Mapping stem volume in fast-growing eucalypt plantations: integrating spectral, textural, and temporal remote sensing information with forest inventories and spatial models

Lívia Lanzi Aló, Guerric Le Maire, Cláudio Roberto Thiersch, Thiago Santos Mota, Rafael Costa Pinheiro, José Raimundo de Souza Passos

HAL Id: hal-04746465

https://hal.inrae.fr/hal-04746465v1

Submitted on 21 Oct 2024

Annals of
Forest Science

Mapping stem volume in fast-growing eucalypt plantations: integrating spectral, textural, and temporal remote sensing information with forest inventories and spatial models

Lívia Lanzi Aló[1], Guerric le Maire[2,3*], Cláudio Roberto Thiersch[4], Thiago Santos Mota[5], Rafael Costa Pinheiro[1] and José Raimundo de Souza Passos[6]

Abstract

Key message Accurate volume estimation in *Eucalyptus* plantation stands was achieved by a linear model using SPOT and Landsat multispectral imagery, specifically texture indices and pixel-scale NDVI time integrals, which reflect the local plantation growth history. Spatial modelling techniques such as Kriging with External Drift and Generalized Additive Model slightly improved predictions by accounting for spatial correlation of volume between sample points.

Context Forest inventories are widely used to estimate stand production. To capture the inherent spatial variability within stands, spatial modelling techniques such as Kriging with External Drift (KED) and the generalized additive model (GAM) have emerged. These models incorporate information on spatial correlation and auxiliary variables that can be obtained from satellite imagery.

Aims Our study explored the use of reflectance data from SPOT and Landsat multispectral imagery. We focused on texture indices and temporal integration of vegetation indices as auxiliary variables in KED and GAM to predict stem volume of fast-growing *Eucalyptus sp.* plantations in Brazil.

Methods The components extracted from the high-resolution SPOT-6 image included spectral band values, band ratio metrics, key vegetation indices (NDVI, SAVI, and ARVI), texture measurements, and indices derived from texture analysis. Additionally, we included the accumulated NDVI time series acquired from Landsat 5, 7, and 8 satellites between the planting date and the forest inventory measurement date.

Results The best linear model of stem volume using remotely sensed predictors gave an R-squared value of 0.95 and a Root Mean Square Error (RMSE) of 12.44 m^3/ha. The R-squared increased to 0.96 and the RMSE decreased to 10.6 m^3/ha when the same predictors were included as auxiliary variables in the KED and GAM spatial models.

Conclusion The linear model using remotely sensed predictors contributed most to volume prediction, but the addition of spatial coordinates in the KED and GAM spatial models improved local volume predictions.

*Correspondence:
Guerric le Maire
guerric.le_maire@cirad.fr
Full list of author information is available at the end of the article

1 Introduction

Short rotation forest plantations are expanding to meet the demand for paper, cellulose, industrial wood (such as particle board) and coal for industry (e.g., iron and steel). In Brazil, there were 9.55 million hectares of tree plantations in 2020, with the majority being eucalypt plantations, covering 7.47 million hectares (IBA 2021). The development of the forestry sector necessitates sustainable management and information about existing forest resources. Stakeholders require monitoring of forest resources for wood production assessment, harvest prevision, sanitary control, carbon accounting, and evaluation of biotic and abiotic damages.

Forest inventories and field surveys remain the primary methods for monitoring forest plantations. Forest inventories involve field measurements of tree diameters and heights of small plots following a plantation sampling design. These measurements are then used in allometric relationships to predict the volume of individual trees, which is subsequently summed for all trees in the plot and divided by the plot area to estimate the volume per unit area. Stand-level values are typically obtained by averaging the plot-level values within the stand. The method assumed that all plots are independent and equally weighted, and therefore do not consider the spatial dependence of field data (Viana et al. 2012; Castillo-Santiago et al. 2013). Additionally, despite the assumption of homogeneity of planted forests, Castillo-Santiago et al. (2013) showed that it was necessary to consider the natural variations that occur in the structure of vegetation. Spatial variability within and between plantation stands, particularly in fast-growing species such as *Eucalyptus* in Brazil, can be influenced by various factors, including different soil conditions, topography, planting dates, genetic materials, management, and past events such as diseases and storms (Marsden et al. 2010a, b; Le Maire et al. 2011a, b, c; Zhou et al. 2013; le Maire et al. 2019). Thus, although the assumption of independence between inventory variables within and between stands is commonly used, knowledge of the spatial variations within a forest (both within and between stands) can improve the prediction of the variable of interest (Zhou et al. 2013).

Various statistical techniques are used to model spatial variations found in the sampled data (Propastin 2012). Geostatistical techniques facilitate the characterization of the spatial variability present in a data structure, including the spatial correlation between sampling units. This information can then be used to improve the prediction of variables of interest at unsampled locations, thereby enhancing the accuracy of stand-scale or regional-scale averages. Numerous studies have used geostatistical interpolation to account for spatial dependence between forest inventory plots. This method falls within the model-based inferential framework described by Ståhl et al. (2016). Geostatistical interpolation was used to predict variables such as volume, biomass, and diameter increments, as well as to examine the spatial continuity of the parameter of interest (Tesfamichael et al. 2009; Viana et al. 2012; Galeana-Pizaña et al. 2014; Scolforo et al. 2016; Silveira et al. 2019).

Correlations may exist between the forest inventory variables, such as tree height and volume, and information extracted from satellite images. This allows for a direct, spatially explicit prediction of these variables at the resolution of the satellite image (Lu 2006; Avitabile et al. 2012; Castillo-Santiago et al. 2013). Satellite images have been widely used to map structural forest parameters, such as above-ground volume and biomass (Zhang et al. 2014; Zhao et al. 2016; Lochhead et al. 2018; Souza et al. 2019). More specifically, some studies used image texture (Avitabile et al. 2012; Rodriguez-Galiano et al. 2012; Dube and Mutanga 2015), image time series (Le Maire et al. 2011a, b, c; Boisvenue et al. 2016), or other types of sensors such as radar (Baghdadi et al. 2015; Zhao et al. 2016).

Geostatistical interpolation techniques, such as Kriging with External Drift (KED, also called Universal Kriging), use ancillary data, including variables derived from satellite imagery, to explain some of the spatial variability observed in the data and complement the interpolation process (Goovaerts 1997). KED combines the strengths of both kriging and regression methods. Kriging models the spatial patterns of the variable of interest, while multivariate regression models the influence of external predictors. This combination allows for more accurate prediction of the variable of interest at unsampled locations, especially when the multivariate regression model provides a precise prediction of the variable of interest. KED is preferred over classical co-kriging when a strong linear relationship exists between the primary and secondary variables (Wackernagel 2003).

The generalized additive model (GAM) is an extension of the generalized linear model. It adds a smoothing function is added to the explanatory variable to make the relationship between the response variable and the auxiliary variables more flexible (Hastie and Tibshirani 1986). The shape of this relationship does not need to be known

Aló *et al. Annals of Forest Science*   (2024) 81:43

Page 3 of 16

beforehand, but it can be estimated through smoothing curves from the dataset. The smoothing functions take into account the geographic coordinates. Compared to KED it handles any non-linear relationships between the auxiliary variables and the spatial values being predicted. GAMs are semi-parametric and multi-dimensional and generally use penalized splines for the smoothing function. When developing ecological models, some studies use GAM to calculate wood stock, forest characteristics, biomass, and other related factors (Ahrens et al. 2020; Alberdi et al. 2020; Fassnacht et al. 2021).

Geostatistical models that incorporate spatial auxiliary variables require known variables at each location of model application. Satellite multispectral images can provide such information, for example, through spectral vegetation indices, which are mathematical formulations that combine reflectance of several spectral bands. Vegetation indices are typically highly correlated with green vegetation within the pixels, while minimizing the contributions of the soil and sun angle (Lu 2006; Cutler et al. 2012). However, vegetation indices often have limited direct spatial correlation with forest inventory variables, indicating the need to consider alternative co-variables to improve the performance of geostatistical models. This is especially true for short rotation plantation forests such as *Eucalyptus* plantations in Brazil. These forests are temporally dynamic, with high growth rates averaging 40 m^3 ha^{-1} $year^{-1}$. They reach canopy height of over 20 m when they are harvested at approximately 7 years old. The aim of this study is to investigate the potential use of image texture computed on very high-resolution satellite images, and historical information about the forest plantation since planting date obtained from satellite image time series, as co-variables in geostatistical models.

KED and GAM geostatistical models were calibrated to predict stem wood volume (m^3 ha^{-1}) in *Eucalyptus* plantations. We added several co-variables computed from satellite images, including band reflectances, simple band ratios, vegetation indices, texture indices, and more complex variables computed from the integration of time series of vegetation indices. The objective was to provide more accurate predictions of wood volume within and between stands.

The specific objectives were (a) to assess the direct correlation between variables derived from satellite imagery and stem volume; (b) to evaluate their effectiveness as auxiliary variables in fitting KED and the GAM geostatistical models for volume prediction; and (c) to compare the volumes predicted by the linear, KED, and GAM models with the volumes observed on randomly selected forest inventory plots not used in model calibration.

2 Material and methods

2.1 Study site and forest inventory data

The study was conducted in *Eucalyptus* sp. stands located in the southern region of Mato Grosso do Sul state, near the city of Ribas do Rio Padro. The area has an average altitude of 374 m and a tropical climate (*Aw* type according to the Köppen classification) characterized by a dry season in the winter with an average precipitation of 1241 mm and an average temperature of 24.2 °C.

The forest inventory data was collected in August 2018 from 211 circular plots of 500 m^2, systematically distributed on a pre-defined grid across the landscape, with one plot every ~722 m in the study area. All plots were geo-referenced using GPS measurements from the central point of each plot. In the study area, all stands were less than 4 years old in August 2018 (Table 1). The inventory consisted of measuring the circumference at breast height (CBH—1.3 m above the ground level) for all trees in the circular inventory plot and the tree height of a central sub-sample of 25% of the trees in the plot. Dominant height was calculated as the mean of the five trees with the largest CBH. Unmeasured tree heights were predicted using the Curtis hypsometric model. Stem volumes of the trees were then predicted by age class using a company-calibrated Schöepfer Fifth-Degree Polynomial volume projection equation specific to the genetic material and calibrated in the same region based on destructive measurements of trees. These equations are generally very precise in these clonal plantations (Scolforo et al. 2019). The plot volume was calculated by adding the volumes of each tree in the plot. The age information was incorporated as a covariable in the linear models (see section "2.4").

Table 1 Summary of planting characteristics and descriptive statistics of average volume data per plot in the study site

	Age (years)	Area (ha)	Nb of clones	Number of plots	Stem volume (m^3 ha^{-1})				
					Minimum	Maximum	Average	SD	CV
By age class	1.5–2.5	7565.88	8	150	1.16	76.50	25.73	15.16	0.59
	2.5–3.5	895.24	7	37	82.41	174.77	121.98	21.61	0.18
	3.5–4.5	970.09	1	23	117.68	249.84	180.40	32.56	0.18
Total	1.5–4.5	9431.21	11	211	1.16	249.84	59.79	58.85	0.98

SD Standard Deviation, *CV* Coefficient of variation, i.e., SD divided by average stem volume

2.2 Image processing

A satellite image from the commercial observation satellite SPOT-6 was acquired in August 2018, during the same period as the forest inventory. The SPOT-6 satellite images have four broadbands (Blue, Green, Red, and Near Infrared) at a spatial resolution of 6 m and a panchromatic band at 1.5 m resolution. The radiometric resolution is 12 bits per pixel. The broadbands were merged with the panchromatic band using a pansharpenning algorithm, resulting in a multispectral image of 1.5 m spatial resolution. The top-of-atmosphere reflectance was used directly in the analysis since only one image was used in this study. There were no clouds in the area of interest. The final image was georeferenced based on ground control points.

A time series of images from Landsat satellites was used, covering the period from January 2000 to August 2018. The time series included 30 m resolution multispectral images from Landsat-5 Thematic Mapper (TM), Landsat-7 Enhanced Thematic Mapper Plus (ETM+), and Landsat-8 Operational Land Imager (OLI). The atmospherically corrected Surface Reflectance images were produced by the United States Geological Survey (USGS). They include information on eventual sensor failure, cloud and cloud shadow masks. To eliminate residual clouds or cloud shadows, the masks were expanded with a supplementary buffer of 150 m (5 pixels). Reflectances from three different Landsat sensors (5, 7, and 8) were inter-calibrated following Roy et al. (2016). The Landsat Surface Reflectance product was obtained from Google Earth Engine (Collection 1 Tier 1 dataset).

The SPOT-6 and Landsat images were resampled to the same resolution of 7.5 m to facilitate post-treatments. This resolution was chosen as it can be easily obtained from the SPOT-6 1.5 m resolution image and from the Landsat 30 m resolution image. Additionally, this resolution is also compatible with the size of an inventory plot ($\sim 3 \times 3$ pixels), and speeds up the texture computation on the SPOT-6 image. The resampling method used was a simple nearest neighbor algorithm.

Five sets of variables were extracted and calculated from SPOT-6 and/or Landsat images. These sets include (i) reflectance of spectral bands, (ii) spectral band ratios, (iii) vegetation indices, (iv) textural indices, and (v) integrals of vegetation indices of satellite time series. Table 2 provides a list of the variables obtained and their descriptions.

The SPOT-6 image was used to compute the four spectral bands, twelve simple ratios of these bands, and three commonly used vegetation indices: NDVI (Tucker 1979), SAVI (Huete 1988), and ARVI (Kaufman and Tanré 1996).

Additionally, textural indices were computed using the gray level co-occurrence matrix (GLCM) proposed by Haralick et al. (1973), which is a widely used method for extracting texture from images. Four second-order texture measurements were obtained, for each spectral band: contrast (CON), correlation (COR), energy (ENE), also known as second angular momentum, and homogeneity (HOM). These measures are commonly used and considered relevant in remote sensing studies (Rodriguez-Galiano et al. 2012). The GLCM was calculated on the resampled 7.5 m resolution SPOT-6 image, using a 3×3 pixel window. This calculation was performed in the four angles of orientation of the neighboring pixel (0°, 45°, 90° and 135°) with an omnidirectional approach to cover the various types of spatial relationships among the pixels. This window definition was also used in the study by Sarker and Nichol (2011). Finally, the average of the values computed in the four directions was determined. The spectral and textural indices were combined to generate indices (Table 2).

The NDVI was calculated for each pixel in the Landsat image time series. The NDVI time series for each Landsat pixel were smoothed over time, using a cubic spline function. This smoothing technique eliminates small variations that occur on a short time scale, which do not reflect changes in the canopy, such as residual atmospheric effects (le Maire et al. 2011a, b, c). The NDVI time series provided several variables: the minimum and maximum values during the rotation, and the integrals of the NDVI smoothed curve between a starting date (e.g., planting date) and an ending date (e.g., inventory date) (Table 2). These variables were selected based on previous studies by Marsden et al. (2010a, b) and le Maire et al. (2011a, b, c). The image components were extracted using R (R Core Team 2021) and the following packages: 'rgdal', 'maptools', 'raster', 'sp' and 'glcm'.

2.3 Linear models for stem volume prediction and variable selection

First, we develop a multilinear regression model to predict stem volume directly from the spatial variables listed in Table 2. Due to the large number of variables (86) and possible multicollinearity, variable selection was necessary to obtain a simpler linear regression model for stem volume prediction. Different regression methods, all including variable selection, were tested (Kuhn 2008): forward stepwise selection (mod_fw), backward stepwise selection (mod_bw), sequential replacement (mod_sw), Sparse Partial Least Squares (mod_spls), Spike and Slab Lasso (mod_sl), Ridge (mod_rd), Lasso (mod_ls), and Elastic Net (mod_en). These methods have different theoretical basis, which may lead to different selected variables and model accuracies, underlining the importance to test a range of models. To evaluate the predictive performance of all tested models to predict plot-level stem volume, the plots were split into two partitions, 80% for training and 20% for validation. The Root Mean Square Error (RMSE), the Adjusted Coefficient of Determination (R^2 adj), and the corrected Akaike Information Criterion (AICc) were

Table 2 Description of the 86 variables computed and extracted from the SPOT-6 and Landsat 5, 7, and 8 satellites images

Abbreviation	Source	Variable type	N	Description of image variables
AGE		Ancillary	1	Age of the stand at the time of the forest inventory
B1, B2, B3, B4	SPOT-6	Spectral bands	4	Blue (B1), green (B2), red (B3), near-infrared (B4) reflectances
$RAT\_B_x\_B_y$	SPOT-6	Simple rations	12	Simple ratio: $RAT\_Bi\_Bj = Bi/Bj$, $i \neq j$, with (Bi,Bj) being all pairs of bands
NDVI	SPOT-6	Vegetation index	1	Normalized Difference Vegetation Index $(B4 - B3)/(B4 + B3)$
SAVI	SPOT-6	Vegetation index	1	Soil Adjusted Vegetation Index $(B4-B1)/(B4 + B1 + L)*(1 + L)$, where $L = 0.5$
ARVI	SPOT-6	Vegetation index	1	Atmosphere Resistant Vegetation Index $(B4 - B3 - \gamma*(B3 - B1))/(B4 + B3 - \gamma*(B3 - B1))$, where $\gamma = 1$
CON_B_x	SPOT-6	Texture index	4	Band's contrast (local variation of the intensity of a pixel and its neighbor): CON_B1 (band 1), CON_B2 (band 2), CON_B3 (band 3), CON_B4 (band 4)
COR_B_x	SPOT-6	Texture index	4	Band's correlation (linear dependence in the reflectance between a pixel and its neighbor): COR_B1 (band 1), COR_B2 (band 2), COR_B3 (band 3), COR_B4 (band 4)
ENE_B_x	SPOT-6	Texture index	4	Band's energy (evaluates the uniformity of the texture): ENE_B1 (band 1), ENE_B2 (band 2), ENE_B3 (band 3), ENE_B4 (band 4)
HOM_B_x	SPOT-6	Texture index	4	Band's homogeneity (measures the regularity present in the image): HOM_B1 (band 1), HOM_B2 (band 2), HOM_B3 (band 3), HOM_B4 (band 4)
$MUL\_B_x\_CON\_B_x$	SPOT-6	Texture index	4	Product of band reflectance and band contrast: MUL_B1_CON_B1 (B1 x CON_B1), MUL_B2_CON_B2 (B2 x CON_B2), MUL_B3_CON_B3 (B3 x CON_B3), MUL_B4_CON_B4 (B4 x CON_B4)
$MUL\_B_x\_COR\_B_x$	SPOT-6	Texture index	4	Product of band reflectance and band correlation: MUL_B1_COR_B1 (B1 x COR_B1), MUL_B2_COR_B2 (B2 x COR_B2), MUL_B3_COR_B3 (B3 x COR_B3), MUL_B4_COR_B4 (B4 x COR_B4)
$MUL\_B_x\_ENE\_B_x$	SPOT-6	Texture index	4	Product of band reflectance and band energy: MUL_B1_ENE_B1 (B1 x ENE_B1), MUL_B2_ENE_B2 (B2 x ENE_B2), MUL_B3_ENE_B3 (B3 x ENE_B3), MUL_B4_ENE_B4 (B4 x ENE_B4)
$MUL\_B_x\_HOM\_B_x$	SPOT-6	Texture index	4	Product of band reflectance and the band homogeneity: MUL_B1_HOM_B1 (B1 x HOM_B1), MUL_B2_HOM_B2 (B2 x HOM_B2), MUL_B3_HOM_B3 (B3 x HOM_B3), MUL_B4_HOM_B4 (B4 x HOM_B4)
$RAT\_B_x\_CON\_B_x$	SPOT-6	Texture index	4	Ratio of band reflectance and band contrast: RAT_B1_CON_B1 (B1/CON_B1), RAT_B2_CON_B2 (B2/CON_B2), RAT_B3_CON_B3 (B3/CON_B3), RAT_B4_CON_B4 (B4/CON_B4)
$RAT\_B_x\_COR\_B_x$	SPOT-6	Texture index	4	Ratio of band reflectance and band correlation: RAT_B1_COR_B1 (B1/COR_B1), RAT_B1_COR_B1 (B2/COR_B2), RAT_B3_COR_B3 (B3/COR_B3), RAT_B4_COR_B4 (B4/COR_B4)
$RAT\_B_x\_ENE\_B_x$	SPOT-6	Texture index	4	Ratio of band reflectance and band energy: RAT_B1_ENE_B1 (B1/ENE_B1), RAT_B2_ENE_B2 (B2/ENE_B2),RAT_B3_ENE_B3 (B3/ENE_B3), RAT_B4_ENE_B4 (B4/ENE_B4)
$RAT\_B_x\_HOM\_B_x$	SPOT-6	Texture index	4	Ratio of band reflectance and band homogeneity: RAT_B1_HOM_B1 (B1/HOM_B1), RAT_B2_HOM_B2(B2/HOM_B2), RAT_B3_HOM_B3 (B3/HOM_B3), RAT_B4_HOM_B4 (B4/HOM_B4)
INT_0_DS	Landsat 5, 7 e 8	Time series	1	Cumulative NDVI from planting date to the last satellite collection date (17/08/2018)
INT_0_DM	Landsat 5, 7 e 8	Time series	1	Cumulative NDVI from planting date to inventory measurement date
INT_0_1	Landsat 5, 7 e 8	Time series	1	Cumulative NDVI from planting date to 1-year-old
INT_0_2	Landsat 5, 7 e 8	Time series	1	Cumulative NDVI from planting date to 2-year-old
INT_0_MIN	Landsat 5, 7 e 8	Time series	1	Cumulative NDVI from planting date to minimum NDVI (minimum NDVI after 2-year-old)
INT_0_MAX	Landsat 5, 7 e 8	Time series	1	Cumulative NDVI from planting date to maximum NDVI
MAX_MIN	Landsat 5, 7 e 8	Time series	1	Difference (INT_0_MAX—INT_0_MIN)
INT_1_DS	Landsat 5, 7 e 8	Time series	1	Cumulative NDVI from 1 year after planting to the last satellite collection date (17/08/2018)
INT_1_DM	Landsat 5, 7 e 8	Time series	1	Cumulative NDVI from 1 year after planting to the inventory measurement date
INT_1_MIN	Landsat 5, 7 e 8	Time series	1	Cumulative NDVI from 1 year after planting to the minimum NDVI (minimum NDVI after 2-year-old)
INT_1_MAX	Landsat 5, 7 e 8	Time series	1	Cumulative NDVI from 1 year after planting to the maximum NDVI
INT_2_DS	Landsat 5, 7 e 8	Time series	1	Cumulative NDVI from 2 years after planting to the satellite collection date (17/08/2018)
INT_2_DM	Landsat 5, 7 e 8	Time series	1	Cumulative NDVI from 2 years after planting to the inventory measurement date
INT_2_MIN	Landsat 5, 7 e 8	Time series	1	Cumulative NDVI from 2 years after planting to the minimum NDVI (minimum NDVI after 2-year-old)
INT_2_MAX	Landsat 5, 7 e 8	Time series	1	Cumulative NDVI from 2 years after planting to the maximum NDVI
NDVI_ 2_MIN	Landsat 5, 7 e 8	Time series	1	Minimum NDVI value after 2-year-old
NDVI_ 2_MAX	Landsat 5, 7 e 8	Time series	1	Maximum NDVI value after 2-year-old
MAX_MIN_NDVI	Landsat 5, 7 e 8	Time series	1	Difference (NDVI_2_MAX—NDVI _2_MIN)

N is the number of variables combinations for a given line in the table

computed to compare the final models. Statistical analyses were conducted using R and the following packages: 'leaps', 'gstat', 'spls', 'pastecs', 'spikeslab', 'caret', and 'measures'.

2.4 Kriging with External Drift (KED) geostatistical model

Kriging with External Drift predicts the variable Z at an unsampled point s by combining a deterministic function $\mu(s)$ which indicates the underlying drift, and a stochastic residual component $\varepsilon(s)$ (Goovaerts 1997; Wackernagel 2003).

$$Z(s) = \mu(s) + \varepsilon(s) \qquad (1)$$

In KED, $\mu(s)$ is modeled by a function of auxiliary variables:

$$\mu(s) = \sum_{l=0}^{L} a_l f_l(X(s)) \qquad (2)$$

where a_l are coefficients to be estimated from the data, f_l is the function of auxiliary variables describing the drift (f_0 is the constant function with value of 1), $X(s)$ is the vector of auxiliary variables at location s, and $L+1$ is the number of functions used in modelling the drift.

Given n observed values, $(s_1),...,Z(s_n)$, at sample points, $s_1, s_2,...,s_n$, we predict $Z^*(s_0)$ at an unsampled location s_0 as:

$$Z^*(s_0) = \sum_{i=1}^{n} \omega_i Z(s_i) \qquad (3)$$

where ω_i is the kriging weight assigned to $Z(s_i)$. These weights are computed by minimizing the estimation error variance subject to the unbiased constraint. This is solved using the Lagrange multiplier method, leading to the KED system of equations:

$$\begin{cases} \sum_{j=1}^{n} \omega_j C(s_i - s_j) + \sum_{l=0}^{L} \lambda_l f_l(X(s_i)) = C(s_i - s_0), & \text{for } i = 1,\ldots,n \\ \sum_{j=1}^{n} \omega_j = 1, \\ \sum_{j=1}^{n} \omega_j f_l(X(s_i)) = f_l(X(s_0)), & \text{for } l = 0,1,\ldots,L \end{cases} \qquad (4)$$

where $C(h)$ is the covariance function of the variable Z in function of the distance h linking any two points s and $s+h$. In Eq. 4, h is written as the difference between two points, e.g., $(s_i - s_j)$ (Goovaerts 1997; Wackernagel 2003). λ_l is a Lagrange multiplier. In the present study, we use the external drift variables $X(s)$ that were selected from Table 2 (see section "2.4").

The covariance function $C(h)$ was obtained from the experimental variogram by fitting a variogram model. The variogram model is a smooth function that is fitted on the experimental variogram obtained from the data. Most variogram models are characterized by three parameters:

the "nugget", reflecting measurement error and unresolved spatial variability; the "sill", representing the spatial variation of the data explained by the variogram model; and the "range", denoting the distance over which the samples exhibit spatial correlation. We tested common variogram functions including exponential (Exp), spherical (Sph), Martén (Mat), Gaussian (Gau) and Martén M. Stein's parameterized (Ste) (mathematical formulations in Stein 1999). The five models tested were fitted to the data using Levenberg–Marquardt fitting with a weighted least square method, and we selected the one with minimum RMSE value. Spatial predictions of stem volume were then made across the entire area at pixel scale by first solving the Eq. 4 for ω and λ through linear algebra. The 'gstat' R package was used for these computations.

2.5 Generalized additive model (GAM) with spatial coordinates

Local predictions were performed using GAM with spatial coordinates. The spatial relationship is represented by the coordinate smoothing function, and non-stationary effects are modeled by interactions between the smoothing function and the auxiliary variables (Nussbaum et al. 2018). The estimate of the local mean stem volume $\widehat{\mu}_s$ at location s follows the equation:

$$g(\widehat{\mu}_s) = \mu_0 + \beta_1 X_s^1 + \cdots + \beta_k X_s^k + f(x_s, y_s) + \varepsilon \qquad (5)$$

where μ_0 is the model intercept, $X_s^1,...,X_s^k$ are local auxiliary variables selected previously (see "2.4"), $\beta_1, \ldots, \beta_k$ are the parameters to be estimated, f is a bivariate smoothing function relating the covariates x_s (longitude) and y_s (latitude) to the response variable, g is an increasing monotonic link function (identity in our case), and ε is the identically distributed random error. We used the 'gam' function of the R package 'mgcv', where the smoothing function f is a penalized regression spline, and where the smoothness criteria of this function is internally optimized using the Generalized Cross Validation (GCV) criterion (Wood 2017). Note that we did not use co-kriging of the residuals in this approach, as spatial location is incorporated in the model through the local smoothing term. Therefore, while our GAM approach may not be considered as a geostatistical technique strictly speaking, it explicitly includes location in the predictions, representing landscape features not captured by other auxiliary variables in the model.

2.6 Analysis of the stem volume predictions generated by the models

To assess the predictive performance of the linear model, KED and GAM, fivefold cross-validation was used. The statistics used to assess the predictive capacity of the

Aló *et al. Annals of Forest Science*   (2024) 81:43

Page 7 of 16

models were the same as those applied in the method of variable selection, i.e., the Root Mean Square Error (RMSE) and the Adjusted Coefficient of Determination (R^2adj). For the KED model and GAM, we also computed the part of the variance explained by the auxiliary variables and the variance explained by the spatial features of the model.

3 Results

3.1 Landsat reflectance, vegetation indices, NDVI time series, and texture indices

Pixel-scale interpolated and smoothed Landsat NDVI time-series were obtained. Figure 1 presents an example of a single pixel that underwent conversion from a pasture to a eucalypt plantation in February 2014 and was 4.5 years old at the time of inventory in August 2018. The time-series clearly illustrates the transition from low NDVI values in the pasture stage to high and stable NDVI values in the eucalypt stage. Seasonal variation is evident during the eucalypt stage, resulting from the contrast between rainy and drought seasons (Fig. 1).

The dataset, described in Table 1, was analyzed using the Spearman correlation method. A correlation matrix was generated (Fig. 4 in Appendix) to identify high correlations, both positive and negative, between spectral bands, vegetation indices, texture, NDVI time series integrals, and stand age. This highlights the need for a variable selection technique to prevent multicollinearity in subsequent linear models. A Spearman correlation was also conducted between stem volume and each variable in Table 1. The results showed that most variables were significantly correlated with stem volume, either positively or negatively. However, there was large variability in the correlation coefficients (Fig. 5 in Appendix). Significant relationships were found between stem volume and tree age (correlation
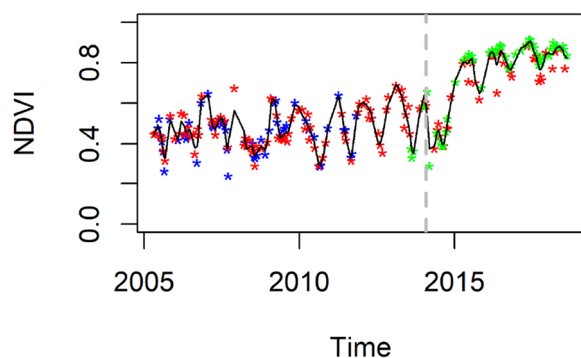
Fig. 1 Pixel-scale NDVI measured by 3 Landsat sensors (blue: Landsat TM 5; Red: Landsat 7 ETM +; Green: Landsat 8 OLI) after pre-treatments described in the main text. The area was a pasture and was planted with *Eucalyptus* in February 2014 (dashed gray line). The black line shows the cubic spline smoothing curve

coefficient of 0.83), some NDVI integrals of Landsat images (INT_0_DM, INT_0_DS and INT_1_DM) with values ranging between 0.86 and 0.89, and SPOT direct spectral bands B1, B2, and B3 with negative relationships of -0.74 to -0.77. However, some vegetation indices, such as the Landsat minimum and maximum NDVI values after 2 years, were not significantly correlated with stem volume.

3.2 Selection of variables

To minimize the impact of multicollinearity in linear regression models, we applied variable selection through different algorithms listed in Section "2.4". Each final model contained between 3 and 7 variables (Table 3).

The models used variables computed from NDVI time series integrals. However, variables such as AGE and B1 to B3 reflectances, which showed high correlation with stem volume (Fig. 5 in Appendix), were not included. Instead, ratios between bands were found in nearly all models, with RAT_B3_B2 (which had a direct correlation coefficient of 0.52 with stem volume) being present in half of the tested models. None of the models selected the NDVI, SAVI, and ARVI vegetation indices. The models included two texture indices: MUL_B1_HOM_B1 (which had a direct correlation of -0.47 with wood volume) in mod_spls and RAT_B2_CON_B2 (which had a correlation coefficient of 0.19) in mod_en. The best model was obtained using the Lasso regression method mod_ls, with a RMSE of 14.03 m^3 ha^{-1}, an R^2 of 0.94 and an AIC of 1399.54 in the calibration data. This model was also the best in the validation dataset, with an RMSE of 12.44 m^3 ha^{-1} and an R^2 of 0.95. The mod_ls used three simple ratio variables and three NDVI integrals variables (Table 4). The variables INT_0_DM was found to be highly significant in the model and was also correlated with AGE (Fig. 4 in Appendix). These variables highlight the importance of spectral information, age, and NDVI fluctuations during the stand growth. The regression and the dispersion of the residuals of the validation data for this model are shown in Fig. 2.

3.3 Analysis of stem volume prediction generated by KED and GAM

The variables selected with mod_ls were implemented as external drift in the KED algorithm. The spatial dependence structure was obtained from the residuals variogram. The best fit of the variogram was with Martén M. Stein's model (Ste) after testing the RMSE of the variogram fit with different model types. The *range* value was 2750 m, the *nugget* was 92.5 m^3.ha^{-1}, and the *sill* was 179 m^3.ha^{-1}. After calibration, the KED was applied to the independent validation dataset, resulting in an R^2 of 0.96 and an RMSE of 10.62 m^3.ha^{-1}. The auxiliary variables

Aló *et al. Annals of Forest Science*   (2024) 81:43

Page 8 of 16

Table 3 Statistics of the multilinear models for stem volume prediction (m^3 ha^{-1}) obtained from different variable selection methods (see Section "2.4")

Method	Selected auxiliary variables	Calibration				Validation	
		RMSE	R$^2_{adj}$	AICc		RMSE	R^2
mod_fw	INT_2_DS, INT_2_MAX, NDVI_2_MIN	20.5	0.88	1519.39		19.1	0.88
mod_bw	RAT_B3_B1, INT_0_MIN, NDVI_2_MIN	31.6	0.72	1666.57		27.3	0.74
mod_sw	MAX_MIN, INT_2_DS, NDVI_2_MIN	21.2	0.87	1531.46		18.5	0.88
mod_spls	MUL_B1_HOM_B1, INT_0_DS, INT_1_MAX, NDVI_2_MAX	18.5	0.90	1487.06		18.1	0.89
mod_sl	RAT_B2_B1, RAT_B3_B2, RAT_B3_B4, MAX_MIN_NDVI, INT_2_MIN	30.2	0.74	1656.22		26.3	0.76
mod_rd	RAT_B3_B2, RAT_B4_B3, INT_2_MIN, MAX_MIN_NDVI	30.5	0.74	1657.00		26.2	0.76
mod_ls	RAT_B1_B2, RAT_B2_B4, RAT_B3_B2, INT_0_DM, INT_0_2, INT_2_MAX	**14.0**	**0.94**	**1399.54**		**12.44**	**0.95**
mod_en	RAT_B1_B2, RAT_B2_B4, RAT_B3_B2, RAT_B2_CON_B2, INT_0_2, INT_1_DM, INT_2_MAX	14.2	0.94	1401.06		14.6	0.93

Statistics are given in the table for the calibration dataset and the validation dataset (20% of the data). In bold the model which gave the best statistics. *mod_fw*: forward stepwise selection, *mod_bw*: backward stepwise selection, *mod_sw*: sequential replacement, *mod_spls*: sparse partial least squares, *mod_sl*: Spike and Slab Lasso; *mod_rd*: Ridge; *mod_ls*: Lasso; *mod_en*: Elastic Net

Table 4 Coefficients and significance levels for variables of the linear model mod_ls and the GAM model

	Linear model mod_ls	GAM
Intercept	− 360.4***	− 379.8***
RAT_B1_B2	236.8***	255.5***
RAT_B2_B4	− 177.5*	− 204.1**
RAT_B3_B2	124.2***	141.3***
INT_0_DM	0.2246***	0.2258***
INT_0_2	0.1327*	0.117*
INT_2_MAX	− 0.0067	− 0.0059

Significance levels: $p < 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

explained approximately 94% of the variance, while the spatial correlation term explained 5%.

The GAM model was also applied to the validation dataset and produced similar results to the KED, with an R^2 of 0.96 and an RMSE of 10.56 m^3.ha^{-1}. In this case, the auxiliary variables accounted for approximately 98% of the variance, while the spatial correlation term accounted for only 1.2%. The coefficients were similar to those in the linear model mod_ls, with only a slight impact on the significance of some variables (Table 4).

3.4 Application of the linear, KED, and GAM models

The Linear, KED, and GAM models used the same set of six variables derived from SPOT-6 and Landsat 5, 7, and 8 images. The KED and GAM models, which incorporated the spatial structure of the data, provided the most accurate predictions in the independent dataset (Fig. 2). The stem volume map was produced by applying the models at a pixel-scale resolution of 7.5 m. This process is illustrated for GAM in a small area in Fig. 3, and a comparison with the linear model and KED is shown in Fig. 6 in Appendix. These comparisons reveal only minor visual differences between the models. Figure 3 shows significant changes in stem volume between
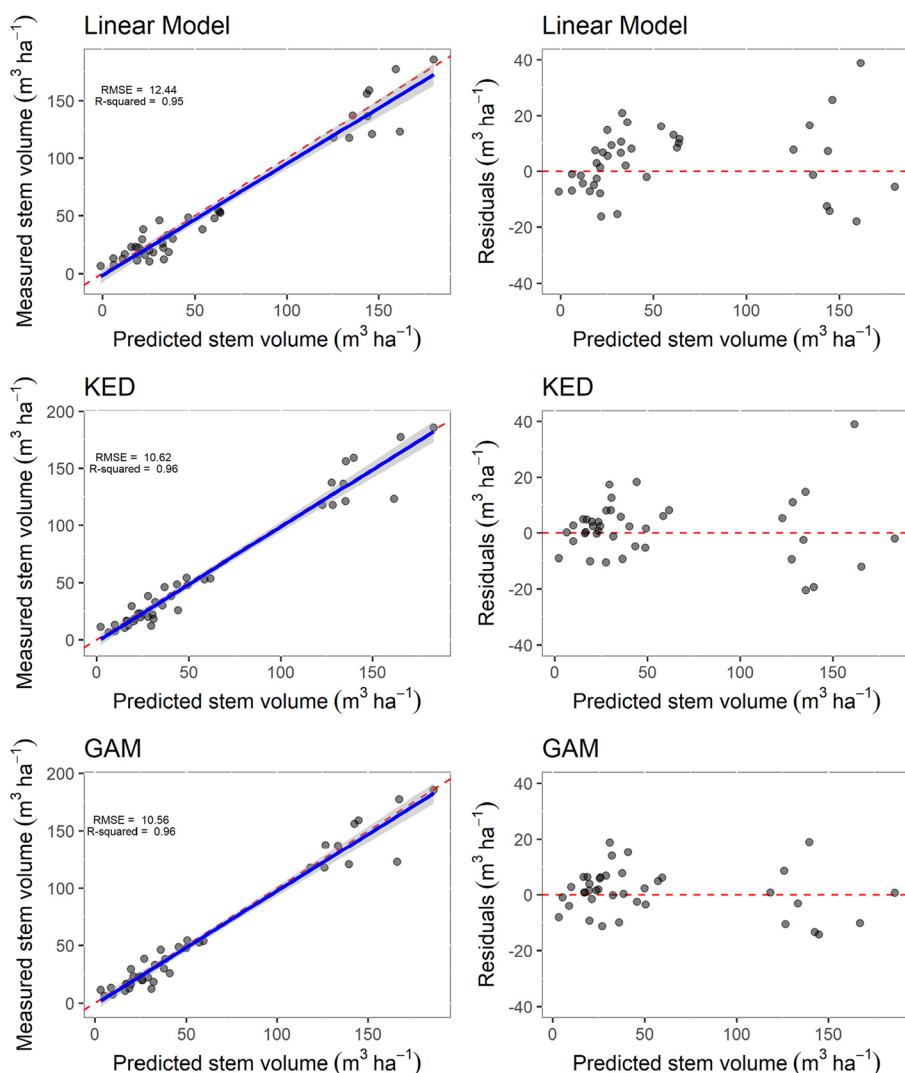
Fig. 2 Dispersion and residual graphs between the predicted and measured stem volume values (left) and Residual Plot (right), obtained on the validation dataset for the Linear, KED, and GAM models, after the mod_ls variable selection method (See Table 3). The regression line with a 95% confidence interval, the RMSE (m3.ha^{-1}), and R-squared are displayed in the left graph

stands, primarily due to differences in management (such as age), as well as within-stand variations.

4 Discussion

4.1 Optimal variable selection for predicting wood volume in linear models

The vegetation indices had a positive and moderate correlation with stem volume, which is consistent with Dos Reis et al. (2018). NDVI and SAVI, in particular, showed correlation coefficients of $R = 0.49$ and $R = 0.47$, respectively. It is important to note that these indices are primarily associated with the green leaf surface within a pixel. During the initial months of the plantation, when the canopy has not yet fully developed, the green leaf surface may be correlated with wood biomass or volume. However, the correlation between NDVI and canopy tends to decrease over time. This is due to

the saturation of NDVI at high values as the canopy matures and closes. This phenomenon has been observed in various plantations after canopy closure (Le Maire et al. 2011a, b, c; Avitabile et al. 2012; Souza et al. 2019). After canopy closure, wood volume continues to increase while leaf area stabilizes or even decreases (le Maire et al. 2019). This is observed in the NDVI time series example shown in Fig. 1.

Texture measurements have been explored as an alternative to vegetation indices for improved discrimination in forest structure (Sarker and Nichol 2011; Dube and Mutanga 2015). However, in the present study, they only showed a moderate correlation with the stem volume. The final models included a combination of texture indices and reflectance bands (Table 3), but not in the best model, mod_ls. Castillo-Santiago et al. (2013) found moderate correlations between texture and biomass, which varied
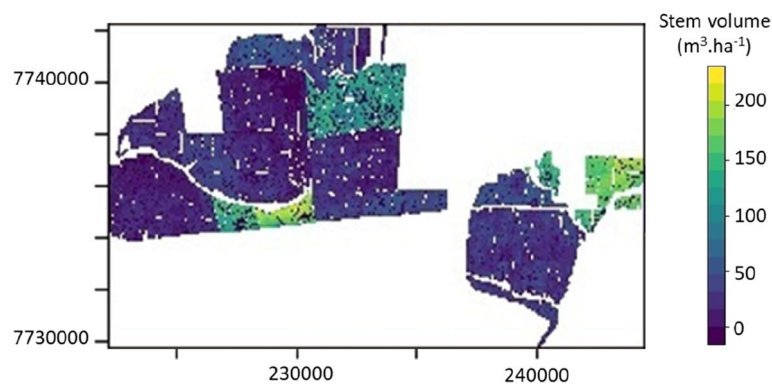
Fig. 3 Example of a map of distribution of the estimated volume (m^3 ha^{-1}) obtained from the GAM model, predicted for August 2018. No data values in white corresponds to other land use around the stand, or to gaps within the stand previously masked

depending on the spectral region. However, Lu (2006) study showed that texture alone may not be sufficient to establish relationships with wood volume estimates. This is primarily because these relationships tend to be site-specific.

In the context of short-rotation planted forests, it is possible to analyze the stand's history using time-series data that extends from the planting date. The variables derived from the Landsat time-series data included the minimum and maximum NDVI values within stand age intervals, or the integration of NDVI over these age intervals. Stand age was included as a distinct variable, obtained directly from the stand manager rather than from remote-sensing. Interestingly, the planting date itself can be obtained from the time-series data (Le Maire et al., 2014). Stand age had a high correlation with stem volume ($R = 0.83$), but was not included in the variable selection process due to its substantial correlation with the integral NDVI variables.

Marsden et al. (2010a, b) and Le Maire et al. (2011a, b, c) used NDVI time-series data derived from MODIS satellite imagery (at a 250-m pixel resolution) to calculate NDVI integrals at the stand level, spanning from planting date to inventory date. They found a strong correlation with stand-scale biomass. This study confirmed the direct correlation, with a R value of 0.89 between volume and INT_O_DM. The correlation between MODIS data integrals and volume estimates was initially established at the stand scale. However, it remained high at the local scale, even with the spatial resolution of Landsat reduced to 30 m and volume estimates based on local inventory plots of 500 m^2. Despite the less frequent data acquisition frequency of Landsat in comparison to MODIS, the strength of this relationship was not diminished.

4.2 KED and GAM geostatistical models for stem volume mapping

The spatial prediction of the stem volume was slightly improved with KED or GAM, compared to the linear model alone. This suggested the presence of spatial patterns within the data that were not adequately accounted for by the remotely sensed data alone. The proportion of variance explained by auxiliary variables was high (94%) compared to that explained by spatial correlation for both KED and GAM. This is consistent with the relatively minor improvement compared to the linear model. Although KED slightly outperformed GAM in capturing spatial variations, it is worth noting that the KED method required more computational resources and the selection of a variogram model. Therefore, we chose to use GAM to produce the final stem volume map, as shown in Fig. 3.

Previous studies have shown that KED is effective in improving volume predictions and spatial representations. For example, Lochhead et al. (2018) found that KED was the most accurate method for predicting various forest attributes in Canada's boreal forest among the tested nonlinear models. Viana et al. (2012) found limited utility of kriging in *Pinus* stands due to low data spatial correlation. However, they emphasized its potential in forests with more sampling units, higher spatial continuity, and by testing diverse auxiliary variables. Other studies have highlighted the advantages of the GAM spatial model in volume or biomass predictions. Alberdi et al. (2020) found that GAM could explain a greater variability in the data and outperformed other methods in estimating wood stock, especially in areas lacking wood stock information. Maack et al. (2016) showed that GAM has potential for large-scale wood volume interpolations by combining inventory data, outperforming models like Random Forest in this context.

The study assessed model uncertainties by testing on an independent subset of data. However, further work is required to improve the analysis of the uncertainty associated with model predictions, and its propagation across scales. Estimating uncertainty is complex, particularly when several models are combined in a hierarchical manner (Saarela et al. 2016), such as a model to predict

individual tree volume (including uncertainty on the tree measurements), a model to predict plot scale volume (including uncertainty from the inventory measurements), and a model to interpolate spatially the plot-scale volume (including uncertainty from the auxiliary data). Saarela et al. (2020) proposed a hierarchical model-based approach that considers multiple sources of uncertainty at two different scales: the tree-level biomass model and models linking the plot-level biomass to auxiliary variables (in their case coming from LiDAR). This approach allowed for the attribution of the uncertainty to the different models, and the generation of uncertainty maps and uncertainty estimates for regional averages. A similar approach could be used to test the hypothesis that model uncertainty in tree-level and plot-level volume predictions only marginally contributes to the final uncertainty on these clonal plantations.

4.3 Model genericity and future directions for improving high-resolution mapping of forest volume

The models were created for eucalypt plantations in a specific pedoclimatic context. The spatial models developed inherently rely on local calibration with inventory data, but the method can be applied to other regions and other plantations. The linear model used to link stem volume to satellite-derived features in the spatial models is specifically designed for short-rotation plantations with dynamic growth. The most important variables are linked to NDVI integrals since planting date. Our study found a robust correlation, resulting in a low RMSE of 12.44 m^3/ha. Marsden et al. (2010a, b) obtained an RMSE of 24 m^3/ha for Eucalyptus plantations, which was almost double of our result. This difference may be due to greater stand variability across a larger area, encompassing diverse pedoclimatic conditions and growth histories. It is worth noting that our linear model's additional features may have improved the predictions compared to their model. Stand age alone could replace NDVI integrals, as discussed before (Dube et al. 2015; Dos Reis et al. 2019). However, this solution reduced the model accuracy, as shown in the present study, and the intra-stand spatial variability of volume relies therefore only on other spatial features of the models (spectral reflectance ratios, etc.). The locally developed approach in our study likely applies to various eucalypt growth conditions in Brazil, with precision subject to local variations. Further exploration is necessary to confirm its applicability in different locations, especially in stands that have experienced damages resulting in local volume heterogeneity.

To improve the precision of high-resolution forest volume mapping, several strategies, and further research may be considered: (i) Incorporating additional information on the 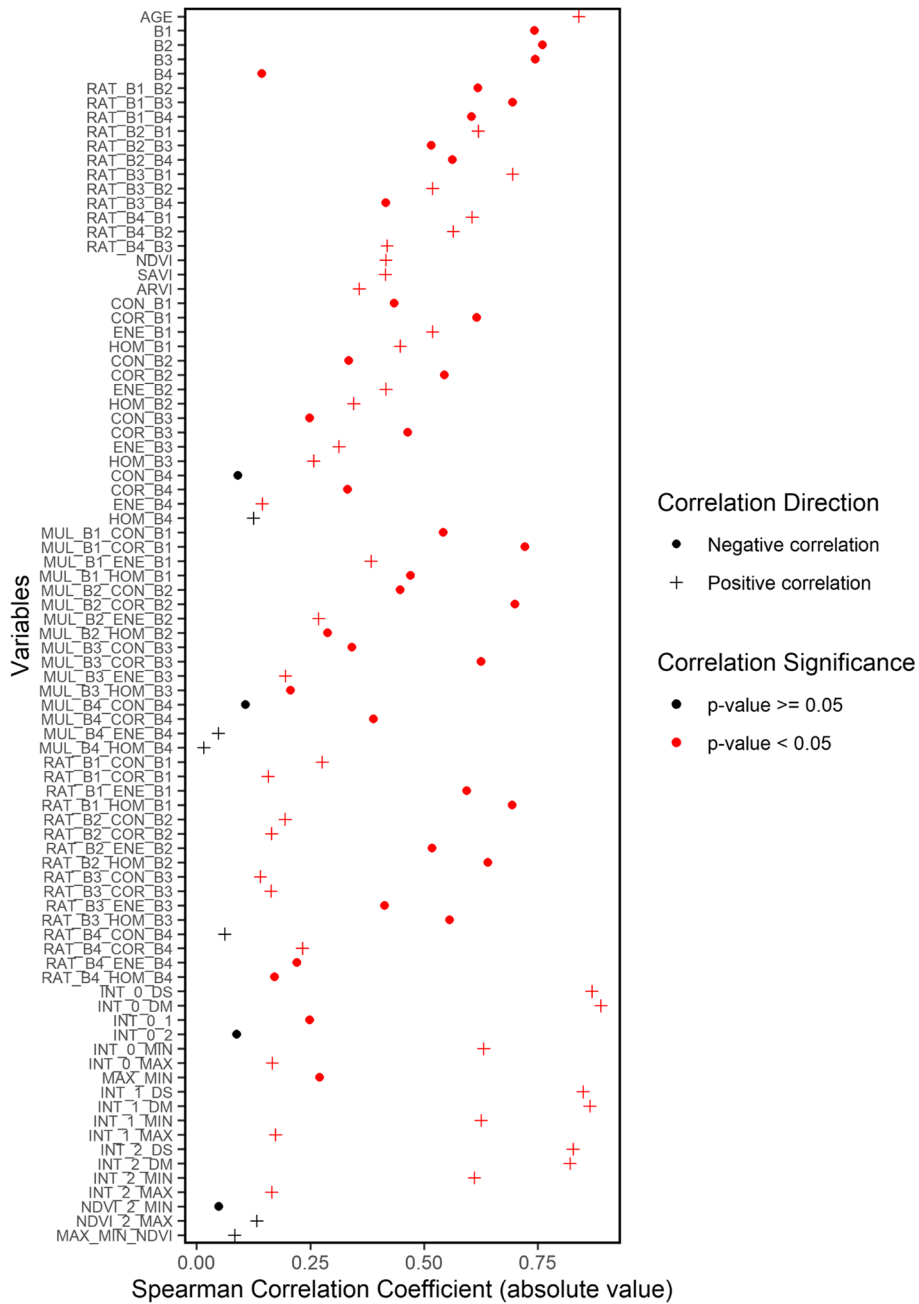stand, such as planting density, management, genotype composition, or environmental variables (e.g. Stape and Alcarde Alvares 2023) to improve volume prediction in the linear model. This can provide more accurate volume prediction combined with the variables coming from satellite images; (ii) Leveraging other types of remote sensing data, such as higher-resolution Sentinel 2 NDVI data (10 m), and radar and lidar, to enhance the volume prediction. These data sources provide different information on the forest stand structure and complement the optical satellite data; (iii) Adding more plots in diverse locations to improve and validate the volume maps.

The method developed in this study has various applications, primarily improving stand volume precision by addressing intra-stand variability that may not be captured by measurements of only a few inventory plots. This method is cost-efficient and scalable to cover thousands of hectares, in contrast to more precise but expensive and local methods based on Lidar acquisition, for instance. The prediction of stem volume, which can be obtained for instance at the first inventory around 2 years of age, can guide the selection of a few additional temporary plots. The method of complementing permanent plots with temporary plots based on satellite analysis was found to be efficient for the National Forest Inventory (Räty and Kangas 2019). In rapidly expanding plantations, this can ensure that inventory plots are more representative of the entire study area. This improves the volume prediction accuracy across different stand ages, particularly at the pre-harvest inventory date.

5 Conclusion

The study proposed a novel approach that combined spectral, textural, and temporal features extracted from both SPOT-6 and Landsat 5, 7, and 8 satellite images, along with spatial models and forest inventory data, to map the stem volume of fast-growing plantations. Out of the 86 remotely sensed variables, only six were selected in a linear model calibrated with the Lasso algorithm. These key variables include integrals of NDVI time-series from planting date to measurement date or to 2 years of age, integral between 2 years of age and the date of maximum NDVI, and spectral reflectance ratios between bands 1 and 2, 2 and 4, and 3 and 2. The linear model alone yielded precise results, with R^2 of ~0.95 and RMSE of ~12.44. The same variables were used as auxiliary variables in KED and GAM models. They provided slight improvements in prediction accuracy, with R^2 of ~0.96 and RMSE of ~10.6 $m^3.ha^{-1}$, and captured finer spatial details, enhancing the robustness and precision of the model's results. These spatial models were applied in the forest area, distinguishing regions with different stand ages and estimating the within-stand variability of wood volume.

Appendix

Fig. 4 Spearman correlation matrix for 62 of the 86 variables listed in Table 2, derived from 211 forest inventories of Eucalyptus stands across various age classes. To enhance interpretability, the matrix displays the absolute values of the correlations. Texture variables from B1 and B2, which showed strong correlations with those from B3, were excluded. The variables are arranged in order to place those with high correlations adjacent to each other.

Fig. 5 Spearman correlation between each of the 86 variables listed in Table 2 and the stem volume. Significant correlation are in red, with plus symbols for positive correlation (*p*<0.05) and dots for negative correlations. Variables are ordered by type of variables, as in Table 2.
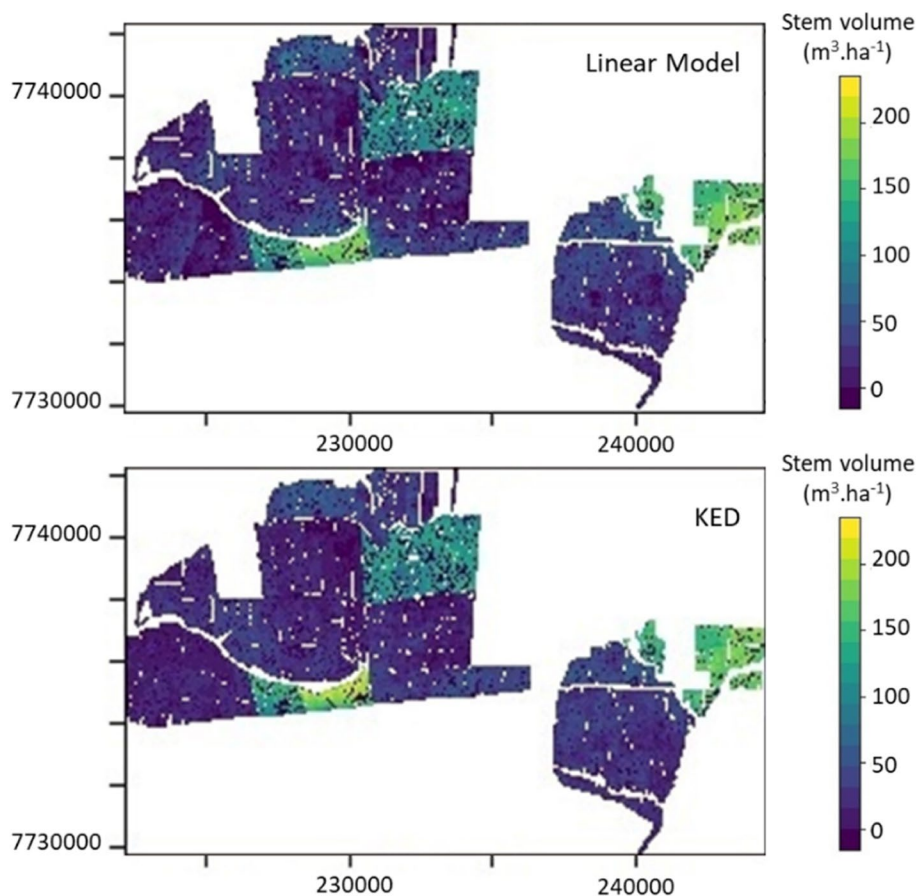
Fig. 6 Example of a map of distribution of the estimated volume (m³ ha-1) obtained from the linear and KED models (see Fig. 3 for GAM), predicted for August 2018. No data values in white corresponds to other land use around the stand, or to gaps within the stand previously masked.

Authors' contributions
LLA: Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing—Original Draft, Visualization. GM: Conceptualization, Methodology, Formal analysis, Resources, Writing—Original Draft, Visualization, Supervision, Project administration, Funding acquisition. CRT: Formal analysis, Writing—Review & Editing. TSM: Resources, Writing—Review & Editing. RCP: Writing—Review & Editing. JRSP: Conceptualization, Methodology, Formal analysis, Resources, Writing—Review & Editing, Supervision, Project administration, Funding acquisition. The authors read and approved the final manuscript.

Availability of data and materials
 The data cannot be made available in a public repository due to proprietary issues. Data sharing agreement can be set up upon request.

Declarations

Ethics approval consent to participate
Not applicable.

Consent for publication
All authors gave their informed consent to this publication and its content.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]School of Agriculture, São Paulo State University (UNESP), Botucatu, SP, Brazil. [2]INRAE, IRD, Institut Agro, CIRAD, Univ Montpellier, Montpellier, France. [3]UMR Eco&Sols, CIRAD, 34398 Montpellier, France. [4]Universidade Federal de São Carlos Campus, Sorocaba, Brazil. [5]Faculty of Technology (FATEC), Botucatu, SP, Brazil. [6]Institute of Biosciences, São Paulo State University (UNESP), Botucatu, SP, Brazil.

Aló *et al. Annals of Forest Science*   (2024) 81:43

Page 15 of 16

References

Ahrens CW, Andrew ME, Mazanec RA, Ruthrof KX, Challis A, Hardy G, Byrne M, Tissue DT, Rymer PD (2020) Plant functional traits differ in adaptability and are predicted to be differentially affected by climate change. Ecol Evol 10:232–248. https://doi.org/10.1002/ece3.5890

Alberdi I, Moreno-Fernández D, Cañellas I, Adame P, Hernández L (2020) Deadwood stocks in south-western European forests: ecological patterns and large scale assessments. Sci Total Environ 747:141237. https://doi.org/10.1016/j.scitotenv.2020.141237

Avitabile V, Baccini A, Friedl MA, Schmullius C (2012) Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda. Remote Sens Environ 117:366–380. https://doi.org/10.1016/j.rse.2011.10.012

Baghdadi N, Le Maire G, Bailly J, Ose K, Nouvellon Y, Zribi M, Lemos C, Hakamada R (2015) Evaluation of ALOS/PALSAR L-band data for the estimation of eucalyptus plantations aboveground biomass in Brazil. Select Topics Appl Earth Observ Remote Sensing, IEEE J 8:3802–3811. https://doi.org/10.1109/jstars.2014.2353661

Boisvenue C, Smiley BP, White JC, Kurz WA, Wulder MA (2016) Integration of Landsat time series and field plots for forest productivity estimates in decision support models. For Ecol Manage 376:284–297. https://doi.org/10.1016/j.foreco.2016.06.022

Castillo-Santiago MÁ, Ghilardi A, Oyama K, Hernández-Stefanoni JL, Torres I, Flamenco-Sandoval A, Fernández A, Mas J-F (2013) Estimating the spatial distribution of woody biomass suitable for charcoal making from remote sensing and geostatistics in central Mexico. Energy Sustain Dev 17:177–188. https://doi.org/10.1016/j.esd.2012.10.007

Cutler MEJ, Boyd DS, Foody GM, Vetrivel A (2012) Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: an assessment of predictions between regions. ISPRS J Photogramm Remote Sens 70:66–77. https://doi.org/10.1016/j.isprsjprs.2012.03.011

Dos Reis AA, Carvalho MC, De Mello JM, Gomide LR, Ferraz Filho AC, Acerbi Junior FW (2018) Spatial prediction of basal area and volume in Eucalyptus stands using Landsat TM data: an assessment of prediction methods. NZ J Forest Sci 48:1. https://doi.org/10.1186/s40490-017-0108-0

Dos Reis AA, Franklin SE, de Mello JM, Acerbi Junior FW (2019) Volume estimation in a Eucalyptus plantation using multi-source remote sensing and digital terrain data: a case study in Minas Gerais State, Brazil. Int J Remote Sens 40:2683–2702. https://doi.org/10.1080/01431161.2018.1530808

Dube T, Mutanga O (2015) Investigating the robustness of the new Landsat-8 Operational Land Imager derived texture metrics in estimating plantation forest aboveground biomass in resource constrained areas. ISPRS J Photogramm Remote Sens 108:12–32. https://doi.org/10.1016/j.isprsjprs.2015.06.002

Dube T, Mutanga O, Abdel-Rahman EM, Ismail R, Slotow R (2015) Predicting Eucalyptus spp. stand volume in Zululand, South Africa: an analysis using a stochastic gradient boosting regression ensemble with multi-source data sets. Int J Remote Sens 36:3751–3772. https://doi.org/10.1080/01431161.2015.1070316

Fassnacht FE, Poblete-Olivares J, Rivero L, Lopatin J, Ceballos-Comisso A, Galleguillos M (2021) Using Sentinel-2 and canopy height models to derive a landscape-level biomass map covering multiple vegetation types. Int J Appl Earth Obs Geoinf 94:102236. https://doi.org/10.1016/j.jag.2020.102236

Galeana-Pizaña JM, López-Caloca A, López-Quiroz P, Silván-Cárdenas JL, Couturier S (2014) Modeling the spatial distribution of above-ground carbon in Mexican coniferous forests using remote sensing and a geostatistical approach. Int J Appl Earth Obs Geoinf 30:179–189. https://doi.org/10.1016/j.jag.2014.02.005

Goovaerts P (1997) Geostatistics for natural resource evaluation. Oxford University Press, USA

Haralick RM, Shanmugam K, Dinstein IH (1973) Textural features for image classification. IEEE Trans Syst Man Cybern SMC 3:610–621. https://doi.org/10.1109/TSMC.1973.4309314

Hastie T, Tibshirani R (1986) Generalized additive models. Stat Sci 1:297–310

Huete AR (1988) A soil-adjusted vegetation index (SAVI). Remote Sens Environ 25:295–309. https://doi.org/10.1016/0034-4257(88)90106-X

IBA (2021) Brazilian tree industry annual report. https://www.ibaorg/datafiles/publicacoes/relatorios/relatorioiba2021-compactadopdf

Kaufman YJ, Tanré D (1996) Strategy for direct and indirect methods for correcting the aerosol effect on remote sensing: from AVHRR to EOS-MODIS. Remote Sens Environ 55:65–79. https://doi.org/10.1016/0034-4257(95)00193-X

Kuhn M (2008) Building predictive models in R using the caret package. J Stat Softw 28:1–26. https://doi.org/10.18637/jss.v028.i05

Le Maire G, Dupuy S, Nouvellon Y, Loos RA, Hakamada R (2014) Mapping short-rotation plantations at regional scale using MODIS time series: Case of eucalypt plantations in Brazil. Remote Sens Environ 152:136–149. https://doi.org/10.1016/j.rse.2014.05.015

Le Maire G, Guillemot J, Campoe OC, Stape J-L, Laclau J-P, Nouvellon Y (2019) Light absorption, light use efficiency and productivity of 16 contrasted genotypes of several Eucalyptus species along a 6-year rotation in Brazil. For Ecol Manage 449:117443. https://doi.org/10.1016/j.foreco.2019.06.040

Le Maire G, Marsden C, Nouvellon Y, Grinand C, Hakamada R, Stape J-L, Laclau J-P (2011) MODIS NDVI time-series allow the monitoring of Eucalyptus plantation biomass. Remote Sens Environ 115:2613–2625

Le Maire G, Marsden C, Verhoef W, Ponzoni FJ, Lo Seen D, Bégué A, Stape J-L, Nouvellon Y (2011) Leaf area index estimation with MODIS reflectance time series and model inversion during full rotations of Eucalyptus plantations. Remote Sens Environ 115:586–599. https://doi.org/10.1016/j.rse.2010.10.004

Lochhead K, LeMay V, Bull G, Schwab O, Halperin J (2018) Multivariate estimation for accurate and logically consistent forest-attributes maps at macroscales. Can J for Res 48:345–359. https://doi.org/10.1139/cjfr-2017-0221

Lu D (2006) The potential and challenge of remote sensing-based biomass estimation. Int J Remote Sens 27:1297–1328. https://doi.org/10.1080/01431160500486732

Maack J, Lingenfelder M, Weinacker H, Koch B (2016) Modelling the standing timber volume of Baden-Württemberg—a large-scale approach using a fusion of landsat, airborne LiDAR and National Forest Inventory data. Int J Appl Earth Obs Geoinf 49:107–116. https://doi.org/10.1016/j.jag.2016.02.004

Marsden C, Le Maire G, Stape JL, Lo Seen D, Roupsard O, Cabral O, Epron D, Nascimento Lima AM, Nouvellon Y (2010) Relating MODIS vegetation index time-series with structure, light absorption and stem production of fast-growing Eucalyptus plantations. For Ecol Manage 259:1741–1753. https://doi.org/10.1016/j.foreco.2009.07.039

Nussbaum M, Spiess K, Baltensweiler A, Grob U, Keller A, Greiner L, Schaepman ME, Papritz A (2018) Evaluation of digital soil mapping approaches with large sets of environmental covariates. Soil 4:1–22. https://doi.org/10.5194/soil-4-1-2018

Propastin P (2012) Modifying geographically weighted regression for estimating aboveground biomass in tropical rainforests by multispectral remote sensing data. Int J Appl Earth Obs Geoinf 18:82–90. https://doi.org/10.1016/j.jag.2011.12.013

R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Räty M, Kangas AS (2019) Effect of permanent plots on the relative efficiency of spatially balanced sampling in a national forest inventory. Ann for Sci 76:20. https://doi.org/10.1007/s13595-019-0802-6

Rodriguez-Galiano VF, Chica-Olmo M, Abarca-Hernandez F, Atkinson PM, Jeganathan C (2012) Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. Remote Sens Environ 121:93–107. https://doi.org/10.1016/j.rse.2011.12.003

Roy DP, Kovalskyy V, Zhang HK, Vermote EF, Yan L, Kumar SS, Egorov A (2016) Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. Remote Sens Environ 185:57–70. https://doi.org/10.1016/j.rse.2015.12.024

Saarela S, Holm S, Grafström A, Schnell S, Næsset E, Gregoire TG, Nelson RF, Ståhl G (2016) Hierarchical model-based inference for forest inventory utilizing three sources of information. Ann for Sci 73:895–910

Saarela S, Wästlund A, Holmström E, Mensah AA, Holm S, Nilsson M, Fridman J, Ståhl G (2020) Mapping aboveground biomass and its prediction uncertainty using LiDAR and field data, accounting for tree-level allometric and LiDAR model errors. Forest Ecosystems 7:43. https://doi.org/10.1186/s40663-020-00245-0

Sarker LR, Nichol JE (2011) Improved forest biomass estimates using ALOS AVNIR-2 texture indices. Remote Sens Environ 115:968–977

Scolforo HF, McTague JP, Burkhart H, Roise J, Carneiro RL, Stape JL (2019) Generalized stem taper and tree volume equations applied to eucalyptus

of varying genetics in Brazil. Can J for Res 49:447–462. https://doi.org/10.1139/cjfr-2018-0276

Scolforo HF, Scolforo JRS, de Mello JM, de Mello CR, Morais VA (2016) Spatial interpolators for improving the mapping of carbon stock of the arboreal vegetation in Brazilian biomes of Atlantic forest and Savanna. For Ecol Manage 376:24–35. https://doi.org/10.1016/j.foreco.2016.05.047

Silveira EMO, Espírito Santo FD, Wulder MA, Acerbi Júnior FW, Carvalho MC, Mello CR, Mello JM, Shimabukuro YE, Terra MCNS, Carvalho LMT, Scolforo JRS (2019) Pre-stratified modelling plus residuals kriging reduces the uncertainty of aboveground biomass estimation and spatial distribution in heterogeneous savannas and forest environments. For Ecol Manage 445:96–109. https://doi.org/10.1016/j.foreco.2019.05.016

Souza GSAD, Soares VP, Leite HG, Gleriani JM, Do Amaral CH, Ferraz AS, Silveira MVDF, Santos JFCD, Velloso SGS, Domingues GF, Silva S (2019) Multi-sensor prediction of Eucalyptus stand volume: a support vector approach. ISPRS J Photogramm Remote Sens 156:135–146. https://doi.org/10.1016/j.isprsjprs.2019.08.002

Ståhl G, Saarela S, Schnell S, Holm S, Breidenbach J, Healey SP, Patterson PL, Magnussen S, Næsset E, McRoberts RE, Gregoire TG (2016) Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation. Forest Ecosystems 3:5. https://doi.org/10.1186/s40663-016-0064-9

Stape J, Alcarde Alvares C (2023) Physiographic position drives Eucalyptus productivity in Mato Grosso do Sul, Brazil. Série Técnica IPEF 26:289–293. https://doi.org/10.18671/sertec.v26n48.057

Stein ML (1999) Interpolation of spatial data: some theory for kriging. Springer, New York

Tesfamichael SG, Ahmed F, Van Aardt JAN, Blakeway F (2009) A semi-variogram approach for estimating stems per hectare in Eucalyptus grandis plantations using discrete-return lidar height data. For Ecol Manage 258:1188–1199. https://doi.org/10.1016/j.foreco.2009.06.016

Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens Environ 8:127–150

Viana H, Aranha J, Lopes D, Cohen WB (2012) Estimation of crown biomass of Pinus pinaster stands and shrubland above-ground biomass using forest inventory data, remotely sensed imagery and spatial prediction models. Ecol Model 226:22–35. https://doi.org/10.1016/j.ecolmodel.2011.11.027

Wackernagel H (2003) Multivariate geostatistics: an introduction with applications. Springer, Berlin Heidelberg

Wood SN (2017) Generalized Additive Models: An Introduction with R, Second Edition (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

Zhang G, Ganguly S, Nemani RR, White MA, Milesi C, Hashimoto H, Wang W, Saatchi S, Yu Y, Myneni RB (2014) Estimation of forest aboveground biomass in California using canopy height and leaf area index estimated from satellite data. Remote Sens Environ 151:44–56. https://doi.org/10.1016/j.rse.2014.01.025

Zhao P, Lu D, Wang G, Liu L, Li D, Zhu J, Yu S (2016) Forest aboveground biomass estimation in Zhejiang Province using the integration of Landsat TM and ALOS PALSAR data. Int J Appl Earth Obs Geoinf 53:1–15. https://doi.org/10.1016/j.jag.2016.08.007

Zhou J, Proisy C, Descombes X, Le Maire G, Nouvellon Y, Stape J-L, Viennois G, Zerubia J, Couteron P (2013) Mapping local density of young Eucalyptus plantations by individual tree detection in high spatial resolution satellite images. For Ecol Manage 301:129–141. https://doi.org/10.1016/j.foreco.2012.10.007

Publisher's Note