Floresta Ambient., Rio de Janeiro, 2024; 31(3): e20240034
https://doi.org/10.1590/2179-8087-FLORAM-2024-0034
ISSN 2179-8087 (online)

ORIGINAL ARTICLE – Forest Management

Regression Kriging in the Productive Capacity of Planted Forests

Emanuel José Gomes de Araújo[1]
Eduardo Vinicius da Silva[1]
Pedro Vaz[1]
Marco Antonio Monte[1]
Vinicius Augusto Morais[2]
Rafaella de Angeli Curto[1]
Danilo Henrique dos Santos Ataide[1]
Thais Cristian Rosa[3]

[1]Universidade Federal Rural do Rio de Janeiro (UFRRJ), Seropédica, RJ, Brasil.
[2]Universidade do Estado de Mato Grosso (UNEMAT), Alta Floresta, MT, Brasil.
[3]Vallourec Florestal, Curvelo, MG, Brasil.

Abstract

The objective of this study was to evaluate the classification of productive capacity in eucalyptus stands using regression kriging. The study was carried out in 62 stands with clonal plantations of Eucalyptus sp. Data were collected at 24, 36, 48, 60, 72 and 84 months in 170 sampling units of 400 m². The variables diameter at 1.30 m above the ground (DBH), total height (Ht) were measured and, subsequently, the average height of the dominant trees (Hd) was obtained. The classification of productive capacity was carried out using the algebraic difference method, with the models: Schumacher, Chapman and Richards and Bailey and Clutter. The site index was interpolated by applying ordinary kriging and regression kriging. The results indicated that the site index showed spatial dependence in all adjustments. It is concluded that regression models for estimating the site index can be used in combination with regression kriging techniques.

Keywords: Spatial dependence, Geostatistics, Modeling, Site, Production.

1. INTRODUCTION

The classification of productive capacity aims to describe the wood production potential of a specific location and species and can be carried out using indirect or direct classification methods. Indirect methods combine factors such as macroclimate, topography, and soil characteristics to classify site quality without an established stand, allowing for species and management practice recommendations for future projects (Scolforo, 2006). Direct methods allow for the representation of productive potential through a quantitative value, known as the site index, wich is an indicator of site quality, used in forest management, economic viability assessment, and in growth and yield models of forest plantations (Castro et al. 2015; Campos & Leite, 2017; Miranda et al. 2021).

The application of direct methods in the classification of productive capacity is more efficient than indirect methods (Oliveira et al., 2008; Leite et al., 2011), as it uses regression models that relate dendrometric variables to the age of the stand. Among the variables that can be used as indicators of site quality are volume, basal area, total height, or the average height of the dominant trees in the stand (Scolforo, 2006). However, the average height of the dominant trees is considered the variable that provides the most consistent results (Scolforo & Machado, 1988), as it is less influenced by stand density and shows a significant correlation with wood volume (Husch et al., 2003).

In the classification of productive capacity, it is important to consider the spatial structuring of the variable used as the site indicator. Geostatistical techniques are widely used to describe the spatial continuity structure of dendrometric variables in commercial stands, particularly for the species *Eucalyptus* sp. (Mello et al., 2005; Ortiz et al., 2006; Rufino et al., 2006; Mello et al., 2009; Guedes et al., 2012; Guedes et al., 2015; Lundgren et al., 2017). These studies demonstrate the strong spatial dependence that the average height of dominant trees exhibits in commercial stands of *Eucalyptus* sp. Therefore, it is important that the classification of productive capacity considers the spatial structuring of the site index. Among the existing geostatistical methods, ordinary kriging is the most commonly used option, as it generates unbiased estimates for unsampled locations using parameters obtained from semivariographic analysis (Isaaks & Srivastava, 1989; Yamamoto & Landim, 2013; Ataíde et al., 2020).

Another technique that stands out among spatial interpolators is regression kriging (Keskin & Grunwald, 2018). This technique is considered a hybrid interpolator, combining regression models with geostatistics. In forestry studies, this technique has been used to map certain forest attributes over large areas, such as aboveground biomass (Scolforo et al., 2015; Scolforo et al., 2016; Morais et al., 2020). Regression kriging can also be applied in the classification of productive capacity in forest stands, in the spatialization of variables such as dominant height, volume, and mean annual increment (Guedes et al., 2015), as well as the site index (Palmer et al., 2009), demonstrating improved accuracy of spatial estimates compared to other interpolators.

When methods that relate dominant height and stand age in the classification of productive capacity are combined with regression kriging techniques, it becomes possible to predict the site index in unsampled locations, along with the associated uncertainty levels. This is essential for mapping the productive potential of forest stands, generating important spatial information for management, allowing for planning and localized silvicultural interventions according to the mapped site, which enables the reduction of costs and time in these practices. Therefore, the objective of this study was to evaluate the classification of productive capacity in eucalyptus stands using regression kriging.

2. MATERIALS AND METHODS

2.1. Study area

The study was conducted in 2,119 hectares planted with the same clone of *Eucalyptus* sp., located in the municipality of Bocaiúva, Minas Gerais, Brazil (Figure 1). The region's climate is classified as Aw according to the Köppen classification, characterized by a humid tropical savanna climate with dry winters and rainy summers (Alvares et al., 2013). The area has an average altitude of 820 m, with annual average precipitation and temperature of 1,246 mm and 24°C, respectively, and the predominant soils are Dark Red Latosol or Dystrophic Yellow Latosol (Caldeira et al., 2005).

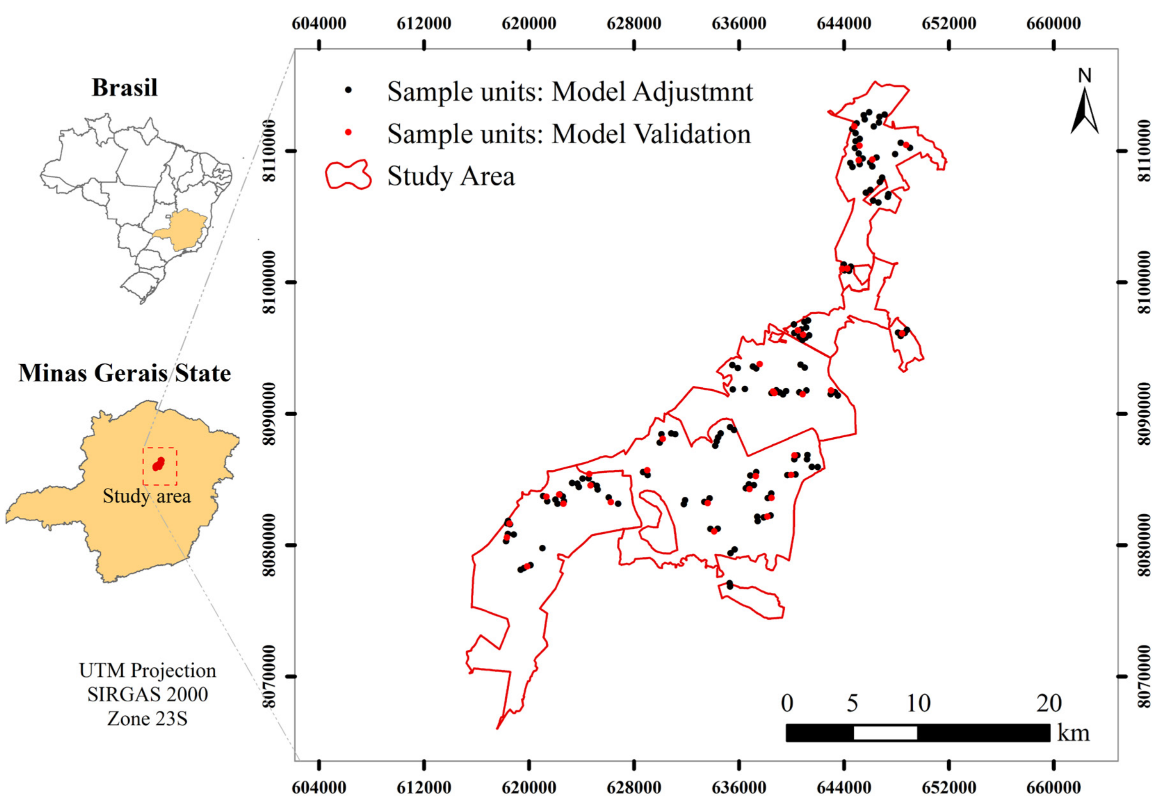

Figure 1. Study area and distribution of sample units in clonal eucalyptus plantations, in the municipality of Bocaiúva, MG, Brazil.

2.2. Data collection and description

The data were collected in 170 sample units, each square with an area of 400m², randomly allocated within 62 stands of the plantations, all with the same clone and initial spacing. The sampling intensity used was one sample unit per 12.46 hectares (0.32% of the total area). In all sample units, the diameter at 1.30 m above ground (DBH) and total height (Ht) were measured, and subsequently, the average height of the dominant trees (Hd) was obtained, following Assmann's concept (1970), at ages 24, 36, 48, 60, 72, and 84 months. The sample units were georeferenced using navigation GPS to obtain their central geographic coordinates and were then divided into two groups: 136 were used for fitting the models of productive capacity classification, and 34 were used for predictive validation, with at least one unit selected for validation per stand, when possible. The division of data into a 70:30 ratio for training and validation of models has been reported to assess predictive capacity of mathematical models in forest studies (Castro et al. 2015; Silveira et al. 2019; Souza et al. 2019). For the variables DBH, Ht, and Hd, maximum, minimum, mean, standard deviation, and coefficient of variation in percentage were obtained. Additionally, the Shapiro-Wilk normality test (α = 5%) was conducted for these dendrometric variables from 24 to 84 months of age.

2.3. Classification of productive capacity

Productive capacity was evaluated using the algebraic difference method, with a reference age of 72 months, The productive capacity was evaluated using the algebraic difference method, with a reference age of 72 months, selected for being close to the technical harvesting age for eucalyptus (Campos & Leite, 2017).

Three regression models were fitted to estimate dominant height as a function of age (Table 1). The models were rearranged to isolate the site variable as an independent variable, allowing for the estimation of the site index, as described by Campos & Leite (2017).

Table 1. Regression models used for site index estimation.

Model	Author	Rearrangement
$Ln(Hd_i) = \beta_0 + \beta_1\left(\frac{1}{I}\right) + \varepsilon_i$	Schumacher (1939)	$S = Hd\, e^{\left[\hat{\beta}_1\left(\frac{1}{I} - \frac{1}{I_{ref}}\right)\right]}$
$Hd_i = \beta_1\left(1 - e^{-\beta_2 I}\right)^{\left(\frac{1}{1-\beta_3}\right)} + \varepsilon_i$	Chapman and Richards (1959)	$S = Hd\left[\frac{1 - e^{(-\hat{\beta}_1 I_{ref})}}{1 - e^{(-\hat{\beta}_1 I)}}\right]^{\hat{\beta}_2}$
$Ln(Hd_i) = \beta_0 + \beta_1\left(\frac{1}{I}\right)^{\beta_2} + \varepsilon_i$	Bailey and Clutter (1974)	$S = \frac{Hd}{e^{\hat{\beta}_1\left(I^{-\hat{\beta}_2} - I_{ref}^{-\hat{\beta}_2}\right)}}$

Ln = natural logarithm; Hd_i = average height of dominant trees (m); β_0, β_1, β_2, and β_3 = model parameters; I = stand age (months); S = site index (m); I_{ref} = reference age (months); e = exponential; ε_i = residual.

All analyses were conducted in R software (R Core Team, 2015), where the fittings were performed using the nlstools package (Baty et al., 2015).

The fitted equations were evaluated using the adjusted coefficient of determination, absolute and percentage standard error of estimation, root mean square error, percentage mean prediction error, and graphical analysis of normalized residuals, in addition to comparing estimated and observed values (Campos & Leite, 2017). We also assessed coefficient significance using the t-test and checked residual normality using the Shapiro-Wilk test, both at a 5% significance level.

The spatial continuity structure of the site index was evaluated through the experimental semivariogram (Equation 1). For this purpose, theoretical models including exponential (Equation 2), spherical (Equation 3), and Gaussian (Equation 4) were fitted using the Maximum Likelihood Method, employing the GeoR package (Ribeiro Júnior & Diggle, 2001) in the R software (R Core Team, 2015).

$$\hat{\gamma}(h) = \frac{1}{2N(h)}\sum_{i=1}^{N(h)}\left[Z(x_i) - Z(x_i + h)\right]^2 \quad \text{Eq. 1}$$

$$\hat{\gamma}(h) = C_0 + C\left[1 - e^{\left(-\frac{h}{a}\right)}\right] \quad \text{Eq. 2}$$

$$\hat{\gamma}(h) = C_0 + C\left[1{,}5\frac{h}{a} - 0{,}5\left(\frac{h}{a}\right)^3\right] \quad \text{se: } h < a \quad \text{Eq. 3}$$

$$\hat{\gamma}(h) = C_0 + C \quad \text{se: } h \geq a$$

$$\hat{\gamma}(h) = C_0 + C\left[1 - e^{\left(-\frac{h}{a}\right)^2}\right] \quad \text{Eq. 4}$$

Where: $\hat{\gamma}(h)$ = semivariance; $Z(x_i)$ = value of the regionalized variable at point x; $Z(x_i + h)$; = value of the variable at point $x + h$; $N(h)$; = number of pairs separated by distance h; C_0; = nugget effect C; = contribution; a = range.

The criteria used for selecting the best semivariogram model included: Akaike Information Criterion (AIC), standard deviation of reduced errors, reduced mean error, and cross-validation. The Spatial Dependency Index (SDI) was obtained for each of the theoretical models fitted to the semivariograms, with classification as weak spatial dependence (SDI < 0.25), moderate (0.25 < SDI < 0.75), and strong (SDI > 0.75) (Zimback, 2003).

Having established the spatial dependence of the site index, ordinary kriging (Equation 5) was performed to obtain spatial estimates at unsampled locations. For each method of spatialization and site index correction, interpolation was conducted using 4, 8, 12, and 16 neighboring points, corresponding to the closest sample units to the point of interest being estimated.

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \quad \text{Eq. 5}$$

Where: $\hat{Z}(x_0)$ = estimate at point x_0; $Z(x_i)$ = estimate at point x_i; n = number of sample points; λ_i = weight assigned to the point by kriging.

The site index values, corrected by ordinary kriging considering different numbers of neighbors, were compared with those obtained from regression models to assess the performance of kriging estimates and the effect of the number of neighbors used in spatialization. For this purpose, were used the values of root mean square error, percentage mean prediction error, and Pearson correlation coefficients (Equation 6), Willmott's concordance index (Equation 7), and performance index (Equation 8).

$$r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \quad \text{Eq. 6}$$

$$d = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(|\hat{y}_i - \bar{y}| + |y_i - \bar{y}|)^2} \quad \text{Eq. 7}$$

$$c = dr \quad \text{Eq. 8}$$

Where: r = Pearson correlation coefficient; d = Willmott's concordance index; c = performance index; y_i = i-th value of the site index estimated by regression models; $\bar{y}$ = mean of the site index estimated by regression models; $\hat{y}_i$ = i-th value of the site index estimated by ordinary kriging; $\bar{\hat{y}}$ = mean of the site index estimated by ordinary krigin.

The best regression model was selected to perform two-stage regression kriging: first, the normality of residuals estimated by the regression models was verified using the Shapiro-Wilk test (α=5%). Subsequently, the spatial dependence structure of these residuals was analyzed, and their spatialization was conducted using ordinary kriging. Once the spatial dependence of the residuals was confirmed, regression kriging as described by Odeh et al. (1994) was applied. For this purpose, interpolation maps obtained by ordinary kriging of the variable of interest and residuals, both in *raster* format with a pixel size of 20 m x 20 m, were summed, generating estimates using the regression kriging technique (Equation 9).

$$\hat{Z} = \hat{Z}r + \hat{\varepsilon} \quad \text{Eq. 9}$$

Where: $\hat{Z}$ = variable resulting from regression kriging type B; $\hat{Z}r$ = study variable estimated by the regression model and interpolated by ordinary kriging; $\hat{\varepsilon}$ = residual generated by the regression model and interpolated by ordinary kriging.

3. RESULTS AND DISCUSSION

3.1. Descriptive statistics

Descriptive statistics and the normality test of the variables diameter at breast height (DBH), total height (Ht), and average height of dominant trees (Hd) are presented in Table 2. The Shapiro-Wilk normality test (α=5%) indicated rejection of the normality assumption for the data at the initial ages (24, 36, and 48 months), showing that in the early growth phase of the stand, the analyzed variables exhibit asymmetric distribution, with a higher proportion of thinner and shorter trees. This relationship was also observed by Castro et al. (2016) when fitting a generalized gamma function to diameter distribution data from a eucalyptus stand in Bahia. At the initial ages assessed, the curve showed asymmetry, with a greater number of individuals in the smaller diameter classes; however, over time, the diameter distribution became similar to a normal distribution.

In the initial phase of the stand, there is also greater variation in these attributes, noticeable through the coefficient of variation, which is higher and reduces with age progression, tending to stabilize after 48 months. This characteristic reflects competition for resources, where in the initial phase, competition is more intense, resulting in these greater variations (Fox et al., 2001). The variables DBH, Ht, and Hd assumed an approximately normal distribution from 60 months onward.

Table 2. Descriptive statistics and Shapiro-Wilk normality test results for dendrometric variables measured at ages 24 to 84 months in eucalyptus stands.

Age (months)	Variable	Minimum	Maximum	$\bar{x}$	*s*	CV%	*p*
24	DBH (cm)	5.32	11.11	9.5	1.24	13.40	$1.61x10^{-08}$*
	Ht (m)	5.52	14.29	11.39	1.75	15.33	$7.01x10^{-06}$*
	Hd (m)	6.20	14.92	12.01	1.67	13.93	$3.27x10^{-05}$*
36	DBH (cm)	8.32	14.03	12.29	0.97	7.88	$2.52x10^{-06}$*
	Ht (m)	10.70	21.53	17.37	1.98	11.39	$1.83x10^{-05}$*
	Hd (m)	12.12	22.48	18.19	1.96	10.78	$2.73x10^{-05}$*
48	DBH (cm)	9.62	15.87	13.68	1.00	7.32	$6.29x10^{-03}$*
	Ht (m)	13.18	24.77	21.08	2.22	10.55	$8.37x10^{-04}$*
	Hd (m)	14.52	26.60	22.01	2.17	9.87	$3.05x10^{-02}$*
60	DBH (cm)	11.22	16.71	14.35	0.96	6.70	$2.96x10^{-01}$
	Ht (m)	14.94	27.54	22.98	2.18	9.50	$6.07x10^{-02}$
	Hd (m)	16.56	29.22	24.10	2.17	9.01	$4.50x10^{-01}$
72	DBH (cm)	11.82	17.52	14.81	1.03	6.96	$8.45x10^{-01}$
	Ht (m)	17.38	30.05	24.38	2.40	9.84	$6.40x10^{-01}$
	Hd (m)	17.74	32.28	25.62	2.44	9.52	$4.91x10^{-01}$
84	DBH (cm)	12.45	17.72	15.10	1.00	6.60	$8.93x10^{-01}$
	Ht (m)	18.20	31.32	25.29	2.45	9.70	$8.00x10^{-01}$
	Hd (m)	19.18	33.64	26.62	2.41	9.05	$9.50x10^{-01}$

$\bar{x}$= mean; *s* = standard deviation; CV% = coefficient of variation (%); *p* = Shapiro-Wilk normality test p-value (α = 5%); * = rejection of the normality hypothesis for the data.

3.2. Regression model fitting for productive capacity classification

The coefficients of the productive capacity classification models were significant by the t-test ($p < 0.05$) for all adjustments performed (Table 3). Additionally, the adjusted coefficient of determination ($R^2_{aj.}$) values were greater than 0.80, and the standard error of estimate (S_{yx}%) values were less than 10%, indicating the high quality of the fitted equations.

Table 3. Parameters and fitting statistics of the models to estimate dominant height as a function of age, using the algebraic difference (AD) method.

Model	$\hat{\beta}_1$	$\hat{\beta}_2$	$R^2_{aj.}$	S_{yx}%	RMSE	bias%
Schumacher	23.03*		0.840	6.440	1.498	0.657
Chapman and Richards	0.036*	1.19*	0.818	6.858	1.594	2.913
Bailey and Clutter	-14.86*	0.81*	0.843	6.363	1.479	0.158

$\hat{\beta}_i$= coefficients of the models; $R^2_{Aj.}$= adjusted coefficient of determination; S_{yx}% = standard error of estimate in percentage; RMSE = root mean square error; bias% = mean prediction error in percentage; * = significant value at 5% significance level by t-test.

In the evaluation of the models, all showed good results in the fittings, with slight superiority for the Bailey and Clutter model, mainly because it presented the smallest absolute values of , RMSE and bias%, com 6.363, 1.479 e 0.158, respectively. Similar results were obtained by Kitikidou et al. (2011) when fitting the Bailey and Clutter models and the Chapman and Richards models in a *Pinus brutia* stand using the algebraic difference method. The similarity of the fitted models is supported by the graphical distribution of normalized residuals (Figure 2). Similar outcomes were also reported by Pissinin & Schneider (2017), applying the guide curve and algebraic difference methods in a eucalyptus stand, where the algebraic difference method showed homogeneous distribution of residuals and lower errors.

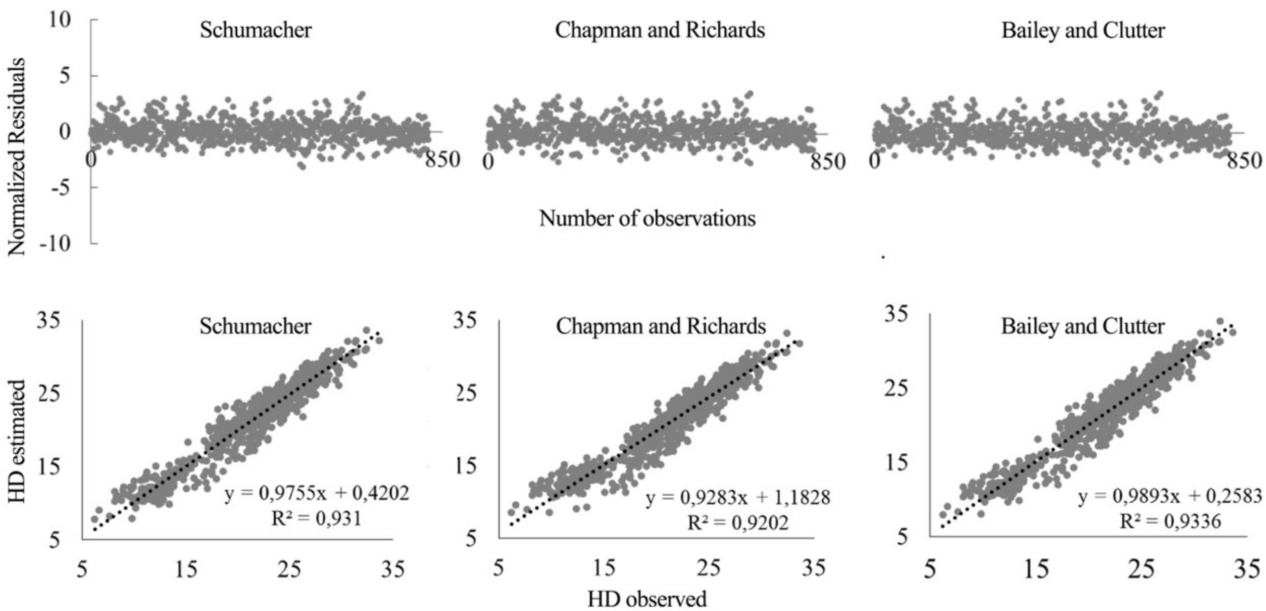

Figure 2. Graphs of normalized residuals and estimated dominant height as a function of observed height for the fitted equations.

When evaluating the trend lines of the graphs of estimated values versus observed values, all showed proximity to a 45° slope and intersected at the origin of the Cartesian axis (Figure 2). Values close to 1 were verified for the slope coefficients (B_1) and determination coefficients (R^2), exceeding 0.84, and low values for the intercept coefficients (B_0), less than 3.21. The determination coefficient describes precision, while the intercept and slope coefficients describe the horizontal offset and slope of the trend line, respectively, related to the accuracy of the estimates. Ideal values for the slope coefficient and determination coefficient of the trend line are close to 1. With these values, the determination coefficient indicates higher precision, and the slope coefficient indicates greater accuracy, represented graphically by a trend line slope close to 45° (B_1=1). Furthermore, for the trend line analysis, ideally, it should not show displacement on the Cartesian plane, meaning the intercept coefficient should be zero (B_0=0). These values indicate greater similarity between the estimates and observed values of the variable of interest, thus demonstrating higher precision and accuracy of the method and the fitted model. Therefore, the Bailey and Clutter model demonstrated superiority over the others, as it showed a slope coefficient of 0.9893, determination coefficient of 0.9336, and the lowest intercept coefficient of 0.2583.

3.3. Classification of productive capacity using geostatistics

The average height of dominant trees generally shows moderate to strong spatial dependence for different species, such as *Eucalyptus* sp. (Mello et al., 2005; Guedes et al., 2015), *Tectona grandis* (Pelissari et al., 2015), and *Pinus* sp. (Zech et al., 2018). Previous studies have confirmed this trend, demonstrating the consistency of spatial dependence across various species. In the present study, the spatial dependence of the site index ranged from moderate to strong, with high SDI% values, regardless of the regression model used. These results are consistent with existing literature and indicate that the use of geostatistical interpolators is viable for generating spatial estimates of the site index at different evaluated ages.

It was observed that the gaussian, exponential, and spherical models provided good fits to the experimental semivariogram of the site index variable. The selection of the best model was based on the lowest AIC value as the primary criterion, following the approach outlined by Burnham & Anderson (2002). Models are considered statistically different if the difference in AIC values is greater than 2. Thus, it was found that most models did not exhibit statistical differences, and overall, the Gaussian model showed the best fits across most ages and procedures for site index classification. This decision was made based on the smallest standard error of the reduced error (S_{er}).

The highest nugget effect values for the selected models were found at 24, 36, and 48 months of age (Table 4). A similar characteristic was observed by Guedes et al. (2015) when assessing the spatial dependence structure of dominant height, volume, and annual mean increment (AMI) in a eucalyptus stands. In their study, the authors noted that the nugget effect was also higher in the initial measurement taken at 2.7 years of age. They emphasize that the nugget effect represents random error that cannot be explained by spatial analysis, often attributed to the proximity between sample units and high data variability in the initial phase of plantation establishment.

Table 4. Parameters, adjustment statistics, and spatial dependence index.

Model	Age	Theoretical Model	C_0	C	a	$\overline{ER}$	S_{er}	AIC	SDI %
Schumacher	24	Gaussian	3.24	13.62	2,242	0.0073	1.031	666.5	80.77
	36	Gaussian	1.80	5.43	1,002	0.0043	1.016	607.9	75.05
	48	Exponencial	1.75	3.28	2,530	0.0019	1.042	564.4	65.24
	60	Gaussian	1.13	3.30	1,112	0.0003	1.036	540.8	74.49
	72	Gaussian	1.29	4.39	1,470	-0.0022	1.039	553.6	77.31
	84	Gaussian	1.41	3.81	1,754	-0.0026	1.025	541.0	72.97
Chapman and Richards	24	Gaussian	2.44	9.31	2,247	0.0072	1.038	623.0	79.25
	36	Gaussian	1.65	5.03	968	0.0048	1.016	598.7	75.24
	48	Exponential	1.66	3,.25	2,428	0.0021	1.042	560.7	66.20
	60	Gaussian	1.11	3.34	1,121	0.0000	1.036	539.9	75.13
	72	Gaussian	1.28	4.60	1,464	-0.0020	1.039	556.1	78.19
	84	Gaussian	1.44	4.09	1,729	-0.0024	1.024	546.7	73.96
Bailey and Clutter	24	Gaussian	3.27	13,39	2,255	0.0072	1.033	666.1	80.37
	36	Gaussian	1.91	5.73	1,010	0.0044	1.016	615.0	74.97
	48	Exponential	1.84	3.31	2,643	0.0017	1.042	567.6	64.28
	60	Gaussian	1.15	3.25	1,103	0.0000	1.037	541.4	73.89
	72	Gaussian	1.29	4.25	1,474	-0.0021	1.040	551.8	76.67
	84	Gaussian	1.39	3.65	1.770	-0.0026	1.025	537.5	72.38

C_0: the nugget effect; C the contribution; a: range; $\overline{ER}$: the mean reduced error; S_{er}: standard deviation of the reduced error; AIC: Akaike Information Criterion; SDI%: spatial dependence index in percentage.

Among the models with the best fit, the gaussian model at ages 36, 60, 72, and 84 months exhibited range values ranging from 968 to 1,770 m. However, at 48 months, where the exponential models showed superior fit, the range values were greater than 2,248 meters. This information indicates the average distance over which the site index correlates spatially (Webster & Oliver, 2007). Some studies conducted in eucalyptus plantations have reported range values ranging from 150 to 450 meters for variables such as volume, total height, and dominant height (Mello et al., 2005; Mello et al., 2009; Guedes et al., 2015). However, these studies were conducted in areas smaller than 1,100 hectares, using higher sampling intensity and systematic sampling. The results obtained in the present study demonstrate that it is possible to capture the spatial structure of site index using geostatistics in extensive areas, with lower sampling intensity and using simple random sampling. Therefore, this information is highly relevant for defining sampling methods and processes in forest inventories for geostatistical studies (Mello et al., 2005; Pelissari et al., 2014), potentially reducing costs associated with lower sampling intensity.

3.3.1. Spatialization of the site index via regression krigin

After confirming the strong spatial dependence of the site index, ordinary kriging was performed to estimate and spatialize this variable in the unsampled locations of the stands. Consequently, the site index values generated by the regression models were also estimated using ordinary kriging. These two procedures were carried out using 4, 8, 12, and 16 neighboring sample units.

At the initial ages of the stands, where there is greater heterogeneity among the sample units, a larger difference was observed between the values estimated by ordinary kriging compared to those estimated by the models, increasing the standard deviation of the site index estimates. The mean prediction error percentage showed negative values, highlighting that regardless of the number of neighbors, the application of ordinary kriging to correct the site index obtained by regression models tends to overestimate the variable on average.

When the average distance between 4 and 8 neighbors was examined, values of 1,754 and 2,827 meters, respectively, were found. These values are similar to the range of the selected theoretical semivariogram models, which vary between 968 and 2,242 meters. However, when analyzing 12 and 16 neighbors, the values of 3,780 and 4,378 meters, respectively, were found, demonstrating that the average distance between neighbors exceeds the distance at which the site index exhibits spatial structuring. This information is of great importance, especially for determining the number of neighbors used in the interpolation.

Webster & Oliver (2007) described that the weight each neighboring point receives when estimating an unsampled point is related to the distance between them, meaning that more distant points receive lower weights, while closer points receive higher weights. The same authors highlighted that there are two ways to determine which points will be influential: by determining the number of neighboring points or through a predetermined distance, which will be the radius of influence of the point to be estimated. If neighboring points do not fit these criteria, they receive a weight of zero and do not influence the estimation of the point.

Another important piece of information related to the number of neighbors is the smoothing effect, which according to Yamamoto & Landim (2013), is a disadvantage of the ordinary kriging interpolator. In this case, the effect is greater on extreme values, tending to overestimate the lower values and underestimate the higher values. This effect is amplified when using large numbers of neighbors with varying distances. Therefore, Yamamoto (2005) recommends using the smallest possible number of neighboring points to avoid the smoothing effect of ordinary kriging. Based on these results, it was found that interpolation by ordinary kriging performed with 4 or 8 neighbors reduces the influence of distant points where spatial correlation is weak or non-existent.

To perform regression kriging, the normality and spatial structure of the residuals at 72 months of age were evaluated for the Schumacher, Chapman and Richards, and Bailey and Clutter models (Table 5). The residuals from each model showed normality according to the Shapiro-Wilk test (α=5%), with p-values greater than 0.05. In the analysis of semivariograms of the residuals, the spherical theoretical model exhibited the best parameters across all fits, with spatial dependence index values exceeding 90%, indicating strong spatial dependence (Zimback, 2003). Consequently, it was possible to spatialize the error obtained from the regression models by applying ordinary kriging.

Table 5. Parameters, fit statistics, and spatial dependence index of theoretical semivariogram models adjusted for residuals from algebraic difference models.

Mod (S)	Mod (sem)	C_0	C	a	$\overline{ER}$	S_{er}	AIC	SDI %
S	Esf	0.10	1.20	1,810.13	-0.0044	1.0460	382.6	92.09
	Exp	0.00	1.27	733.33	-0.0037	1.0107	385.3	100.00
	Gaus	0.32	0.97	870.70	-0.0090	1.0723	383.8	75.41
CR	Esf	0.08	1.25	1,821.52	-0.0039	1.0512	383.4	93.75
	Exp	0.00	1.30	760.64	-0.0044	1.0080	386.4	100.00
	Gaus	0.31	1.01	884.30	-0.0112	1.0614	384.7	76.47
BC	Esf	0.11	1.18	1,795.08	-0.0021	1.0490	382.4	91.68
	Exp	0.00	1.26	725.59	-0.0035	1.0085	384.9	100.00
	Gaus	0.32	0.95	865.28	-0.0102	1.0579	383.5	75.01

Mod (S): site index models; S: Schumacher; CR: Chapman and Richards; BC: Bailey and Clutter; Mod (sem): theoretical semivariogram models; Esf: spherical model; Exp: exponential model; Gaus: Gaussian model; Co: nugget effect; C: contribution; a: range; $\overline{ER}$: mean reduced error; S_{er}: standard deviation of reduced error; AIC: Akaike information criterion; SDI%: spatial dependence index in percentage. Lines filled with gray color refer to the best-fit model.

The procedure of regression kriging was carried out by adding the values of site index interpolated by ordinary kriging to the regression model errors, after these were spatialized (Figures 3 and 4). The regression kriging maps were generated in raster format, requiring the conversion of the site index and residual maps to raster format first, followed by the addition of these maps.

The use of 8 neighboring pairs resulted in a slight smoothing effect on the maps, both for the site index and the interpolated error, influencing the outcome of regression kriging. This relationship can be observed in the maximum and minimum values present in the maps, where there is a reduction in their amplitude approaching the average of the variables. Through the interpolation of residuals, it was possible to identify areas with higher estimation errors and whether there is a tendency to overestimate or underestimate the site index. It is noticeable that areas with higher site indices exhibit larger errors, whereas areas with lower indices have smaller errors, which is related to the smoothing effect. The smoothing effect is a characteristic commonly observed in ordinary kriging; high values are generally underestimated, while low values tend to be overestimated (Yamamoto & Landim, 2013).

Thus, regression kriging partially corrected the overestimations and underestimations generated by ordinary kriging. A similar result was found by Knotters et al. (1995), where regression kriging reduced the smoothing effect compared to ordinary kriging and cokriging. The authors also emphasized that this benefit of regression kriging is related to its contribution to the estimate generated by the models.

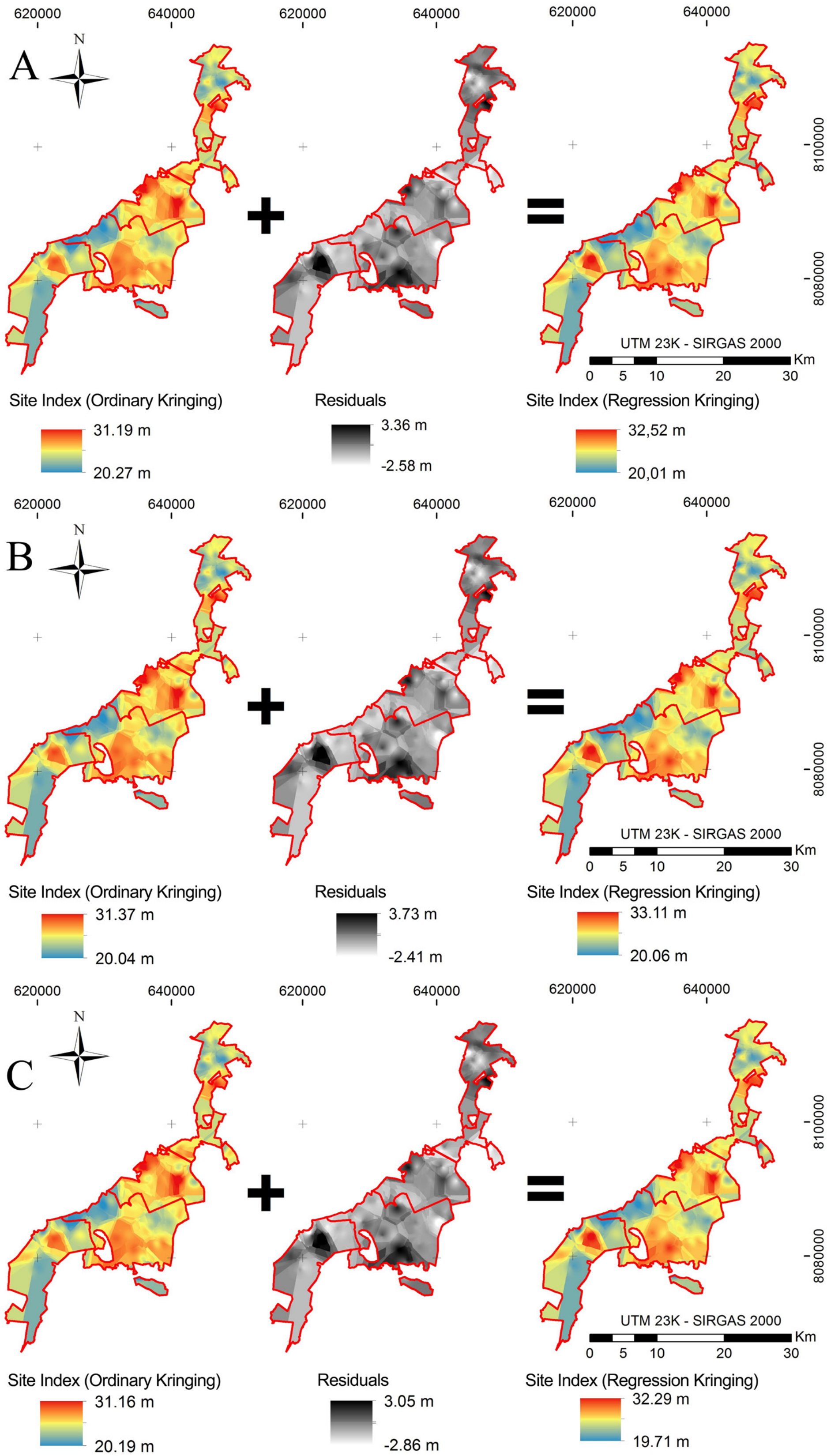

Figure 3. Spatialization of the site index using 4 neighbors in regression kriging for the Schumacher model (A), Chapman and Richards model (B), and Bailey and Clutter model (C), adjusted by the algebraic difference method, for 72-month-old eucalyptus stands.

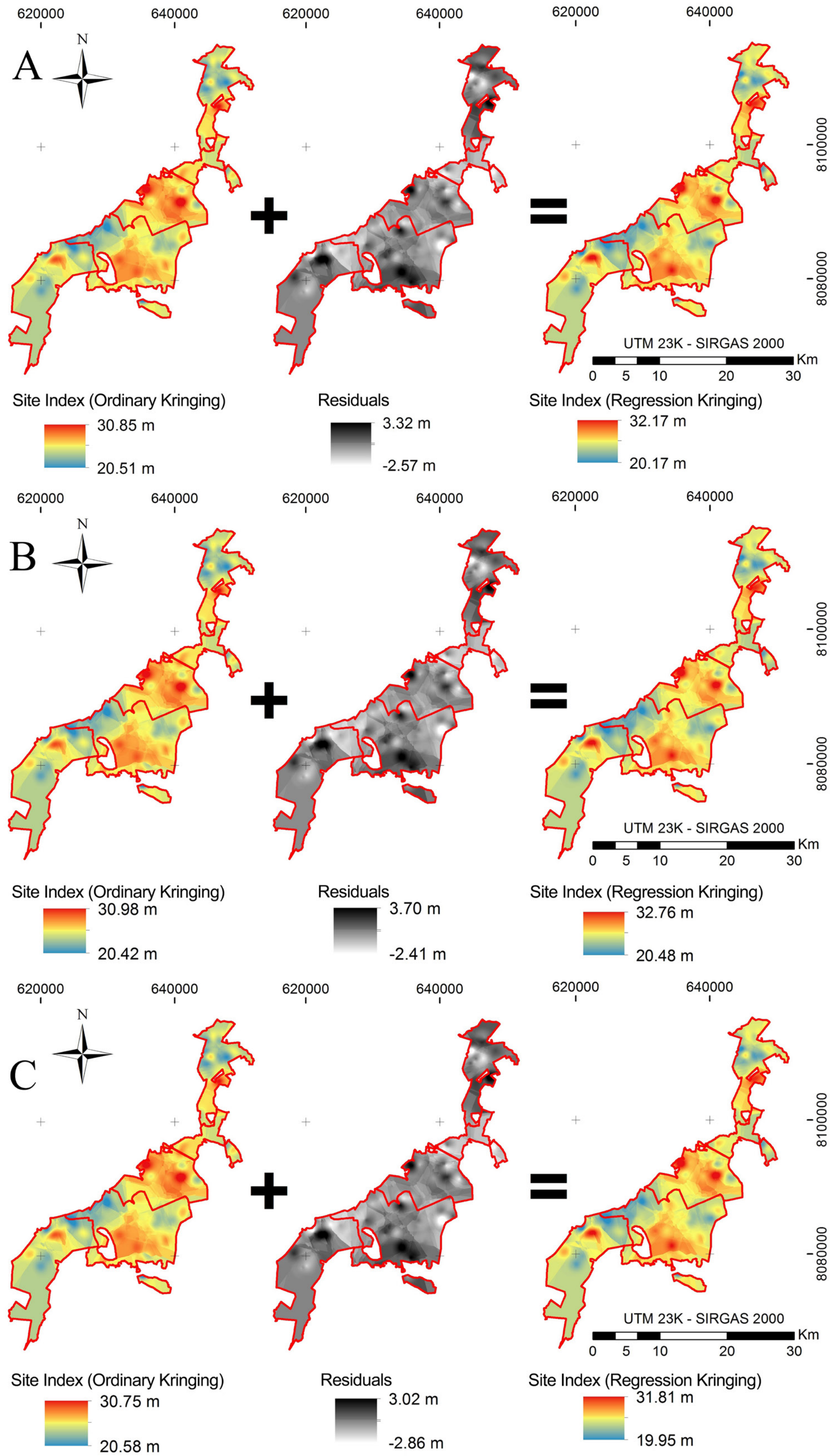

Figure 4. Spatialization of the site index using 8 neighbors in regression kriging for the Schumacher model (A), Chapman and Richards model (B), and Bailey and Clutter model (C), adjusted by the algebraic difference method, for 72-month-old eucalyptus stands.

3.3.2. Validation of regression kriging

When analyzing the relationship between estimated values of the site index by regression kriging and the dominant height of validation units (Figure 5), it was possible to observe improved precision and accuracy provided by regression kriging. The slope of the trend line in the validation plots for regression kriging reached values between 0.84 and 0.94, which were higher than those achieved by ordinary kriging, ranging between 0.6 and 0.7. There was also a reduction in the intercept of the trend lines, indicating that this technique improved the accuracy of the estimates. Similar to ordinary kriging, the use of 8 neighbors resulted in inferior results regardless of the model used.

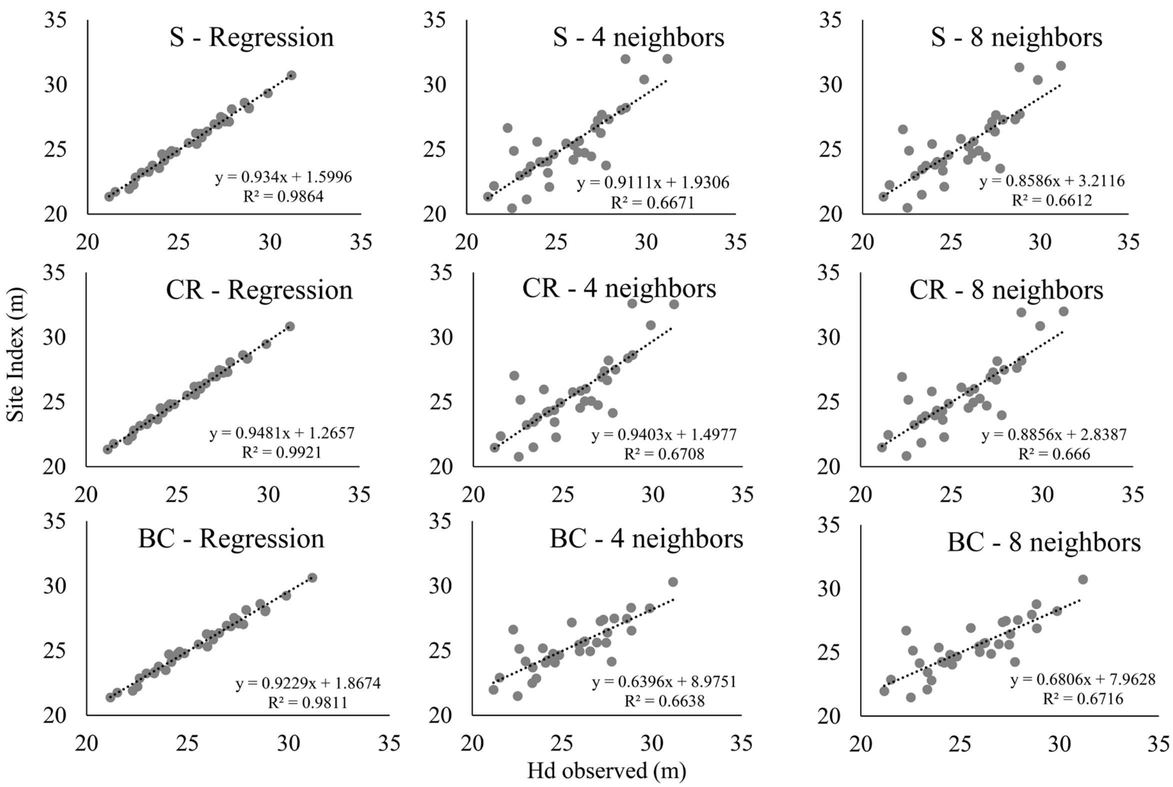

Figure 5. Validation of site index estimates by regression models and regression kriging using 4 and 8 neighbors, at 72 months of age. Where S, CR, and BC represent the Schumacher, Chapman and Richards, and Bailey and Clutter models, respectively.

The regression models outperformed regression kriging in estimating the site index in validation units. The Chapman and Richards model was the most efficient, with a coefficient of determination of 0.9921, slope of the trend line (coefficient angular) of 0.9481, and intercept of the trend line (coefficient linear) of 1.2657. Despite the regression models being more accurate in estimating the site index in validation units, regression kriging with 4 neighbors proved efficient in spatializing the productive potential, surpassing ordinary kriging.

4. CONCLUSIONS

The site index showed strong spatial dependence across all ages when estimated by the Schumacher, Chapman and Richards, and Bailey and Clutter models using the Algebraic Difference method. Among the models tested, Bailey and Clutter's model produced the most accurate site index estimates.

The application of regression kriging technique was effective in spatializing and estimating the site index in stands, with higher precision and accuracy observed when using 4 neighboring points at ages older than 60 months.

Regression models for site index estimation can be applied in conjunction with regression kriging techniques, enabling refined classification of productive capacity and obtaining the site index in unsampled locations.

SUBMISSION STATUS

Received: 25 Jun. 2024

Accepted: 30 Set. 2024

Associate editor: Fernando Gomes

CORRESPONDENCE TO

Eduardo Vinicius da Silva

BR465, KM 07, 23890-000, Seropédica, RJ, Brasil.

e-mail: eduvini@gmail.com

AUTHORS' CONTRIBUTIONS

Emanuel José Gomes de Araújo: Conceptualization (lead), Data curation (lead), Formal analysis (lead), Funding acquisition (lead), Investigation (lead), Methodology (lead), Project administration (lead), Resources (lead), Supervision (lead), Validation (lead), Writing - original draft (lead), Writing - review & editing (lead).

Eduardo Vinicius Silva: Conceptualization (lead), Formal analysis (lead), Investigation (lead), Resources (lead), Writing - original draft (lead), Writing - review & editing (lead).

Pedro Vaz: Conceptualization (lead), Data curation (lead), Formal analysis (lead), Investigation (lead), Methodology (lead), Writing - original draft (lead), Writing - review & editing (lead).

Marco Antonio Monte: Conceptualization (lead), Formal analysis (lead), Investigation (lead), Methodology (lead), Resources (lead), Supervision (lead), Validation (lead), Writing - original draft (lead), Writing - review & editing (lead).

Vinicius Augusto Morais: Conceptualization (lead), Formal analysis (supporting), Investigation (supporting), Resources (supporting), Validation (supporting), Writing - original draft (supporting), Writing - review & editing (supporting).

Rafaella de Angeli Curto: Conceptualization (lead), Formal analysis (lead), Investigation (lead), Supervision (lead), Validation (lead), Writing - original draft (lead), Writing - review & editing (lead).

Danilo Henrique dos Santos Ataide: Conceptualization (lead), Data curation (lead), Formal analysis (lead), Investigation (lead), Methodology (lead), Resources (lead), Writing - original draft (lead), Writing - review & editing (lead).

Thais Cristian Rosa: Conceptualization (lead), Data curation (lead), Formal analysis (lead), Investigation (supporting), Validation (lead), Writing - original draft (supporting), Writing - review & editing (supporting).

REFERENCES

Alvares CA, Stape JL, Sentelhas PC, Gonçalves JLM, Sparovek G. Köppen›s climate classification map for Brazil. Meteorologische Zeitschrift 2013; 22(6): 711-728.

Assmann E. The principles of forstry yield study. Oxford: Pergamon Press; 1970.

Ataíde DHS, Araújo EJG, Monte MA, Mendonça BAF, Silveira CE, Albuquerque AM, Silva LC. Métodos geoestatísticos e determinísticos na espacialização da altura dominante em povoamento de eucalipto. Revista de Ciências Agrárias 2020; 43(1): 98-108.

Baty F, Ritz C, Charles S, Brutsche M, Flandrois JF, Muller MLD. A Toolbox for Nonlinear Regression in R: The Package nlstools. Journal of Statistical Software 2015; 66(5): 1-21.

Burnhan KP, Anderson DR. Model selection and multimodel inference. 2 ed. New York: Springer; 2002.

Castro RVO, Cunha AB, Silva LV, Leite HG, Silva AAL. Modelagem do crescimento e produção para um povoamento de Eucalyptus utilizando dois métodos para quantificação do índice de local. Scientia Forestalis 2015;43:83-90.

Caldeira MA, Zanetti R, Moraes JC, Zanuncio JC. Spatial distribution of leaf-cutting ant nests (Hymenoptera: Formicidae) in eucalyptus plantations. Cerne 2005; 11(1): 34-39.

Campos JCC, Leite HG. Mensuração Florestal: perguntas e respostas. 5 ed. atualizada e ampliada. Viçosa: Editora UFV; 2017.

Castro RVO, Araújo Júnior CA, Leite HG, Castro AFNM, Nogueira GS, Costa ALS. Função Gama Generalizada para descrever a distribuição diamétrica de um povoamento de eucalipto. Revista Floresta 2016; 46(1): 67-73.

Fox JC, Ades PK, Bi H. Stochastic structure and individual tree growth models. Forest Ecology and Management 2001; 154: 261-276.

Guedes ICL, Mello JM, Mello CR, Oliveira AD, Silva ST, Scolforo JRS. Técnicas geoestatísticas e interpoladores espaciais na estratificação de povoamentos de *Eucalyptus* sp. Ciência Florestal 2012; 22(3): 541-550.Guedes ICL, Mello JM, Silveira EMO, Mello CR, Reis AA, Gomide LR. Continuidade espacial de características dendrométricas em povoamentos clonais de *Eucalyptus* sp. avaliada ao longo do tempo. Cerne 2015; 21(4): 527-534.

Husch B, Beers TW, Kershaw Jr. JA. Forest mensuration. 4ª ed. Hoboken, New Jersey: John Wiley & Sons; 2003.

Isaaks EH, Srivastava RM. Applied geostatistics. New York: Oxford University Press; 1989.

Keskin H, Grunwald S. Regression kriging as a workhorse in the digital soil mapper's toolbox. Geoderma 2018; 326: 22-41.

Kitikidou K, Bountis D, Milios E. Site index models for calabrian pine (*Pinus brutia* Ten.) in Thasos island, Greece. Ciência Florestal 2011; 21(1): 125-131.

Knotters M, Brus DJ, Oude Voshaar JH. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. Geoderma 1995; 67(3-4): 227-246.

Leite H, Castro R., Silva A, Binoti D, Castro AF, Binoti M. Classificação da capacidade produtiva de povoamentos de eucalipto utilizando diâmetro dominante. Silva Lusitana 2011; 19(2): 181-195.

Lundgren WJC, Silva JAA, Ferreira RLC. Estimação do volume de eucaliptos por krigagem e cokrigagem no Semiárido Pernambucano. Floresta e Ambiente 2017; 24.

Mello JM, Batista JLF, Oliveira MS, Ribeiro Júnior PJ. Estudo da dependência espacial de características dendrométricas para *Eucalyptus grandis*. Cerne 2005; 11(2): 113-126.

Mello JM, Diniz FS, Oliveira AD, Mello CR, Scolforo JRS, Acerbi Júnior FW. Continuidade espacial para características dendrométricas (número de fustes e volume) em plantios de *Eucalyptus grandis*. Revista Árvore 2009; 33(1): 185-194.

Miranda ROV, Figueiredo Filho A, Costa EA, Fiorentin LD, Kohler SV, Ebling ÂA. Guide curve and equation of differences methods in site classification and their relationship in the description of height in Pinus taeda L. Stands. Scientia Forestalis/Forest Sciences. 2021;49(131):1-12.

Morais VA, Ferreira GWD, Mello JM, Silva CA, Mello CR, Araújo EJG, et al. Spatial distribution of soil carbon stocks in the Cerrado biome of Minas Gerais, Brazil. Catena 2020; 185.

Odeh IOA, Mcbratney AB, Chittleborough DJ. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma 1994; 63(3-4): 197-214.

Oliveira MLR, Leite HG, Nogueira GS, Garcia SLR, Souza AL. Classificação da capacidade produtiva de povoamentos não desbastados de clones de eucalipto. Pesquisa Agropecuária Brasileira 2008; 43(11): 1559-1567.

Ortiz JL, Vettorazzi CA, Couto HTZ, Gonçalves JLM. Relações espaciais entre o potencial produtivo de um povoamento de eucalipto e atributos do solo e do relevo. Scientia Forestalis 2006; (72): 67-79.

Palmer DJ, Höck BK, Kimberley MO, Watt MS, Lowe DJ, Payn TW. Comparison of spatial prediction techniques for developing *Pinus radiata* productivity surfaces across New Zealand. Forest Ecology and Management 2009; 258(9): 2046-2055.

Pelissari AL, Figueiredo Filho A, Caldeira SF, Machado SA. Geoestatística aplicada ao manejo de povoamentos florestais de teca, em períodos pré-desbaste seletivo, no estado do Mato Grosso. Revista Brasileira de Biometria 2014; 32(3): 430-444.

Pelissari AL, Caldeira SF, Figueiredo Filho A, Machado SA. Propostas de mapeamentos da capacidade produtiva de sítios florestais por meio de análises geoestatísticas. Scientia Forestalis 2015; 43(107): 601-608.

Pissinin LZ, Schneider PR. Tendências de crescimento em altura dominante de dois clones de *Eucalyptus saligna* Smith implantados em distintas condições de solo. Ciência Florestal 2017; 27(1): 263-275.

R Core Team. R: A language and environment for statistical computing. Vienna, Austria. R Foundation for Statistical Computing, 2015.

Ribeiro Júnior PJ, Diggle PJ. GeoR: a package for geostatistical analysis. RNEWS 2001; 1(2): 15-18.

Rufino TMC, Thiersch CR, Ferreira SO, Kanegae Júnior H, Fais D. Uso da Geoestatística no estudo da relação entre variáveis dentrométricas de povoamentos de Eucalyptus sp. e atributos do solo. Ambiência 2006; 2(1): 83-93.

Scolforo JRS, Machado SA. Curvas de índice de sítio para plantações de *Pinus taeda* nos estados do Paraná e Santa Catarina. Revista Floresta 1988; 17: 140-158.

Scolforo JRS. Biometria florestal: modelos de crescimento e produção florestal. Lavras: UFLA/FAEPE; 2006.

Scolforo HF, Scolforo JRS, Mello CR, Mello JM, Ferraz Filho AC. Spatial distribution of aboveground carbon stock of the arboreal vegetation in Brazilian Biomes of Savanna, Atlantic Forest and Semi-arid woodland. PLOS ONE 2015; 10(6): 1-20.

Scolforo HF, Scolforo JRS, Mello JM, Mello CR, Morais VA. Spatial interpolators for improving the mapping of carbon stock of the arboreal vegetation in Brazilian biomes of Atlantic forest and Savanna. Forest Ecology and Management 2016; 376: 24-35.

Silveira EMO, Reis AA, Terra MCNS, Withey KD, Mello JM, Acerbi-Júnior FW, Ferraz Filho AC, MELLO CR. Spatial distribution of wood volume in Brazilian savannas. Anais da Academia Brasileira de Ciencias 2019;91:1-12.

Souza GSA de, Soares VP, Leite HG, Gleriani JM, do Amaral CH, Ferraz AS, Silveira MVF, Santos FC, Velloso SGS, Domingues GF, Silva S. Multi-sensor prediction of *Eucalyptus* stand volume: A support vector approach. ISPRS Journal of Photogrammetry and Remote Sensing 2019;156:135-46.

Webster R, Oliver MA. Geostatistics for environmental scientists. 2. ed. WestSussex: John Wiley & Sons Ltd; 2007.

Yamamoto JK. Correcting the smoothing effect of ordinary kriging estimates. Mathematical Geology 2005; 37(1): 69-94.

Yamamoto JK, Landim PMB. Geoestatísitica: conceitos e aplicações. 1. ed. São Paulo: Oficina de Textos; 2013.

Zech FD, Refosco JC, Yoshizumi LT, Abreu ROS. Uso de interpoladores espaciais na estratificação estatística de *Pinus taeda*. Scientia Forestalis 2018; 46(117): 87-96.

Zimback CRL. Geoestatística. 1. ed. Botucatu: UNESP, 2003.