RESEARCH ARTICLE   **Open Access**

CrossMark

Spatial prediction of basal area and volume in *Eucalyptus* stands using Landsat TM data: an assessment of prediction methods

Aliny Aparecida dos Reis[1*], Mônica Canaan Carvalho[1], José Marcio de Mello[1], Lucas Rezende Gomide[1], Antônio Carlos Ferraz Filho[2] and Fausto Weimar Acerbi Junior[1]

Abstract

Background: In fast-growing forests such as *Eucalyptus* plantations, the correct determination of stand productivity is essential to aid decision making processes and ensure the efficiency of the wood supply chain. In the past decade, advances in remote sensing and computational methods have yielded new tools, techniques, and technologies that have led to improvements in forest management and forest productivity assessments. Our aim was to estimate and map the basal area and volume of *Eucalyptus* stands through the integration of forest inventory, remote sensing, parametric, and nonparametric methods of spatial prediction.

Methods: This study was conducted in 20 5-year-old clonal stands (362 ha) of *Eucalyptus urophylla* S.T.Blake x *Eucalyptus camaldulensis* Dehnh. The stands are located in the northwest region of Minas Gerais state, Brazil. Basal area and volume data were obtained from forest inventory operations carried out in the field. Spectral data were collected from a Landsat 5 TM satellite image, composed of spectral bands and vegetation indices. Multiple linear regression (MLR), random forest (RF), support vector machine (SVM), and artificial neural network (ANN) methods were used for basal area and volume estimation. Using ordinary kriging, we spatialised the residuals generated by the spatial prediction methods for the correction of trends in the estimates and more detailing of the spatial behaviour of basal area and volume.

Results: The ND54 index was the spectral variable that had the best correlation values with basal area ($r = -0.91$) and volume ($r = -0.52$) and was also the variable that most contributed to basal area and volume estimates by the MLR and RF methods. The RF algorithm presented smaller basal area and volume errors when compared to other machine learning algorithms and MLR. The addition of residual kriging in spatial prediction methods did not necessarily result in relative improvements in the estimations of these methods.

Conclusions: Random forest was the best method of spatial prediction and mapping of basal area and volume in the study area. The combination of spatial prediction methods with residual kriging did not result in relative improvement of spatial prediction accuracy of basal area and volume in all methods assessed in this study, and there is not always a spatial dependency structure in the residuals of a spatial prediction method. The approaches used in this study provide a framework for integrating field and multispectral data, highlighting methods that greatly improve spatial prediction of basal area and volume estimation in *Eucalyptus* stands. This has potential to support fast growth plantation monitoring, offering options for a robust analysis of high-dimensional data.

Keywords: Forest inventory, Machine learning algorithms, Multiple linear regression, Random forest, Support vector machine, Artificial neural networks

---

* Correspondence: alinyreis@hotmail.com
[1]Department of Forest Science, Federal University of Lavras – UFLA, PO Box 3037, Lavras, Minas Gerais 37200-000, Brazil
Full list of author information is available at the end of the article

dos Reis *et al. New Zealand Journal of Forestry Science* (2018) 48:1

Page 2 of 17

Background

The Brazilian forestry sector represents an important share of the products, taxes, jobs, and income generation of the country and accounts for 3.5% of the national GDP (IBÁ 2015). This is in large part due to the successful establishment of fast-grown plantations of *Eucalyptus* species, which currently occupy around 5.6 million hectares (71.9% of the total planted forest area in Brazil) and represent 17% of the harvested wood in the world (IBÁ 2014, 2015).

The *Eucalyptus* genus has more than 500 species, and a subset of these are used in fast-growing plantations (Barrios et al. 2015), commonly located in tropical and sub-tropical regions, and more recently in temperate regions. Spain (González-García et al. 2015), Portugal (Lopes et al. 2009), Uruguay (Barrios et al. 2015), Chile (Watt et al. 2014), South Africa (Dye et al. 2004), Australia (Verma et al. 2014), and the USA (Wear et al. 2015) are some examples of productive *Eucalyptus* plantations in temperate regions that have cutting cycles ranging from 8 to 12 years. In tropical regions such as Brazil, the cutting cycles of *Eucalyptus* plantations range from 5 to 7 years (Guedes et al. 2015, Scolforo et al. 2016).

Timber production is the main ecosystem service of planted forests and the main management objective for these plantations (Gao et al. 2016). In the case of fast-growing plantations, the correct determination of stand productivity is essential to support forest management planning strategies (González-García et al. 2015, Retslaff et al. 2015). Traditionally, productivity assessments of a plantation are carried out based on field measurements of the diameter at breast height (DBH) and tree height via forest inventory. However, in fast-growing plantations, field-based inventory programmes may not be sufficient to capture productivity differences across the entire area, such as those arising from losses due to pest and disease attacks (Coops et al. 2006), or from climatic anomalies (González-García et al. 2015, Scolforo et al. 2016).

In the past decade, advances in geographical information systems (GIS), global positioning systems (GPS), and remote sensing have provided new tools, techniques, and technologies to support forest management. Thus, low-cost and accurate forest productivity assessment can be made, as well as allowing the collection of information in areas not sampled by forest inventory (Morgenroth and Visser 2013). The analysis of remote sensing information combined with field data has been used by several authors to fill the information gap left by data collected only in the field (Watt et al. 2016, Boisvenue et al. 2016, Moreno et al. 2016, Fayad et al. 2016, Vicharnakorn et al. 2014). Ponzoni et al. (2015) used data collected from Landsat 5 thematic mapper (TM) images for spectral-temporal characterisation of *Eucalyptus* canopies. Berra et al. (2012) estimated the volume of a *Eucalyptus* plantation in the southern region of Brazil from Landsat 5 TM images. Canavesi et al. (2010) used hyperspectral data from the Hyperion EO-1 sensor for the volume estimation of *Eucalyptus* plantations under different relief conditions. The results found by these authors corroborate the potential use of data collected by remote sensing to estimate the productivity of *Eucalyptus* plantations.

In parallel to the advances in remote sensing, computational techniques, such as machine learning algorithms (MLA), have been increasingly used to model spectral and biological data. These techniques overcome the difficulties of classical statistical methods such as spatial correlation, non-linearity of data, and overfitting (Were et al. 2015). In addition, these algorithms allow the use of categorical data, with statistical noise and incomplete data, and therefore are able to address needs under different dataset scenarios (Breiman 2001).

Several studies have shown the superiority of machine learning algorithms in relation to classical statistics in several areas, such as in forest management. For instance, Ahmed et al. (2015) modelled a Landsat time-series data structure in conjunction with LiDAR data and found that the random forest algorithm achieved better results than multiple regression for all forest classes. In another study, García-Gutiérrez et al. (2015) found that machine learning algorithms (mainly support vector machine) were superior for modelling a range of forest variables (viz., aboveground biomass, basal area, dominant height, mean height, and volume) compared with multiple linear regression. Machine learning algorithms have also been shown to provide an economical and accurate way to estimate aboveground biomass in forests from Landsat satellite images (Wu et al. 2016). These studies highlight the benefits of applying more robust techniques in solving problems previously resolved by traditional statistical modelling.

In this context, the aims of this study were: (i) to estimate and map basal area and volume of a *Eucalyptus* plantation through the integration of forest inventory, remote sensing, and parametric and nonparametric methods of spatial prediction; (ii) to compare the performance of machine learning algorithms (random forest, support vector machine, and artificial neural networks) with the linear regression model; and (iii) to assess the improvement in basal area and volume estimation with the addition of residual kriging in spatial prediction methods.

Methods

Study area

The study area is located in Minas Gerais state, the fourth largest state in Brazil, with an area of 586,521 km^2. Minas Gerais state has the largest area occupied by plantations of the *Eucalyptus* genus in the country (1,400,232 ha),

dos Reis *et al. New Zealand Journal of Forestry Science* (2018) 48:1

Page 3 of 17

corresponding to 25.2% of Brazilian *Eucalyptus* plantations. The wood from these plantations is mainly used for the production of charcoal, as well as pulp, lumber, and panels (IBÁ 2015).

The *Eucalyptus* clonal stands under study are located in Lagoa Grande municipality, in the northwest of Minas Gerais state (lat. 17° 43′ 00″ S–17° 44′ 00″ S, long. 46° 32′ 00″ W–46° 33′ 00″ W, elevation 560 m a.s.l.) (Fig. 1). According to the Köppen climatic classification system, the climate in this region is Aw, classified as a tropical savanna climate, with drier months during the winter, high annual precipitation in the summer and average temperature of all months greater than 18 °C (Alvares et al. 2013). The average annual rainfall and the average monthly rainfall of the dry and wet seasons are 1430, 8, and 257 mm, respectively.

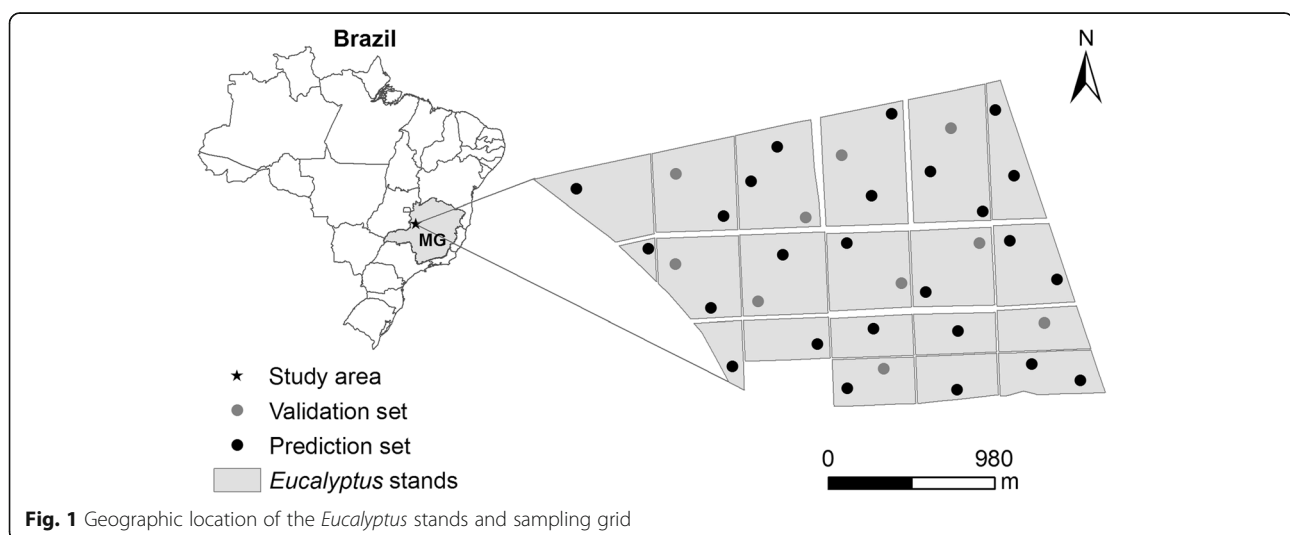### Field data description and sampling

This study was undertaken in a set of 20 clonal stands of *Eucalyptus urophylla* S.T.Blake x *Eucalyptus camaldulensis* Dehnh, totalling an area of 362.2 ha. These stands were planted in April and May 2004, with initial spacing of either 3×2 m or 3×3 m. The forest inventory was carried out in June and July 2009 on a set of 35 georeferenced square plots of 400 m^2. The plots were georeferenced in the field with GPS (Garmin 60CSx, Garmin Ltd., Olathe, Kansas, USA). The sampling procedure adopted was systematic, allocating approximately one plot per 10 ha of forest. In each plot, the diameter at breast height (DBH) of all stems was measured, as well as the total height of the first 15 trees with normal stems (without bifurcation or any other defect) and height of dominant trees (the 100 largest diameter trees per hectare). Descriptive statistics of the variables collected in the field are shown in Table 1. Estimates of basal area

($m^2\ ha^{-1}$), and total stem volume ($m^3\ ha^{-1}$) were obtained from the information collected in the plots.

Remote sensing data and processing

Spectral data were obtained from a Landsat 5 TM satellite image, with spatial resolution of 30 m, on the date of June 25, 2009, corresponding with field data collection, in orbit 220, point 072, in bands TM1 (0.45–0.52 μm), TM2 (0.52–0.60 μm), TM3 (0.63–0.69 μm), TM4 (0.76–0.90 μm), TM5 (1.55–1.75 μm), and TM7 (2.18–2.35 μm). The Landsat 5 TM Surface Reflectance Climate Data Record (CDR) was used, which is a Landsat Level-2A product generated by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Masek et al. 2006) obtained from the USGS (United States Geological Survey) database (USGS 2017). These images already contain radiometric calibration, and geometric and atmospheric corrections.

In addition, vegetation indices using the red, near infrared and short wave infrared spectral bands of Landsat 5 TM (Table 2) were calculated, as described by Lu et al. (2004) and Ponzoni et al. (2012). The normalised difference vegetation index (NDVI) is the most widely used vegetation index for retrieval of forest biophysical parameters (Rouse et al. 1973, Lu et al. 2004). The soil-adjusted vegetation index (SAVI) and modified soil-adjusted vegetation index (MSAVI) are soil adjusted vegetation indices used to reduce the effect of soil background reflectance (Qi et al. 1994). The enhanced vegetation index (EVI) was developed to optimise the vegetation signal, correcting reflected light distortions caused by particulate matter suspended in the air, as well as by influence of background data under the vegetation canopy (Justice et al. 1998). The global environment monitoring index (GEMI) minimises atmospheric effects, similar to the EVI and minimises

Fig. 1 Geographic location of the *Eucalyptus* stands and sampling grid

Table 1 Descriptive statistics of the variables collected in the field

Statistic	DBH	H	Hd
Minimum	11.98	16.98	19.40
Maximum	15.45	24.63	26.38
Mean	14.02	21.18	22.98
Standard deviation	0.85	2.33	1.92

DBH diameter at breast height (cm), *H* total height (m), *Hd* dominant height (m)

observational angular effects in the observed vegetation index signal (Pinty and Verstraete 1992).

Dataset integration

The choice of an appropriate pixel size is one of the issues to be considered when using remote sensing data to estimate dendrometric characteristics. Due to easy accessibility and affordability, a number of studies have employed Landsat images and found statistically significant correlations between remotely sensed data and dendrometric characteristics using ground plots ranging from 315 to 2500 m² (Dube and Mutanga 2015, López-Sánchez et al. 2014, Zhang et al. 2014, López-Serrano et al. 2016).

Although the size of a single plot (20 × 20 m) in this study does not cover a Landsat pixel, we considered that a plot represents an area larger than its size. As the sampling design was one plot per hectare, we ensured that each plot matched with the reference pixel in order to extract reliable data.

Spatial modelling and prediction methods
Exploratory data analysis

Spectral response was extracted from the Landsat TM bands and vegetation indices from the geographical coordinates of the forest inventory plots. Thus, the plot database was composed of basal area (m² ha⁻¹), volume (m³ ha⁻¹), spectral band values, and vegetation index values. The total database (35 plots) was systematically

divided into two datasets: prediction or fitting set (70% of the database) and validation set (30% of the database). Therefore, 25 plots were used for basal area and volume predictions, and 10 plots were used for validation of the different approaches to estimate basal area and volume in the *Eucalyptus* stands under study.

Pearson correlation analysis was carried out among basal area, volume, values of spectral bands, and vegetation indices. From these correlations, the relationship between the dendrometric characteristics of *Eucalyptus* stands and its spectral response in Landsat images was explored.

Multiple linear regression (MLR) analysis

Basal area and volume estimation were accomplished through MLR analysis. A stepwise variable elimination method was used in conjunction with the Akaike information criterion (AIC) to select only those spectral variables that "best" explained basal area and volume variation. The residuals from regression models were analysed to assess the existence of trends in the errors. The variance inflation factor (VIF) was used to detect possible correlations between explanatory variables (multicollinearity). The adopted VIF cutoff value was 10.

Random forest (RF)

The RF algorithm, initially proposed by Breiman (2001), is an ensemble method that generates a set of individually trained decision trees and combines their results. The greatest advantage of these decision trees as regression methods is that they are able to accurately describe complex relationships among multiple variables, and by aggregating these decision trees, more accurate solutions are generated (Gleason and Im 2012). In addition to these characteristics, RF is an easy parameterisation method (Immitzer et al. 2012). This method has shown great potential in regression studies with integration of spectral data, in some cases generating better results

Table 2 Vegetation indices used in the spectral characterisation of the *Eucalyptus* stands

Vegetation indices	Formulation	Reference
NDVI	$(TM4 - TM3)/(TM4 + TM3)$	Rouse et al. (1973)
ND53	$(TM5 - TM3)/(TM5 + TM3)$	Huete et al. (2002)
ND54	$(TM5 - TM4)/(TM5 + TM4)$	Huete et al. (2002)
ND57	$(TM5 - TM7)/(TM5 + TM7)$	Huete et al. (2002)
SAVI	$[(TM4 - TM3)/(TM4 + TM3 + 0.5)].(1.5)$	Huete (1988)
MSAVI	$\left[(2TM4 + 1) - \sqrt{(2TM4 + 1)^2 - 8(TM4 - 2TM3)} \right]/2$	Qi et al. (1994)
EVI	$2.5 \times [(TM4 - TM3)/(TM4 + 6TM3 - 7.5TM1 + 1)]$	Justice et al. (1998)
GEMI	$n(1 - 0.25n). [(TM3 - 0.125)/(1 - TM3)]$ $$n = \frac{2(TM4^2 - TM3^2) + 1.5TM4 + 0.5TM3}{TM4 + TM3 + 0.5}$$	Pinty and Verstraete (1992)

TM thematic mapper, *ND* normalised difference, *NDVI* normalised difference vegetation index, *SAVI* soil-adjusted vegetation index, *MSAVI* modified soil-adjusted vegetation index, *EVI* enhanced vegetation index, *GEMI* global environment monitoring index

dos Reis *et al. New Zealand Journal of Forestry Science* (2018) 48:1

Page 5 of 17

than conventional techniques (Stojanova et al. 2010, Dube et al. 2014, García-Gutiérrez et al. 2015, Görgens et al. 2015, Wu et al. 2016). The RF algorithm fitted in this study is implemented in the open-source software WEKA 3.8 (Frank et al. 2016). Tests were carried out with the exchange of tree numbers and attribute numbers to be drawn. Then, 20 trees with 10 attributes to be drawn by node for basal area and 80 trees and 11 attributes for volume were fixed.

Support vector machine (SVM)

SVMs operate by assuming that each set of inputs will have a unique relation to the response variable and that the grouping and the relation of these predictors to one another is sufficient to identify rules that can be used to predict the response variable from new input sets. For this, SVMs project the input space data into a feature space with a much larger dimension, enabling linearly non-separable data to become separable in the feature space. This method has been successfully used in forestry classification problems (Huang et al. 2008, Shao and Lunetta 2012) and more recently in regression problems with the use of spectral data (García-Gutiérrez et al. 2015, Wu et al. 2016). The Kernel function used in the present study was the Gaussian or radial basis function (RBF). The algorithm used is implemented in WEKA 3.8 software under the sequential minimal optimization (SMO) function. Values of parameters C and σ (bandwidth or influence range of each training point in the RBF) were tested within the interval $(10^i)_{i = -3, -2, -1, 0, 1, 2, 3}$, where the least squared mean error configuration was chosen for application. For basal area and volume, selected C and σ values were 10 and 0.1, and 100 and 0.01, respectively.

Artificial neural networks (ANNs)

ANNs are a parallel-distributed information processing system that simulates the working of neurons in the human brain, being able to learn from examples. Artificial neural networks are widely used to model complex and non-linear relations between inputs and outputs or to determine patterns in data (Diamantopoulou 2012). The use of this technique in conjunction with remote sensing data is consolidated in several studies (Cluter et al. 2012, García-Gutiérrez et al. 2015, Rodriguez-Galiano et al. 2015, Were et al. 2015). We used the ANN obtained by running the Multilayer Perceptron function (of the multilayer perceptron type) provided by WEKA 3.8 software. The training of neural networks occurred through the back-propagation algorithm, which fit the weights of all the layers of the network from the backpropagation of the error, obtained in the output layer. The weights updating was carried out according to the error, learning rate, and momentum terms (Delta rule). The sigmoidal

activation function was employed in all neurons. Determined by previous tests, ANNs were structured with 14 neurons in the input layer (number of variables), 1 neuron in the hidden layer, and 1 neuron in the output layer, corresponding to estimated basal area or volume. The learning rate, the momentum term, and iteration numbers were fixed at 0.3, 0.5, and 500 for basal area, and 0.2, 0.7, and 500 for volume, respectively.

Relative importance evaluation

The variable importance was assessed for each model with a removal-based approach in order to avoid the limited interpretability of the MLA and to verify how each independent variable contribute to the performance of machine learning algorithms (RF, SVM, and ANN). All algorithms were adjusted n times, with n being the number of available variables. At each time, one variable was removed from the training set and then the root mean square error (RMSE) of the algorithm was quantified. At the end, the obtained errors were normalised by the ratio of the largest RMSE so that they were between 0 and 1 and multiplied by 100 (Were et al. 2015). The variable that results in the highest RMSE when removed from the database is the variable with the highest relative importance within the model. This methodology was chosen because it can be consistently applied to all algorithms, allowing comparisons of variable contribution between the methods.

Geostatistical modelling of prediction methods errors

Spatial prediction methods capture the average behaviour of the main variable, allowing the identification of its general spatial behaviour, without detailing more specific areas or regions. For details of specific regions, estimates obtained exclusively from the auxiliary variables need to be corrected. Thus, residuals generated by spatial prediction methods (MLR, RF, SVM, and ANN) were used for the correction of trends in the estimates and for detailing the spatial behaviour of the main variables (basal area and volume) using ordinary kriging. The interpolated values of the residuals were then added to the estimates of the spatial prediction methods (MLR, RF, SVM, and ANN). Thus, we obtained the basal area and volume estimates corrected by the ordinary kriging of the residuals for each spatial prediction method.

For the application of ordinary kriging to the spatial prediction method residuals, we considered the stationarity presupposition of the intrinsic hypothesis (Journel and Huijbregts 1978), through fitting of theoretical functions to experimental semivariogram models. Spherical, exponential, and Gaussian models were fitted to the semivariogram of the residuals from each spatial prediction method using weighted least squares. The semivariogram parameters (nugget (τ^2), sill (σ^2), and range (ϕ))

dos Reis *et al. New Zealand Journal of Forestry Science* (2018) 48:1

Page 6 of 17

were calculated from the best fitted models, which provided information about the spatial structure as well as input parameters for the kriging interpolation. The nugget represents the minimum semivariance among different sampling intervals. Nugget values greater than zero represent a combination of experimental error and of unresolved spatial variability occurring at scales smaller than inter-sampling lag distance. Sill is the plateau reached by the values of semivariance and indicates the amount of variation than can be explained by the spatial structure of the data. Range is the distance at which the semivariogram reaches the plateau, indicating the distance which values are spatially correlated. The evaluation of the performance of each semivariogram model and the selection of the best models were based on cross-validation, which estimates the reduced average error (RAE) and the standard deviation of the reduced average error (SRE) (Yamamoto and Landim 2013).

Validation and assessment of the prediction methods

The different approaches to basal area and volume estimation of *Eucalyptus* stands were evaluated by comparing the basic statistics of the predicted maps (mean and standard deviation) with the estimates obtained from the forest inventory, and through the discrepancies between observed and predicted values in the fitting and validation datasets. These discrepancies were evaluated using the mean error (ME), the mean absolute error (MAE), and the root mean square error (RMSE), as described in Eqs. 1–3.

$$ME = \frac{1}{N} \sum_{i=1}^{N} (X_i - \hat{X}_i) \tag{1}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |X_i - \hat{X}_i| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (X_i - \hat{X}_i)^2} \tag{3}$$

where N is the number of values in the dataset; $\hat{X}_i$ is the estimated value of the main variable; X_i is the observed value in the prediction and validation sets.

The relative improvement (RI) achieved by residual kriging for a particular spatial prediction method was calculated by comparing the change in RMSE when the residual kriging was applied using Eq. 4.

$$RI = \frac{RMSE_{spm} - RMSE_{spm\text{-}RK}}{RMSE_{spm}} \times 100\% \tag{4}$$

where $RMSE_{spm}$ is the root mean square error of a spatial prediction method, $RMSE_{spm\text{-}RK}$ is the root mean square error of the spatial prediction method when residual kriging is added to this method.

Data analysis for this study was performed using the following software: R (R Core Team 2016) with the geoR package (Ribeiro Júnior and Diggle 2001), WEKA 3.8 (Frank et al. 2016), and ArcGis version 10.1 (Esri 2010) with Geostatistical Analyst extension (Esri 2010).

Results

Descriptive statistic of the measured basal area and volume

Basal area ranged from 10.07 to 21.63 m^2 ha^{-1}, with average of 16.86 m^2 ha^{-1} and standard deviation of 2.4 m^2 ha^{-1} (Table 3). The average volume was 169.34 m^3 ha^{-1} with a standard deviation of 29.66 m^3 ha^{-1} and range from 95.80 up to 213.85 m^3 ha^{-1}. Basal area had a lower coefficient of variation (CV = 14.26%) compared to volume (CV = 17.51%), demonstrating a considerable homogeneity of this dendrometric characteristic in the evaluated *Eucalyptus* stands.

Correlation among basal area, volume, spectral bands, and vegetation indices

The correlation between plot basal area and the different spectral bands and their ratios (Table 4) ranged from − 0.91 (ND54) to 0.15 (TM2). The SAVI, MSAVI, GEMI, and EVI were also highly correlated with basal area ($r > 0.85$). The correlation between plot volume and the spectral bands and ratios ranged from − 0.52 (ND54) to − 0.02 (TM2). The NDVI ($r = 0.49$) and SAVI ($r = 0.47$) also had high correlations with volume, but these were lower in magnitude when compared with those for basal area. Many of the spectral bands and ratios were also highly correlated with each other ($r > 0.90$), which can be considered a drawback due possible to multicollinearity problems in linear regression models.

Table 3 Descriptive statistics obtained from forest inventory processing using the estimators of simple random sampling (SRS)

Estimators	Basal area		Volume	
	(m^2)[a]	(m^2 ha^{-1})	(m^3)[a]	(m^3 ha^{-1})
Minimum	0.91	10.07	8.62	95.80
Maximum	1.95	21.63	19.25	213.85
Mean	1.52	16.86	15.24	169.34
Standard deviation	0.22	2.4	2.67	29.66
Coefficient of variation (%)	14.26		17.51	
Sampling error (%)	4.89		6.00	
Total confidence interval	5807.9–6405.0		57,652.7–65,018.7	

[a]Estimates obtained for an area of 900 m^2 (corresponding to the area of each pixel of the Landsat images)

dos Reis *et al. New Zealand Journal of Forestry Science* (2018) 48:1

Page 7 of 17

Table 4 Pearson's correlation coefficient (*r*) among basal area, volume, and spectral data for the *Eucalyptus* stands

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. G	1.00															
2. V	0.70*	1.00														
3. TM1	− 0.24^{ns}	0.10^{ns}	1.00													
4. TM2	0.15^{ns}	− 0.02^{ns}	0.59*	1.00												
5. TM3	− 0.20^{ns}	− 0.10^{ns}	0.80*	0.72*	1.00											
6. TM4	0.82*	0.41*	− 0.05^{ns}	0.43*	0.12^{ns}	1.00										
7. TM5	− 0.66*	− 0.36^{ns}	0.53*	0.31^{ns}	0.68*	− 0.40*	1.00									
8. TM7	− 0.68*	− 0.40*	0.56*	0.32^{ns}	0.66*	− 0.42*	0.90*	1.00								
9. NDVI	0.83*	0.49*	− 0.53*	− 0.13^{ns}	− 0.55*	0.75*	− 0.78*	− 0.82*	1.00							
10. ND53	− 0.60*	− 0.32^{ns}	− 0.29^{ns}	− 0.50*	− 0.37^{ns}	− 0.66*	0.43*	0.31^{ns}	− 0.31^{ns}	1.00						
11. ND54	− 0.91*	− 0.52*	0.31^{ns}	− 0.09^{ns}	0.30^{ns}	− 0.86*	0.80*	0.78*	− 0.93*	0.65*	1.00					
12. ND57	0.45*	0.27^{ns}	− 0.49*	− 0.28^{ns}	− 0.49*	0.27^{ns}	− 0.50*	− 0.82*	0.60*	0.00^{ns}	− 0.48*	1.00				
13. SAVI	0.88*	0.47*	− 0.23^{ns}	0.25^{ns}	− 0.12^{ns}	0.97*	− 0.57*	− 0.60*	0.89*	− 0.57*	− 0.94*	0.41*	1.00			
14. MSAVI	0.88*	0.45*	− 0.36^{ns}	0.13^{ns}	− 0.27^{ns}	0.92	− 0.65*	− 0.67*	0.94*	− 0.50*	− 0.95*	0.46*	0.99*	1.00		
15. GEMI	0.86*	0.45*	− 0.14^{ns}	0.34^{ns}	0.00^{ns}	0.99*	− 0.49*	− 0.52*	0.83*	− 0.62*	− 0.91*	0.35^{ns}	0.99*	0.96*	1.00	
16. EVI	0.87*	0.42*	− 0.41*	0.12^{ns}	− 0.28^{ns}	0.92*	− 0.64*	− 0.67*	0.94*	− 0.48*	− 0.94*	0.47*	0.98*	1.00*	0.96*	1.00

V volume (m^3 ha^{-1}), *G* basal area (m^2 ha^{-1}), *TM* thematic mapper, *ND* normalised difference, *NDVI* normalised difference vegetation index, *SAVI* soil-adjusted vegetation index, *MSAVI* modified soil-adjusted vegetation index, *GEMI* global environment monitoring index, *EVI* enhanced vegetation index
^{ns}Not significant at 5%; *significant at 5%

Spatial prediction of basal area and volume by MLR, RF, SVM, and ANN

The spectral data examined had several significant correlations with the basal area and volume data (Table 4). However, they contributed in a reduced form to the regression models due to multicollinearity problems, which resulted in final regression models with few significant explanatory variables (Table 5). The basal area model only included the ND54 vegetation index (Table 5), while the volume model included the TM1 band and NDVI. The coefficient of determination was high for the basal area model ($R^2 = 0.81$), but was much lower for the volume model ($R^2 = 0.37$).

In the case of basal area and volume predictions using machine learning algorithms, the increases in RMSEs when the predictors were excluded one by one from the SVM, ANN, and RF models are shown in Fig. 2. The variable ranking by relative importance differed for each algorithm. The ND54 index, chosen for basal area model by the MLR, also had the greatest effect on the accuracy of the RF model, both for basal area and volume. The TM2 band had the highest relative importance for the ANN and SVM models of both basal area and volume. The TM1 band, selected by the MLR for volume estimation, also had high importance in the ANN and SVM models of volume.

Comparisons of measured values and estimated values of basal area (Fig. 3) showed that basal area was underestimated by the ANN model (Fig. 3d). The model fitted using the RF algorithm produced values of basal area that were in closer agreement with measured values (Fig. 3b). Similar results were seen for the volume models, but with a slight overestimation for the plots with small volumes and an underestimation of the plots with high volumes. The model fitted using ANN algorithm did not produce estimates of volume that were consistent with measured values (Fig. 3h). The models fitted using the MLR and SVM (Fig. 3e, g) algorithms produced predicted values that were more closely related to the measured values than those from the ANN algorithm.

Prediction and validation sets of basal area and volume were compared by means of Student's *t* test, in order to check if they provided unbiased subsets of the original

Table 5 Regression model fitted for basal area and volume estimation for the *Eucalyptus* stands

Model	β_0	β_1	β_2	R^2_{aj}	S_{xy}	S_{xy} (%)
G = β_0 + β_1ND54	0.78***	− 1.80***	–	0.81	0.09	5.76
V = β_0 + β_1NDVI + β_2TM1	− 24.11*	42.69***	241.61*	0.37	2.01	13.08

G basal area (m^2), *V* volume (m^3), β_0, β_1, and β_2 coefficients, R^2_{aj} adjusted coefficient of determination, S_{xy} residual standard error, *TM* thematic mapper, *ND* normalised difference, *NDVI* normalised difference vegetation index
***Significant at 1%; *significant at 10%
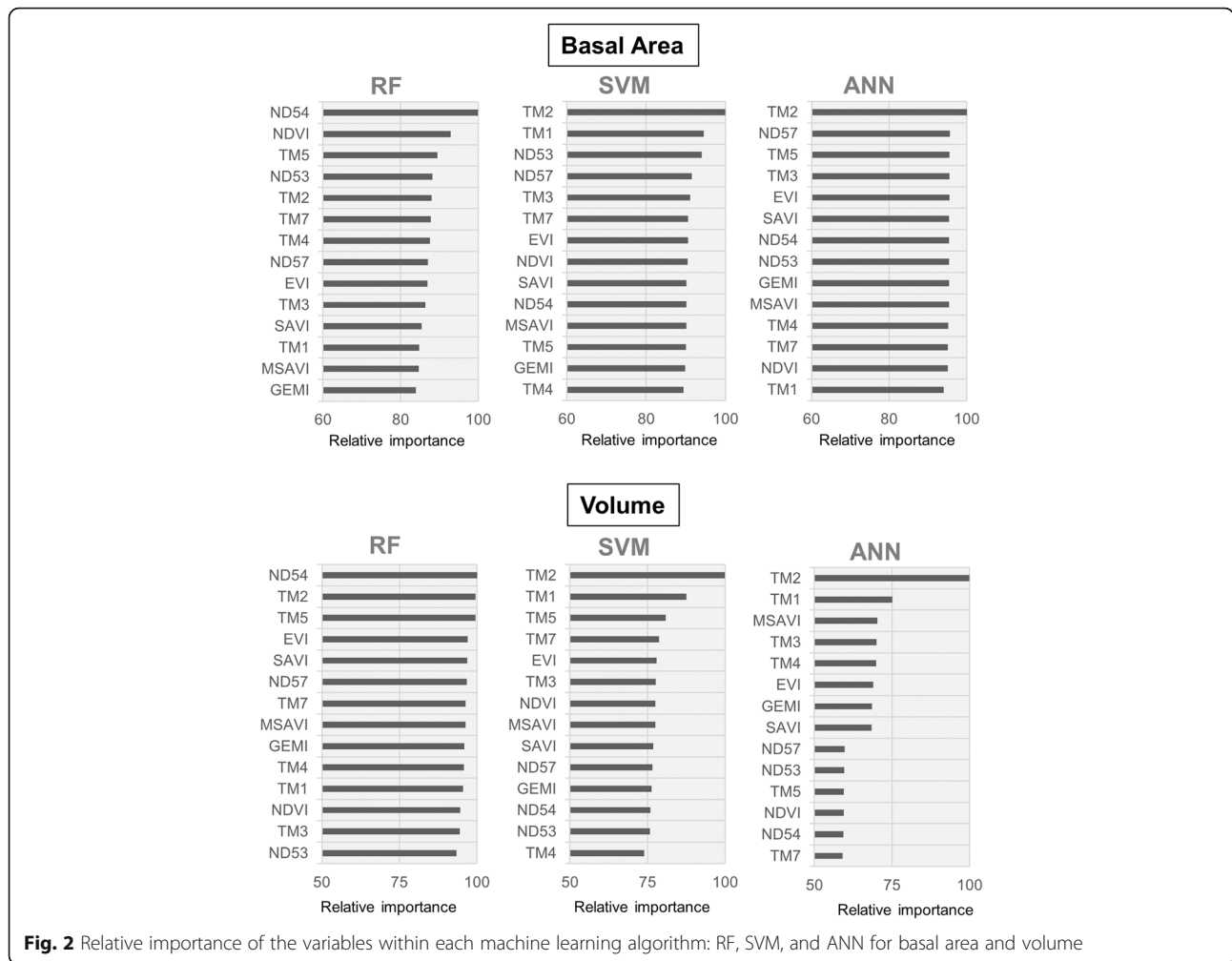
Fig. 2 Relative importance of the variables within each machine learning algorithm: RF, SVM, and ANN for basal area and volume

data (Viana et al. 2012). Average basal area (17.03 m^2 ha^{-1}) and volume (171.10 m^3 ha^{-1}) obtained from the prediction set did not statistically differ from average basal area (16.45 m^2 ha^{-1}) and volume (164.92 m^3 ha^{-1}) obtained from the validation set, considering two-tailed Student's t test (Basal area: $t = 0.629^{ns}$, df = 33, p value = 0.533; volume: $t = 0.550^{ns}$, df = 33, p value = 0.585).

The evaluation of spatial prediction methods, based on prediction and validation sets, was done by comparing the statistics presented in Eqs. 1 through 4 (Table 6). The mean error (ME) should ideally be close to zero if the prediction method is unbiased, and the values of this parameter suggested that all predictions generated impartial estimates when evaluated from both prediction and validation sets. Both the MAE and RMSE showed that basal area estimates were more accurate than volume estimates for all spatial prediction methods. The MAE and RMSE results obtained from the validation set demonstrated that there were no significant differences among the MLR, RF, SVM, and ANN for basal area estimates. For the volume estimates, the models fitted by
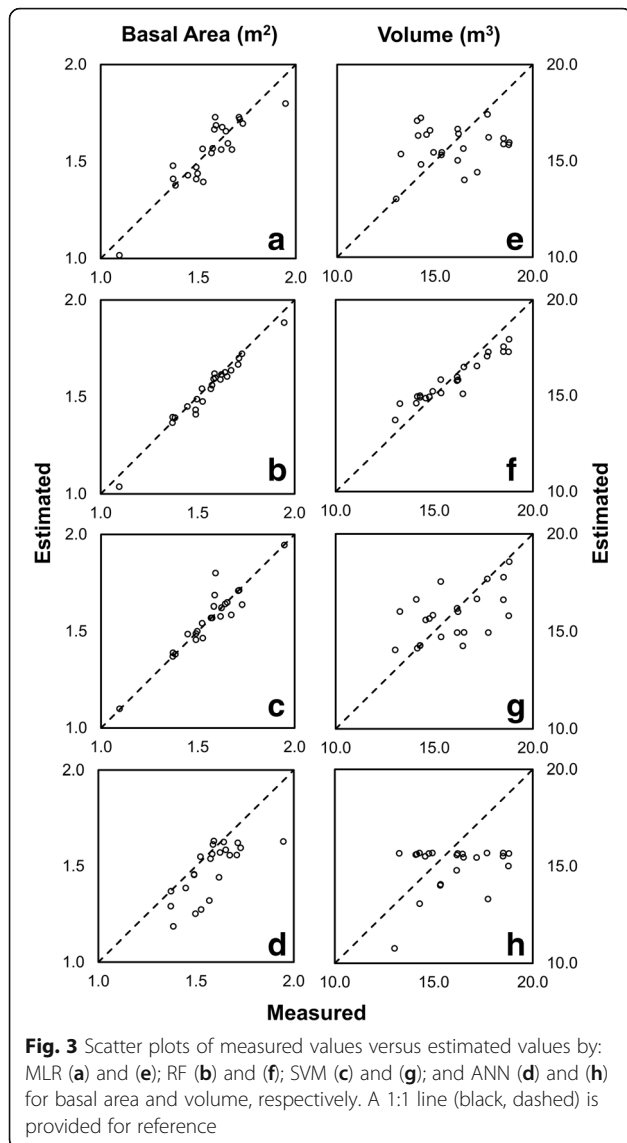
SVM had the best performance and MLR the poorest performance.

Geostatistical modelling of prediction method errors

The semivariogram models were selected based on RAE and SRE values close to 0 and 1, respectively (Yamamoto and Landim 2013). The experimental semivariograms constructed from the residuals of the basal area and volume prediction methods had a spatial dependence structure defined in six of the eight analysed situations (Fig. 4 and Table 7). The volume residuals from MLR and ANN methods had a pure nugget effect, i.e. no spatial dependence structure. This result indicated a random spatial distribution of the residuals in these two situations.

The residuals of the spatial prediction methods that had defined spatial dependence structures (Fig. 4) were interpolated using ordinary kriging, and their estimates were added to basal area and volume estimates of the respective spatial prediction methods. The relative improvement (RI) of the addition of basal area residual

Fig. 3 Scatter plots of measured values versus estimated values by: MLR (**a**) and (**e**); RF (**b**) and (**f**); SVM (**c**) and (**g**); and ANN (**d**) and (**h**) for basal area and volume, respectively. A 1:1 line (black, dashed) is provided for reference

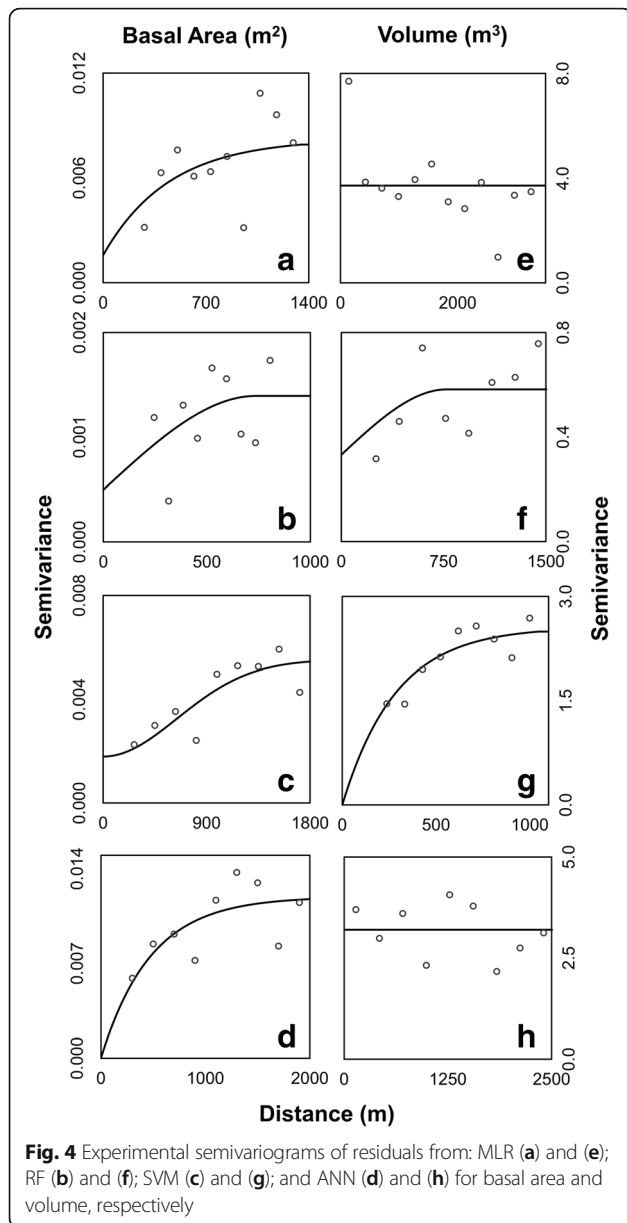**Table 6** Prediction methods evaluation using the prediction and validation sets for the *Eucalyptus* stands

Method	Statistic	Basal area error (m^2)		Volume error (m^3)	
		Prediction set	Validation set	Prediction set	Validation set
MLR	ME	0.00	− 0.05	0.00	− 0.74
	MAE	0.07	0.09	1.56	2.08
	RMSE	0.08	0.14	1.89	2.48
	RMSE (%)	5.50	9.42	12.27	16.72
RF	ME	0.01	− 0.03	0.08	− 0.90
	MAE	0.03	0.09	0.62	1.63
	RMSE	0.04	0.14	0.73	2.21
	RMSE (%)	2.48	9.54	4.77	14.91
SVM	ME	− 0.01	− 0.05	0.00	− 0.66
	MAE	0.04	0.09	1.19	1.59
	RMSE	0.06	0.14	1.60	2.02
	RMSE (%)	4.14	9.39	10.41	13.58
ANN	ME	0.09	0.03	0.94	0.45
	MAE	0.10	0.09	1.70	1.68
	RMSE	0.14	0.13	1.98	2.05
	RMSE (%)	8.87	8.52	12.88	13.82

MLR multiple linear regression, *RF* random forest, *SVM* support vector machine, *ANN* artificial neural networks, *ME* mean error, *MAE* mean absolute error, *RMSE* root mean square error

kriging by the ANN method was 25%, i.e. there was a reduction from 8.52 to 6.37% in the RMSE (Table 8). For the RF method, the RMSE increased from 9.54 to 10.08%, which corresponds to a 5.7% increase in the error of the basal area estimates by kriging of the residuals. For the volume, the addition of residual kriging improved the precision of SVM estimates and reduced the precision of the RF estimates.

Mapping of basal area and volume for *Eucalyptus* stands
Basal area and volume estimates obtained by different spatial prediction methods (Table 9) had average values very close to each other, and were in agreement with the forest inventory estimates (Table 3). Only the ANN method generated underestimated values for both basal area and volume, so that the total values of basal area

and volume were not within the confidence interval generated by the forest inventory.

Maps showing the spatial distribution of basal area and volume identified the same areas with high and low productivity, regardless of the spatial prediction method (Figs. 5 and 6). The maps obtained by ANN had a smaller difference between maximum and minimum estimated values for basal area and volume, while the mapping obtained from the SVM models had a greater difference between these values. MLR and RF methods provided similar estimates in the basal area and volume mapping.

The addition of residual kriging in the basal area and volume mapping (Fig. 7) resulted in a greater difference between maximum and minimum estimated values in all spatial prediction methods. For ANN, residual kriging resulted in estimates that were more in agreement with the field observations, correcting the basal area underestimation behaviour for the *Eucalyptus* stands under study. However, the addition of residual kriging to the models fitted by RF and SVM methods did not result in significant differences in basal area and volume mapping, and also led to increases in estimation errors in non-sampled areas in the field (Table 8).

dos Reis *et al. New Zealand Journal of Forestry Science* (2018) 48:1

Page 10 of 17

Fig. 4 Experimental semivariograms of residuals from: MLR (**a**) and (**e**); RF (**b**) and (**f**); SVM (**c**) and (**g**); and ANN (**d**) and (**h**) for basal area and volume, respectively

Discussion

Remote detection of forest canopies is complex due to the size, shape, and dielectric properties of its scatter elements (leaves, branches, and stems) (Galeana-Pizaña et al. 2014). The spatial diversity of forest canopies makes the relationship between forest parameters and remote sensing data a major challenge, although several studies have already demonstrated correlation between spectral data and forest characteristics of interest (Stojanova et al. 2010, Viana et al. 2012, Castillo-Santiago et al. 2013, Fayad et al. 2016, Gao et al. 2016). For instance, plantations comprised of different *Eucalyptus* species may have very similar values of basal area and volume, but have different spectral characteristics due to differences in spectral behaviour of the species that form the canopies. Also, according to Ponzoni et al. (2015), the canopy reflectance of older *Eucalyptus* plantations (between 4 and 6 years) tend to contain a greater contribution from green leaves and a lower contribution from shadows, the background, and from dry branches inside the canopies than the canopy reflectance of young *Eucalyptus* plantations (< 4 years). Thus, the canopy reflectance of older *Eucalyptus* plantations generated highest correlations with bands of the infrared region of the electromagnetic spectrum and, therefore, with vegetation indices that include these bands in their compositions (Ponzoni et al. 2015). These results are consistent with the best correlations found in this study among the infrared bands, vegetation indices derived from these bands, basal area, and volume. This same behaviour was observed in the studies of Gebreslasie et al. (2008), Canavesi et al. (2010), Berra et al. (2012), and Pacheco et al. (2012).

Basal area was more strongly correlated with the spectral data because this variable is derived from only the diameter of the trees, which is directly related to size of the tree canopies, and determines the canopy reflectance (Ponzoni et al. 2012). On the other hand, volume is derived from the diameter, form factor, and height of the trees. Height estimates are obtained from empirical equations that add errors during the volume estimation

Table 7 Nugget (τ^2), sill (σ^2), and range (φ) parameters for the selected semivariance function models for each of the variables in study

Variables	Residual	Selected model	τ^2	σ^2	φ (m)	RAE	SRE
Basal area	MLR	Exponential	0.0016	0.0067	1350	− 0.0092	1.0818
	RF	Spherical	0.0004	0.0009	737	− 0.0079	1.0586
	SVM	Gaussian	0.0017	0.0037	1577	0.0089	0.9610
	ANN	Exponential	0.0000	0.0119	1430	− 0.0303	1.1393
Volume	MLR	Exponential	PNE	PNE	PNE	PNE	PNE
	RF	Spherical	0.3316	0.2505	773	− 0.0051	1.0258
	SVM	Exponential	0.0000	2.5582	858	− 0.0039	0.9958
	ANN	Exponential	PNE	PNE	PNE	PNE	PNE

MLR multiple linear regression, *RF* random forest, *SVM* support vector machine, *ANN* artificial neural networks, *RAE* reduced average error, *SRE* standard deviation of the reduced average error, *PNE* pure nugget effect

Table 8 Prediction methods with addition of the residual estimation by ordinary kriging using the prediction and validation sets for the *Eucalyptus* stands

Method	Statistic	Basal area error (m^2)		Volume error (m^3)	
		Prediction set	Validation set	Prediction set	Validation set
MLR-RK	ME	0.00	− 0.03	–	–
	MAE	0.03	0.09	–	–
	RMSE	0.04	0.14	–	–
	RMSE (%)	2.80	9.30	–	–
	RI	49.09	1.27	–	–
RF-RK	ME	0.01	− 0.05	0.00	− 1.03
	MAE	0.04	0.10	0.63	1.70
	RMSE	0.05	0.15	0.77	2.26
	RMSE (%)	3.08	10.08	5.02	15.25
	RI	− 24.19	− 5.66	− 5.24	− 2.28
SVM-RK	ME	0.01	− 0.03	− 0.32	− 0.57
	MAE	0.05	0.10	0.80	1.22
	RMSE	0.06	0.15	1.11	1.74
	RMSE (%)	4.09	9.83	7.19	11.72
	RI	1.21	− 4.69	30.93	13.70
ANN-RK	ME	0.02	− 0.06	–	–
	MAE	0.04	0.06	–	–
	RMSE	0.09	0.09	–	–
	RMSE (%)	5.79	6.37	–	–
	RI	34.72	25.23	–	–

MLR multiple linear regression, *RF* random forest, *SVM* support vector machine, *ANN* artificial neural networks, *RK* residual estimation by ordinary kriging, *ME* mean error, *MAE* mean absolute error, *RMSE* root mean square error, *RI* relative improvement

process. This acts to reduce the strength of relationships between volume and variables obtained from remotely sensed images. The ND54 index was the spectral variable that had the strongest correlation with basal area ($r = − 0.91$) and volume ($r = − 0.52$). However, it was also significantly correlated with the other spectral variables. During multiple linear regression analysis, the fact that two or more explanatory variables are highly correlated may generate multicollinearity problems in the fitted models, since one of the

regression assumptions is that no linear relationship may exist between any independent variables or linear combinations of these (Montgomery et al. 2006).

For the MLR method, the best volume estimation model was obtained from the TM1 band and the NDVI (Table 5), yet was only able to explain approximately 37% of the variation in this stand attribute. Conversely, the best model for basal area estimation used the ND54 index as the predictor variable and was able to explain

Table 9 Statistics of basal area and volume maps estimated by spatial predictions methods MLR, RF, SVM, and ANN

| Method	Basal area (m^2)					Volume (m^3)				
	Min	Max	Mean	Standard deviation	Total estimate	Min	Max	Mean	Standard deviation	Total estimate
MLR	0.62	1.83	1.52	0.20	6151.9	4.51	19.99	15.30	2.30	61,739.5
MLR-RK	0.65	1.93	1.52	0.21	6141.0	–	–	–	–	–
RF	0.96	1,89	1.51	0.17	6101.5	9.26	18.08	15.27	1.81	61,600.1
RF-RK	0.93	1.93	1.53	0.17	6166.6	9.02	18.37	15.36	1.91	61,965.7
SVM	0.88	2.12	1.57	0.18	6326.2	1.36	19.64	15.31	2.57	61,760.7
SVM-RK	0.76	2.10	1.56	0.19	6284.2	1.10	21.78	15.29	2.92	61,683.8
ANN	0.97	1.65	1.42	0.22	5715.3	8.32	15.68	13.93	2.70	56,223.8
ANN-RK	0.90	1.94	1.50	0.23	6070.3	–	–	–	–	–

Min minimum value, *Max* maximum value, *MLR* multiple linear regression, *RF* random forest, *SVM* support vector machine, *ANN* artificial neural networks, *RK* residual estimation by ordinary kriging

dos Reis *et al. New Zealand Journal of Forestry Science* (2018) 48:1

Page 12 of 17
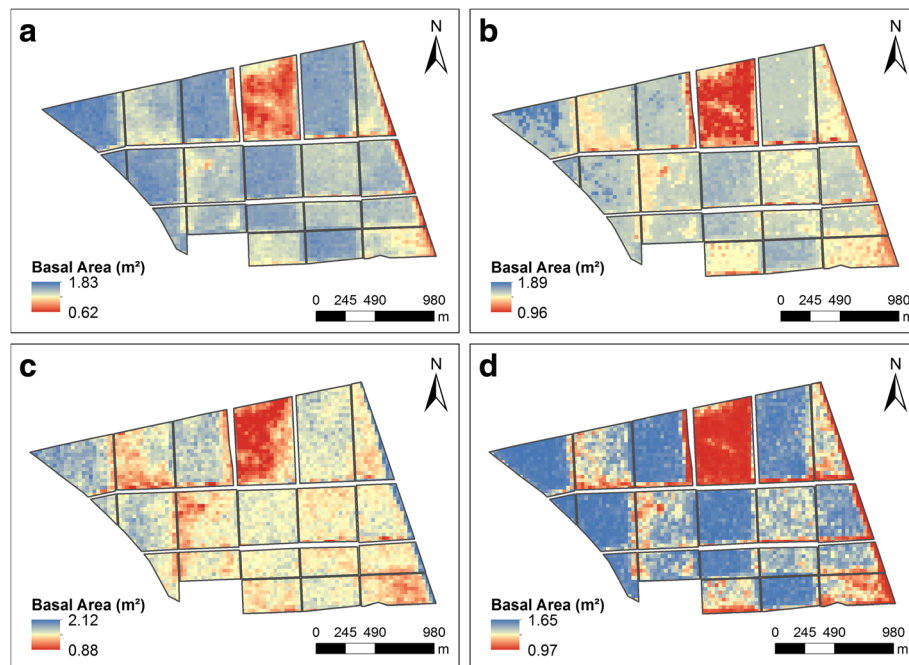
Fig. 5 Spatial distribution of the basal area in *Eucalyptus* stands, estimated by: MLR (**a**), RF (**b**), SVM (**c**), and ANN (**d**)

more than 80% of the variation in this attribute, confirming the explanatory power of spectral data for basal area estimation in *Eucalyptus* stands. Gebreslasie et al. (2010) assessed the suitability of both visible and shortwave infrared ASTER data and vegetation indices for estimating forest structural attributes of *Eucalyptus* species in southern KwaZulu Natal, South Africa. These authors applied a MLR using MSAVI and band 3 as predictor variables and were able to explain slightly more of the variation in basal area ($R^2 = 0.67$) than volume ($R^2 = 0.65$). Although the MLR model for volume does not have a high coefficient of determination, the spectral data can efficiently
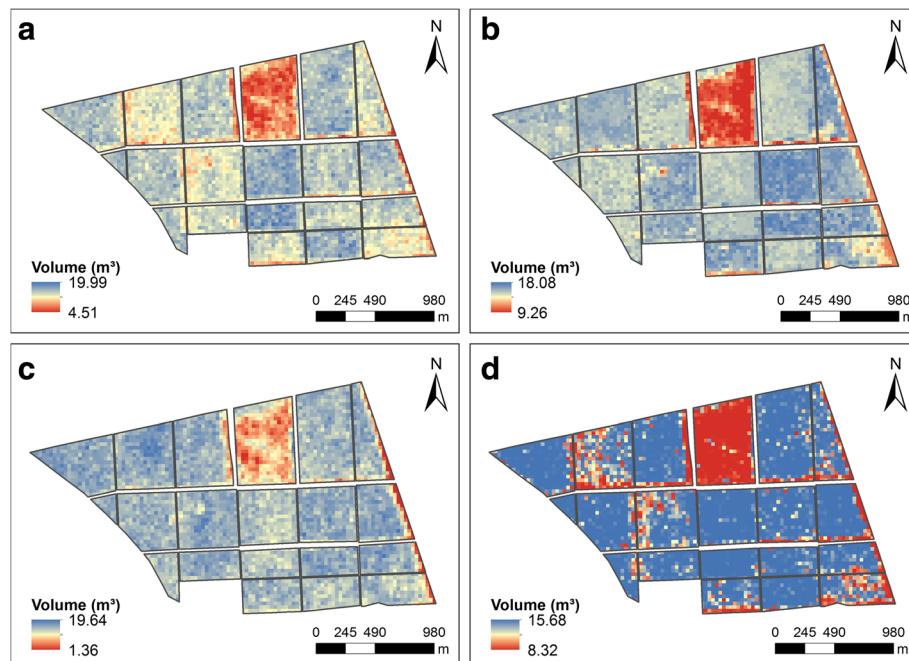
Fig. 6 Spatial distribution of the volume in *Eucalyptus* stands, estimated by: MLR (**a**); RF (**b**); SVM (**c**); and ANN (**d**)
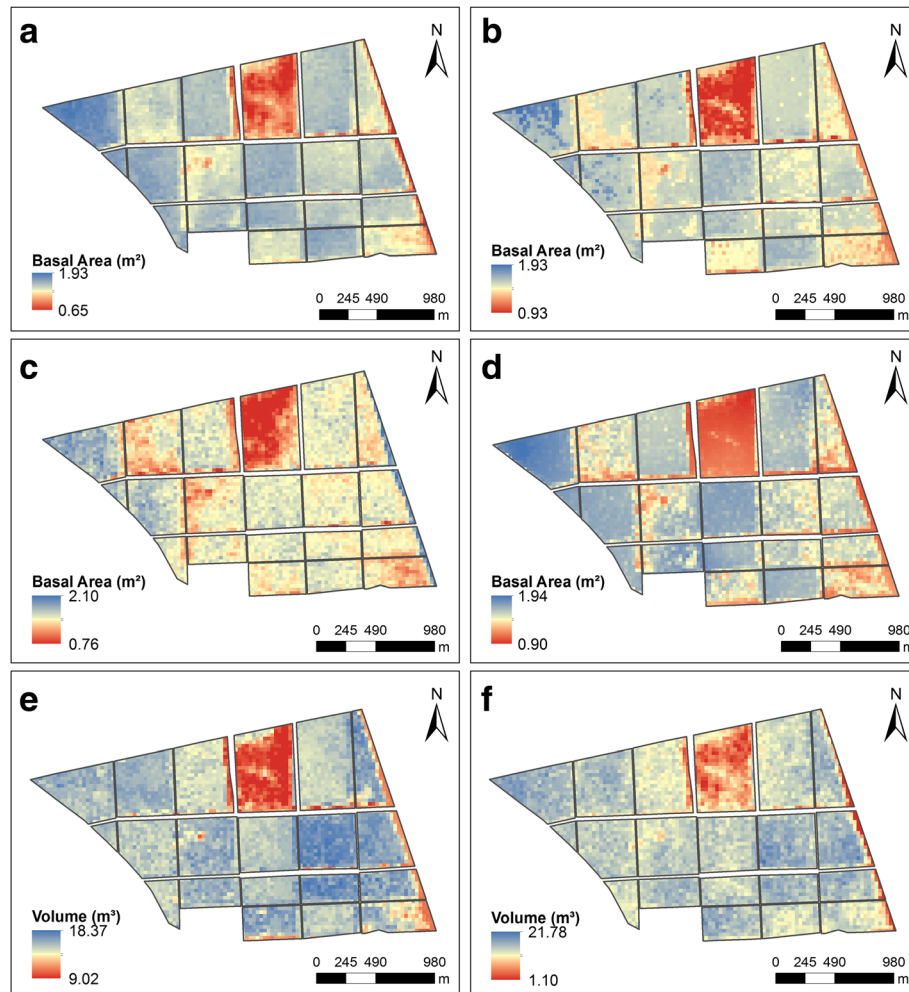
Fig. 7 Spatial distribution of the basal area in *Eucalyptus* stands estimated by: MLR (**a**): RF (**b**); SVM (**c**); and ANN (**d**) with addition of the residual estimation by ordinary kriging; and for volume estimated by RF (**e**) and SVM (**f**) with addition of the residual estimation by ordinary kriging

explain the volumetric variations in non-sampled areas in the field. In a similar study for *Eucalyptus* stands located in the southern region of Brazil, Berra et al. (2012) concluded that spectral data obtained from Landsat images were efficient in mapping the volume in the study area, even when the regression models did not present high coefficients of determination ($R^2 < 0.70$).

Divergence among variables that were deemed important between the different methods was observed with the machine learning algorithms. For basal area modelling, the ND54 index and NDVI had a higher importance value for RF. Statistically, these indices had high correlation values with the variable of interest ($r = -0.91$ and 0.83, respectively) and high multicollinearity ($r = -0.93$). The ND54 index also was the variable that most contributed to the volume estimate by the RF method. The fact that the explanatory variables are correlated does not affect the performance of these algorithms. These methods do not rely on underlying assumptions about the data, which

allows them to work with all available explanatory variables, without loss of information in the process of variable selection and reduction (Görgens et al. 2015). For the models fitted using ANN and SVM algorithms, the TM2 band was the most important predictor variable for basal area and volume. The linear correlation between this variable and basal area and volume is low to non-existent ($r = 0.15$ and -0.02, respectively). However, this band is usually applied in vegetation vigour assessment (Meng et al. 2009), a characteristic that is indirectly related to volume and basal area, and which may explain the greater contribution of the TM2 band in the ANN and SVM algorithms, since trees that are more vigorous tend to have higher values of basal area and volume.

The models of basal area and volume developed by the RF algorithm had smaller errors compared with those developed by other machine learning algorithms and MLR. The performance of this algorithm has been proven in many modelling and remote sensing studies

dos Reis *et al. New Zealand Journal of Forestry Science* (2018) 48:1

Page 14 of 17

(Lafiti et al. 2010, Rodriguez-Galiano et al. 2015, Wu et al. 2016). In the study by Shataee et al. (2012), volume prediction models developed by RF performed better than those developed using k-nearest neighbour (k-NN) and SVM. Employing ASTER satellite data, the relative RMSE obtained for all three volume models was higher than for the models developed in our study: 28.54% for k-NN, 25.86% for SVM, and 26.86% for RF, and only the RF algorithm produced unbiased volume estimations. For basal area, RF produced models with lower RMSE (18.39%) when compared with SVM (RMSE = 19.35%) and k-NN (RMSE = 20.20%); however, only k-NN was able to generate unbiased estimation compared with the other two algorithms used.

One of the positive features of RF is that it achieves satisfactory performance even with a limited number of samples and with many independent variables (attributes), as in the case of this current study. It is an ensemble method, which combines several regression trees to generate an average estimate, in which different attributes are used in each tree, making the results take into account the information of all available attributes. Stojanova et al. (2010) also concluded that ensemble methods (RF) were significantly better in height and canopy cover modelling using remote sensing data than single- and multi-target regression trees. The ANN and SVM algorithms also have proven good performance and robustness in several studies (e.g. Shao and Lunetta 2012, Were et al. 2015). However, the parameterisation of these methods is laborious, and they are very sensitive to the variation of input parameters, with ANN being more sensitive than other methods (Rodriguez-Galiano et al. 2015). This same behaviour was observed in this study, where the use of a restricted dataset by ANN resulted in estimates that were not compatible with the forest inventory estimates (Tables 3 and 9).

The addition of residual kriging in spatial prediction methods did not necessarily result in relative improvements in the estimation of these methods. In the case of MLR and ANN methods, residual kriging contributed to better accuracy of the basal area estimates. These results are consistent with the results of Dai et al. (2014), who reported that the combination of the residual kriging with artificial neural networks provides an improvement in the estimate accuracy of the variables of interest. The combination of MLR with residual kriging also provided improvements in estimates in the studies of Viana et al. (2012), Castillo-Santiago et al. (2013), and Galeana-Pizaña et al. (2014). For basal area and volume estimation, the addition of residual kriging in the RF and SVM methods resulted in a lower precision of the estimates. Hybrid methods are advantageous in the ability to use spatial information (ordinary kriging of residuals) and non-spatial information (multiple linear regression

analysis and machine learning algorithms). However, in some situations, hybrid methods provide less-accurate estimates in regions where the data collected in the field are sparse (Palmer et al. 2009).

The high growth rate of *Eucalyptus* stands in Brazil reinforces the importance of robust methods that consider auxiliary information in the process of estimating variables of interest, such as basal area and volume. The methodologies presented here are powerful tools for estimating basal area and volume from spectral data obtained from Landsat 5 TM or from other multispectral optical sensors. According to Görgens et al. (2015), machine learning algorithms can continuously learn from new data and keep all the accumulated knowledge of previous datasets. This fact allows the implementation of these algorithms in other situations where only limited amounts of data are available. The use of all auxiliary variables in the estimation process is another advantage over traditional regression methods, since machine learning algorithms are not restricted by correlation between input variables, thus avoiding the loss of important information in the estimation process of the variable of interest. Nevertheless, these methods have as disadvantage the transparency of the resulting models, so an alternative to overcome this obstacle is the evaluation of the relative importance of the explanatory variables. Furthermore, the causal relation between inputs and outputs of the estimation process is not clear, which implies a limited biological interpretation (Aertsen et al. 2010, Özçelik et al. 2013).

The results from the current study do need to be interpreted cautiously, as they are limited to a homogenous and relatively small study area. While this work uses a small number of plots, it represents the sampling intensity adopted by most Brazilian forestry companies, i.e. one plot (usually 200–500 m^2 in size) for each 10 ha of *Eucalyptus* plantation (Raimundo et al. 2017, Scolforo et al. 2016) and the results from this research showcase the importance of using remotely sensed data and robust prediction methods for basal area and volume estimation. The data used here were also from a relatively old sensor, Landsat 5 TM, and a study by Fassnacht et al. (2014) concluded that predictor data (sensor) type is the most important factor for the accuracy of biomass estimates and that the prediction method had a substantial effect on accuracy and was generally more important than the sample size. Fassnacht et al. (2014) also suggested that choosing the appropriate statistical method may be more effective than obtaining additional field data for obtaining good biomass estimates.

Considering the cost of improving accuracy of timber production estimates by field measurements in *Eucalyptus* stands, it seems sensible to invest in further studies that

dos Reis *et al. New Zealand Journal of Forestry Science* (2018) 48:1

Page 15 of 17

focus on more test sites and a wider range of sensor systems (particularly RADAR and LIDAR). This would further increase our understanding of the role of the statistical model set-up in remote sensing-based estimates of forest variables in *Eucalyptus* stands. Further studies could also investigate whether other prediction methods, such as non-linear regression or partial least squares regression (PLSR) approaches, alter our findings. The integration of additional predictors (e.g. topographic information or climate variables) would be a further possible extension of our work.

Conclusions

Machine learning algorithms, particularly the random forest (RF) and support vector machine (SVM) algorithms, were able to develop models that estimate basal area and volume in *Eucalyptus* stands using spectral data collected from Landsat 5 TM images. The artificial neural network (ANN) method did not perform well in this context, due in part to the limited data availability.

Random forest was the best method of spatial prediction and mapping of basal area and volume in *Eucalyptus* stands in Minas Gerais state. However, due to the close performance to the support vector machine and multiple linear regression methods, we propose that both methods should be tested and then the best result applied for spatial prediction of basal area and volume in other regions with *Eucalyptus* stands. The approaches used in this study provide a framework for integrating field and multispectral data, highlighting methods that greatly improve spatial prediction of basal area and volume estimation in *Eucalyptus* stands. Although the sensor TM of Landsat satellites is no longer operational, the concepts presented in this study are expected to be consistent regardless of the sensor. Thus, the approach used in this study can be more broadly applied to basal area and volume estimation in *Eucalyptus* stands using the new optical sensors such as Landsat 8 OLI and Sentinel-2.

The combination of spatial prediction methods with residual kriging should be used with caution, since the relative improvement of spatial prediction accuracy of basal area and volume did not occur in all methods, and there is not always a spatial dependency structure in the residuals of a spatial prediction method.

Abbreviations
AIC: Akaike information criterion; ANN: Artificial neural networks; EVI: Enhanced vegetation index; G: Basal area; GDP: Gross domestic product; GEMI: Global environment monitoring index; GIS: Geographical information systems; GPS: Global positioning systems; MAE: Mean absolute error; ME: Mean error; MLA: Machine learning algorithms; MLR: Multiple linear regression; MSAVI: Modified soil-adjusted vegetation index; ND: Normalised difference; NDVI: Normalised difference vegetation index; PNE: Pure nugget effect; R^2_{aj}: Adjusted coefficient of determination; RAE: Reduced average error; RBF: Radial basis function; RF: Random forest; RI: Relative improvement; RK: Residual estimation by ordinary kriging; RMSE: Root mean square error; SAVI: Soil-adjusted vegetation index; SMO: Sequential minimal optimization; SRE: Standard deviation of the reduced average error; SVM: Support vector machine; S_{xy}: Residual standard error; TM: Thematic mapper; USGS: United States Geological Survey; V: Volume; VIF: Variance inflation factor

Authors' contributions
All authors contributed substantially to the work reported here. AAR, MCC, LRG, and ACFF analysed and interpreted the data. ARR and MCC wrote the manuscript. JMM, ACFF, LRG, and FWAJ reviewed and edited the manuscript. All authors read and approved the final manuscript.

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

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Department of Forest Science, Federal University of Lavras – UFLA, PO Box 3037, Lavras, Minas Gerais 37200-000, Brazil. [2]CPCE, Federal University of Piauí, BR 135 - km 3, Bom Jesus, Piauí 64900-000, Brazil.

References
Aertsen, W, Kint, V, Van Orshoven, J, Özkan, KA, Muys, B. (2010). Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecological Modelling*, 221, 1119–1130. https://doi.org/10.1016/j.ecolmodel.2010.01.007.

Ahmed, OS, Franklin, SE, Wulder, MA, White, JC. (2015). Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the random forest algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, 89–101. https://doi.org/10.1016/j.isprsjprs.2014.11.007.

Alvares, CA, Stape, JL, Sentelhas, PC, Gonçalves, JLM, Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, 6, 711–728. https://doi.org/10.1127/0941-2948/2013/0507.

Barrios, PG, Bidegain, MP, Gutiérrez, L. (2015). Effects of tillage intensities on spatial soil variability and site-specific management in early growth of *Eucalyptus grandis*. *Forest Ecology and Management*, 346, 41–50. https://doi.org/10.1016/j.foreco.2015.02.031.

Berra, EF, Brandelero, C, Pereira, RS, Sebem, E, Goergen, LCG, Benedetti, ACP, Lippert, DB. (2012). Estimativa do volume total de madeira em espécies de eucalipto a partir de imagens de satélite Landsat. *Ciência Florestal*, 22(4), 853–864. https://doi.org/10.5902/198050987566.

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. *Forest Ecology and Management*, 376, 284–297. https://doi.org/10.1016/j.foreco.2016.06.022.

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Canavesi, V, Ponzoni, FJ, Valeriano, MM. (2010). Estimativa de volume de madeira em plantios de *Eucalyptus* spp. utilizando dados hiperespectrais e dados topográficos. *Revista Árvore*, 4(3), 539–549. https://doi.org/10.1590/S0100-67622010000300018.

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

dos Reis *et al. New Zealand Journal of Forestry Science* (2018) 48:1

Page 16 of 17

Cluter, 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 Journal of Photogrammetry and Remote Sensing*, *70*, 66–77. https://doi.org/10.1016/j.isprsjprs.2012.03.011.

Coops, NC, Johnson, M, Wulder, MA, White, JC. (2006). Assessment of QuickBird high spatial resolution imagery to detect red attack damage due to mountain pine beetle infestation. *Remote Sensing of Environment*, *103*(1), 67–80. https://doi.org/10.1016/j.rse.2006.03.012.

Dai, F, Zhou, Q, Lv, Z, Wang, X, Liu, G. (2014). Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan plateau. *Ecological Indicators*, *45*, 184–194. https://doi.org/10.1016/j.ecolind.2014.04.003.

Diamantopoulou, MJ. (2012). Assessing a reliable modeling approach of features of trees through neural network models for sustainable forests. *Sustainable Computing: Informatics and Systems*, *2*, 190–197. https://doi.org/10.1016/j.suscom.2012.10.002.

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 Journal of Photogrammetry and Remote Sensing*, *108*, 12–32. https://doi.org/10.1016/j.isprsjprs.2015.06.002.

Dube, T, Mutanga, O, Adam, E, Ismail, R. (2014). Intra-and-inter species biomass prediction in a plantation forest: Testing the utility of high spatial resolution Spaceborne multispectral RapidEye sensor and advanced machine learning algorithms. *Sensors*, *14*, 15348–15370. https://doi.org/10.3390/s140815348.

Dye, PJ, Jacobs, S, Drew, D. (2004). Verification of 3-PG growth and water-use predictions in twelve *Eucalyptus* plantation stands in Zululand, South Africa. *Forest Ecology and Management*, *193*, 197–218. https://doi.org/10.1016/j.foreco.2004.01.030.

Environmental Systems Research Institute (2010). *ArcGIS desktop: Release 10.1*. Redlands: ESRI.

Fassnacht, FE, Hartig, F, Latifi, H, Berger, C, Hernández, J, Corvalán, P, Koch, B. (2014). Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. *Remote Sensing of Environment*, *154*, 102–114. https://doi.org/10.1016/j.rse.2014.07.028.

Fayad, I, Baghdadi, N, Guitet, S, Bailly, JS, Hérault, B, Gond, V, Hajj, ME, Minh, DHT. (2016). Aboveground biomass mapping in French Guiana by combining remote sensing, forest inventories and environmental data. *International Journal of Applied Earth Observation and Geoinformation*, *52*, 502–514. https://doi.org/10.1016/j.jag.2016.07.015.

Frank, E, Hall, MA, Witten, I (2016). The WEKA workbench [online appendix]. In I Witten, E Frank, M Hall, C Pal (Eds.), *Data mining: Practical machine learning tools and techniques*, (4th ed.,). Burlington: Morgan Kaufmann.

Galeana-Pizaña, JM, López-Caloca, A, López-Quiroza, 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. *International Journal of Applied Earth Observation and Geoinformation*, *30*, 179–189. https://doi.org/10.1016/j.jag.2014.02.005.

Gao, T, Zhu, J, Deng, S, Zheng, X, Zhang, J, Shang, G, Huang, L. (2016). Timber production assessment of a plantation forest: An integrated framework with field-based inventory, multi-source remote sensing data and forest management history. *International Journal of Applied Earth Observation and Geoinformation*, *52*, 155–165. https://doi.org/10.1016/j.jag.2016.06.004.

García-Gutiérrez, J, Martínez-Álvarez, F, Troncoso, A, Riquelme, JC. (2015). A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. *Neurocomputing*, *167*, 24–31. https://doi.org/10.1016/j.neucom.2014.09.091.

Gebreslasie, MT, Ahmed, FB, Aardt, JAN. (2008). Estimating plot-level forest structural attributes using high spectral resolution ASTER satellite data in even-aged *Eucalyptus* plantations in southern KwaZulu-Natal, South Africa. *Southern Forests*, *70*(3), 227–236. https://doi.org/10.2989/SF.2008.70.3.6.667.

Gebreslasie, MT, Ahmed, FB, Aardt, JAN. (2010). Predicting forest structural attributes using ancillary data and ASTER satellite data. *International Journal of Applied Earth Observation and Geoinformation*, *12S*, S23–S26. https://doi.org/10.1016/j.jag.2009.11.006.

Gleason, CJ, & Im, J. (2012). Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sensing of Environment*, *125*, 80–91. https://doi.org/10.1016/j.rse.2012.07.006.

González-García, M, Hevia, A, Majada, J, Anta, RC, Barrio-Anta, M. (2015). Dynamic growth and yield model including environmental factors for *Eucalyptus nitens* (Deane & Maiden) maiden short rotation woody crops in Northwest Spain. *New Forests*, *46*, 387–407. https://doi.org/10.1007/s11056-015-9467-7.

Görgens, EB, Montaghi, A, Rodriguez, LCE. (2015). A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. *Computers and Electronics in Agriculture*, *116*, 221–227. https://doi.org/10.1016/j.compag.2015.07.004.

Guedes, ICL, Mello, JM, Silveira, EMO, Mello, CR, Reis, AA, Gomide, LR. (2015). Spatial continuity of dendrometric characteristics of clonal cultivated *Eucalyptus* sp. throughout the time. *Cerne*, *21*(4), 527–534. https://doi.org/10.1590/01047760201521041824.

Huang, C, Song, K, Kim, S, Townshend, JRG, Davis, P, Masek, JG, Goward, SN. (2008). Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote Sensing of Environment*, *112*, 970–985. https://doi.org/10.1016/j.rse.2007.07.023.

Huete, A, Didan, K, Miura, T, Rodriguez, EP, Gao, X, Ferreira, LG. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*, 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2.

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

Immitzer, M, Atzberger, C, Koukal, T. (2012). Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data. *Remote Sensing*, *4*, 2661–2693. https://doi.org/10.3390/rs4092661.

Indústria Brasileira de Árvores (2014). *Anuário estatístico da indústria brasileira de árvores: ano base 2014*. Brasília: IBA.

Indústria Brasileira de Árvores (2015). *Anuário estatístico da indústria brasileira de árvores: ano base 2015*. Brasília: IBA.

Journel, AG, & Huijbregts, CJ (1978). *Mining geostatistics*. London: Academic.

Justice, CO, Vermote, E, Townshend, JRG, Defries, R, Roy, DO, Hall, DK, Salomonson, VV, Privette, JL, Riggs, G, Strahler, A, Lucht, W, Myneni, RB, Knyazikhin, Y, Running, SW, Nemani, RR, Wan, Z, Huete, AR, Leeuwen, VV, Wolfe, RE, Giglio, L, Muller, J, Lewis, P, Barnsley, MJ. (1998). The moderate resolution imaging spectroradiometer (MODIS): Land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing*, *36*(4), 1228–1249. https://doi.org/10.1109/36.701075.

Lafiti, H, Nothdurft, A, Koch, B. (2010). Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: Application of multiple optical/LiDAR-derived predictors. *Forestry*, *83*(4), 395–407. https://doi.org/10.1093/forestry/cpq022.

Lopes, DM, Aranha, JT, Walford, N, O'Brien, J, Lucas, N. (2009). Accuracy of remote sensing data versus other sources of information for estimating net primary production in *Eucalyptus globulus* Labill. and *Pinus pinaster* Ait. ecosystems in Portugal. *Canadian Journal of Remote Sensing*, *35*(1), 37–53. https://doi.org/10.5589/m08-078.

López-Sánchez, CA, García-Ramírez, P, Resl, R, José, C, Hernández-Díaz, JC, López-Serrano, PM, Wehenkel, C. (2014). Modelling dasometric attributes of mixed and uneven-aged forests using Landsat-8 OLI spectral data in the Sierra Madre Occidental, Mexico. *iForest*, *10*, 288–295. https://doi.org/10.3832/ifor1891-009.

López-Serrano, PM, Corral-Rivas, JJ, Díaz-Varela, RA. (2016). Evaluation of radiometric and atmospheric correction algorithms for aboveground forest biomass estimation using Landsat 5 TM data. *Remote Sensing*, *8*(5), 369. https://doi.org/10.3390/rs8050369.

Lu, D, Mausel, P, Brondízio, E, Moran, E. (2004). Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. *Forest Ecology and Management*, *198*, 149–167. https://doi.org/10.1016/j.foreco.2004.03.048.

Masek, JG, Vermote, EF, Saleous, NE, Wolfe, R, Hall, FG, Huemmrich, KF, Gao, F, Kutler, J, Lim, TK. (2006). A Landsat surface reflectance dataset for North America, 1990 – 2000. *IEEE Geoscience and Remote Sensing Letters*, *3*(1), 68–72. https://doi.org/10.1109/LGRS.2005.857030.

Meng, Q, Cieszewski, C, Madden, M. (2009). Large area forest inventory using Landsat ETM+: A geostatistical approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*, 27–36. https://doi.org/10.1016/j.isprsjprs.2008.06.006.

Montgomery, DC, Peck, EA, Vining, GG (2006). *Introduction to linear regression analysis*. New York: Wiley.

Moreno, A, Neumann, M, Hasenauer, H. (2016). Optimal resolution for linking remotely sensed and forest inventory data in Europe. *Remote Sensing of Environment*, *183*, 109–119. https://doi.org/10.1016/j.rse.2016.05.021.

Morgenroth, J, & Visser, R. (2013). Uptake and barriers to the use of geospatial technologies in forest management. *New Zealand Journal of Forestry Science*, *43*(16), 1–9. https://doi.org/10.1186/1179-5395-43-16.

Özçelik, R, Diamantopoulou, MJ, Crecente-Campo, F, Eler, U. (2013). Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models. *Forest Ecology and Management*, *306*, 52–60. https://doi.org/10.1016/j.foreco.2013.06.009.

Pacheco, LRF, Ponzoni, FJ, Santos, SB, Andrades Filho, CO, Mello, MP, Campos, RC. (2012). Structural characterization of canopies of *Eucalyptus* spp. using radiometric data from TM/Landsat 5. *Cerne*, *18*(1), 105–116. https://doi.org/10.1590/S0104-77602012000100013.

Palmer, DJ, Höck, BK, Kimberley, MO, Watt, MS, Lowe, DJ, Payn, TW. (2009). Comparison of spatial prediction techniques for developing *Pinus radiata* productivity surfaces across New Zealand. *Forest Ecology and Management*, *258*(9), 2046–2055. https://doi.org/10.1016/j.foreco.2009.07.057.

Pinty, B, & Verstraete, MM. (1992). GEMI: A non-linear index to monitor global vegetation from satellites. *Vegetatio*, *101*(1), 15–20. https://doi.org/10.1007/BF00031911.

Ponzoni, FJ, Pacheco, LRF, Santos, SB, Andrades Filho, CO. (2015). Caracterização espectro-temporal de dosséis de *Eucalyptus* spp. mediante dados radiométricos TM/Landsat 5. *Cerne*, *21*(2), 267–275. https://doi.org/10.1590/01047760201521021457.

Ponzoni, FJ, Shimabukuro, YE, Kuplich, TM (2012). *Sensoriamento Remoto da Vegetação*, (2nd ed.,). São Paulo: Oficina de Textos.

Qi, J, Chehbouni, A, Huete, AR, Kerr, YH, Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, *48*, 119–126. https://doi.org/10.1016/0034-4257(94)90134-1.

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Raimundo, MR, Scolforo, HF, Mello, JM, Scolforo, JRS, McTague, JP, Reis, AA. (2017). Geostatistics applied to growth estimates in continuous forest inventories. *Forest Science*, *63*(1), 29–38. https://doi.org/10.5849/FS.2016-056.

Retslaff, FAS, Figueiredo Filho, A, Dias, AN, Bernett, LG, Figura, MA. (2015). Curvas de sítio e relações hipsométricas para *Eucalyptus grandis* na Região dos Campos Gerais, Paraná. *Cerne*, *2*(2), 199–207. https://doi.org/10.1590/01047760201521021349.

Ribeiro Júnior, PJ, & Diggle, PJ. (2001). GeoR: A package for geostatistical analysis. *R-NEWS*, *1*(2), 15–18.

Rodriguez-Galiano, V, Castillo, MS, Chica-Olmo, M, Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, *71*, 804–818. https://doi.org/10.1016/j.oregeorev.2015.01.001.

Rouse, J, Haas, R, Schell, J, Deering, D, Harlan, J (1973). *Monitoring the vernal advancements and retrogradation (greenwave effect) of nature vegetation. NASA/GSFC final report*. Greenbelt: NASA.

Scolforo, HF, Castro Neto, F, Scolforo, JRS, Burkhart, H, McTague, JP, Raimundo, MR, Loos, RA, Fonseca, S, Sartório, RC. (2016). Modeling dominant height growth of *Eucalyptus* plantations with parameters conditioned to climatic variations. *Forest Ecology and Management*, *380*, 182–195. https://doi.org/10.1016/j.foreco.2016.09.001.

Shao, Y, & Lunetta, RS. (2012). Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. *ISPRS Journal of Photogrammetry and Remote Sensing*, *70*, 78–87. https://doi.org/10.1016/j.isprsjprs.2012.04.001.

Shataee, S, Kalbi, S, Fallah, A, Pelz, D. (2012). Forest attribute imputation using machine-learning methods and ASTER data: Comparison of k-NN, SVR and random forest regression algorithms. *International Journal of Remote Sensing*, *33*, 6254–6280. https://doi.org/10.1080/01431161.2012.682661.

Stojanova, D, Panov, P, Gjorgjioski, V, Kobler, A, Džeroski, S. (2010). Estimating vegetation height and canopy cover from remotely sensed data with machine learning. *Ecological Informatics*, *5*, 256–266. https://doi.org/10.1016/j.ecoinf.2010.03.004.

United States Geological Survey (2017). Landsat imagery. Available online at: https://earthexplorer.usgs.gov. Accessed Jan 2017.

Verma, NK, Lamb, DW, Reid, N, Wilson, B. (2014). An allometric model for estimating DBH of isolated and clustered *Eucalyptus* trees from measurements of crown projection area. *Forest Ecology and Management*, *326*, 125–132. https://doi.org/10.1016/j.foreco.2014.04.003.

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. *Ecological Modelling*, *226*, 22–35. https://doi.org/10.1016/j.ecolmodel.2011.11.027.

Vicharnakorn, P, Shrestha, RP, Nagai, M, Salam, AP, Kiratiprayoon, S. (2014). Carbon stock assessment using remote sensing and forest inventory data in Savannakhet, Lao PDR. *Remote Sensing*, *6*, 5452–5479. https://doi.org/10.3390/rs6065452.

Watt, MS, Dash, JP, Watt, P, Bhandari, S. (2016). Multi-sensor modelling of a forest productivity index for radiata pine plantations. *New Zealand Journal of Forestry Science*, *46*, 9. https://doi.org/10.1186/s40490-016-0065-z.

Watt, MS, Rubilar, R, Kimberley, MO, Kriticos, DJ, Emhart, V, Mardones, O, Acevedo, M, Pincheira, M, Stape, J, Fox, T. (2014). Using seasonal measurements to inform ecophysiology: Extracting cardinal growth temperatures for process-based growth models of five *Eucalyptus* species/crosses from simple field trials. *New Zealand Journal of Forestry Science*, *44*, 9. https://doi.org/10.1186/s40490-014-0009-4.

Wear, DN, Dixon IV, E, Abt, RC, Singh, N. (2015). Projecting potential adoption of genetically engineered freeze-tolerant *Eucalyptus* in the United States. *Forest Science*, *61*(3), 466–480. https://doi.org/10.5849/forsci.14-089.

Were, K, Bui, DT, Dick, OB, Singh, BR. (2015). A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecological Indicators*, *52*, 394–403. https://doi.org/10.1016/j.ecolind.2014.12.028.

Wu, C, Shen, H, Shen, A, Deng, J, Gan, M, Zhu, J, Xu, H, Wang, K. (2016). Comparison of machine-learning methods for above-ground biomass estimation based on Landsat imagery. *Journal of Applied Remote Sensing*, *10*, 3. https://doi.org/10.1117/1.JRS.10.035010.

Yamamoto, JK, & Landim, PMB (2013). *Geoestatística: conceitos e aplicações*. São Paulo: Oficina de Textos.

Zhang, J, Huang, S, Hogg, EH, Lieffers, V, Qin, Y, He, F. (2014). Estimating spatial variation in Alberta forest biomass from a combination of forest inventory and remote sensing data. *Biogeosciences*, *11*, 2793–2808. https://doi.org/10.5194/bg-11-2793-2014.