Forestry: An International Journal of Forest Research, 2026, **99**, 1–17
https://doi.org/10.1093/forestry/cpaf080
Advance access publication date 2 December 2025
Original Article

Combining multitemporal remote sensing data and repeated ground observations from the Spanish National Forest Inventory to estimate site index and diameter distribution for uneven-aged management of *Juniperus thurifera L.* forests

Francisco Mauro[1,*], J. Bernardo González-Mesquida[1], Manuel Gomez-Roux[1], Darío Domingo-Ruiz[1,2], Cristina Gómez[1,3,4], Lorena Caiza-Morales[1], Francisco Rodríguez-Puerta[1], Beatriz Águeda[1], Bryce Frank[5], Steven Filippelli[6], Andrew T. Hudak[7], Hailemariam Temesgen[8], Patrick A. Fekety[6], José Antonio Manzanera[9], David Candel-Pérez[1,10]

[1]iuFOR, EiFAB, University of Valladolid, Campus Duques de Soria, sn, 42005 Soria, Spain
[2]GEOFOREST-IUCA, Department of Geography, University of Zaragoza, Pedro Cerbuna 12, 50009 Zaragoza, Spain
[3]Department of Geography, University of Valladolid., Campus Duques de Soria, sn, 420045 Soria, Spain
[4]Department of Geography and Environment, School of Geoscience, University of Aberdeen, Aberdeen AB24 3UE, Scotland, United Kingdom
[5]USDA Forest Service, Pacific Northwest Research Station, 3625 93rd Avenue SW, Olympia, Washington 98512, USA
[6]Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, Colorado 80523-1499, United States
[7]USDA Forest Service, Rocky Mountain Research Station, Moscow, Idaho 83843, United States
[8]Department of Forest Engineering, Resources and Management, Oregon State University, Corvallis Oregon 97333, United States
[9]Universidad Politécnica de Madrid. Research Group SILVANET. C. de José Antonio Novais, 10, 28040 – Madrid, Spain
[10]Sección de Desarrollo Territorial Sostenible, NASUVINSA, Av. de San Jorge, 8, 31012 Pamplona, Navarra
[*]Corresponding author. iuFOR, EiFAB, University of Valladolid, Campus de Soria, sn, 420045 Soria, Spain. E-mail: francisco.mauro@uva.es

Abstract

Spanish juniper, *Juniperus thurifera L*, forests are typically managed as uneven-aged forests in the Iberian range, which requires information about site index (SI) classes and diameter at breast height (dbh) distributions. The objective of this study is to evaluate and compare the performance of remote sensing models using multitemporal airborne laser scanning (ALS) data, Landsat metrics of state and change along with topographic, geographic and climatic predictors for the prediction of SI classes and diameter distributions. The best model for SI classes was a logistic model that showed an overall accuracy of 83%, a balanced accuracy of 77% and a kappa index of 0.52. The best model for diameter distributions consisted of an ensemble of imputation models for separate diameter classes and had a root mean squared Mahalanobis distance of 0.49, and class-wise coefficients of determination (R2) ranging from 64.95% for the 5–15 cm diameter class, to 89.63% for the 25–35 cm diameter class. Models were used to predict diameter distributions and SI classes, and these predictions were compared to the reference diameter distributions provided by a local silvicultural model. Deviations with respect to a reference silvicultural model were summarized in polygons of the Spanish Forest Map for the study area. These deviations facilitate the identification of areas requiring silvicultural interventions and highlight the value of integrating remote sensing technologies with ground-based inventories with repeated measurements for the adaptive management and conservation of *Juniperus thurifera* forests.

Keywords: site index; diameter distributions; airborne laser scanning; uneven-aged forests; Spanish juniper

Introduction

Endemic to southwestern Europe and northwestern Africa, the Spanish juniper (*Juniperus thurifera L.*), forms distinctive forests in Spain, France, Morocco, and Algeria. Owing to their uniqueness, these forests are protected under European Union regulations. One of the most significant areas within the species´ range is the high-elevation plateau between the Iberian and Central ranges. The decline of traditional land-use practices in this region, driven by rural abandonment, has facilitated an increase in Spanish juniper recruitment on formerly managed lands (Olano et al. 2008). However, several factors threaten to offset this recruitment. Shrub encroachment and interspecific competition are driving the replacement of Spanish juniper by other species such as *Quercus ilex* L., *Quercus faginea* Lam., and *Pinus nigra* Ait. (Olano et al. 2008; Martínez-(Martínez-Rodrigo et al. 2025). Moreover, rising temperatures and increasing aridity are expected to further limit regeneration potential and compromise the survival of mature trees (Acuña-Míguez et al. 2020). Together, these threats suggest that a contraction of the species' distribution is a plausible future

Handling editor: Dr. Fabian Fassnacht
Received 8 March 2025; **revised** 11 October 2025; **accepted** 29 October 2025

scenario. Consequently, proactive forest management interventions are essential to secure the long-term conservation of Spanish juniper forests.

On the high-elevation plateau between the Iberian and the Central ranges, Spanish juniper forests have traditionally been managed as uneven-aged forests for fodder production and construction timber. From an operational perspective, effective silvicultural planning in such systems requires two key types of information to ensure forest stability. First, managers need reference models that define the ideal or target condition for a specific forest type, derived from empirical studies and long-term observations. Second, they require an accurate assessment of the current forest condition. By comparing the present state with its target condition, managers can design tailored silvicultural treatments to guide the system toward the desired state (von Gadow and Puumalainen 2000, p. 331).

A reference model for these uneven-aged Spanish juniper forests exists in the region (Barrio de Miguel 2006). This model provides a description of the ideal state of the forest expressed as a target diameter at breast height (dbh) distribution across three site index (SI) classes and includes harvest prescriptions spaced 20 years in time and post-harvest diameter distributions once the ideal state is reached. For effective application of this model, forest managers need information on the SI class and current diameter distribution within stands and comparable management units.

Traditionally, SI and diameter distributions have been estimated through ground-based field inventories, which are both costly and time-intensive. These methods make limited use of the extensive spatially explicit auxiliary information now available and face significant limitations in mapping the spatial distribution of these variables. Previous studies have demonstrated that by integrating supervised learning techniques with remote sensing data it is possible to precisely estimate both SI (Gatziolis 2007, Penner et al. 2023) and diameter distributions (Bollandsås and Næsset 2007, Maltamo et al. 2007). However, precision of these methods depends heavily on factors such as the availability of auxiliary information and the choice of modeling techniques.

SI is the most widely used descriptor of site productivity and is a necessary input for numerous tree- and stand-level growth models. For even-aged forests, SI is typically defined as a dominant height value at a reference stand age. Multiple studies attempted to obtain spatially explicit predictions of SI, from topographic and climate variables (Gustafson et al. 2003, McKenney and Pedlar 2003). However, these models lacked the spatial resolution needed to inform operational forest management decisions. With the advent of airborne laser scanning (ALS) technologies in forest inventories, fine scale (i.e. 20–30 m resolution) determinations of SI became possible. Owing to the high correlation of ALS data with dominant height, early studies showed that when age is known, SI can be reliably predicted at fine scales using data from a single ALS acquisition (Gatziolis 2007).

However, stand age may be unknown in naturally regenerated areas and in areas without precise management records. Furthermore, for uneven-aged stands, there is no clear concept of stand age as multiple cohorts coexist in the same place, and the standard definition of SI as a dominant height at a given age is not meaningful. To overcome this problem in the model from Barrio de Miguel (2006) for uneven-aged Spanish juniper forests, SI is defined as average annual basal area increment. Multiple authors have shown that by comparing rates of change in ALS metrics to modeled growth curves, SI can be determined for even aged forests when no information about stand age is available (Guerra-Hernández et al. 2021, Penner et al. 2023). This age-independent approach is an encouraging advancement that poses questions about the potential use of multitemporal ALS data to estimate SI in uneven-aged Spanish juniper forests.

Diameter distributions provide information about the stocking levels in a forest stand, disaggregated by diameter classes, therefore informing about the tree size variability. Stocking levels in the diameter distribution can be expressed as frequencies, basal areas or volumes per diameter class. This information is necessary for precise appraisals of forest products but also to guide management activities in uneven-aged forests. Multiple techniques have been employed to model diameter distributions using auxiliary variables, particularly since the early 2000s when ALS data and the area-based approach (ABA) became operational (Næsset 2002). For ABA, regression models to predict percentiles of the diameter distribution—i.e. specific values of tree height below which a given percentage of trees fall (e.g. the 95th percentile)—have been widely used (Bollandsås and Næsset 2007, Maltamo et al. 2007, Gobakken and Næsset 2012), but models for abundance by diameter classes have also been tested (Finley et al. 2014). For continuous distributions, parameter recovery and parameter prediction methods are abundant in the literature (Arias-Rodil et al. 2018). However, other parametric alternatives based on generalized linear models (Breidenbach et al. 2008, Räty et al. 2021) or based on geometrical models for the interception of ALS returns have shown accurate results for both diameter distributions and tree-height-distributions (Magnussen et al. 1999, Mehtätalo and Nyblom 2012, Spriggs et al. 2015). Finally, ABA methods based on k-nearest neighbor (kNN) imputation techniques have been shown to be effective at predicting tree-lists, i.e. tables that mimic the structure of a set of plot level measurements that allow recovering discrete distributions with arbitrary diameter classes (Packalén and Maltamo 2007, Hudak et al. 2014, Mauro et al. 2019, Pukkala et al. 2024). Individual tree detection (ITD) methods have also received important attention. While they are more computationally intensive than ABA methods, under certain conditions ITD methods allow predicting diameter distributions with good accuracy. Factors such as the density of ALS returns or the sparseness of the forest have an impact on the results of ITD methods, and best performances are typically obtained in sparse forests and with return densities >8 returns per square meter (Sparks et al. 2022).

In 2023, the Spanish National Plan for Aerial Orthophotography (Plan Nacional de Ortofortografía Aerea, PNOA) released the final project area of the second round of ALS acquisitions, encompassing the high-elevation plateau between the Iberian and Central ranges. This development opens new opportunities for area-wide implementation of remote sensing-based forest inventories. In this study, we explore this potential using Spanish juniper forests as a case study, given their high ecological value and distinctive structure, which differs markedly from many other forest types where such approaches have already been tested. By integrating newly available ALS data with Landsat time series, as well as topographic, geographic, and climate predictors, we aimed to develop models for predicting diameter distributions and SI classes in Spanish juniper forests across this region. These predictions can provide decision-support tools for local managers, particularly in applying the silvicultural model proposed by Barrio de Miguel (2006), and inform sustainable management strategies for these unique forests. The primary objective of this study

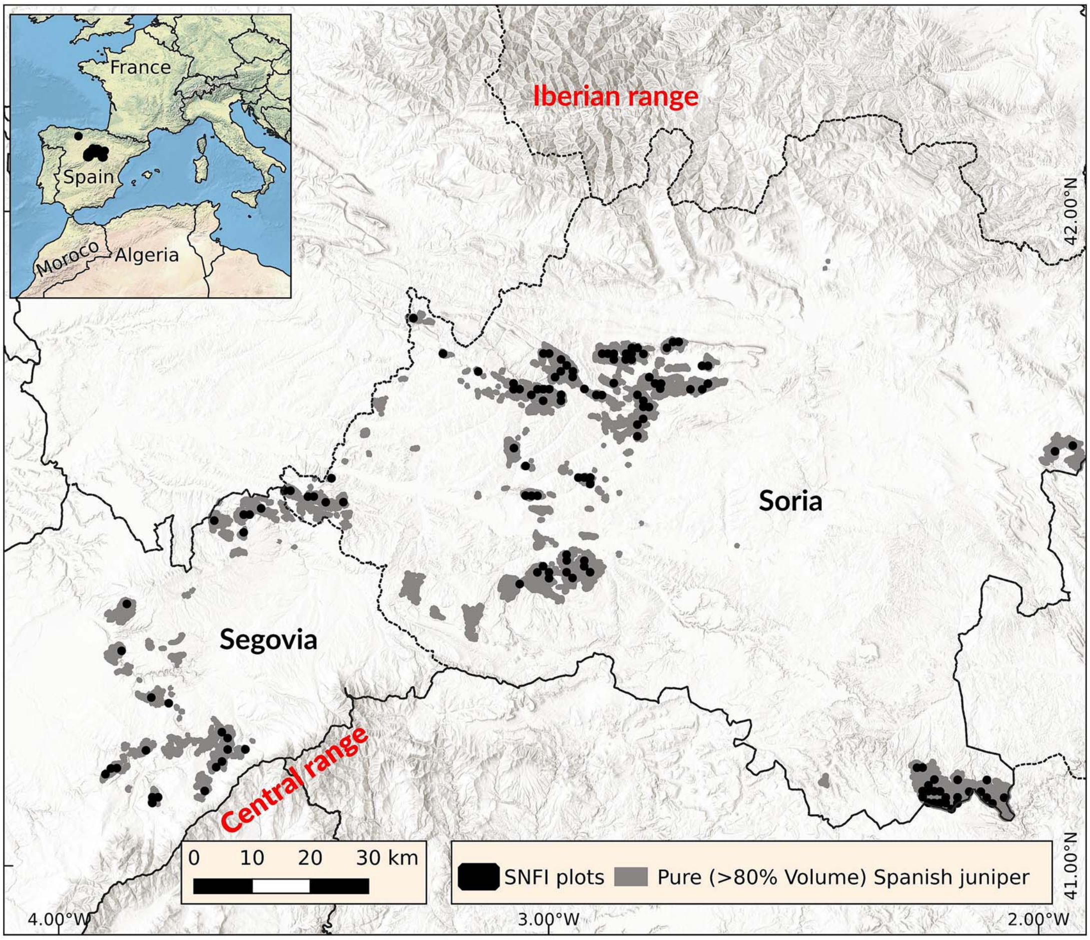

Figure 1. Location of the study area. SFM polygons with at least 80% dominance of Juniperus thurifera. Plot locations of the Spanish National Forest Inventory encompassed within the study area displayed in black.

was to develop, quantitatively evaluate, and compare alternative models for predicting SI classes and diameter distributions. As a secondary objective, we used the best-performing models to generate spatially explicit predictions of these variables, which were then analyzed to assess the alignment of current forest conditions with the target states defined in the aforementioned silvicultural model.

Material and methods

Study area

The study area encompasses polygons from the Spanish Forest Map (SFM) within the provinces of Soria and Segovia, where Spanish juniper occupies at least 80% of the polygon area (see Forest Map of Spain (2025) for details on the SFM). This set comprises 880 polygons that make a total area of ∼33 000 hectares (see Supplementary material Fig. S1, for photos illustrating the forest structures typically found in these polygons). On average, selected polygons are slightly larger than typical stands in the study area, with an average size of 38 hectares, but are highly variable in size, i.e. the coefficient of variation of the size of the polygons is 152.25%. Elevations in these areas range from 868 to 1659 m a.s.l, with predominantly alkaline shallow soils. The climate is characterized by hot, dry summers and cold winters, with annual precipitation ranging from 371 mm to 738 mm.

Table 1. Synthesis of plots employed in this study for modeling diameter distributions and SI. "All plots" represent plots from the 4th SNFI cycle and "Remeasured" represents plots in both the 3rd and 4th SNFI measurement cycles.

		n			
All plots		128			
Remeasured		112	SI class		
	Disturbed	26	Low	Medium	High
	Undisturbed	86	67	6	13

Ground observations and response variables

We selected all plots from the fourth cycle of the Spanish National Forest Inventory (SNFI) within the selected SFM polygons. In total, 128 plots measured during the fourth cycle (i.e. the measurement years were 2018 and 2019) were chosen, of which 112 had also been measured in the previous SNFI cycle during 2003 and 2004 (Table 1). Plots where field crews recorded harvest or thinning activities and those showing a decrease in basal area between the third and fourth SNFI measurements were labeled as "disturbed". The remaining plots were classified as "undisturbed". Secondary species were excluded from the analysis, as their volume averaged <3% and are not typically considered when making forest management decisions in these areas. In >75% of the plots, Spanish juniper accounted for >95% of the total plot volume.

Table 2. Mean and standard deviation of Spanish juniper volume per hectare and number of trees per plot in each diameter classes and for all diameter classes combined.

Diameter class		Mean volume (m^3/ha)	Standard deviation volume (m^3/ha)	Mean trees per plot	Standard deviation trees per plot
1st class	V_{5-15}	6.62	7.73	2.88	2.71
2nd class	V_{15-25}	10	8.65	4	3.37
3rd class	V_{25-35}	8.12	11.44	2.57	3.44
4th class	$V_{>35}$	5.38	12.278	1.21	2.53
All classes		30.12	24.5	10.66	7.2

The silvicultural model from Barrio de Miguel (2006) establishes three SI classes (i.e. low, medium, and high) from average annual basal area increments. These classes are separated using the thresholds of 0.17 m^2/(ha year) and 0.21 m^2/(ha year) that were derived from an empirical relationship based on ground observations from local management projects. Disturbed plots were not considered in the determination of SI classes because basal area changes in these plots are not the result of site factors alone. For each undisturbed plot with repeated measurements, we calculated the basal area increment of Spanish juniper by taking the difference between basal area measurements from the current and previous SNFI measurement cycles. SNFI plots are structured as nested concentric plots. Trees with dbh larger than 42.5 cm are measured in a 25 m radius plot, trees with dbh in the interval (42.5 cm, 22.5 cm) are measured in a 15 m radius plot, trees with dbh in the interval (22.5 cm, 12.5 cm) in are measured in a 10 m radius plot and trees with dbh 12.5 or smaller are measured in a 5 m radius plot. The i^{th} plot's basal area (G) for each SNFI cycle was calculated by summing the cross-sectional area of each Spanish juniper tree and adjusting it to a per unit area basis using the trees' expansion factors dictated by the trees' dbh. Basal area changes were computed as the sum, adjusted with the corresponding expansion factor, of the basal area increments for trees that were measured in the first field visit that remained alive in the second one minus mortality. Annualized basal area changes, ΔG_i, were computed by dividing basal area change by the number of years elapsed between SNFI measurements. The thresholds from the reference silvicultural model were applied to ΔG_i to classify each plot's SI class. Due to the predominance of plots in the *Low* SI class and the scarcity of medium SI plots, medium and high SI classes were combined for modeling purposes.

For all plots measured during the current SNFI cycle, we generated a discrete volume-weighted diameter distribution using diameter classes aligned with those in the Spanish juniper silvicultural model. The diameter distribution for the i^{th} SNFI plot, $\mathbf{y}_i$,is represented by a 4D vector, as shown in Equation (1). The first, second, third and fourth diameter classes, $V_{i,5-15}$, $V_{i,15-25}$, $V_{i,25-35}$ and $V_{i,>35}$, were respectively calculated by scaling to a per hectare basis, individual-tree volumes provided by SNFI and based on local allometric equations and then summing the volumes within each diameter class. Table 2 presents the mean and standard deviation of total Spanish juniper volume per unit area, as well as Spanish juniper volume per unit area and by diameter class.

$$\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \end{pmatrix} = \begin{pmatrix} V_{i,5-15} \\ V_{i,15-25} \\ V_{i,25-35} \\ V_{i,>35} \end{pmatrix} \quad (1)$$

Auxiliary information for area-based approach models

We compiled a spatial database of auxiliary variables potentially correlated with both diameter distributions and SI. This database includes variables from two separate ALS acquisitions, spectral information from Landsat image time series, and a range of topographic, geographic, and climatic descriptors. A brief overview of the auxiliary variables in each category is provided in the following subsections, with a comprehensive list available in Supplementary material, Tables S1 to S4).

Airborne laser scanning predictors of state and change

Available ALS data from the first and second nation-wide ALS acquisition campaigns performed by PNOA were used to obtain predictors of state and change. Hereafter we will use PNOA1 and PNOA2 to refer to the first and second ALS acquisition campaigns that took place between August and December 2010 and November 2017 and October 2019, respectively. The nominal pulse density of PNOA1 was 0.5 points (i.e. first returns) per square meter and the nominal return density for PNOA2 ranged from 1 to 4 points per square meter (Table 3).

For each ALS dataset we filtered out noise and overlap points, obtained digital terrain models (DTMs), and normalized point clouds using the lidR R package (Roussel et al. 2020). The normalization was performed using the vendor-provided ground classification, and the *normalize_height* and *knnidw* functions implemented in the lidR package. Normalized clouds from PNOA1 and PNOA2 were used to compute the ALS metrics described in Supplementary material Table S2, at a 30 m resolution, following methods described by Hudak et al. (2020). Considering the variable radius design of the SNFI plots, i.e. there is not a unique plot area, the 30 m resolution was selected as an intermediate option that had the advantage of matching the resolution of the Landsat imagery used in the study. Metrics from the PNOA2 acquisition were used as auxiliary variables in modeling diameter distributions and were considered proxies describing current structure. For each metric we obtained a rate of change per year, dividing the difference between PNOA2 and PNOA1 values by the number of years elapsed between acquisitions. These metrics were considered proxies of structural change and were used for SI modeling.

Spectral predictors of state and change

Spectral information corresponding to the third and fourth SNFI measurement cycles were obtained by creating cloud-free time series of Landsat 5, 7, and 8 summer images (i.e. June to September) for the 1984–2023 period using Collection 2 Level 2 Tier 1 surface reflectance data in Google Earth Engine, GEE (Gorelick et al. 2017). Yearly composites were obtained using medoids, then yearly values of the normalized difference vegetation index (NDVI) were used to temporally segment each

Table 3. ALS data collection specifications of the Spanish National Plan for Aerial Orthophotography (PNOA). PNOA 1, refers to the first nationwide ALS campaign and PNOA 2 to the second.

ALS acquisition	Years	Sensor	Scan frequency	Return density
PNOA 1.	2010	LEICA ALS50	> 45 kHz	0.5 points/m^2
PNOA 2. North block	2017–2018	RIEGL LMS-Q1560		4 points/m^2
PNOA 2. South block	2017–2018	LEICA ALS80		1 point/m^2

pixel time series using the GEE implementation of the LandTrendr algorithm (Kennedy et al. 2010, Kennedy et al. 2018). Finally, linear segments fitted by LandTrendr between identified breakpoints were used to remove noise that can be attributed to small interannual variations within segments. These smoothed time series were used to compute spectral values for all bands and indices indicated in Supplementary material Table S2, for the years 2004 and 2019. These years respectively accounted for 57% and 93% of the ground measurements of the third and fourth SNFI cycles in the study area. Spectral change metrics were derived by calculating the difference between the smoothed values from the third and fourth SNFI cycles and then converting these differences into annual rates of change.

Topographic and geographic predictors

DTMs obtained from the most recent ALS acquisition were used to derive a set of topographic and geographic metrics at 30 m resolution. These predictors (Supplementary material Table S3) include fundamental descriptors such as elevation, slope, aspect, latitude and longitude, as well as metrics capturing terrain roughness and variability across multiple scales to reflect topographic influences on vegetation growth and forest structure.

Climate predictors

We obtained the 19 bioclimatic variables and 30-year mean values for the period, 1970–2000, for vapor pressure and solar radiation provided by WorldClim2 at 30 arc second resolution (Fick and Hijmans 2017). All climate descriptors (Supplementary material Table S4) were resampled to the 30 m grid used for ALS and topographic metrics through bilinear interpolation. These climate variables are expected to correlate with both SI and diameter distributions.

Computation of area-based approach predictors for Spanish National Forest Inventory plots

ALS metrics for each plot were derived by clipping the normalized point clouds from PNOA1 and PNOA2 with a 16.93 m radius circle centered on each plot, ensuring alignment with the 900 m^2 area used for gridded metrics. Normalization of PNOA1 and PNOA2 was performed with their respective DTMs. Then ALS metrics for state and change detailed in previous subsections were calculated. Spectral metrics, spectral change metrics, topographic, geographic, and climate metrics were obtained by retrieving grid values at the plot centers for the variables described in previous subsections.

Individual tree detection

For each plot, we generated a 1 m resolution digital canopy height model (CHM) with the PNOA2 normalized point cloud using all points within a 100 m buffer around the plot center to avoid edge effects. The CHM was created with the pitfree algorithm by Khosravipour et al. (2014), as implemented with default parameters in the lidR R package (Roussel and Auty 2017). Tree tops were identified using the local maximum filter algorithm by Popescu and Wynne (2004). In the next step, we selected all treetops within a 16.93 m radius around the plot center, aligning with the area of 900 m^2 used for all other auxiliary variables, and generated a histogram of tree-top height with 1 m interval width.

Table 4. Volume before and after harvest operations spaced 20 years apart according to the silvicultural model from Barrio de Miguel (2006). The medium–high class is the average of the medium and high SI classes originally considered in the model.

SI class	Diameter class	Before V(m^3/ha)	After V(m^3/ha)
Low	V_{5-15}	2.37	2.37
	V_{15-25}	6.35	4.94
	V_{25-35}	9.77	9.31
	$V_{>35}$	5.21	1.19
Medium–high	V_{5-15}	4.16	4.16
	V_{15-25}	11.565	8.81
	V_{25-35}	17.79	17.11
	$V_{>35}4$	13.425	5.37

Site index class modeling

To model SI classes, only remeasured and undisturbed plots were used (Table 1). Disturbed plots were not considered in the modeling because in these plots basal area changes are decoupled from site factors. We used ALS predictors of state and change. The ALS auxiliary information only covers part of the time between SNFI measurements; thus, ALS proxies of state and change were complemented with Landsat proxies of state and change, and topographic, geographic, and climate metrics. We considered four model types to predict SI classes:

a) The first type was a non-parametric random forest classification model, fitted using the randomForest algorithm in R (Liaw and Wiener 2002). All auxiliary variables were initially considered; however, variable selection was conducted using the Variable Selection Using Random Forest, VSURF, algorithm (Genuer et al. 2022) to refine the model. For both, the variable selection and the fitting of the final model, we used the default values of ntree and mtry for classification settings. In this case, the value of ntree was 500 and mtry was 10.
b) The second type was a logistic generalized linear model with auxiliary variables selected using a backwards elimination stepwise starting with a saturated model, i.e. a model containing all predictors, followed by a manual addition/removal of variables supervised by the modeler to further reduce the predictors set. Thresholds for distinguishing between low and medium–high si classes were determined by maximizing the kappa index.
c) The third type was a non-parametric random forest regression model for changes in basal area ΔG_i that implemented

the variable selection of the classification model based on VSURF using default values of ntree and mtry for regression settings. In this case, ntree was 500 and mtry was 29.

d) The fourth model type is a linear regression model with auxiliary variables selected using the regsubsets function from the leaps R package (Lumley 2020). This function generated a list of 25 candidate models, each including up to 5 auxiliary variables. Models with a variance inflation factor (VIF) >10 or with coefficients not significant at a 0.05 level were discarded and from the remaining models, the one maximized the adjusted R^2 was selected.

For the models c and d, we transformed predictions of ΔG_i into predicted SI classes using the thresholds provided by the silvicultural model. Then, for all models we computed class-wise omission and commission errors, overall accuracy (OA), balanced accuracy (BA), sensitivity, specificity and the kappa index to assess model performance. Finally, models were ranked based on BA to identify the best-performing one.

Diameter distribution modeling

To model diameter distributions providing volume per diameter class, four types of models were compared. The first three, a multivariate linear regression model, a random forest model and kNN imputation model, are pure ABA methods, while the fourth category combines an ITD step with an ABA calibration using neural networks to transform the output of the ITD method into a diameter distribution. Models incorporated ALS state predictors, spectral state predictors, topographic, geographic, and climate metrics. Climate metrics were included as potential predictors for diameter distributions because climate has an impact on SI, which indirectly determines the target diameter distributions in the silvicultural model from Barrio de Miguel (2006). Before any model selection, we removed highly correlated predictors from the auxiliary information database using a 0.95 Pearson correlation threshold.

Multivariate linear regression model

The first type of ABA model used to predict diameter distributions consisted of a multivariate linear regression model that linked volume by diameter classes with auxiliary variables. To construct the multivariate model, we started fitting separate regression models for each diameter class. For the j^{th} diameter class, the initial model took the following form

$$y_{ij} = \mathbf{x}^{t}_{ij}\boldsymbol{\beta}_j + \varepsilon_{ij} \tag{2}$$

Where y_{ij} is the volume per hectare for the j^{th} diameter class in the i^{th} plot, $\mathbf{x}_{ij}$ is the vector of auxiliary variables for the plot, t the transpose operator, $\boldsymbol{\beta}_j$ a vector of model coefficients for the j^{th} class, and ε_{ij} the model error for the i^{th} plot when considering the j^{th} diameter class. Model error variance was allowed to increase with the predicted value $\mathbf{x^t}_{ij}\boldsymbol{\beta_j}$ using an exponential variance function that depended on a parameter α_j that controlled the rate at which model error variance grew.

$$Var\left(\varepsilon_{ij}\right) = \sigma_j^2 e^{\left(2\alpha_j \mathbf{x}^t_{ij}\boldsymbol{\beta}_j\right)} \tag{3}$$

Auxiliary variables were selected using the regsubsets method, as previously outlined for ΔG_i.Once the initial models were fit, we obtained the final multivariate linear model that accounted for potential correlations in errors across diameter classes. Auxiliary variables for each diameter class were those selected in the preliminary univariate models. Similarly, error heteroscedasticity was modeled using the same variance function that was used for the preliminary univariate models. Finally, multivariate model errors were assumed to be uncorrelated between plots regardless of the diameter class but potentially correlated across diameter classes within plots, Equation (4). Note that in Equation 4, plots are represented by subindexes i and p that range from 1 to 128 (i.e. the number of plots used to model diameter distributions) and diameter classes by subindexes j and l that range from 1 to 4 (i.e. the number of modeled diameter classes).

$$Corr\left(\varepsilon_{ij}, \varepsilon_{pl}\right) = \begin{cases} 0 & if\ i \neq p \\ -1 \leq \rho_{jl} = \rho_{lj} \leq 1, & if\ j \neq l\ and\ i = p \\ \rho_{jl} = \rho_{lj} = 1, & if\ j = l\ and\ i = p \end{cases} \tag{4}$$

Final coefficients were estimated through maximum likelihood using the nlme R package.

Random forest model

The second type of ABA models used to predict diameter distributions comprised a set of separate random forest models, one for each diameter class. Auxiliary variables for each diameter class were selected using the VSURF algorithm using default values of ntree and mtry for regression setting, in this case ntree was 500 and mtry was 29.

k-nearest neighbor imputation

The third type of models to predict diameter distributions employed separate kNN imputation models for each diameter class. For each diameter class we tested ten combinations of number of neighbors, k, and distance metrics to select neighbors. The number of neighbors, k, was allowed to vary between 1 and 5, and the distance metrics to select neighbors were the most similar neighbor (*msn*) distance and the *random forest* distance, both implemented in the yaImpute package (Crookston and Finley 2007). A variable selection step was conducted for each of the ten combinations of distance and k using a genetic algorithm outlined in Supplementary material section 3, that restricted the maximum number of predictors per model to five. The resulting models for each diameter class, j, were ranked based on their root mean squared error ($RMSE_j$):

$$RMSE_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{ij} - \hat{y}_{ij}\right)^2} \tag{5}$$

Finally, the model with the lowest $RMSE_j$ was selected to make predictions for the j^{th} diameter class.

Neural network calibration of individual tree detection output

To transform the histogram of tree top heights obtained with the method described in 2.3.6 into a diameter distribution we developed a set of four neural network models, one per diameter class, with a common structure. Each network had two branches that converged into a common block that generated predictions of volume in the diameter class. The first branch received as input the histogram of tree top heights and consisted of a single dense layer with six neurons and sigmoid activation functions to ensure outcomes were bound between zero and one. The vector of ABA

Table 5. Accuracy assessment metrics for the SI-class models. Sens: sensitivity, spec: specificity, Obs: observed, Low: low SI class, MH: medium–high SI class.

Model	**Confusion matrix**					
Random forest classification		Reference				Kappa
			Low	MH	Class error	0.41
	Prediction	Low	61	13	0.17	
		MH	6	6	0.50	
	Class error	0.09	0.68			
		Sens	Spec	BA	OA	
		0.91	0.31	0.61	0.78	
Logistic generalized linear model		Reference				Kappa
			Low	MH	Class error	0.47
	Prediction	Low	67	12	0.15	
		MH	0	7	0.00	
	Class error	0.00	0.63			
		Sens	Spec	BA	OA	
		0.34	1	0.68	0.86	
Random forest regression		Reference				Kappa
			Low	MH	Class error	0.32
	Prediction	Low	58	11	0.16	
		MH	8	8	0.50	
	Class error	0.12	0.42			
		Sens	Spec	BA	OA	
		0.88	0.42	0.65	0.77	
Linear regression		Reference				Kappa
			Low	MH	Class error	0.52
	Prediction	Low	58	6	0.09	
		MH	9	13	0.41	
	Class error	0.13	0.31			
		Sens	Spec	BA	OA	
		0.87	0.68	0.77	0.83	

Table 6. Coefficients of the linear regression model for annualized change in basal area, ΔG_i, used to predict SI classes.

Coefficient	**Estimate**	**Std error**	**P-value**	**Residual variance**	R^2
Intercept	−3.483	0.82	<.01	0.08	42.53%
Proportion of first returns >4 m	0.34	0.09	<.01		
Max temperature of warmest month	−0.026	0.013	.04		
Mean solar radiation	2.84×10^{-4}	6.13×10^{-5}	<.01		
ΔProportion of first returns between 2 m and 4 m	−0.27	0.04	<.01		
ΔProportion of returns between 2 m and 4 m	2.49	0.04	<.01		

auxiliary variables was fed into the second branch, consisting of two dense layers with 24 and 9 neurons, respectively. The first layer used a leaky rectified linear unit (ReLU) activation function, while the second layer employed a sigmoid activation. The outcomes from both branches were then concatenated into a 12D vector, which was passed to a sequence of two dense layers. The first layer had eight nodes and ReLU activations, and the output layer featured a single neuron with a softplus activation function to ensure that the output prediction remained positive as they represented volumes for diameter classes (see Supplementary material Fig. S2). Neural networks for each diameter class were implemented using the R interface for the *tensorflow* and *keras* libraries (Allaire and Chollet 2024, Allaire and Tang 2024) and were trained for 2500 epochs, using a 50% batched split between training and validation datasets.

Comparison of models for diameter distributions

To identify the best model, we used the root mean squared Mahalanobis distance between predictions and observations, *RMSMD*, as the ranking criterion. The *RMSMD* is calculated as follows:

$$RMSMD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\left(\mathbf{y}_i - \hat{\mathbf{y}}_i\right)^t \boldsymbol{\Sigma}^{-1}\left(\mathbf{y}_i - \hat{\mathbf{y}}_i\right)\right]} \tag{6}$$

Where $\hat{\mathbf{y}}_i$ is the predicted diameter distribution vector for the i^{th} plot, n is the number of plots used in the modeling, and $\boldsymbol{\Sigma}^{-1}$ is the inverse of the variance–covariance matrix of the components of the vectors $\mathbf{y}_i$. Each entry of $\boldsymbol{\Sigma}$ is calculated as a sample variance or covariance and once all entries of $\boldsymbol{\Sigma}$ are calculated, the matrix is inverted.

In addition to *RMSMD*, diameter-class $RMSE_j$ and coefficients of determination, R_j^2, and other indicators of similarity between predicted and observed diameter distributions such as the average Reynolds error index, *REI*, (Reynolds et al. 1988) and the average relative Reynolds error index, *RREI*, proposed by Packalén and Maltamo (2008) were also computed to assist in the evaluation

Table 7. Coefficients of the multivariate regression models for volume by diameter classes.

Variable	Coefficient	Value	P-value	σ_j	α_j	$Corr(\varepsilon_{ij},\varepsilon_{il})$
$j=1$	Intercept	2.38	.04	4.70	0.06	$Corr(\varepsilon_{i1},\varepsilon_{i2})=0.02$
V_{5-15}	Proportion of first returns between 1 m and 2 m	0.76	<.01			$Corr(\varepsilon_{i1},\varepsilon_{i3})=0.09$
						$Corr(\varepsilon_{i1},\varepsilon_{i4})=-0.03$
$j=2$	Intercept	−17.37	.01	2.93	0.06	$Corr(\varepsilon_{i2},\varepsilon_{i3})=-0.04$
V_{15-25}	Proportion of first returns >4 m	55.08	<.01			$Corr(\varepsilon_{i2},\varepsilon_{i4})=-0.17$
						$Corr(\varepsilon_{i3},\varepsilon_{i4})=-0.07$
	Proportion of first returns between 2 m and 4 m	0.36	<.01			
	Proportion of first returns between 8 m and 10 m	−1.16				
	Min temperature of coldest month	−3.06	.01			
	Mean temperature of coldest quarter	3.02	.04			
$j=3$	Intercept	−20.96	.03	3.28	0.05	
V_{25-35}	Proportion of first returns between 6 m and 8 m	1.60	<.01			
	Average topographic position index	0.10	.02			
	Isothermality	0.56	.02			
$j=4$	Intercept	5.32	<.01	3.27	0.08	
$V_{>35}$	Proportion of first returns between 1 m and 2 m	−0.52	<.01			
	Proportion of first returns between 8 m and 10 m	2.46	<.01			

of competing models.

$$REI = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{4}\left|y_{ij}-\hat{y}_{ij}\right| \tag{7}$$

$$RREI = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\sum_{j=1}^{4}\left|y_{ij}-\hat{y}_{ij}\right|}{\sum_{j=1}^{4}y_{ij}}\right) \tag{8}$$

Departure from reference diameter distribution model

To assess the overall state of the Spanish juniper dominated SFM polygons in the study area with respect to the reference model, we first used the European Disturbance Atlas from Viana-Soto and Senf (2025) to mask all areas subject to disturbances in the period 2003 to 2019. Following Riofrío et al. (2023), SFM polygons with >40% of their area classified as disturbed were removed from the evaluation of the departure from the silvicultural model. For all other polygons, predictions of the SI class and the diameter distribution were obtained for all undisturbed pixels using the best model for each response. These predictions were used to calculate differences with respect to the silvicultural model by Barrio de Miguel (2006) (Table 4) as follows. For pixels where the predicted volume for a given diameter class exceeded the reference value before harvest operations, we determined the difference between the current volume and the before-harvest reference volume for the diameter class. Conversely, for plots where the volume for a given diameter class was below the reference value after harvest operations, we calculated the difference between the after-harvest reference volume and the predicted volume for the diameter class. In instances where the predicted volume for a given diameter class was between the before- and after-harvest reference values of the silvicultural model, we assigned a difference of zero for that diameter class. For each SFM polygon we calculated the average value of the before and after-harvest reference volume. Finally, we calculated the average difference with respect to the silvicultural model for all pixels within each polygon and compared the distribution of those differences to the differences obtained for the set of undisturbed SNFI plots.

To facilitate evaluating the differences with respect to the silvicultural model, we represented the SFM polygons and SNFI plots in a cartesian space where the x-axis, I_1, is the sum of the differences for the first and second diameter classes, i.e. younger trees, and the y-axis, I_2, is the sum of the differences for the third and fourth diameter classes, i.e. mature trees. This representation allows identifying areas with an excess or deficit of younger or mature trees. In addition, we computed the proportion of SNFI plots in each quadrant of the cartesian space $I_1 x\, I_2$ and computed the confidence interval, (CI), for the proportion of plots in each quadrant as

$$CI = \hat{p} \pm 2\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} \tag{9}$$

where $\hat{p}$ is the proportion of plots in the considered quadrant and n is the number of undisturbed SNFI plots. Finally, we compared these intervals to the proportion of SFM polygons area in each quadrant.

Results

Site index class modeling

The linear regression model for ΔG_i resulted in an OA 0.83 and a BA of 0.77. This model outperformed the logistic generalized linear model, the random forest classification model and the random forest regression model (Table 5). Therefore, this model was selected to predict SI classes in further stages of our analysis.

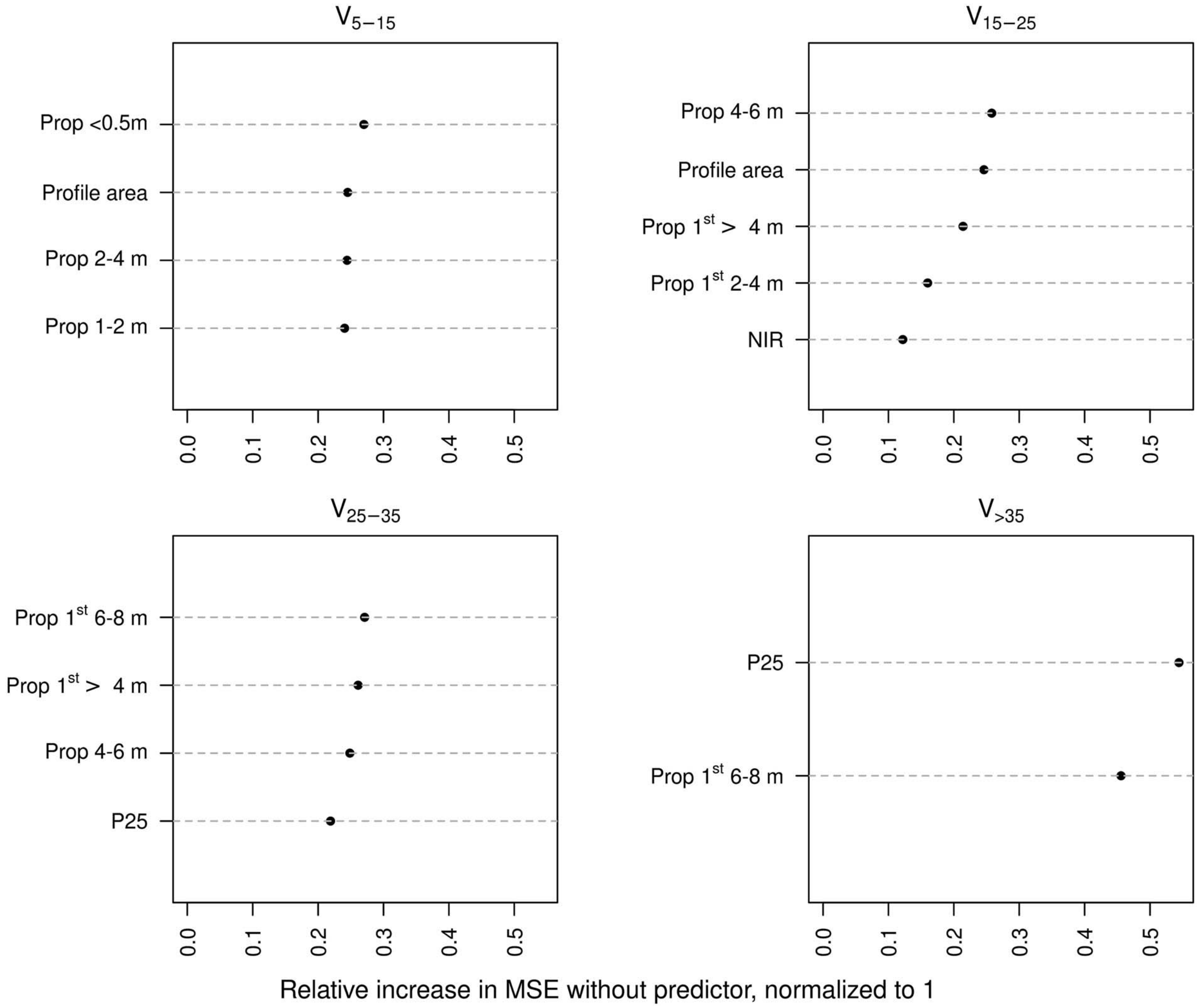

Figure 2. Importance of predictors used in random forest models for estimating volume across diameter classes. Prop <0.5 m: Proportion of returns <2 m, Prop 0.5–1: proportion of returns between 0.5 m and 1 m, Prop 1–2: proportion of returns between 1 m and 2 m, Prop 4–6: proportion of returns between 2 m and 4 m, Prop 1^{st} > 4 m: proportion of first returns >4 m, Prop 1^{st} 2–4: proportion of first returns between 2 m and 4 m. Prop 1^{st} 6–8: proportion of first returns between 6 m and 8 m. P25 25th percentile of return heights, NIR: near infrared band. $V_{i,5-15}$, $V_{i,15-25}$, $V_{i,25-35}$, and $V_{i,>35}$ respectively indicate the models for the first, second, third and fourth diameter classes of the silvicultural model for Spanish juniper.

The parametric regression model for ΔG_i included one ALS metric (the proportion of first returns between >4 m), two climate predictor (the mean solar radiation and the maximum temperature of the warmest month), and two changes in ALS metrics (the change in the proportion of returns between 2 m and 4 m and the change in the proportion of first returns between 2 m and 4 m). All coefficients were statistically significant at the 0.05 level (Table 6) for this model that had a coefficient of determination of 42.53%.

Diameter distribution models

Multivariate linear regression model

The multivariate linear regression model incorporated between two and four predictors for each diameter class, with ALS metrics providing the proportion of first returns across various height intervals consistently present in all diameter classes. Notably, climate metrics and topographic metrics emerged as significant predictors in the models for the second and third diameter classes (Table 7). The error correlations between different diameter classes were all negative except for the correlations between errors for the first diameter class and the second and third classes, yet the magnitudes of all correlations were small, consistently remaining below 0.17. This suggests that total volume predictions tend to experience a compensatory effect, where over- or underestimations for one diameter class are balanced by errors of the opposite sign in other diameter classes. However, due to the small magnitudes of these correlations, the compensation effect is expected to be minimal. For the multivariate linear regression model, *RMSMD* was 0.82, *REI*was 17.39 m^3ha^{-1} and *RREI* was 0.33.

Random forest models

After variable selection was applied, random forest models for diameter classes relied on ALS predictors, except for the second diameter class that included a climate predictor and a Landsat predictor (Fig. 2). However, these non-ALS predictors held relatively low importance, ranking fourth and fifth out of five predictors, respectively. The set of random forest models achieved a *RMSMD* of 0.83, *REI*was 17.89 m^3ha^{-1} and *RREI* was 0.32.

k-nearest neighbor imputation

The best model configurations for imputation models always used the random forest distance, but the best number of neighbors, k,

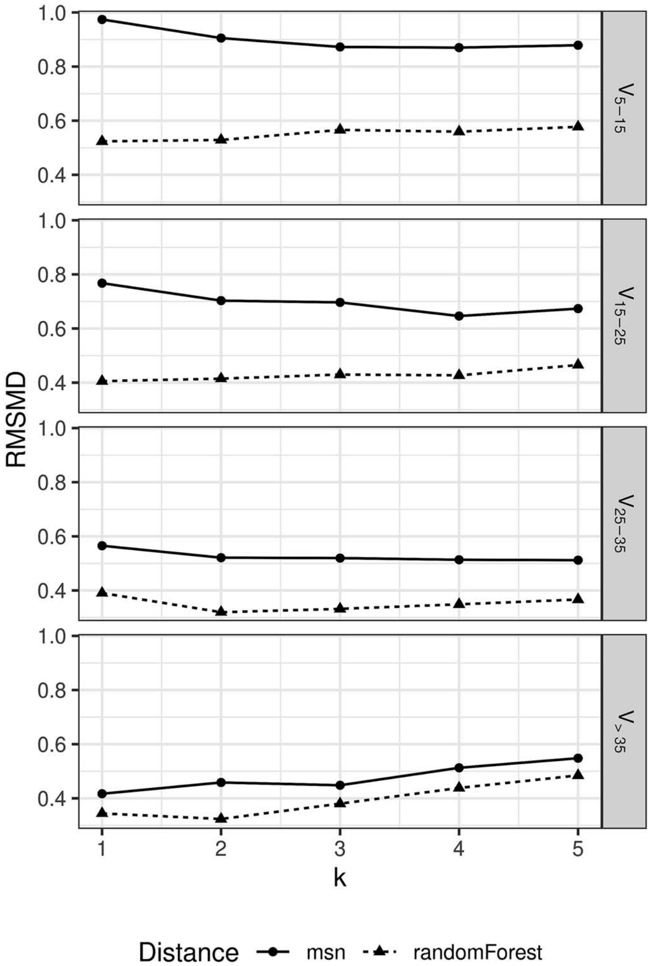

Figure 3. Relative mean square Mahalanobis distance between predicted and observed volume by diameter classes against the number of neighbors used by the imputation model. Circles and solid lines are imputation models using the most similar neighbor distance (msn), triangles connected by dashed lines are imputation models using the random forest distance. $V_{i,5-15}$, $V_{i,15-25}$, $V_{i,25-35}$, and $V_{i,>35}$, respectively, indicate the models for the first, second, third and fourth diameter classes of the silvicultural model for Spanish juniper.

changed across diameter classes, ranging from one neighbor for the first and second diameter classes to two neighbors for the other two diameter classes (Fig. 3). The most important metrics for all responses were ALS metrics, however, a difference with respect to the random forest models was that imputation models included a larger number of spectral, topographic and climate metrics, with at least one metric from these groups present in all models (Fig. 4). The value of *RMSMD* was 0.49, *REI*was 9.29 $m^3ha^{-1}$and *RREI* was 0.19.

Neural network calibration of ITD.

Neural networks used all the available auxiliary information, and no predictor was discarded. This strategy achieved an *RMSMD* was 0.79, *REI* was 16.65 m^3ha^{-1} and *RREI* was 0.31.

Model comparison

The best modeling strategy for diameter distributions was the set of kNN imputation models that had the lowest values for a *RMSMD*, *REI* and *RREI*. In addition, this modeling strategy showed the largest coefficients of determination for all diameter classes (Fig. 5). The next strategy was the calibration of an ITD output with a neural network, that ranked second in terms of *RMSMD*, *REI* and *RREI* and showed coefficients of determination for the third and fourth diameter class >70%. However, differences in *RMSMD*, *REI* and *RREI* between this method, the multivariate linear model and the set of random forest models never reached 8%. Furthermore, the neural network calibration of an ITD output had the lowest R^2 values for first and second diameter classes and failed at predicting volume in the first diameter class showing R^2 values below zero that indicate that predictions are worse than a sample mean. Finally, for all modeling strategies, coefficients of determination for the first diameter class were poorer and consistently increased from the first to the third diameter class (Fig. 5).

Departure from the reference diameter distribution model

Models selected in the previous stage to predict SI classes and diameter distributions were applied to all pixels in the study area. SFM polygons with >40% of their area classified as disturbed, according to the disturbance map from Viana-Soto and Senf (2025), were excluded from the evaluation of departures from the silvicultural model. These disturbed polygons represented only 0.2% of the total area. For the remaining polygons, we calculated the differences between the predicted diameter distributions and the target distributions defined by the silvicultural model. This approach provides a spatially explicit tool that managers can use to identify deficits or surpluses of volume in specific diameter classes and adjust management actions accordingly. This is illustrated in an example for two SFM polygons in Fig. 6. For the first polygon, polygon 654 397, there is a deficit in the first three diameter classes and an excess of volume in the fourth diameter class, which indicates that trees of large dimensions could be extracted to create openings for trees of smaller size. For the second polygon, polygon 692 827, Spanish juniper is colonizing lands bordering agricultural fields. The lack of trees in the third and fourth diameter classes indicate that it is necessary to let the trees of smaller dimension grow until the third and fourth diameter classes are appropriately stocked.

When the differences with respect to the silvicultural model for SFM polygons were represented in a cartesian space I_1x I_2 (Fig. 7) it was possible to observe that the proportion of SFM polygons and SNFI plots assigned to each quadrant were distributed in a similar manner. However, the model tended to underestimate the proportion of areas in the upper left and lower right quadrants (i.e. $I_1 < 0$ and $I_2 > 0$ and $I_1 > 0$ and $I_2 < 0$). For the other two combinations, i.e. deficit or excess of volume in the first and second and third and fourth diameter classes (i.e. $I_1 < 0$ and $I_2 < 0$ or $I_1 > 0$ and $I_2 > 0$), the proportions of area predicted by the model were within confidence intervals, CI, provided by the SNFI plots (Fig. 7). When considering the proportion of the landscape with a deficit of volume of younger trees ($I_1 < 0$) and older trees ($I_2 < 0$), the model predicted that 18.83% and 88.51% of the landscape were under these conditions. The proportions of landscape under these conditions based on the SNFI plots were 30.23%, and 76.74%, respectively.

Discussion

Models for disturbances and site index

The best model for predicting SI classes was primarily driven by changes in ALS metrics and climate metrics, confirming the critical role of multitemporal ALS data for assessing site productivity across extensive landscapes (Guerra-Hernández et al. 2021,

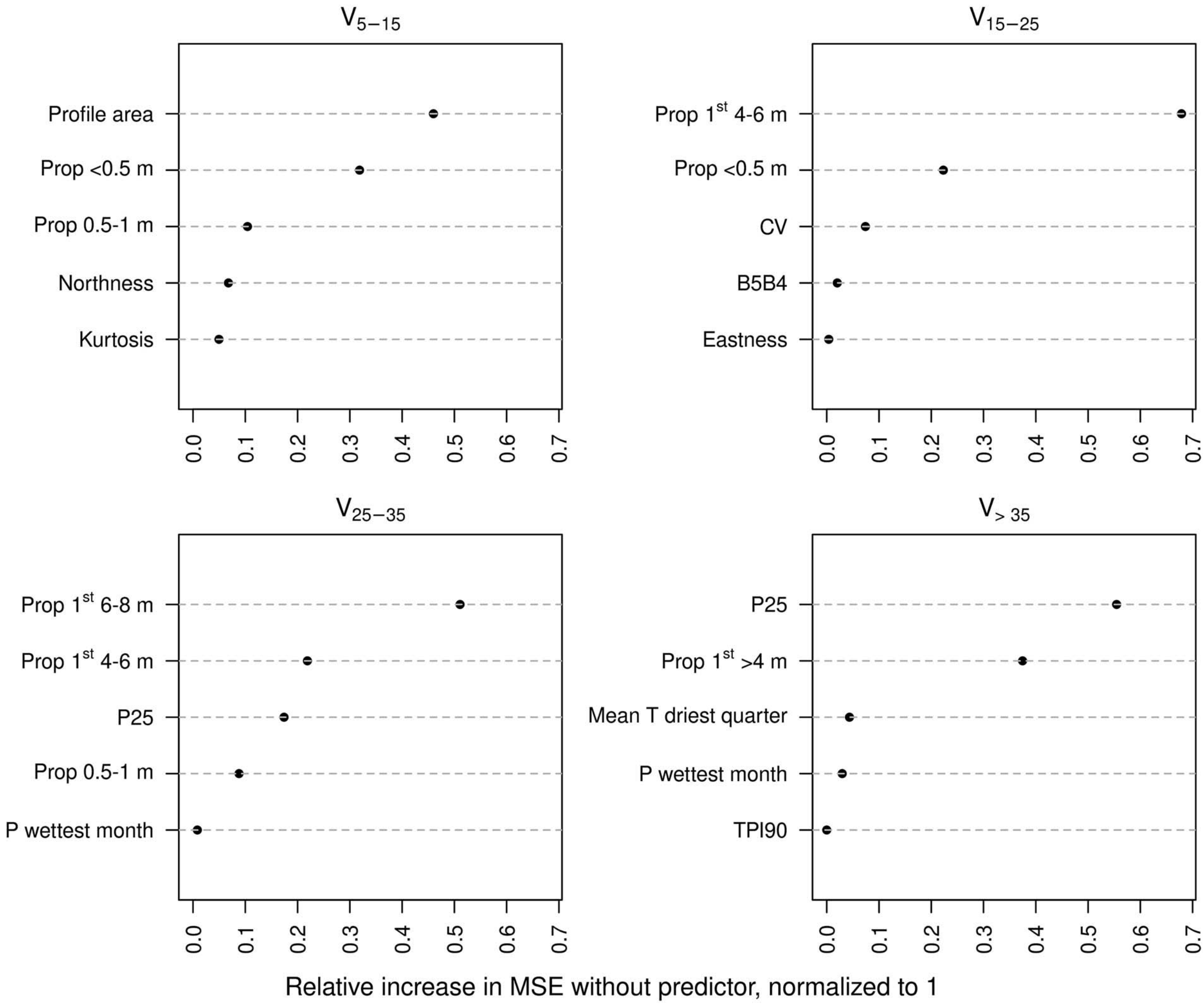

Figure 4. Importance of predictors used in imputation models for estimating volume across diameter classes. Prop <0.5 m: proportion of returns <2 m, Prop 0.5–1: proportion of returns between 0.5 m and 1 m, Prop 1st > 4 m: proportion of first returns >4 m, Prop 1st 4–6: proportion of first returns between 4 m and 6 m, Prop 1st 6–8: proportion of first returns between 6 m and 8 m P25 25th percentile of return heights. CV: coefficient of variation or return height. T: Temperature, P precipitation. NIR: Near infrared band. B5B4: Ratio of Landsat bands 5 and 4. TPI90 topographic position index 90 m. $V_{i,5-15}$, $V_{i,15-25}$, $V_{i,25-35}$, and $V_{i,>35}$, respectively, indicate the models for the first, second, third and fourth diameter classes of the silvicultural model for Spanish juniper.

Penner et al. 2023, Riofrío et al. 2023). A notable innovation of our study is its focus on a species that typically forms uneven-aged structures, where trees of different age cohorts coexist. Unlike traditional SI definitions that rely on dominant height, the SI metric used for Spanish juniper—similar to recently proposed measures of productivity for uneven-aged and mixed stands—is based on ΔG (Dănescu et al. 2017). Numerous studies have established strong correlations between ALS metrics and dominant height (e.g. Næsset 2002, González-Ferreiro et al. 2012), which are fundamental to estimating SI in even-aged forests. Our findings show that multitemporal ALS data can also effectively predict site productivity based on ΔG for uneven aged-forests. This suggests that alternative proxies for site productivity based on average tree slenderness (Alonso Ponce 2008), that combine information about tree diameters and heights, could similarly be estimated using ALS data.

Models for diameter distributions

The most effective approach for predicting diameter distributions was the set of separate kNN imputation models for each diameter class, aligning with previous research that highlights the suitability of imputation-based methods for diameter distribution prediction (Räty et al. 2021). Consistent with prior studies, kNN imputation models using the random forest distance metric outperformed other distance metrics, such as the msn distance, based on weighted linear combinations of auxiliary information (Mauro et al. 2019). This superior performance is likely due to the random forest algorithm's flexibility and its capacity to capture nonlinear relationships between auxiliary variables. Recent advances by Paramasivam et al. (2023) and Sani et al. (2017) propose kNN imputation methods that leverage neural networks. Given the recent success of neural networks in solving complex prediction challenges, these methods appear promising for enhancing diameter distribution predictions from remote sensing data. Moreover, kNN methods based on neural networks offer the advantage of calculating distances between prediction and training instances, which is particularly valuable when kNN imputations are used to reconstruct tree lists (Packalén and Maltamo 2007).

The calibration of an ITD output using neural networks failed at predicting volume for the first diameter class. However, it

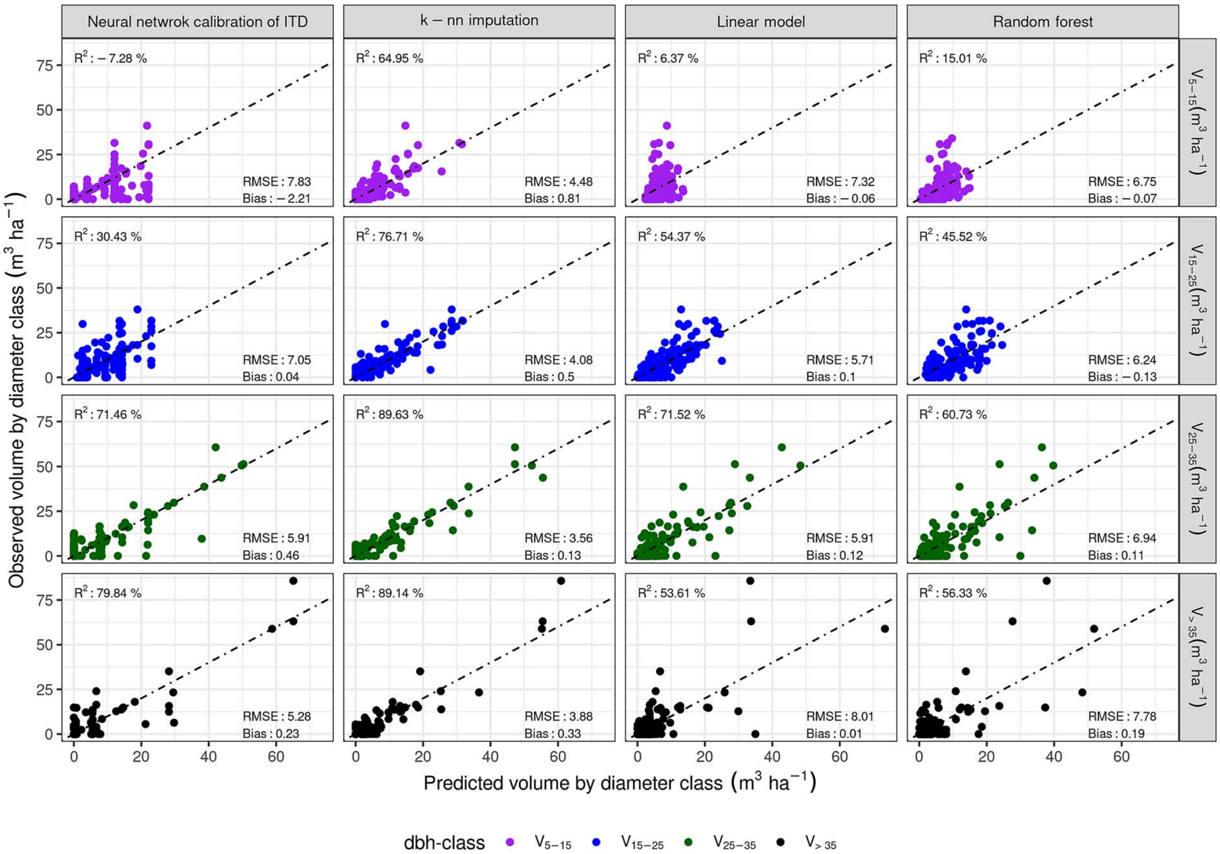

Figure 5. Observed vs predicted volume of Spanish juniper by diameter classes for the eight alternative modeling strategies to predict diameter distributions of Spanish juniper. RMSE: root mean square error, R^2: coefficient of determination.

ranked second best in terms of *RMSMD*, *REI*, and *REI*, and provided coefficients of determination for the third and fourth diameter classes that were similar or better than those obtained with the multivariate linear model and the set of random forest models. This indicates that this is a promising new method for modeling diameter distributions, combining key strengths of both ABA and ITD methodologies. Traditional ITD methods generally involve an unsupervised tree detection step, followed by tree-level allometry calculations to predict diameters and/or volumes based on the heights of the detected treetops. However, Vauhkonen et al. (2014) highlighted several limitations of this approach: uncertainties in treetop height measurements can propagate through allometric calculations, and omissions of understory trees often reduces accuracy, especially for smaller diameter classes. In addition, ITD methods typically underutilize ground data, incorporating it primarily in the validation phase rather than during model fitting. To address this issue, Xu et al. (2014) proposed calibrating ITD outputs using stem-mapped plots and results from a pure ABA approach, creating a supervised process. Our approach for calibrating an ITD output using neural networks follows a similar strategy, incorporating field data in the model fitting phase but simplifies the model output that does not include tree positions and attributes. However, this simpler output satisfies the information needs of multiple management tasks in which exact tree locations are unnecessary and has the advantage of allowing for the use of simpler ground plots without mapped tree coordinates. Another advantage of our approach is that it bypasses the need for individual tree level allometries. Instead, the neural network learns to translate a histogram of tree top heights directly into a diameter distribution using the training data, making the process less complex. Finally, it is important to highlight that further testing with larger datasets and under more diverse conditions is necessary for this approach. Spanish juniper forests typically present sparse structures which are advantageous to control omission errors of ITD algorithms. This seemed to compensate for the low density of the ALS data used in this study that was far from 8 returns per square meter recommended by Sparks et al. (2022) for ITD methods.

Räty (2020) proposed combining top-performing models to reconstruct different segments of diameter distributions. In our study, the set of kNN models implemented this approach and outperformed all other alternatives in terms of *RMSMD* and *REI*. However, differences with other methods were substantial for the lower diameter classes but became consistently smaller when considering larger diameter classes (Fig. 5). The better performance observed for larger diameter classes with all methods seems to be related to two factors. On one hand, previous studies have shown that precision tends to degrade when estimating the lower part of the diameter distribution because smaller trees are frequently hidden by larger trees (e.g. Gobakken and Næsset 2004). On another hand, the variable radius design of the SNFI plots causes that smaller diameter classes are sampled less intensively, which also seem to contribute to their larger errors. Regardless of these factors, the fact that differences in performance between different methods tend to disappear indicates that hybrid strategies including other modeling alternatives for some diameter classes, could potentially improve the performance of the kNN methods.

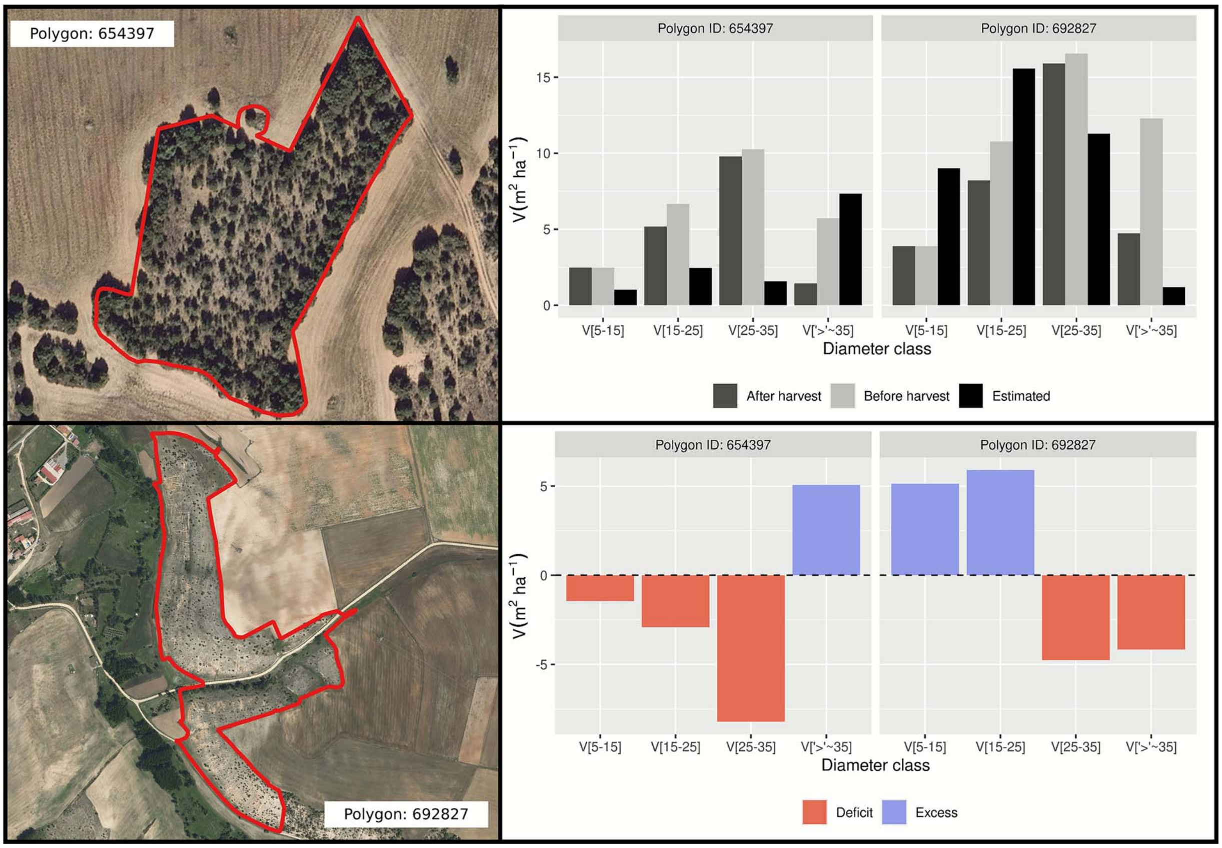

Figure 6. Example of two SFM polygons where the differences with respect to the target state proposed by the silvicultural model from Barrio de Miguel (2006) suggest different management actions. Left, orthophoto of the selected polygon. Upper right, average of before- and after-harvest values for the stand and average predicted diameter distribution. Lower right, average of the pixel level differences with respect to the silvicultural model.

However, using the set of kNN models offers distinct advantages over alternatives mixing modeling techniques. Unlike the other modeling alternatives, that provide volume estimates limited to the four dbh classes, the set of kNN models can reconstruct full tree lists (Packalén and Maltamo 2007). Such reconstructed tree lists could be used to derive additional parameters, to customize the diameter class intervals (Mauro et al. 2019) or to serve as inputs for growth and yield simulators (Fekety et al. 2020), such as FVS (Dixon 2002), medfate (De Cáceres et al. 2022), or SIMANFOR (Bravo et al. 2012) in order to further refine forest management activities. Consequently, the comprehensive capabilities of the kNN model set make it preferable for broader forest management applications.

Departures from the reference state

Visualization of the SNFI plots on the I_1x I_2plane shows that approximately three quarters of Spanish juniper dominated areas in our study region have an excess of trees in the first and second diameter classes relative to the silvicultural targets of Barrio de Miguel (2006) (i.e. $I_1 > 0$), while a slightly larger proportion display a deficit of volume in the third and fourth diameter classes (i.e. $I_2 < 0$). At the landscape scale, 53.49% of the territory is characterized by an overabundance of young trees and shortage of mature trees. These findings, based solely on SNFI sample data, indicate that management actions targeted at increasing trees of larger dimensions are needed in these forests. The models and indices developed in this study provide a practical, spatially explicit framework for addressing these challenges by enabling the rapid identification of SFM polygons where silvicultural interventions—particularly those aimed at promoting the recruitment and growth of intermediate and large-diameter trees—are most urgently needed. Although these model-based predictions inherently involve uncertainty, they offer a scalable approach for integrating field data, remote sensing, and ecological knowledge into adaptive forest management (Fig. 6).

Auxiliary information and Spanish National Forest Inventory

The availability of extensive nationwide auxiliary data, including two ALS acquisitions, Landsat time series, and climate variables, was one of the cornerstones of this study. However, the SNFI's spatially balanced sample of permanent plots was equally important. Without these plots, the value of the auxiliary data would be substantially reduced. As with other national forest inventories, the SNFI, with its extensive coverage, balanced distribution, and plot permanence, ensures representative samples in large areas and enables effective modeling and monitoring of forest attributes from remote sensing data. While local inventory data can support the mapping of structural attributes such as total volume, biomass, and dominant height (Mauro et al. 2017, Domingo et al. 2018, Fekety et al. 2020, Frank et al. 2020) as well as more disaggregated attributes like diameter distributions (Mauro et al. 2019, Gorgoso-Varela et al. 2021), these measurements are often derived from temporary plots, limiting their utility for tracking site productivity over time in areas without age records or with uneven-aged structures where age cannot be clearly stablished. Long-term research data could be used for mapping productivity

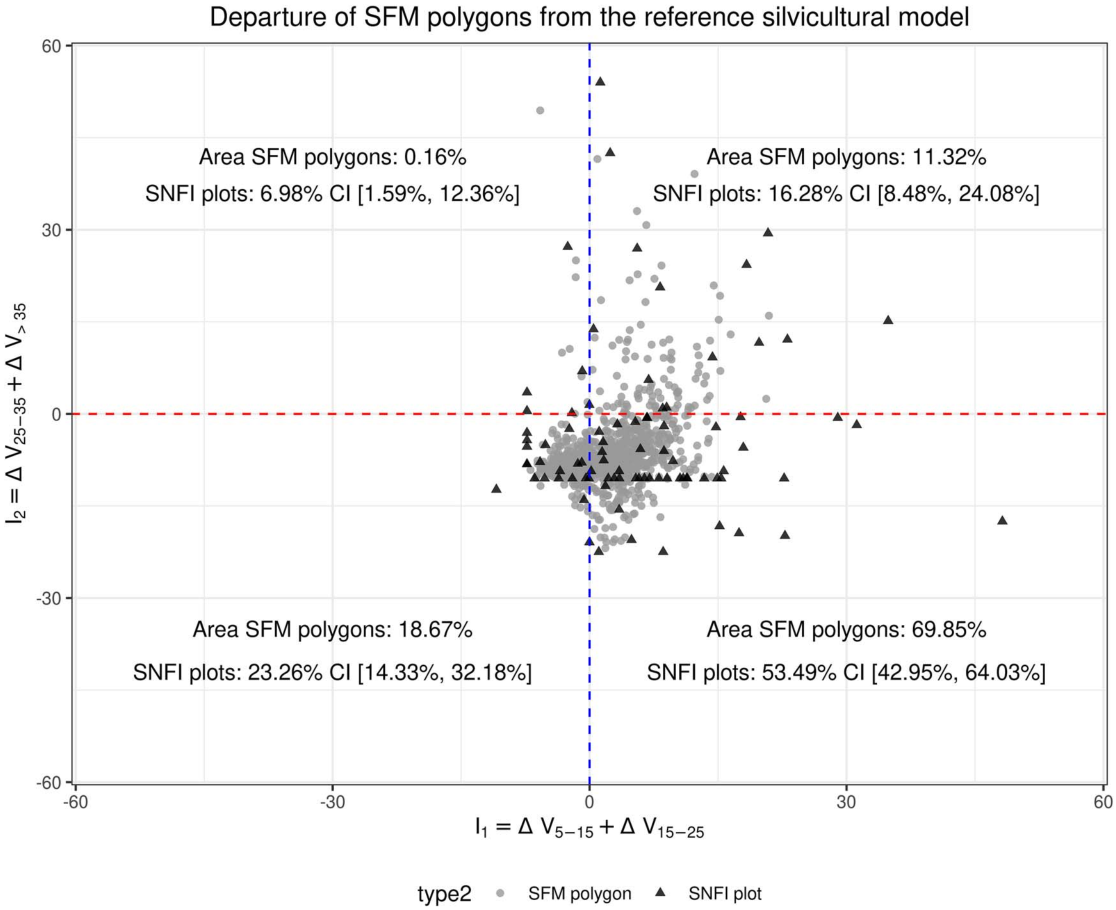

Figure 7. Representation of plots of the Spanish National Forest Inventory (black) and polygons of the SFM, dominated by Spanish juniper (gray). The x-axis is the sum of differences with respect to the silvicultural model from Barrio de Miguel (2006) for the first and second diameter class. The y-axis is the sum of differences with respect to the silvicultural model from Barrio de Miguel (2006) for the third and fourth diameter class. For polygons, differences are based on predicted volumes by diameter class and SI class. For plots, differences are determined from observations of volume by diameter class and SI class.

(Castaño-Santamaría et al. 2023), yet such datasets are rare and generally lack the spatial coverage and balance of NFI samples. This highlights the unique value of long-term measurement programs implemented by NFIs in studies focused on mapping site productivity where consistent, wide-ranging, and repeatable data are essential.

Conclusions

This study underscores the effectiveness of combining ground observations from the SNFI with extensive auxiliary data sources to enhance forest management actions in Spanish juniper forests. Our results show that by combining multitemporal ALS data, Landsat imagery, and metrics on topography, geography, and climate with repeated SNFI measurements it is possible to accurately estimate SI classes and diameter distributions in Spanish juniper forest. A parametric linear regression model was the most accurate alternative for estimating SI classes. For diameter distributions, the best estimates were obtained using separate kNN imputation models for each diameter class. These models provide estimates that support the application of a reference silvicultural model for Spanish juniper, offering forest managers valuable insights for informed decision-making in the conservation and sustainable management of forests of this species.

Acknowledgements

The authors wish to thank editor and three anonymous reviewers for their helpful comments. We also wish to thank the Spanish National Forest Inventory and the National Plan of Orthophotography for their efforts collecting and maintaining the open datasets used in this study.

Author contributions

Francisco Mauro (Conceptualization, Data curation, Formal analysis, Funding acquisition, Resources, Software, Writing—original draft), Cristina Gómez (Investigation, Writing—original draft, Writing—review & editing), Manuel Gomez-Roux (Writing—review & editing), Lorena Caiza-Morales (Software, Writing—review & editing), Francisco Rodríguez-Puerta (Writing—review & editing), Beatriz Águeda (Writing—review & editing), Steven Filippelli (Investigation, Software, Writing—review & editing), Patrick Fekety (Methodology, Writing—review & editing), Andrew Hudak (Methodology, Writing—review & editing), Hailemariam Temesgen (Methodology, Writing—review & editing), Jose Antonio Manzanera (Writing—review & editing), and David Candel-Pérez (Funding acquisition, Investigation, Project administration, Writing—review & editing)

Supplementary data

Supplementary data are available at *Forestry* online.

Conflict of interest: The authors declare that there is no conflict of interest.

Funding

This work was supported by the Ministry of Science and Innovation and the Spanish Research Agency through the project TLM PROJECT [PID2022-140104OA-I00 MICIU/AEI/ 10.13039/501100011033 to F.M.] and by the Maria Zambrano's Excellence Program from the Ministry of Science and Innovation—University of Valladolid-European Union-NextGeneration EU [E-42-2022-0000233, to F.M.]. J. Bernardo González-Mesquida has been funded under the UVa 2022 predoctoral contract call, co-financed by Banco Santander. Manuel Gómez Roux was funded by a predoctoral contract (PREP2022-000579), funded by MCIN/AEI/10.13039/501100011033.

Data availability

Airborne laser scanning data used in this study is publicly available at https://pnoa.ign.es/. Landsat imagery can be accessed using the Google Earth Engine platform. Climate data was obtained from WorldClim2 https://www.worldclim.org/data/index.html. The Spanish Forest Map is publicly available at https://www.miteco.gob.es/es/cartografia-y-sig/ide/descargas/biodiversidad/mfe.html. Ground observations from the Spanish National Forest Inventory are publicly available at https://www.miteco.gob.es/es/biodiversidad/temas/inventarios-nacionales/inventario-forestal-nacional.html, however, plot locations are considered confidential.

References

Acuña-Míguez B, Valladares F, Martín-Forés I. Both mature patches and expanding areas of Juniperus thurifera forests are vulnerable to climate change but for different reasons. *Forests* 2020;**11**:1–13. https://doi.org/10.3390/f11090960.

Allaire JJ, Chollet F. *Keras: R Interface to "Keras"*, 2024. R package version 2.15.0. https://CRAN.R-project.org/package=keras.

Allaire JJ, Tang Y. *Tensorflow: R Interface to "TensorFlow"*, 2024. R package version 2.16.0.90. https://github.com/rstudio/tensorflow.

Alonso, Ponce R. Autoecología paramétrica de Juniperus thurifera L. en Castilla y León. Doctoral thesis. Technical University of Madrid. 2008. https://doi.org/10.20868/UPM.thesis.965.

Arias-Rodil M, Diéguez-Aranda U, Álvarez-González JG. *et al.* Modeling diameter distributions in radiata pine plantations in Spain with existing countrywide LiDAR data. *Ann Forest Sci* 2018;**75**:36. https://doi.org/10.1007/s13595-018-0712-z.

Barrio, de Miguel JM. Propuesta de Modelo de selvicultura de masas irregulares de sabina albar (Juniperus Thurifera L.). In-proceedings: *III Coloquio Internacional sobre sabinares y enebrales (género Juniperus): Ecología y gestión forestal sostenible*. Valladolid (Spain), Portal Forestal Castilla y Leon. 2006. https://pfcyl.es/sites/default/files/eventos/adjuntos/73.pdf. (11 November 2025, date last accessed).

Bollandsås OM, Næsset E. Estimating percentile-based diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data. *Scand J For Res* 2007;**22**:33–47. https://doi.org/10.1080/02827580601138264.

Bravo F, Rodriguez F, Ordóñez C. A web-based application to simulate alternatives for sustainable forest management: SIMANFOR. *Forest Syst* 2012;**21**:4–8. https://doi.org/10.5424/fs/2112211-01953.

Breidenbach J, Gläser C, Schmidt M. Estimation of diameter distributions by means of airborne laser scanner data. *Can J For Res* 2008;**38**:1611–20. https://doi.org/10.1139/x07-237.

Castaño-Santamaría J, López-Sánchez CA, Ramón Obeso J. *et al.* Development of a site form equation for predicting and mapping site quality. A case study of unmanaged beech forests in the Cantabrian range (NW Spain). *For Ecol Manage* 2023;**529**:120711. https://doi.org/10.1016/j.foreco.2022.120711.

Dixon Gary E. Essential FVS: A user's guide to the Forest Vegetation Simulator. *Internal Rep.* Fort Collins, CO: U. S. Department of Agriculture, Forest Service, Forest Management Service Center. 238p. 2002. (Revised: September 23, 2025) https://www.fs.usda.gov/sites/default/files/forest-management/essential-fvs.pdf.

Crookston NL, Finley AO. yaImpute: an R package for kNN imputation. *J Stat Softw* 2007;**23**:1–16. https://doi.org/10.18637/jss.v023.i10.

Dănescu A, Albrecht AT, Bauhus J. *et al.* Geocentric alternatives to site index for modeling tree increment in uneven-aged mixed stands. *For Ecol Manage* 2017;**392**:1–12. https://doi.org/10.1016/j.foreco.2017.02.045.

De Cáceres M, Molowny-Horas R, Cabon A. *et al.* MEDFATE 2.8.1: a trait-enabled model to simulate Mediterranean forest function and dynamics at regional scales. *Geoscientific Model Development Discussions* 2022;**2022**:1–52. https://doi.org/10.5194/gmd-2022-243.

Domingo D, Lamelas MT, Montealegre AL. *et al.* Estimation of Total biomass in Aleppo pine Forest stands applying parametric and nonparametric methods to low-density airborne laser scanning data. *Forests* 2018;**9**:1–17. https://doi.org/10.3390/f9040158.

Fekety PA, Crookston NL, Hudak AT. *et al.* Hundred year projected carbon loads and species compositions for four National Forests in the northwestern USA. *Carbon Balance Manag* 2020;**15**:5. https://doi.org/10.1186/s13021-020-00140-9.

Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol* 2017;**37**:4302–15. https://doi.org/10.1002/joc.5086.

Finley AO, Banerjee S, Weiskittel AR. *et al.* Dynamic spatial regression models for space-varying forest stand tables. *Environmetrics* 2014;**25**:596–609. https://doi.org/10.1002/env.2322.

Forest Map of Spain. *Inventario español del Patrimonio Natural y la Biodiversidad. Ministerio para la Transición Ecológica y el Reto Demográfico. [WWW Document]*, 2025 URL https://iepnb.es/en/thematic-areas/forestry/forest-map-of-spain (accessed 7/7/25).

Frank B, Mauro F, Temesgen H. Model-based estimation of forest inventory attributes using Lidar: a comparison of the area-based and semi-individual tree crown approaches. *Remote Sens* 2020;**12**:1–24. https://doi.org/10.3390/rs12162525.

von Gadow K, Puumalainen J. Scenario planning for sustainable forest management. In: von Gadow K, Pukkala T, Tomé M, (eds.). *Sustainable Forest Management*. Netherlands, Dordrecht: Springer, 2000, 319–56.

Gatziolis D. LIDAR-derived site index in the US pacific northwest–challenges and opportunities. *Proc SilviLaser* 2007;136–142.

Genuer R, Poggi J-M, Tuleau-Malot C. *VSURF: Variable Selection Using Random Forests*, 2022. R package version 1.2.0. https://CRAN.R-project.org/package=VSURF.

Gobakken T, Næsset E. Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. *Scand J For Res* 2004;**19**:529–42. https://doi.org/10.1080/02827580410019454.

González-Ferreiro E, Diéguez-Aranda U, Miranda D. Estimation of stand variables in Pinus radiata D. Don plantations using different LiDAR pulse densities. *Forestry* 2012;**85**:281–92. https://doi.org/10.1093/forestry/cps002.

Gorelick N, Hancher M, Dixon M. *et al.* Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sens Environ* 2017;**202**:18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Gorgoso-Varela JJ, Ponce RA, Rodríguez-Puerta F. Modeling diameter distributions with six probability density functions in Pinus halepensis mill. Plantations using low-density airborne laser scanning data in Aragón (Northeast Spain). *Remote Sens* 2021;**13**:1–17. https://doi.org/10.3390/rs13122307.

Guerra-Hernández J, Arellano-Pérez S, González-Ferreiro E. *et al.* Developing a site index model for P. Pinaster stands in NW Spain by combining bi-temporal ALS data and environmental data. *For Ecol Manage* 2021;**481**:118690. https://doi.org/10.1016/j.foreco.2020.118690.

Gustafson EJ, Lietz SM, Wright JL. Predicting the spatial distribution of aspen growth potential in the upper Great Lakes region. *For Sci* 2003;**49**:499–508. https://doi.org/10.1093/forestscience/49.4.499.

Hudak AT, Fekety PA, Kane VR. *et al.* A carbon monitoring system for mapping regional, annual aboveground biomass across the northwestern USA. *Environ Res Lett* 2020;**15**:095003. https://doi.org/10.1088/1748-9326/ab93f9.

Hudak AT, Haren AT, Crookston NL. *et al.* Imputing Forest structure attributes from stand inventory and remotely sensed data in western Oregon, USA. *For Sci* 2014;**60**:253–69. https://doi.org/10.5849/forsci.12-101.

Kennedy RE, Yang Z, Cohen WB. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—temporal segmentation algorithms. *Remote Sens Environ* 2010;**114**:2897–910. https://doi.org/10.1016/j.rse.2010.07.008.

Kennedy RE, Yang Z, Gorelick N. *et al.* Implementation of the LandTrendr algorithm on Google earth engine. *Remote Sens* 2018;**10**:1–10. https://doi.org/10.3390/rs10050691.

Khosravipour A, Skidmore AK, Isenburg M. *et al.* Generating pit-free canopy height models from airborne Lidar. *Photogramm Eng Remote Sens* 2014;**80**:863–72. doi:10.14358/PERS.80.9.863.

Liaw A, Wiener M. Classification and regression by randomForest. *R News* 2002;**2**:18–22.

Lumley T. Leaps: regression subset selection. 2020. R package version 3.2. https://CRAN.R-project.org/package=leaps.

Magnussen S, Eggermont P, LaRiccia VN. Recovering tree heights from airborne laser scanner data. *For Sci* 1999;**45**:407–22. https://doi.org/10.1093/forestscience/45.3.407.

Maltamo M, Suvanto A, Packalén P. Comparison of basal area and stem frequency diameter distribution modelling using airborne laser scanner data and calibration estimation. *For Ecol Manage* 2007;**247**:26–34. https://doi.org/10.1016/j.foreco.2007.04.031.

Martínez-Rodrigo R, Porto-Rodríguez JC, Sabín P. *et al.* Análisis de la dinámica forestal de la provincia de Soria utilizando el IFN3 y el IFN4. In *Proceedings: IX Congreso Forestal Español*, 2025, Gijón. https://9cfe.congresoforestal.es/wp-content/uploads/2025/9cfe-1905.pdf. (11 November 2025, date last accessed).

Mauro F, Frank B, Monleon VJ. *et al.* Prediction of diameter distributions and tree-lists in southwestern Oregon using LiDAR and stand-level auxiliary information. *Can J For Res* 2019;**49**:775–87. https://doi.org/10.1139/cjfr-2018-0332.

Mauro F, Monleon VJ, Temesgen H. *et al.* Analysis of spatial correlation in predictive models of forest variables that use LiDAR auxiliary information. *Can J For Res* 2017;**47**:788–99. https://doi.org/10.1139/cjfr-2016-0296.

McKenney DW, Pedlar JH. Spatial models of site index based on climate and soil properties for two boreal tree species in Ontario, Canada. *For Ecol Manage* 2003;**175**:497–507. https://doi.org/10.1016/S0378-1127(02)00186-X.

Mehtätalo L, Nyblom J. A model-based approach for airborne laser scanning inventory: application for square grid spatial pattern. *For Sci* 2012;**58**:106–18. https://doi.org/10.5849/forsci.10-023.

Næsset E. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens Environ* 2002;**80**:88–99. https://doi.org/10.1016/S0034-4257(01)00290-5.

Olano JM, Rozas V, Bartolomé D. *et al.* Effects of changes in traditional management on height and radial growth patterns in a Juniperus thurifera L. woodland. *For Ecol Manage* 2008;**255**:506–12. https://doi.org/10.1016/j.foreco.2007.09.015.

Packalén P, Maltamo M. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sens Environ* 2007;**109**:328–41. https://doi.org/10.1016/j.rse.2007.01.005.

Packalén P, Maltamo M. Estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. *Can J For Res* 2008;**38**:1750–60. https://doi.org/10.1139/X08-037.

Paramasivam K, Sindha MM, Balakrishnan SB. KNN-based machine learning classifier used on deep learned spatial motion features for human action recognition. *Entropy* 2023;**25**:1–15. https://doi.org/10.3390/e25060844.

Penner M, Woods M, Bilyk A. Assessing site productivity via remote sensing—age-independent site index estimation in even-aged forests. *Forests* 2023;**14**:1–14. https://doi.org/10.3390/f14081541.

Popescu SC, Wynne RH. Seeing the trees in the Forest. *Photogramm Eng Remote Sens* 2004;**70**:589–604. https://doi.org/10.14358/PERS.70.5.589.

Pukkala T, Aquilué N, Just A. *et al.* Developing kNN forest data imputation for Catalonia. *Journal of Forestry Research* 2024;**35**:80. https://doi.org/10.1007/s11676-024-01735-5.

Räty J. Prediction of diameter distributions in boreal forests using remotely sensed data. *Dissertationes Forestales*; University of Eastern Finland. 2020;1–47. https://doi.org/10.14214/df.294.

Räty J, Astrup R, Breidenbach J. Prediction and model-assisted estimation of diameter distributions using Norwegian national forest inventory and airborne laser scanning data. *Can J For Res* 2021;**51**:1521–33. https://doi.org/10.1139/cjfr-2020-0440.

Reynolds MR, Burk TE, Huang W-C. Goodness-of-fit tests and model selection procedures for diameter distribution models. *For Sci* 1988;**34**:373–99. https://doi.org/10.1093/forestscience/34.2.373.

Riofrío J, White JC, Tompalski P. *et al.* Modelling height growth of temperate mixedwood forests using an age-independent approach and multi-temporal airborne laser scanning data. *For Ecol Manage* 2023;**543**:121137. https://doi.org/10.1016/j.foreco.2023.121137.

Roussel J-R, Auty D. lidR: airborne LiDAR data manipulation and visualization for forestry applications. 2017.

Roussel J-R, Coops NC, Tompalski P. *et al.* lidR: an R package for analysis of airborne laser scanning (ALS) data. *Remote Sens Environ* 2020;**251**:112061. https://doi.org/10.1016/j.rse.2020.112061.

Sani S, Wiratunga N, Massie S. *Learning Deep Features for kNN-Based Human Activity Recognition*, In *Proceedings: 25th International conference on case-based reasoning (ICCBR 2017)*. 2017;95–103. http://ceur-ws.org/Vol-2028/paper9.pdf.

Sparks AM, Corrao MV, Smith AMS. Cross-comparison of individual tree detection methods using low and high pulse density airborne laser scanning data. *Remote Sens* 2022;**14**:1–19. https://doi.org/10.3390/rs14143480.

Spriggs RA, Vanderwel MC, Jones TA. *et al.* A simple area-based model for predicting airborne LiDAR first returns from stem diameter distributions: an example study in an uneven-aged, mixed temperate forest. *Can J For Res* 2015;**45**:1338–50. https://doi.org/10.1139/cjfr-2015-0018.

Vauhkonen J, Packalen P, Malinen J. *et al.* Airborne laser scanning-based decision support for wood procurement planning. *Scand J For Res* 2014;**29**:132–43. https://doi.org/10.1080/02827581.2013.813063.

Viana-Soto A, Senf C. The European Forest disturbance atlas: a forest disturbance monitoring system using the Landsat archive. *Earth Syst Sci Data* 2025;**17**:2373–404. https://doi.org/10.5194/essd-17-2373-2025.

Xu Q, Hou Z, Maltamo M. *et al.* Calibration of area based diameter distribution with individual tree based diameter estimates using airborne laser scanning. *ISPRS J Photogramm Remote Sens* 2014;**93**:65–75. https://doi.org/10.1016/j.isprsjprs.2014.03.005.

Forestry: An International Journal of Forest Research, 2026, **99**, 1–17
https://doi.org/10.1093/forestry/cpaf080
Original Article