, P10, and P90 were calculated.

Table 2. Multispectral indices calculated from WorldView-3 bands.

Variable	Description	Formulas	Source
$NDVI_{1A}$	Normalized Difference Vegetation Index	$\frac{B5-B7}{B5+B7}$	[42]
$NDVI_{1B}$	Normalized Difference Vegetation Index	$\frac{B6-B7}{B6+B7}$	[42]
$NDVI_2$	Normalized Difference Vegetation Index	$\frac{B5-B6}{B5+B6}$	[43]
MCARI	Modified Chlorophyll Absorption Ratio Index	$[(B6-B5)-0.2*(B6-B4)]*\left(\frac{B6}{B5}\right)$	[44]
NDGI	Normalized Difference Generic Index	$\frac{B7-B6}{B7+B6}$	[45]
LCI_{NIR1}	Leaf Chlorophyll Index	$\frac{B7-B6}{B7+B5}$	[46]
LCI_{NIR2}	Leaf Chlorophyll Index	$\frac{B8-B6}{B8+B5}$	[46]
EVI	Enhanced Vegetation Index	$\frac{2.5*(B8-B5)}{(B8+(6*B5))-((7.5*B2)+1)}$	[47]
MSAVI	Modified Soil-Adjusted Vegetation Index	$\frac{2*B8+1-\sqrt{(2*B8+1)^2-8*(B8-B5)}}{2}$	[48]
IRECI	Inverted Red-Edge Chlorophyll Index	$\frac{B7-B5}{\frac{B6}{B5}}$	[49]
GSI	Gap Soil Index	$\frac{B4-B5}{B4+B5}$	This study
YNDVI	Yellow Normalized Difference Vegetation Index	$\frac{B8-B4}{B8+B4}$	This study

Table 3. Multispectral indices calculated from Sentinel-2 bands.

Variable	Description	Formulas	Source
$NDVI_{1A}$	Normalized Difference Vegetation Index	$\frac{B8-B4}{B8+B4}$	[42]
$NDVI_{1B}$	Normalized Difference Vegetation Index	$\frac{B7-B4}{B7+B4}$	[42]
$NDVI_2$	Normalized Difference Vegetation Index	$\frac{B6-B5}{B6+B5}$	[43]
MCARI	Modified Chlorophyll Absorption Ratio Index	$[(B5-B4)-0.2*(B5-B3)]*\left(\frac{B5}{B4}\right)$	[44]
NDGI	Normalized Difference Generic Index	$\frac{B8-B5}{B8+B5}$	[45]
LAI	Leaf Area Index	Neural network inversion of PROSAIL model (Biophysical Processor of SNAP version 10.0.0)	[50,51]
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation		
FCOVER	Vegetation cover fraction		
LCI	Leaf Chlorophyll Index	$\frac{B8-B5}{B8+B4}$	[46]
EVI	Enhanced Vegetation Index	$\frac{2.5*(B8-B4)}{(B8+(6*B4))-((7.5*B2)+1)}$	[47]
MSAVI	Modified Soil-Adjusted Vegetation Index	$\frac{2*B8+1-\sqrt{(2*B8+1)^2-8*(B8-B4)}}{2}$	[48]
IRECI	Inverted Red-Edge Chlorophyll Index	$\frac{B8-B4}{\frac{B6}{B5}}$	[49]

2.3.4. RGB-NIR Orthophotographs

RGB-NIR aerial orthophotographs were obtained from the PNOA program (https://centrodedescargas.cnig.es/CentroDescargas/home, accessed on 25 October 2026). These images provide a planimetric accuracy better than 0.5 m (RMSE), a spatial resolution of 0.25 m, and contain four spectral bands: red, green, blue, and near-infrared. From these images, we derived the Canopy Cover Fraction (CCF) through spectral delineation of tree canopies. For this, an NDVI threshold of ≥0.3 was applied to the orthophoto mosaic, and the result was reclassified to isolate pixels corresponding to tree cover, following a spectral-index-based approach consistent with previous studies [19].

2.4. Statistical Modeling

The aim of the modeling was to: (1) spatially predict the Old-Growth Index (OGI) across the study area, and (2) assess whether the information derived from aerial orthophotographs and satellite imagery adds value to the three-dimensional canopy structure measured by ALS. To achieve this, four predictor variable sets were constructed:

- ALS set: The 43 ALS metrics (Table 1).
- ALS + CCF set: The ALS metrics plus the canopy cover fraction (CCF) derived from the RGB–NIR orthophotographs (44 predictors).
- ALS + CCF + WorldView-3 (WVIEW3set): The ALS + CCF set plus the 36 spectral variables derived from WorldView-3 (12 vegetation indices expressed as mean, P10 and P90; Table 2), yielding 80 predictors in total.
- ALS + CCF + Sentinel-2 (SENT2set): The ALS + CCF set plus the 36 spectral variables derived from Sentinel-2 (12 vegetation indices expressed as mean, P10 and P90; Table 3), yielding 80 predictors in total.

The first two sets contain no satellite data and serve as a baseline reference to assess the contribution of satellite-derived variables. Each of the four predictor sets was modeled using three statistical approaches: Multiple Linear Regression (MLR), Random Forest (RF) and Support Vector Regression (SVR).

Modeling was conducted through two complementary approaches. First, final models were fitted to the entire dataset of 30 plots to assess goodness-of-fit to the training data and to identify the selected variables (for MLR) or key predictors based on variable importance (for RF and SVR). Second, predictive generalization performance was evaluated using repeated nested cross-validation. In each repetition, the 30 plots were randomly divided into five outer folds. Each fold was held out in turn as the test set (6 plots), and the remaining 24 plots formed the training set. Within each training set, an inner 5-fold cross-validation was used to select variables (MLR and RF) or to tune hyperparameters (RF and SVR), ensuring that the test plots were never used to build or tune the model. The model selected in the inner loop was fitted to the 24 training plots and used to predict the 6 test plots. The procedure was repeated 10 times with different random assignments of plots to folds (5 outer folds × 10 repetitions = 50 outer evaluations per model). R^2 (the squared Pearson correlation coefficient between observed and predicted OGI values) was computed in each outer test fold and is reported as mean ± SD across the 50 evaluations.

2.4.1. Multiple Linear Regression (MLR)

In an initial step, the predictor variables were partitioned into subsets of variables providing similar information and therefore exhibiting high internal correlation. The 43 ALS metrics were partitioned into nine groups (Table S2, Supplementary Material). The partitions of the Sentinel-2 and WorldView-3 spectral variables are given in Tables S3 and S4, respectively. Subsequently, all possible regression models were constructed under two constraints: each model could include at most one variable from each subset and no more than five variables in total. These restrictions ensured computational feasibility and reduced the risk of overfitting. Additionally, to avoid multicollinearity, models with a Variance Inflation Factor (VIF) greater than 5 were discarded. The best-performing models were selected according to the Akaike Information Criterion (AIC). All analyses were conducted in R version 4.4.2 (R Core Team, 2024) [52] using a custom script. This approach was preferred over principal component analysis because it retains the original variables, thus preserving their interpretability while reducing the number of candidate predictors and controlling collinearity [53].

In the repeated nested cross-validation, variable selection was performed within each inner fold by greedy forward selection with AIC, instead of the exhaustive search. The

evaluation metrics of the MLR models were the R^2 of the nested cross-validation, which estimates generalization, and the adjusted R^2 and RMSE of the models fitted to the 30 plots, which describe the fit to the training data and not generalization.

2.4.2. Random Forest (RF)

A two-stage variable selection procedure was implemented. First, predictors exhibiting very high pairwise correlation ($r > 0.95$) were removed to reduce multicollinearity, consistent with established practices [54]. Second, a preliminary Random Forest model (1500 trees) was fitted to identify the most influential variables using the Percentage Increase in Mean Squared Error (%ΔMSE) metric, from which the eight most important predictors were retained. This procedure reduced the initial sets of 43 (ALSset), 44 (ALS + CCF), 80 (SENT2set) and 80 (WVIEW3set) variables to 8 final predictors for each set.

The mtry parameter (the number of predictors randomly sampled at each split) of the final model was optimized by 5-fold cross-validation, across a predefined range of values, following established methodologies [55]. The number of trees was fixed at 1500 after graphically verifying the stabilization of the Out-of-Bag (OOB) error, as recommended by Breiman [56]. Variable importance was assessed using the %ΔMSE, which is widely regarded as a robust indicator of predictor relevance in remote sensing applications [54].

For the repeated nested cross-validation, both the variable selection and the hyperparameter tuning (mtry, searched between 2 and 8) were performed within each inner fold, following the same two-stage procedure described above, to avoid information leakage. A fixed random seed (set.seed(123)) was used for reproducibility. The evaluation metrics of the RF models were the R^2 of the nested cross-validation, which estimates generalization, and the Out-of-Bag (OOB) R^2 and RMSE, estimated from the bootstrap samples not included in each tree, which describe the fit to the 30 plots and not generalization.

2.4.3. Support Vector Regression (SVR)

SVR modeling was conducted using a Radial Basis Function (RBF) kernel following [57,58]. To ensure model robustness and prevent overfitting, the key hyperparameters (the regularization parameter (C) and the kernel sigma parameter (σ)) were optimized through 5-fold cross-validation, selecting the combination that minimized the RMSE from a grid of 16 combinations (C = 0.1, 1, 10 and 100; σ = 0.001, 0.01, 0.1 and 1). All predictors were standardized prior to analysis. To assess the relative contribution of predictors within the SVR framework, permutation importance was computed following [59]. This method quantifies the increase in prediction error that occurs when the values of each predictor are randomly permuted, thereby providing an estimate of the variable's influence on model accuracy.

In the repeated nested cross-validation, C and σ were tuned within each inner 5-fold loop using the same grid of 16 combinations, and no variable selection was performed (all predictors were used). The evaluation metrics of the SVR models were the R^2 of the nested cross-validation, which estimates generalization, and the R^2 and RMSE of the trained model, which describe the fit to the 30 plots and not generalization.

2.5. Spatial Prediction

The final model of each combination of algorithm and predictor set was trained on all 30 plots, using the optimal hyperparameters (mtry for RF; C and σ for SVR) selected by 5-fold cross-validation on the 30 plots. The nested cross-validation was used only to estimate generalization performance. The prediction variability and the Domain of Applicability Index described below were calculated only for the combination of predictor set and algorithm with the highest nested cross-validation R^2. The OGI was predicted at

50 × 50 m resolution in the area where the Spanish Forest Map (MFE) indicates that *P. nigra* is dominant (>80%).

2.5.1. Prediction Variability

To estimate prediction variability, a bootstrap procedure was applied to the combination of predictor set and modeling approach with the highest nested cross-validation R^2. The model was retrained 30 times (B = 30 bootstrap replicates) using bootstrap samples of the 30 plots, keeping the optimal hyperparameters fixed. For each pixel, the standard deviation (SD) of the 30 predictions was calculated. Because this SD reflects the sensitivity of the model to the training sample and not the error against field observations, it is referred to as prediction variability.

2.5.2. Domain of Applicability Index (DAI)

The Domain of Applicability Index (DAI) was used to assess to what extent the predictor values of each pixel fall within the range of conditions represented by the 30 training plots. For each pixel, the DAI is the proportion of the predictors of the model whose values lie within the range observed in the training plots, in line with range-based approaches to detect extrapolation [60,61]. Because the top-performing SVR model uses all 80 predictors from the WVIEW3set, DAI values vary in discrete steps of 1.25% (1/80). Consequently, pixels falling within the 80–100% DAI category can contain between 0 and 16 predictors outside the univariate calibration range, and only pixels with a DAI of 100% represent full univariate bounding. It was calculated only for the combination of predictor set and modeling approach with the highest nested cross-validation R^2, and grouped into five classes of 20% range.

The complete methodological workflow is summarized in Figure 2.

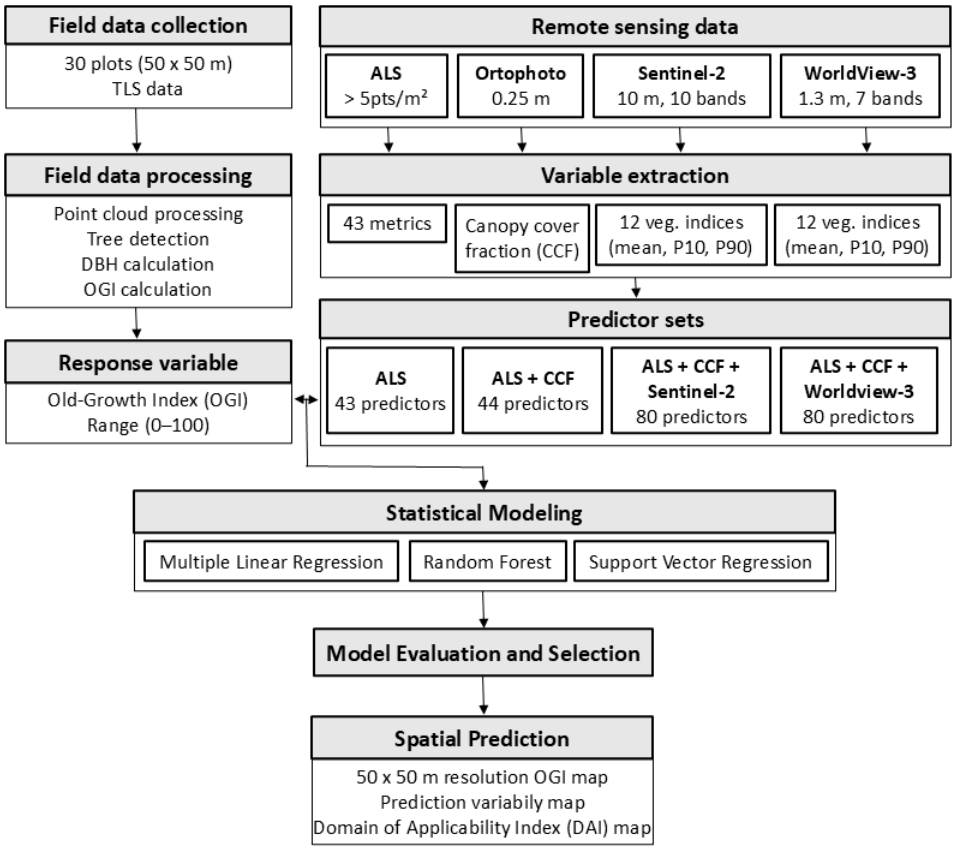


Figure 2. Methodological workflow of the study.

3. Results

3.1. Structural Characterization of Young and Old Plots

Older stands showed higher structural heterogeneity, reflected in a higher DBH standard deviation (Table 4) and a greater density of large trees (>50 cm DBH). Total tree density was higher in younger stands, whereas basal area was similar between both stand types. The mean OGI was 68.12 for older stands and 17.29 for younger stands. The distribution of OGI values across the 30 field plots is shown in Figure 3.

Table 4. Structural variable values for old and young stands of *Pinus nigra*.

Structural Variables	Old Forest (n = 15)	Young Forest (n = 15)
Mean tree diameter (mDBH, cm)	44.78 (7.42)	25.55 (6.31)
Diameter standard deviation (STDDBH)	21.92 (5.92)	9.84 (2.71)
Basal area (m^2/ha)	29.12 (9.26)	35.11 (12.14)
Basal area of trees > 50 cm DBH (m^2/ha)	21.48 (5.77)	2.57 (2.40)
Basal area of trees > 70 cm DBH (m^2/ha)	10.78 (7.48)	0.48 (1.06)
Basal area of trees > 100 cm DBH (m^2/ha)	0.56 (1.48)	0.00 (0.00)
Density (trees/ha)	171.73 (114.49)	741.60 (650.59)
Density of trees > 50 cm DBH (trees/ha)	58.93 (20.92)	9.60 (8.53)
Density of trees > 70 cm DBH (trees/ha)	19.73 (13.98)	1.07 (2.37)
Density of trees > 100 cm DBH (trees/ha)	0.53 (1.41)	0.00 (0.00)
Gini coefficient (GC)	0.45 (0.16)	0.40 (0.12)

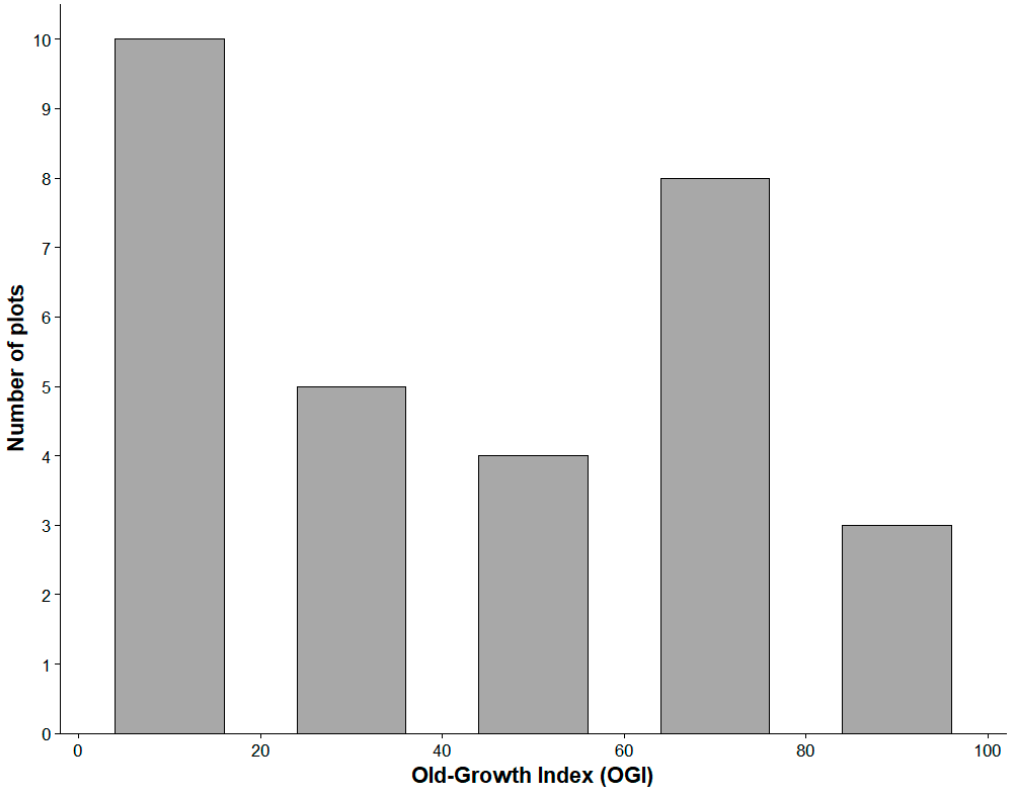


Figure 3. Distribution of OGI values across the 30 field plots.

3.2. Multiple Linear Regression (MLR)

The performance of the best Multiple Linear Regression models for each predictor set is summarized in Tables 5–7, showing the five regressions with the lowest AIC. The WVIEW3set slightly outperformed the SENT2set-based model, with lower AIC and higher adjusted R^2 (Adj. R^2), although differences between both predictor sets were relatively small.

Table 5. Performance metrics and selected predictor variables for the top-five Multiple Linear Regression models using the ALS + CCF predictor set. Adj.R^2: Adjusted R-squared; RMSE: Root Mean Square Error; AIC: Akaike Information Criterion.

Model	Adj.R^2	RMSE	AIC	Predictor Variables and Sign
1	0.813	11.99	154.35	−hmode, −hskew, +h99, −ARAM.TFR, +CCF
2	0.809	12.10	154.90	+hCV, +h99, +CRR, −ARAM.TFR, +CCF
3	0.806	12.20	155.39	−hmode, +hkurt, +h95, −ARAM.TFR, +CCF
4	0.805	12.24	155.58	−hmax, +h90, −PARA3, +PERAMO, +CCF
5	0.804	12.25	155.65	−hmode, +hkurt, +h95, −ARA3.TFR, +CCF

Table 6. Performance metrics and selected predictor variables for the top-five Multiple Linear Regression models using the WVIEW3set predictor set. Adj.R^2: Adjusted R-squared; RMSE: Root Mean Square Error; AIC: Akaike Information Criterion.

Model	Adj.R^2	RMSE	AIC	Predictor Variables and Sign
1	0.881	9.57	140.83	$+h_{95}$, −ARA3.TFR, +CCF, $-EVI_{p10}$, $+GSI_{p10}$
2	0.878	9.68	141.48	$+h_{99}$, −ARA3.TFR, +CCF, $+EVI_{p90}$, $+GSI_{p10}$
3	0.876	9.77	142.06	$+h_{99}$, −ARA3.TFR, +CCF, $+EVI_{p90}$, $-YNDVI_{_p90}$
4	0.874	9.82	142.38	$-h_{mode}$, $+h_{95}$, −ARA3.TFR, +CCF, $+GSI_{p10}$
5	0.871	9.95	143.17	$-h_{min}$, $+h_{95}$, −ARAM.TFR, +CCF, $+GSI_{p10}$

Table 7. Performance metrics and selected predictor variables for the top-five Multiple Linear Regression models using the SENT2set predictor set. Adj.R^2: Adjusted R-squared; RMSE: Root Mean Square Error; AIC: Akaike Information Criterion.

Model	Adj.R^2	RMSE	AIC	Predictor Variables and Sign
1	0.852	10.67	147.32	$-h_{mode}$, $+h_{90}$, −PARA3, $+EVI_{p90}$, $+IRECI_{p90}$
2	0.851	10.69	147.48	$-h_{mode}$, $+h_{95}$, −PARA3, $+LAI_{mean}$, $+IRECI_{p90}$
3	0.848	10.79	148.04	$-h_{mode}$, $+h_{90}$, −PARA3, +CCF, +MCARI
4	0.844	10.93	148.79	$-h_{mode}$, $+h_{95}$, −PARA3, +FAPAR, $+IRECI_{p90}$
5	0.843	10.99	149.11	$-h_{mode}$, $+h_{95}$, −PARA3, +CCF, +MCARI

For the ALS + CCF set, all five best models included CCF and an upper height percentile (h90, h95 or h99) with positive coefficients, and a metric describing the proportion of returns above 3 m or above the mean height (ARAM.TFR, ARA3.TFR or PARA3) with a negative coefficient. The most frequent height (hmode) had a negative coefficient in three of the models.

The best MLR models integrated structural and spectral predictors to estimate the OGI. For the WVIEW3set, models relied on upper height percentiles (h_{95}/h_{99}) and the Canopy Cover Fraction (CCF), both showing a positive relationship with OGI. The density metric

ARA3.TFR (and ARAM.TFR in model 5) had negative coefficients, while spectral indices such as GSI_p10 and EVI_p90 were selected with positive coefficients. Conversely, EVI_p10 and YNDVI_p90 exhibited negative coefficients.

For the SENT2set, the best models consistently included positive coefficients for upper height percentiles (h_{90}/h_{95}) and spectral indices like EVI_p90 and IRECI_p90. In contrast, metrics related to the most frequent height (h_{mode}) and the percentage of returns above 3 m (PARA3) were selected with negative coefficients. The CCF also appeared as a positive predictor in some of the best models for this predictor set.

3.3. Random Forest (RF)

The Random Forest models included mainly LiDAR metrics of vertical structure among the most important predictors of the OGI. The OOB R^2 of the final models, fitted to the 30 plots, was 0.671 (RMSE = 15.88) for ALS + CCF, 0.678 (RMSE = 15.71) for WVIEW3set and 0.658 (RMSE = 16.20) for $SENT2_{set}$. The variable importance of the three models is shown in Figure 4. For the ALS + CCF model, the most important variables were ARA3.TFR, hIQR, hAAD and hLkur, and CCF was not among the eight variables retained. The WVIEW3set model relied mainly on ARA3.TFR, related to the density of vegetation above 3 m, and vertical structure descriptors h_{IQR}, h_{AAD} and h_{Lkur}. The spectral index LCI_{NIR2_p10}, sensitive to low chlorophyll areas, and upper canopy height percentiles (h_{90}, h_{99}) were also among the most important variables.

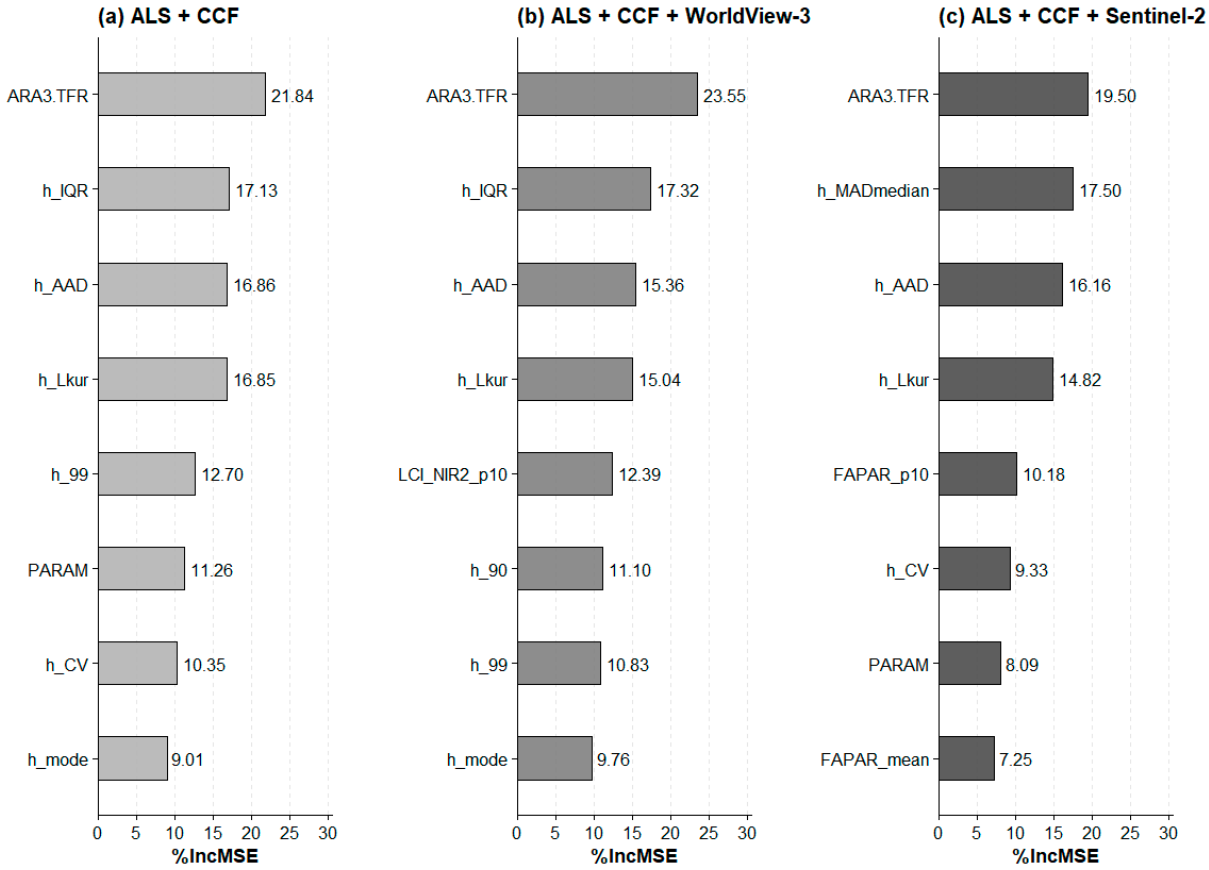


Figure 4. Variable importance (%IncMSE, percentage increase in out-of-bag mean squared error after permuting each predictor) of the eight variables retained in the final Random Forest models fitted to the 30 plots: (**a**) ALS + CCF, (**b**) ALS + CCF + WorldView-3 and (**c**) ALS + CCF + Sentinel-2.

For the SENT2set combination, key variables included ARA3.TFR and vertical dispersion metrics $h_{MADmedian}$ and h_{AAD}. The distribution shape metric h_{Lkur} was also important. These were complemented by the spectral indices $FAPAR_{p10}$ and $FAPAR_{mean}$.

3.4. Support Vector Regression (SVR)

The optimized SVR models showed different performance across predictor sets. For the fit to the 30 plots, the SENT2set model gave the highest R^2 (0.986; RMSE = 3.29), followed by the ALS + CCF model (0.972; RMSE = 4.62) and the WVIEW3set model (0.903; RMSE = 8.52).

The SVR models identified distinct sets of influential predictors in each predictor set. For the ALS + CCF model, the most important variables were CCF and the canopy cover metric CC, followed by ARA3.TFR and ARAM.TFR. For the WVIEW3set model, the most important variables were the spectral indices EVI_{mean} and EVI_{p10}, together with CCF. LiDAR metrics describing the shape and dispersion of the canopy height distribution (h_{kurt} and h_{IQR}) were also among the most important.

For the SENT2set model, the most important variables were vertical density metrics PARAM and ARAM.TFR, alongside the horizontal canopy cover CCF. Spectral indices $MCARI_{p90}$ and $IRECI_{p90}$ were also among the most important variables. The ten most important predictors of each model are shown in Figure 5, and the permutation importance of all spectral predictors is shown in Figure S1 (Supplementary Material).

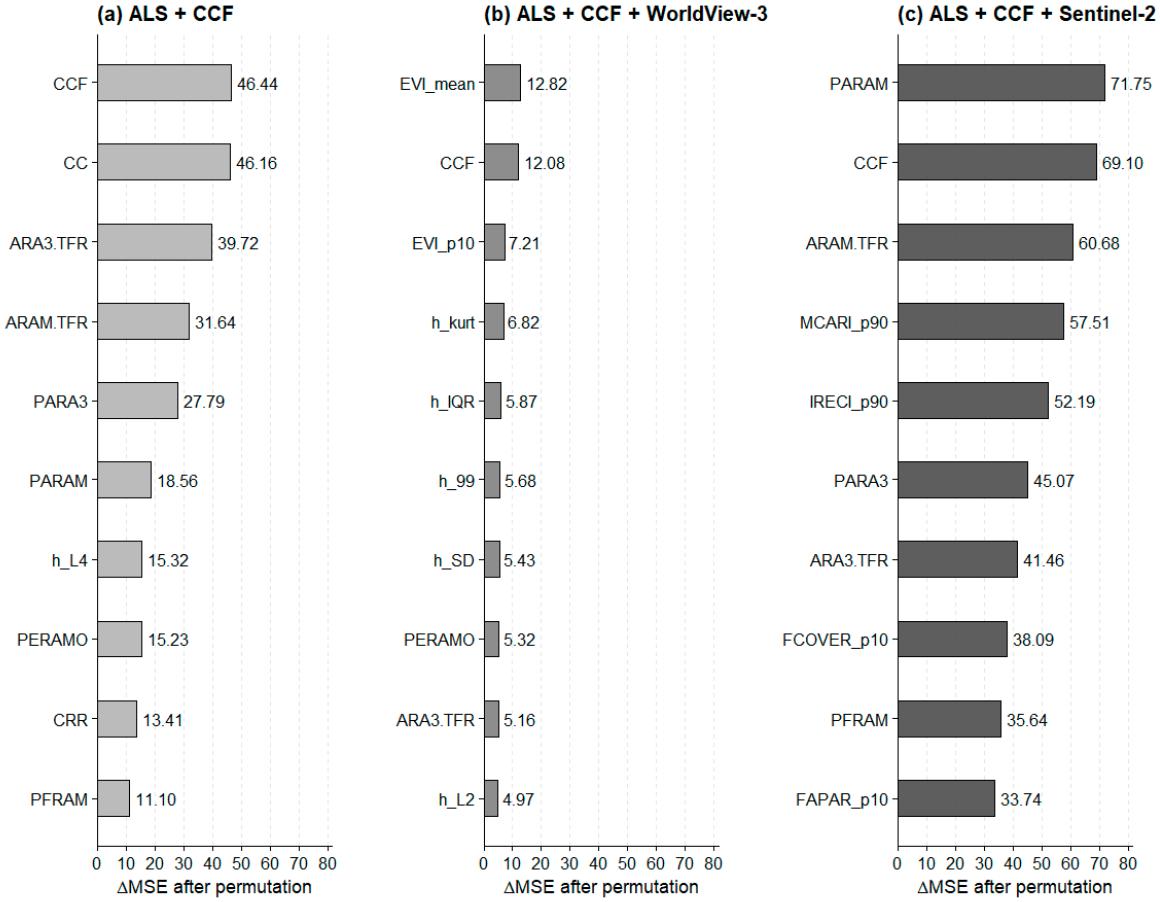


Figure 5. Variable importance (ΔMSE, mean percentage increase in mean squared error after permuting each predictor) of the top-10 predictors in the final SVR models fitted to the 30 plots: (**a**) ALS + CCF, (**b**) ALS + CCF + WorldView-3 and (**c**) ALS + CCF + Sentinel-2.

3.5. Global Comparison and Residual Analysis

In the repeated nested cross-validation (Figure 6), the highest mean nested R^2 was achieved by SVR using the ALS + CCF + WorldView-3 set (0.790), which was similar to that obtained with SVR and ALS + CCF (0.770). The difference was smaller than the SD of these two models (0.138 to 0.162), indicating that commercial satellite imagery was not essential to achieve similar predictive performance. Across the four predictor sets, SVR and RF consistently yielded higher nested R^2 values than MLR. Most of the predictive information was contributed by ALS, which alone produced a nested R^2 of 0.710 with SVR and 0.714 with RF. Adding CCF further increased the nested R^2 of SVR (0.770) and MLR (from 0.521 to 0.607), but had no effect on RF (0.714). Adding Sentinel-2 data did not improve predictions; it reduced the nested R^2 of SVR (0.644) and RF (0.678) compared to ALS + CCF, and remained virtually unchanged for MLR (0.609 compared with 0.607). For SVR, the R^2 of the fit to the 30 plots was substantially higher than the nested R^2 across all three sets, with the difference being largest for the SENT2set (0.986 vs. 0.644), compared to ALS + CCF (0.972 vs. 0.770) and WVIEW3set (0.903 vs. 0.790).

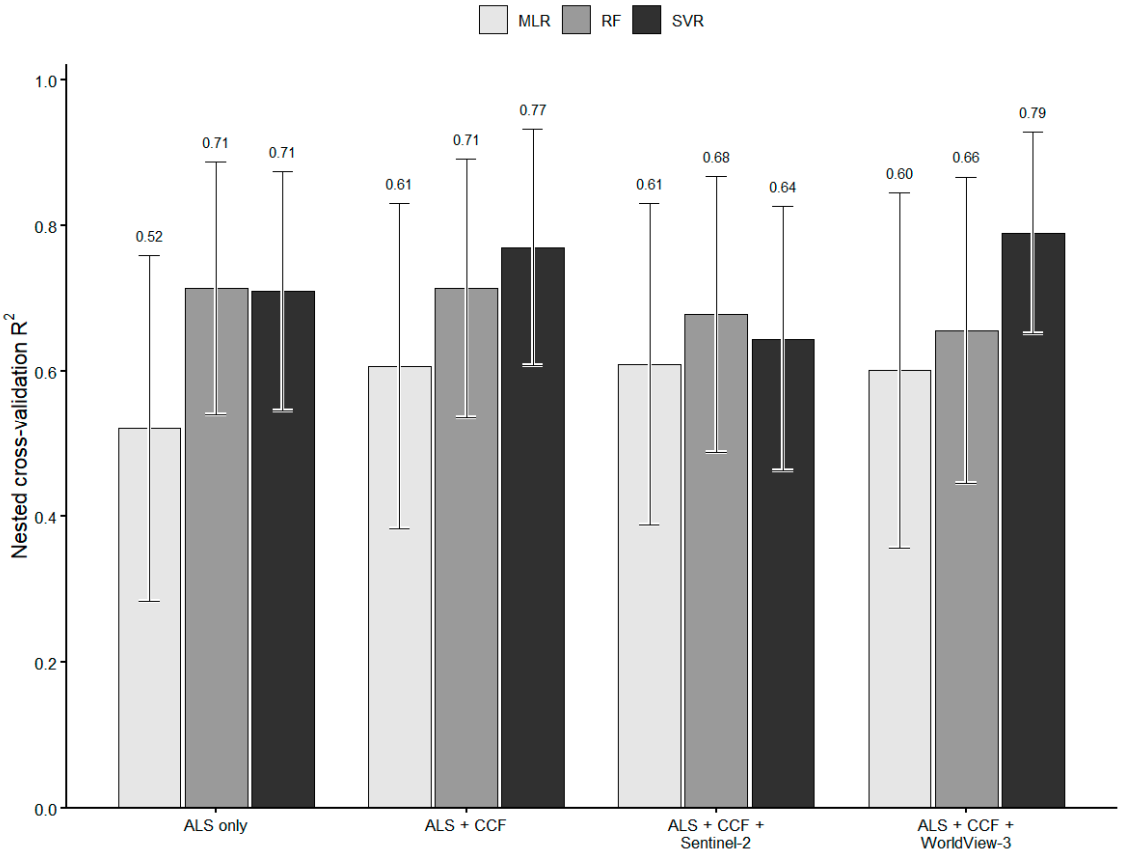


Figure 6. Predictive performance (repeated nested cross-validation R^2) across predictor sets and modeling algorithms. Bars represent mean R^2 values, and error bars indicate ± 1 standard deviation (SD) across the 50 outer evaluations (5 folds × 10 repetitions). MLR: Multiple Linear Regression; RF: Random Forest; SVR: Support Vector Regression.

Model residuals as a function of predicted OGI values for the final SVR model using the WVIEW3set predictor set are presented in Figure 7. This residual plot is derived from models calibrated on all 30 plots; thus, they characterize in-sample goodness-of-fit rather than predictive generalization performance. Observed versus predicted OGI values across all tested predictor sets and the three modeling approaches are shown in Figure S2. The

mean value of residuals was close to zero for all models (−0.4 to 0.2), and their SD ranged from 3.3 (SVR with SENT2set) to 9.7 (MLR with SENT2set).

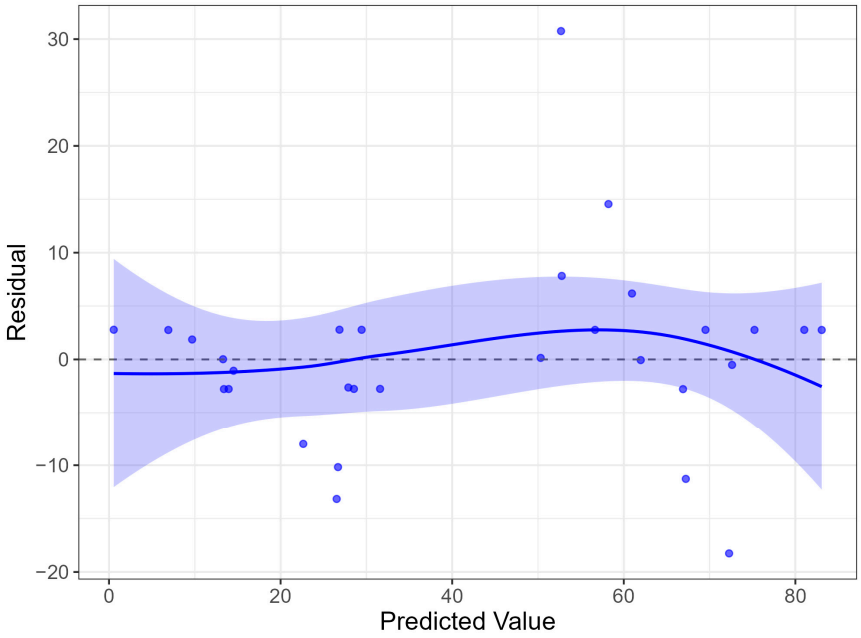


Figure 7. Residuals versus predicted OGI values for the SVR model and the WVIEW3set predictor set. Each point represents a field plot ($n = 30$). The solid line represents a locally estimated scatterplot smoothing (LOESS) curve with a 95% confidence band; the dashed line indicates zero residual.

Across modeling approaches, the variables with the highest importance were predominantly LiDAR-derived metrics of vertical canopy structure (ARA3.TFR, PARAM, ARAM.TFR, h_{IQR}, h_{AAD}, h_{Lkur} and h_{kurt}), while CCF was also among the most important variables in the MLR and SVR models (Tables 5–7; Figure 5). Key spectral variables differed between sensors: chlorophyll-related indices for Sentinel-2 ($IRECI_{p90}$, $MCARI_{p90}$, FAPAR) and EVI-based indices and LCI_{NIR2_p10} for WorldView-3.

3.6. Spatial Prediction of OGI

To facilitate the interpretation and potential practical application of the modeling results, a predictive map of the OGI across the study region was generated using the top-performing model (the highest nested R^2: SVR with WVIEW3set, Figure 8). The spatial resolution was 50 × 50 m, matching the extent of the field sampling unit (0.25 ha).

The Domain of Applicability Index (DAI) was calculated for the SVR model with WorldView3set and mapped across the study area (Figure 9; Table 8). Over 80% of the study area exhibited a DAI > 60% (meaning that at least 48 of the 80 predictors fell within the observed training bounds), although this does not preclude multivariate extrapolation. Conversely, areas with a very low DAI (<20%) were scarce and coincided with locations characterized by very low *Pinus nigra* stand density.

Prediction variability across OGI classes is shown in Figure 10. Estimated as the standard deviation of the bootstrap predictions, prediction variability increased along the OGI gradient, rising from a mean of 7.89 ± 2.62 in the lowest class (0–20) to 11.90 ± 2.80 in the highest class (80–100). Across the entire study area, the overall mean prediction variability was 9.47 ± 3.23 OGI units.

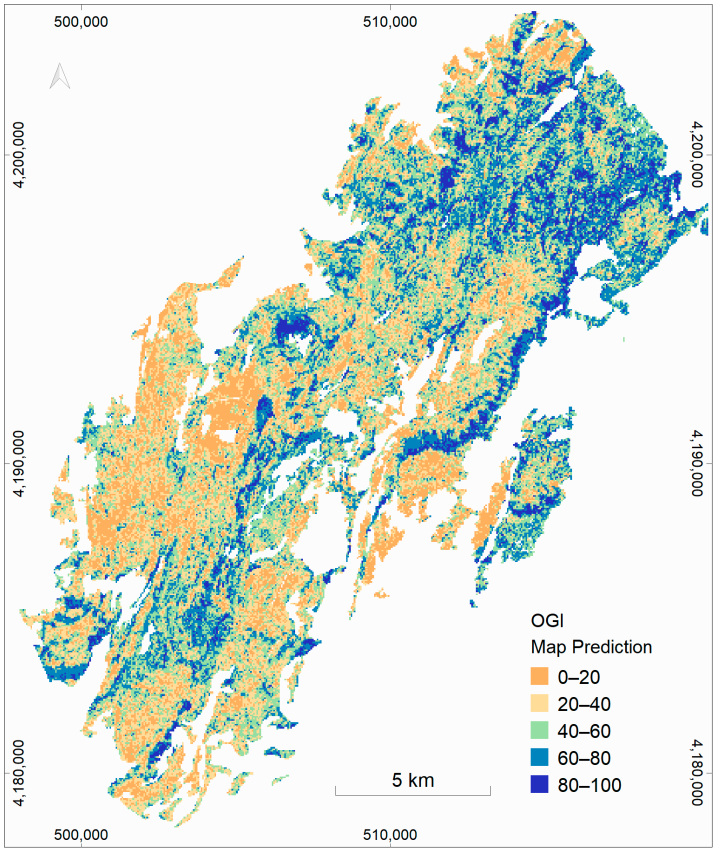


Figure 8. OGI map generated from the Support Vector Regression model with WVIEW3set.

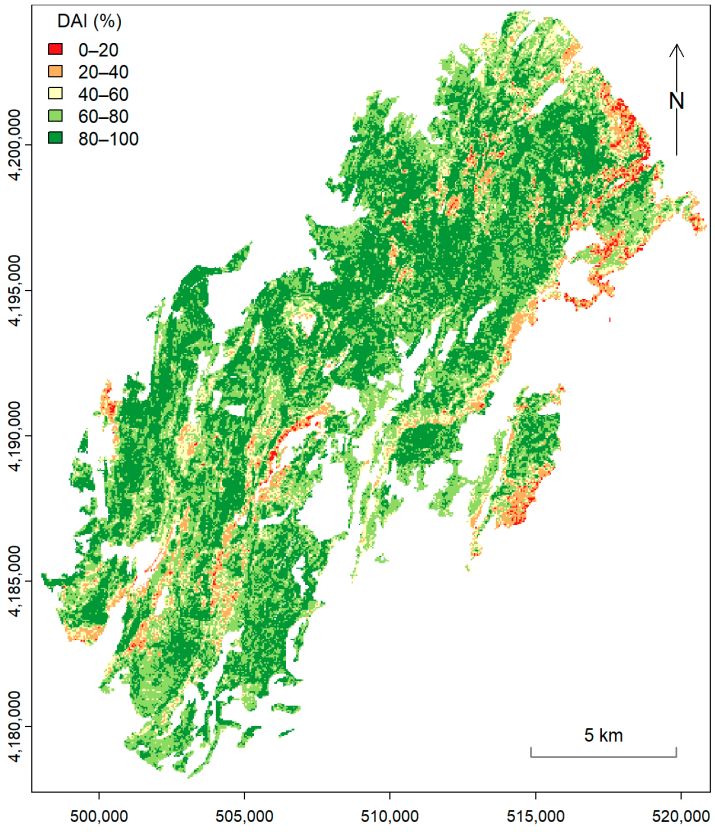


Figure 9. Domain Applicability Index (DAI) for the Support Vector Regression (SVR) model with the WVIEW3set predictor set.

Table 8. Area coverage per class of the Domain of Applicability Index (DAI) for the Support Vector Regression (SVR) model with the WorldView3set predictor set across the study area.

DAI Class	Area (ha)	Area (%)
0–20	213.5	1.0
20–40	1453.3	6.7
40–60	2592.0	11.9
60–80	8092.5	37.3
80–100	9366.5	43.1
Total	21,717.8	100.0

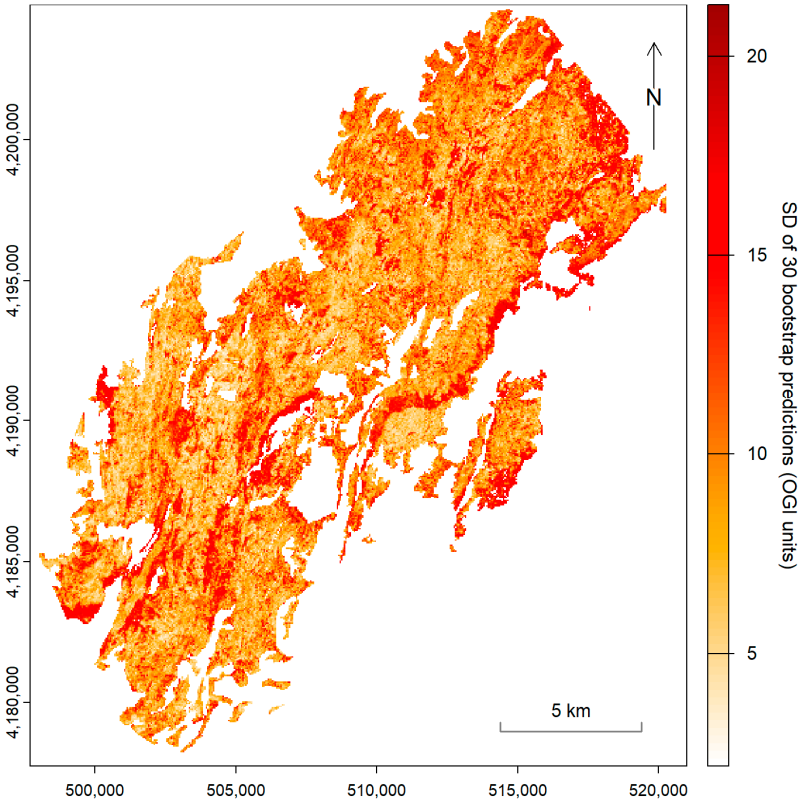


Figure 10. Prediction variability (standard deviation of bootstrap predictions) for the SVR model with the WVIEW3set predictor set.

4. Discussion

4.1. Structural Characterization of Young and Old-Growth Stands

Field characterization confirmed that *Pinus nigra* stands in the study area exhibit wide structural variability, forming a gradient from young regular dense stands to old irregular ones. Young stands are characterized by higher tree density and a simpler, more homogeneous structure, with lower heterogeneity in diameter distribution and a low density of large trees. In contrast, stands classified as old exhibit a distinctive structure with greater structural heterogeneity reflected in a wider diameter distribution and a significantly higher density of large trees. These attributes are consistent with those reported in natural *Pinus nigra* stands with irregular structures and the presence of large trees [17] and align with the structural criteria proposed for identifying OGFs in this species [11,12]. The OGI derived from these structural attributes effectively captured the underlying structural gradient. This multivariate approach is essential because old trees are not necessarily large, as their

physical dimensions are strongly constrained by local growth conditions [1]. Consequently, relying on an integrated OGI provides a more robust and ecologically meaningful indicator than relying solely on isolated structural metrics.

4.2. Contribution of Satellite Imagery to ALS-Based Models

Studies using only optical imagery, such as Sentinel-2, have reported moderate accuracy due to their limitation in capturing vertical canopy complexity. For example, Spracklen and Spracklen [13] achieved an overall accuracy of 85% in classifying OGFs in the Carpathians but observed lower accuracy in mixed forests due to the lack of 3D structural information. Similarly, Erfanifard et al. [27] identified low remote sensing accuracy in Bialowieza Forest stands with complex vertical stratification, such as broadleaved stands and strict reserves. Their findings highlight that optical techniques tend to underestimate species richness in old-growth structures due to the difficulty of capturing the diversity situated beneath the upper canopy layers.

In contrast, several studies report that integrating LiDAR and optical data outperforms unimodal approaches [25]. However, unimodal models may be easier to generalize because they rely on fewer data sources, and ALS data are not available everywhere or at all times. Adiningrat et al. [21] reported an F1-score of 92% using ALS combined with Sentinel-2, compared with 71% using optical data alone, while in a subsequent study [2], they obtained an overall accuracy of 78% in separating old-growth from second-growth stands using LiDAR structural metrics. Similarly, combining ALS with ultra-high-resolution imagery (such as WorldView-3) has shown high performance in delineating complex forest stands and classifying development stages [62], as well as in modeling continuous structural variables such as individual tree biomass [63]. Hirschmugl et al. [1] found that incorporating ALS data increases the overall classification accuracy by an average of 6% relative to the exclusive use of Sentinel-2, achieving accuracies of up to 94% in distinguishing between old-growth and non-old-growth forests. This increase is relevant for operational monitoring, but the acquisition and processing of ALS data remain expensive, which may limit their applicability in regions without national ALS programs.

In our study, the highest nested R^2 was obtained with SVR and WVIEW3set (0.790), but the values for ALS only (0.710) and ALS + CCF (0.770) differed from it by less than the standard deviation across evaluations, while Sentinel-2 gave no clear gain (nested R^2 of 0.609 to 0.678). Our results therefore indicate that most of the predictive information came from the high-density ALS data, and the added value of the satellite imagery could not be demonstrated. The PNOA 2024 LiDAR has a mean density of about 20 points/m^2. This density is an order of magnitude greater than that used by Hevia et al. [12], who worked with the same species and area but with lower-density ALS data (~0.5 points/m^2) and achieved an adjusted R^2 of 0.42. This value describes the fit of the model and is not equivalent to our nested R^2. However, LiDAR density was not experimentally manipulated in our study, and the differences between both studies may also arise from other factors, such as sensor configuration, flight pattern, field plot design or study area characteristics. In LiDAR-based forest inventory, obtaining reliable ALS metrics (especially those sensitive to internal canopy complexity) requires sufficient pulse density, a requirement that is more critical in complex and heterogeneous forest types [64]. Moreover, high-density ALS data are not available in many regions, where low-density or project-based ALS campaigns still prevail, so approaches such as ours cannot rely exclusively on high-density ALS.

4.3. Comparison of Modeling Methods: SVR vs. MLR and RF

In the repeated nested cross-validation, MLR gave the lowest nested R^2 across the four predictor sets (0.521 to 0.609), whereas SVR (0.644 to 0.790) and RF (0.656 to 0.714)

achieved substantially higher values. SVR produced the highest nested R^2 with ALS + CCF and with WVIEW3set, but the differences between SVR and RF (0.004 to 0.134) were smaller than the standard deviation of each algorithm across evaluations (Figure 6); furthermore, RF performed similarly or slightly better with ALS only and with the SENT2set. SVR can model non-linear relationships through kernel functions [58], which is particularly suitable for capturing the structural complexity of OGFs. Furthermore, SVR is designed to maximize generalization capability even with a limited number of training samples [65], an advantage in studies like ours with a moderate plot sample size. The Random Forest model, while offering interpretative transparency through variable importance analysis [56], yielded a lower nested R^2 than SVR with ALS + CCF and WorldView-3, which may be related to the small number of training plots ($n = 30$), since, unlike SVR, RF has no explicit regularization parameter equivalent to C. MLR achieved high adjusted R^2 values for the fit to the 30 plots but the lowest nested R^2, confirming that it is conceptually less suitable due to its linear nature and sensitivity to multicollinearity [66]. The R^2 of the fit of SVR to the 30 plots was clearly higher than its nested R^2, especially with SENT2set (0.986 and 0.644), underscoring that only the nested R^2 should be interpreted as a measure of predictive generalization performance. This lower generalization of the SENT2set models may be related to the lower spatial resolution of Sentinel-2 (10 m), although this was not explicitly tested in this study.

4.4. Analysis and Ecological Interpretation of Key Predictor Variables

Permutation importance analysis in the best model (SVR with the WVIEW3set) identified a reduced set of key predictors of forest maturity, with EVI_{mean}, CCF and EVI_p10 being the most influential, followed by LiDAR metrics of vertical heterogeneity, such as h_{kurt}, h_{IQR} and h_{SD} (Figure 5b).

Ecologically, these variables capture intrinsic attributes of *Pinus nigra* old-growth forests. CCF derived from ultra-high-resolution (0.25 m) orthophotography reflects the horizontal architecture of the canopy [19]. The positive coefficient of CCF in the MLR models (Tables 5–7) indicates that plots with higher OGI exhibit higher CCF, suggesting that the presence of veteran trees with broad crowns maintains high overall coverage in the upper stratum, a characteristic feature of advanced structural maturity [10,23]. This increase in canopy cover with maturity suggests that the crowns of dominant trees compensate for natural gaps in the canopy [67].

EVI is sensitive to canopy structure and photosynthetic productivity [48]. Its importance in our model (specifically EVI_{mean} and EVI_p10) is consistent with studies identifying it as a sensitive indicator of successional stages in old-growth forests where higher values are associated with advanced developmental phases [22]. Furthermore, the relevance of indices sensitive to red-edge spectral information, such as $IRECI_{p90}$ (used in SENT2set), aligns with recent findings identifying these features as essential for mapping structural maturity in European OGFs [21]. For Sentinel-2, MCARIp90, FAPARp10 and FCOVER_p10 were also among the most important spectral variables, and for WorldView-3, IRECI_p10 and GSI_p90 followed EVI (Figures 5 and S1).

Regarding WorldView-3, the selection of the exploratory indices GSI_p10 and YNDVI_p90 in the best MLR models (Table 6), which incorporate the sensor's unique yellow band (Band 4), was not matched in the SVR model, where neither was among the ten most important predictors (Figure 5b). This choice aligns with research highlighting the yellow spectral region as critical for identifying structural heterogeneity in complex forest stands [40]. Specifically, the combination of the yellow band with NIR and red-edge information allows for capturing biochemical traits related to foliar age and complex crown architecture [41,42]. However, the yellow band seems to offer only complementary information, and its overall contribution to the prediction of the OGI was limited compared with

other structural and spectral predictors, so the performance of the WorldView-3 models cannot be attributed to it.

LiDAR-derived metrics of vertical heterogeneity provide direct and quantitative descriptors of structural complexity. The kurtosis of the height distribution (hkurt) describes how the returns are distributed along the vertical profile and has been used as a descriptor of vertical heterogeneity [68]. Old-growth forests show a more complex vertical structure than younger managed forests [14]. Fuhr et al. [69] also found that forest maturity increased with the metrics of height distribution heterogeneity, including kurtosis, which is consistent with the positive coefficient of hkurt in two of our MLR models (Table 5). Similarly, high values of h_{IQR} and h_{SD} reflect an irregular vertical profile resulting from natural regeneration processes and gap dynamics where different age classes coexist [10,70]. The use of these vertical dispersion metrics to capture structural complexity remains robust even under different LiDAR pulse densities [71].

Finally, metrics such as h_{99} (or h_{95} in MLR models) and ARA3.TFR provide information on dominant height and vegetation density in the upper canopy, respectively. The inclusion of upper height percentiles confirms the strong relationship between height and stand age, being a key parameter for modeling maturity using an area-based approach [12]. In line with previous studies in the same area where similar variables such as ARAM.TFR showed negative coefficients [12], our MLR models also show a negative association for ARA3.TFR (Tables 5 and 6). The negative coefficient is consistent with the interpretation that, in the study area, mature stands may be characterized by an upper stratum with large but open crowns that allow most returns to pass through and a weakly represented mid-story.

4.5. Sentinel-2 and National ALS Data for Long-Term Monitoring

Although the WVIEW3set combination reached the highest nested R^2 (0.790 with SVR), the SENT2set models reached nested R^2 values of 0.609 to 0.678 (Figure 6), without a clear gain over the models based on ALS and canopy cover alone (Section 4.2). Sentinel-2 nevertheless offers global, free, and periodic coverage (every 5 days), and PNOA LiDAR is updated approximately every six years. This periodicity would be compatible with the long-term monitoring objectives of the EU Biodiversity Strategy for 2030. However, the added value of Sentinel-2 for monitoring has not been shown here, and the models were tested only within the study area, so our results show the predictive potential of the approach, not its suitability for operational monitoring. Moreover, national ALS programs such as PNOA are not available in all EU countries: some (for example, Spain, Finland and the Netherlands) have systematic national ALS coverage, whereas others rely on project-based or incomplete ALS data, which limits the direct transferability of the approach. Further validation in independent areas is required before operational implementation.

4.6. Implications for Management and Conservation

The ability to generate continuous OGI maps from PNOA ALS, orthophotos and Worldview-3 imagery represents a methodological contribution with future projection. These cartographic products derived from empirical models offer a spatially explicit view of the distribution of structural maturity across the forest landscape and may support managers and conservationists in the study area in: (I) the identification and delineation of stands with a high probability of being old-growth in extensive and difficult-to-access areas; (II) the prioritization of conservation and adaptive management actions, directing efforts and resources towards stands with higher structural maturity value; and (III) the establishment of a baseline to monitor the evolution of these ecosystems in response to disturbances or under different management regimes. However, the prediction variability increases with the OGI class (Figure 10), and the Domain of Applicability Index (DAI) is low

in some areas of the map (Figure 9), so these maps should be interpreted with caution and validated in the field and in independent areas before being used in management decisions.

4.7. Limitations

Several methodological and data-related limitations should be explicitly acknowledged when interpreting the predictive performance and regional cartographic outputs of this study. First, the predictive models were calibrated using a limited sample size of 30 purposively selected plots representing the extremes of the maturity gradient (young vs. old-growth stands), and no independent spatial validation dataset was available. The reported model accuracies are derived from repeated random nested cross-validation. However, because high-OGI plots are inherently clustered in specific relict areas, standard random cross-validation can yield overly optimistic accuracy metrics for distant areas that constitute most of the regional projection, a well-known phenomenon in remote sensing-based ecological mapping [72,73]. Consequently, the mapped accuracies across the broader study area may be inflated. Future research should prioritize field campaigns with independent validation plots stratified across the regional landscape. Alternatively, spatially explicit cross-validation methods, such as nearest-neighbor distance matching (NNDM or kNNDM) [74,75], could be implemented to reflect spatial autocorrelation and prediction distances more faithfully. Nevertheless, as with any spatially buffered resampling, NNDM in contexts with clustered plots and extensive mapping extents may discard substantial fractions of training points per iteration, leading to very small training subsets and potentially unstable models. Furthermore, distance-matching cross-validation provides a more realistic internal error metric rather than a true external independent validation.

Second, the interpretation of the Domain of Applicability Index (DAI) requires careful consideration. The DAI evaluates whether environmental conditions fall within the empirical bounds of the training data on a strictly univariate basis (i.e., checking each predictor independently against its observed minimum and maximum). Therefore, it cannot detect novel, unsampled combinations in multidimensional feature space, nor can it account for internal sampling gaps within the observed univariate ranges; consequently, a high DAI value does not inherently guarantee high prediction accuracy. Furthermore, because the best-performing SVR model relies on the full WorldView-3 set composed of 80 predictors, the DAI takes discrete values in increments of 1.25% (1/80). This means that the uppermost category (80–100% DAI) encompasses pixels where between 64 and 80 predictors lie within the training domain; in other words, a pixel in this class may still have up to 16 predictors extrapolating outside the training envelope. Only pixels reaching an exact DAI of 100% have all 80 predictors completely bounded by the field plots. Thus, the finding that over 80% of the mapped area exhibits a DAI greater than 60% indicates that a majority of the individual variables are within sampled ranges, but it must not be conflated with a guarantee that the regional landscape fully falls within the model's multidimensional domain of applicability.

5. Conclusions

High-density ALS data, either alone or combined with multispectral imagery, provide a promising framework for modeling the Old-Growth Index (OGI). While the Support Vector Regression (SVR) model combining ALS and WorldView-3 achieved the highest predictive accuracy, simpler models relying exclusively on ALS metrics or ALS combined with Canopy Cover Fraction (CCF) yielded statistically comparable performance. Furthermore, integrating Sentinel-2 imagery with ALS provided no clear gain over the ALS-based models. The resulting spatial predictions offer a valuable tool to support forest conservation planning in the study area; however, because prediction variability increases in late-successional stages and the Domain of Applicability Index (DAI) drops in low-density stands, spatial

outputs should be interpreted with caution and validated in independent regions prior to operational implementation.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18213739/s1, Table S1: Geographic coordinates and stand structural characteristics of the 30 field plots; Table S2: Predictor variable partitioning for constrained Multiple Linear Regression (MLR): Airborne Laser Scanning (ALS) metrics and Canopy Cover Fraction (CCF); Table S3: Predictor variable partitioning for constrained Multiple Linear Regression (MLR): Sentinel-2 spectral variables; Table S4: Predictor variable partitioning for constrained Multiple Linear Regression (MLR): WorldView-3 spectral variables. Figure S1: Permutation importance (ΔMSE) of spectral variables in the Support Vector Regression (SVR) models; Figure S2: Observed versus predicted Old-Growth Index (OGI) for the 30 field plots from the final models fitted to the 30 plots, for Multiple Linear Regression (MLR), Random Forest (RF) and Support Vector Regression (SVR).

Author Contributions: C.G.-C.: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing—original draft, Visualization. A.C.C.: Investigation, Writing—review and editing. C.P.-C.: Investigation, Data curation, Writing—review and editing. R.A.-M.: Investigation, Writing—review and editing. J.V.-P.: Conceptualization, Supervision, Funding acquisition, Project administration, Investigation, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by MICIU/AEI/10.13039/501100011033 and FEDER, EU [grant number PID2022-136906OB-C22]. Funding for open access charge: Universidad de Huelva/Research Group RNM-315.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The ALS data are available from the Spanish National Center for Geographic Information (CNIG) at https://centrodedescargas.cnig.es (accessed on 25 October 2026). Sentinel-2 imagery is available from the Copernicus Data Space Ecosystem (https://dataspace.copernicus.eu/, accessed on 25 October 2026). WorldView-3 imagery is commercially licensed and not publicly available. Field data and derived OGI values are available from the corresponding author upon reasonable request.

Acknowledgments: This work has been carried out within the framework of the project PID2022-136906OB-C22 funded by MICIU/AEI/10.13039/501100011033 and FEDER, EU. We thank the EBD-LAST of CSIC and David Aragonés Borrego, Technician at LAST, for the use of the Leica TLS BLK2go LiDAR device, his assistance in the fieldwork, and for making the facilities of the EBD-LAST of CSIC available for subsequent data processing. We also thank the forestry services of the Cazorla, Segura y Las Villas Natural Park for providing the inventory data of the Navahondona forest and their continued collaboration. During the preparation of this work, the authors used AI to improve the language and readability of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hirschmugl, M.; Sobe, C.; Meyer, P.; Kirchmeir, H.; Di Filippo, A.; Vanhaecht, R.; Ivanova, Y.; Vandekerkhove, K. Classification of old-growth beech forest across Europe using Sentinel-2 and airborne laser scanning. *J. Nat. Conserv.* **2026**, *91*, 127202. [CrossRef]
2. Adiningrat, D.P.; Skidmore, A.; Schlund, M.; Wang, T.; Abdullah, H.; Heurich, M. Investigating LiDAR metrics for old-growth beech- and spruce-dominated forest identification in Central Europe. *Remote Sens.* **2025**, *17*, 251. [CrossRef]
3. Sabatini, F.M.; Burrascano, S.; Keeton, W.S.; Levers, C.; Lindner, M.; Pötzschner, F.; Verkerk, P.J.; Bauhus, J.; Buchwald, E.; Chaskovsky, O.; et al. Where are Europe's last primary forests? *Divers. Distrib.* **2018**, *24*, 1426–1439. Correction in *Divers. Distrib.* **2018**, *24*, 1890–1892. https://doi.org/10.1111/ddi.12778. [CrossRef]
4. European Commission. *Guidelines for Defining, Mapping, Monitoring and Strictly Protecting EU Primary and Old-Growth Forests*; Directorate-General for Environment, European Commission: Brussels, Belgium, 2023. Available online: https://environment.ec.europa.eu/publications/guidelines-defining-mapping-monitoring-and-strictly-protecting-eu-primary-and-old-growth-forests_en (accessed on 18 March 2026).

5. Luyssaert, S.; Schulze, E.D.; Börner, A.; Knohl, A.; Hessenmöller, D.; Law, B.E.; Ciais, P.; Grace, J. Old-growth forests as global carbon sinks. *Nature* **2008**, *455*, 213–215. [CrossRef] [PubMed]
6. Barros, L.D.A.; Venter, M.; Elkin, C.; Venter, O. Managing forests for old-growth attributes better promotes the provision of ecosystem services than current age-based old-growth management. *For. Ecol. Manag.* **2022**, *511*, 120130. [CrossRef]
7. Carle, H.; Bauman, D.; Evans, M.N.; Coughlin, I.; Binks, O.; Ford, A.; Bradford, M.; Nicotra, A.; Murphy, H.; Meir, P. Aboveground biomass in Australian tropical forests now a net carbon source. *Nature* **2025**, *646*, 611–618. [CrossRef] [PubMed]
8. Keren, S.; Svoboda, M.; Janda, P.; Nagel, T.A. Relationships between structural indices and conventional stand attributes in an old-growth forest in Southeast Europe. *Forests* **2019**, *11*, 4. [CrossRef]
9. Forzieri, G.; Girardello, M.; Ceccherini, G.; Spinoni, J.; Feyen, L.; Hartmann, H.; Beck, P.S.A.; Camps-Valls, G.; Chirici, G.; Mauri, A.; et al. Emergent vulnerability to climate-driven disturbances in European forests. *Nat. Commun.* **2021**, *12*, 1081. [CrossRef] [PubMed]
10. Spies, T.A.; Franklin, J.F. The structure of natural young, mature, and old-growth Douglas-fir forests in Oregon and Washington. In *Wildlife and Vegetation of Unmanaged Douglas-Fir Forests*; Ruggiero, L.F., Aubry, K.B., Carey, A.B., Huff, M.H., Eds.; General Technical Report PNW-GTR-285; USDA Forest Service, Pacific Northwest Research Station: Portland, OR, USA, 1991; pp. 91–110.
11. Tíscar, P.A.; Lucas-Borja, M.E. Structure of old-growth and managed stands and growth of old trees in a Mediterranean *Pinus nigra* forest in southern Spain. *Forestry* **2016**, *89*, 201–207. [CrossRef]
12. Hevia, A.; Calzado, A.; Alejano, R.; Vázquez-Piqué, J. Identification of old-growth Mediterranean forests using airborne laser scanning and geostatistical analysis. *Remote Sens.* **2022**, *14*, 4040. [CrossRef]
13. Spracklen, B.; Spracklen, D.V. Identifying European old-growth forests using remote sensing: A study in the Ukrainian Carpathians. *Forests* **2019**, *10*, 127. [CrossRef]
14. Spracklen, B.; Spracklen, D.V. Determination of structural characteristics of old-growth forest in Ukraine using spaceborne LiDAR. *Remote Sens.* **2021**, *13*, 1233. [CrossRef]
15. Hirschmugl, M.; Sobe, C.; Di Filippo, A.; Berger, V.; Kirchmeir, H.; Vandekerkhove, K. Review on the possibilities of mapping old-growth temperate forests by remote sensing in Europe. *Environ. Model. Assess.* **2023**, *28*, 761–785. [CrossRef]
16. Carrión, J.S.; Munuera, M.; Dupré, M.; Andrade, A. Abrupt vegetation changes in the Segura Mountains of southern Spain throughout the Holocene. *J. Ecol.* **2001**, *89*, 783–797. [CrossRef]
17. Abellanas, B.; Pérez-Moreno, P.J. Assessing spatial dynamics of a *Pinus nigra* subsp. *salzmannii* natural stand combining point and polygon patterns analysis. *For. Ecol. Manag.* **2018**, *424*, 136–153. [CrossRef]
18. Tíscar, P.A.; Linares, J.C. Structure and regeneration patterns of *Pinus nigra* subsp. *salzmannii* natural forests: A basic knowledge for adaptive management in a changing climate. *Forests* **2011**, *2*, 1013–1030. [CrossRef]
19. Jiménez, A.; Palacios, F.; Vázquez-Piqué, F.J.; Calzado-Carretero, A. Individualización de copas en encinares (*Quercus ilex* L.) mediante el empleo de técnicas de segmentación de imágenes y clasificación por objetos [Individual crown delineation in holm oak stands using image segmentation and object-based classification techniques]. *GeoFocus* **2022**, *29*, 131–144. [CrossRef]
20. Bruening, J.M.; Dubayah, R.O.; Pederson, N.; Poulter, B.; Calle, L. Definition criteria determine the success of old-growth mapping. *Ecol. Indic.* **2024**, *159*, 111709. [CrossRef]
21. Adiningrat, D.P.; Schlund, M.; Skidmore, A.K.; Abdullah, H.; Wang, T.; Heurich, M. Mapping temperate old-growth forests in Central Europe using ALS and Sentinel-2A multispectral data. *Environ. Monit. Assess.* **2024**, *196*, 841. [CrossRef] [PubMed]
22. Govedar, Z.; Arsenić, N. Vegetation indices monitoring by using Copernicus data in the old-growth forests of the Republic of Srpska/Bosnia and Herzegovina. *Front. For. Glob. Change* **2024**, *7*, 1354769. [CrossRef]
23. Räty, J.; Myllymäki, M.; Peltoniemi, M.; Lehtonen, A.; Packalen, P. Mapping old-growth forests using airborne lidar data and satellite images: How do plot size and rarity affect accuracy? *Can. J. For. Res.* **2025**, *55*, 1–15. [CrossRef]
24. de Almeida, C.T.; Galvão, L.S.; Ometto, J.P.H.B.; Jacon, A.D.; Pereira, F.R.D.S.; Sato, L.Y.; Silva-Junior, C.H.L.; Brancalion, P.H.S.; de Aragão, L.E.O.C. Advancing forest degradation and regeneration assessment through light detection and ranging and hyperspectral imaging integration. *Remote Sens.* **2024**, *16*, 3935. [CrossRef]
25. Lalechère, E.; Monnet, J.M.; Breen, J.; Fuhr, M. Assessing the potential of remote sensing-based models to predict old-growth forests on large spatiotemporal scales. *J. Environ. Manag.* **2024**, *351*, 119865. [CrossRef] [PubMed]
26. Birdsey, R.A.; DellaSala, D.A.; Walker, W.; Gorelik, S. How much more carbon could be protected in mature and old-growth forests of the United States? *Biol. Conserv.* **2025**, *306*, 111114. [CrossRef]
27. Erfanifard, Y.; Kraszewski, B.; Lisiewicz, M.; Mielcarek, M.; Czerepko, J.; Kuberski, L.; Sterenczak, K. A new lens on biodiversity assessment: The reliability of high-resolution remote sensing in investigating tree species diversity in old-growth forests. *For. Ecol. Manag.* **2025**, *595*, 122987. [CrossRef]
28. Erfanifard, Y.; Lisiewicz, M.; Sterenczak, K. High resolution remote sensing for biodiversity assessment and monitoring: A case study of dominant tree species in an old-growth forest. *For. Ecol. Manag.* **2024**, *566*, 122094. [CrossRef]
29. Rafieijahed, R.; Eslamdoust, J.; Martin, M. Predicting forest age up to 250 years in old-growth boreal mixedwoods of Western Quebec, Canada, using airborne laser scanning data and modelled tree attributes. *Can. J. Remote Sens.* **2025**, *51*, 1–20. [CrossRef]

30. Tíscar Oliver, P.A. Estructura, Regeneración y Crecimiento de *Pinus nigra* en el Área de Reserva Navahondona-Guadahornillos [Structure, Regeneration, and Growth of *Pinus nigra* in the Navahondona-Guadahornillos Reserve Area]. Ph.D. Thesis, Universidad Politécnica de Madrid, Madrid, Spain, 2004.
31. Alejano, R. Regeneración Natural de *Pinus nigra* Arn. ssp. *salzmannii* en las Sierras Béticas [Natural Regeneration of *Pinus nigra* Arn. ssp. *salzmannii* in the Baetic Ranges]. Ph.D. Thesis, Universidad Politécnica de Madrid, Madrid, Spain, 1997.
32. European Soil Data Centre (ESDAC). *The Soils of Europe Map*; Joint Research Centre, European Commission: Ispra, Italy, 2025. Available online: https://esdac.jrc.ec.europa.eu/ (accessed on 15 March 2026).
33. IUSS Working Group WRB. *World Reference Base for Soil Resources. International Soil Classification System for Naming Soils and Creating Legends for Soil Maps*, 4th ed.; International Union of Soil Sciences (IUSS): Vienna, Austria, 2022.
34. Dlesk, A.; Vach, K.; Šedina, J.; Pavelka, K. Comparison of Leica BLK360 and Leica BLK2GO on chosen test objects. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *XLVI-5/W1-2022*, 77–82. [CrossRef]
35. Hunčaga, M.; Chudá, J.; Tomaštík, J.; Slámová, M.; Koreň, M.; Chudý, F. The comparison of stem curve accuracy determined from point clouds acquired by different terrestrial remote sensing methods. *Remote Sens.* **2020**, *12*, 2739. [CrossRef]
36. Burrascano, S.; Keeton, W.S.; Sabatini, F.M.; Blasi, C. Commonality and variability in the structural attributes of moist temperate old-growth forests: A global review. *For. Ecol. Manag.* **2013**, *291*, 458–479. [CrossRef]
37. Roussel, J.-R.; Auty, D.; Coops, N.C.; Tompalski, P.; Goodbody, T.R.H.; Sanchez Meador, A.; Bourdon, J.-F.; de Boissieu, F.; Achim, A. lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sens. Environ.* **2020**, *251*, 112061. [CrossRef]
38. González-Ferreiro, E.; Arellano-Pérez, S.; Castedo-Dorado, F.; Hevia, A.; Vega, J.A.; Vega-Nieva, D.; Álvarez-González, J.G.; Ruiz-González, A.D. Modelling the vertical distribution of canopy fuel load using national forest inventory and low-density airborne laser scanning data. *PLoS ONE* **2017**, *12*, e0176114. [CrossRef] [PubMed]
39. Aragonés, D.; Rodríguez-Galiano, V.F.; Caparrós-Santiago, J.A.; Navarro-Cerrillo, R.M. Could land surface phenology be used to discriminate Mediterranean pine species? *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *78*, 281–294. [CrossRef]
40. Jackson, C.M.; Adam, E. A machine learning approach to mapping canopy gaps in an indigenous tropical submontane forest using WorldView-3 multispectral satellite imagery. *Environ. Conserv.* **2022**, *49*, 255–262. [CrossRef]
41. Lin, C.-Y.; Lin, C. Using ridge regression method to reduce estimation uncertainty in chlorophyll models based on WorldView multispectral data. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2019), Yokohama, Japan, 28 July–2 August 2019*; IEEE: Piscataway, NJ, USA, 2019; pp. 1777–1780. [CrossRef]
42. Wei, Y.; Lu, M.; Yu, Q.; Li, W.; Wang, C.; Tang, H.; Wu, W. The normalized difference yellow vegetation index (NDYVI): A new index for crop identification by using GaoFen-6 WFV data. *Comput. Electron. Agric.* **2024**, *226*, 109417. [CrossRef]
43. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring the vernal advancement and retrogradation (Greenwave effect) of natural vegetation. In *NASA Special Publication 351: Third Earth Resources Technology Satellite-1 Symposium*; Freden, S.C., Mercanti, E.P., Becker, M.A., Eds.; NASA Goddard Space Flight Center: Washington, DC, USA, 1974; Volume 1, Section A, pp. 309–317.
44. Gitelson, A.; Merzlyak, M.N. Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. *J. Plant Physiol.* **1994**, *143*, 286–292. [CrossRef]
45. Wu, C.; Niu, Z.; Tang, Q.; Huang, W. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agric. For. Meteorol.* **2008**, *148*, 1230–1241. [CrossRef]
46. Pasqualotto, N. Estimación del Índice de Área Foliar Verde y Marrón de Diferentes Cultivos Con Sentinel-2 [Estimation of Green and Brown Leaf Area Index of Different Crops Using Sentinel-2]. Master's Thesis, Universitat de València, Valencia, Spain, 2018.
47. Cardim, M.; Lima, F. Development of Multispectral Indices for Organic Fertilization Monitoring in Tomato Plants at Early Stages. Master's Thesis, Universitat Politècnica de València, Valencia, Spain, 2019. Available online: https://riunet.upv.es/handle/10251/129471 (accessed on 15 July 2026).
48. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]
49. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [CrossRef]
50. Frampton, W.J.; Dash, J.; Watmough, G.; Milton, E.J. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS J. Photogramm. Remote Sens.* **2013**, *82*, 83–92. [CrossRef]
51. Weiss, M.; Baret, F.; Jay, S. *S2ToolBox Level 2 Products: LAI, fAPAR, fCover—Algorithm Theoretical Basis Document (ATBD)*, version 2.1; European Space Agency (ESA): Paris, France, 2021.
52. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2024. Available online: https://www.R-project.org/ (accessed on 1 July 2026).
53. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.G.; Gruber, B.; Lafourcade, B.; Leitão, P.J.; et al. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2013**, *36*, 27–46. [CrossRef]

54. Oliveira, S.; Oehler, F.; San-Miguel-Ayanz, J.; Camia, A.; Pereira, J.M. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *For. Ecol. Manag.* **2012**, *275*, 117–129. [CrossRef]
55. Forkuor, G.; Hounkpatin, O.K.L.; Welp, G.; Thiel, M. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: A comparison of machine learning and multiple linear regression. *PLoS ONE* **2017**, *12*, e0170478. [CrossRef] [PubMed]
56. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
57. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
58. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [CrossRef]
59. Fisher, A.; Rudin, C.; Dominici, F. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.* **2019**, *20*, 1–81.
60. Cobos, M.E.; Owens, H.L.; Soberón, J.; Peterson, A.T. Detailed multivariate comparisons of environments with mobility oriented parity. *Front. Biogeogr.* **2024**, *17*, e132916. [CrossRef]
61. Elith, J.; Kearney, M.; Phillips, S. The art of modelling range-shifting species. *Methods Ecol. Evol.* **2010**, *1*, 330–342. [CrossRef]
62. Ozkan, U.Y.; Demirel, T.; Ozdemir, I.; Saglam, S.; Mert, A. Examining LiDAR–WorldView-3 data synergy to generate a detailed stand map in a mixed forest in the north-west of Turkey. *Adv. Space Res.* **2020**, *65*, 2608–2621. [CrossRef]
63. Zhao, Y.; Ma, Y.; Quackenbush, L.J.; Zhen, Z. Estimation of individual tree biomass in natural secondary forests based on ALS data and WorldView-3 imagery. *Remote Sens.* **2022**, *14*, 271. [CrossRef]
64. White, J.C.; Tompalski, P.; Vastaranta, M.; Wulder, M.A.; Saarinen, N.; Stepper, C.; Coops, N.C. *A Model Development and Application Guide for Generating an Enhanced Forest Inventory Using Airborne Laser Scanning Data and an Area-Based Approach*; Information Report FI-X-018; Canadian Forest Service, Natural Resources Canada: Victoria, BC, Canada, 2017; pp. 1–49. [CrossRef]
65. Vapnik, V.N. *The Nature of Statistical Learning Theory*, 2nd ed.; Springer: New York, NY, USA, 2000.
66. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*, 2nd ed.; Springer Texts in Statistics; Springer: New York, NY, USA, 2021. [CrossRef]
67. Rugani, T.; Diaci, J.; Hladnik, D. Gap dynamics and structure of two old-growth beech forest remnants in Slovenia. *PLoS ONE* **2013**, *8*, e52641. [CrossRef] [PubMed]
68. Hevia, A.; Álvarez-González, J.G.; Ruiz-Fernández, E.; Ruiz-González, A.D.; González-Ferreiro, E. Modelling canopy fuel and forest stand variables and characterizing the influence of thinning in the stand structure using airborne LiDAR. *Rev. Teledetec.* **2016**, *45*, 43–58. [CrossRef]
69. Fuhr, M.; Lalechère, E.; Monnet, J.M.; Bergès, L. Detecting overmature forests with airborne laser scanning (ALS). *Remote Sens. Ecol. Conserv.* **2022**, *8*, 731–743. [CrossRef]
70. Barros, L.D.A.; Elkin, C. An index for tracking old-growth value in disturbance-prone forest landscapes. *Ecol. Indic.* **2021**, *121*, 107175. [CrossRef]
71. Fischer, F.J.; Jackson, T.; Vincent, G.; Jucker, T. Robust characterisation of forest structure from airborne laser scanning—A systematic assessment and sample workflow for ecologists. *Methods Ecol. Evol.* **2024**, *15*, 1873–1888. [CrossRef]
72. Ploton, P.; Mortier, F.; Réjou-Méchain, M.; Barbier, N.; Picard, N.; Rossi, V.; Dormann, C.; Cornu, G.; Viennois, G.; Bayol, N.; et al. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **2020**, *11*, 4540. [CrossRef] [PubMed]
73. Meyer, H.; Pebesma, E. Machine learning-based global maps of ecological variables and the challenge of assessing them. *Nat. Commun.* **2022**, *13*, 2208. [CrossRef] [PubMed]
74. Milà, C.; Mateu, J.; Pebesma, E.; Meyer, H. Nearest neighbour distance matching Leave-One-Out Cross-Validation for map validation. *Methods Ecol. Evol.* **2022**, *13*, 1304–1316. [CrossRef]
75. Linnenbrink, J.; Milà, C.; Ludwig, M.; Meyer, H. kNNDM CV: K-fold nearest-neighbour distance matching cross-validation for map accuracy estimation. *Geosci. Model. Dev.* **2024**, *17*, 5897–5912. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.