

## GEOSTATISTICS AS A TOOL TO REDUCE THE SAMPLING EFFORT IN FOREST INVENTORIES

Myrcia Minatti<sup>1\*</sup>, Carlos Roberto Sanquetta<sup>1</sup>, Sylvio Péllico Neto<sup>1</sup>,  
Ana Paula Dalla Corte<sup>1</sup>, Vinicius Costa Cysneiros<sup>1</sup>

SAP 25601 Received: 13/08/2020 Accepted: 24/01/2021

Sci. Agrar. Parana., Marechal Cândido Rondon, v. 20, n. 2, apr./jun., p. 174-179, 2021

**ABSTRACT** - Geostatistics is one of the tools applied to investigate the spatial variability of forests to reduce costs and recognize the best productivity areas for planning. This study aimed to test the performance of geostatistical techniques in reducing the sampling effort in forest inventories. For this purpose, we used the height of dominant trees as a discriminator of the homogeneous strata to obtain a better representation of the productivity within the forest stands. We carried out the study in *Pinus taeda* L. stands in the Center-South of Paraná, Brazil, by using plots from a forest inventory allocated with the systematic process. Then, we tested three models to determine the site curves (Schumacher, Chapman-Richards 2, and 3 coefficients) with the thirty-seventh year being the reference age. To model the spatial patterns of the dominant height, we used the ordinary kriging, and, after that, we generated the thematic maps of the site classes. Similarly, we used the indicator kriging which allowed obtaining the probabilities of high, medium, and low productivity sites. The processing of the stratified sampling, with the support of the visual interpretation of the images, allowed us to define five strata according to productivity. Results showed that ordinary kriging is effective in defining the productivity classes. Along with geostatistical techniques, it produces more homogeneous strata and reduces the errors of the forest inventory. Moreover, the best-selected model was the Chapman-Richards (3 coefficients) for the site curves. The exponential model was the best model to identify the best areas of the probability of occurrence of sites with higher productivity. The efficiency of indicative kriging generated thematic maps to delimit the likely locations of the most promising sites. Overall, geostatistics proved to be efficient concerning error when compared to simple random sampling.

**Keywords:** site index curves, indicator kriging, ordinary kriging.

## GEOESTATÍSTICA APLICADA À REDUÇÃO DO ESFORÇO AMOSTRAL EM INVENTÁRIOS FLORESTAIS

**RESUMO** - A geoestatística é uma das ferramentas utilizadas para investigar a variabilidade espacial das florestas, com a finalidade em diminuir custos e observar as melhores áreas de produtividade para o planejamento. O trabalho teve como objetivo testar o desempenho de técnicas geoestatísticas visando à redução do esforço amostral em inventários florestais. Para tanto, utilizou-se, como fonte de modelagem, a variável classe de sítios (altura das árvores dominantes) como discriminadora de estratos homogêneos na floresta, visando melhor representação da produtividade dentro de talhões florestais. O estudo foi realizado no Centro-Sul do Estado do Paraná, em plantios de *Pinus taeda* L. Para tanto, foram utilizadas parcelas de inventário florestal alocadas com o processo sistemático. Foram testados três modelos para determinar as curvas de sítios (Schumacher, Chapman-Richards 2 e 3 coeficientes), tendo o trigésimo sétimo ano como a idade de referência. Para modelar os padrões espaciais da altura dominante foi utilizada a krigagem ordinária, sendo gerados, em seguida, os mapas temáticos de limites das classes de sítios. A krigagem indicatriz permitiu obter as probabilidades de ocorrência de sítios de alta, média e baixa produtividade. O processamento da amostragem estratificada, com o suporte de interpretação visual das imagens, permitiu definir cinco estratos em função da produtividade. Os resultados mostraram o melhor modelo selecionado foi o de Chapman-Richards (3 coeficientes), para as curvas de sítio. O modelo exponencial foi o melhor modelo para identificar as melhores áreas de probabilidade de ocorrência de sítios com maior produtividade. A eficiência da krigagem indicatriz gerou mapas temáticos para delimitar os locais com probabilidade dos sítios mais promitente. A geoestatística se mostrou eficiente em relação ao erro comparada com a amostragem aleatória simples.

**Palavras-chave:** índices de sítio, krigagem indicatriz, krigagem ordinária.

### INTRODUCTION

Forest plantations are the main source of raw material of the Brazilian forestry's production chain. Among the species, *Pinus* spp. stands have been an

alternative to reduce the exploitation of native forests in recent decades in a renewable and sustainable way. According to IBÁ (2019), pine plantations occupy 1.6 million hectares in Brazil and they are concentrated in

<sup>1</sup>Universidade Federal do Paraná (UFPR), Campus III, Curitiba, Brasil. E-mail: [myrciaminatti@gmail.com.br](mailto:myrciaminatti@gmail.com.br). \*Corresponding author.

Paraná (42%) and Santa Catarina (34%), followed by Rio Grande do Sul (12%) and São Paulo (8%).

Geostatistics differs from traditional statistics in considering the variability according to the distances between the measured points. This idea served as the basis for the development of Matheron's Theory of Regionalized Variables (1963), which is a numerical spatial function that varies from one place to another with some continuity and whose variation cannot be represented by a simple mathematical function. The technique uses spatial interpolation for projecting data, being called kriging. The semivariogram can be defined as a statistical method that allows studying the natural dispersion, the degree of continuity of the data sampled in the field, and the spatial-dependence structure of georeferenced data (YAMAMOTO; LANDIM, 2013).

Moreover, ordinary kriging (OK), which is a data interpolation method, uses the spatial dependence between neighboring samples showed in the semivariogram to estimate values at any position within the analyzed space to which the semivariogram model was fitted without bias and with minimum variance. This estimator is nothing more than a weighted average of the observed values.

Indicatrix kriging (IK) is a non-parametric type of conditional ordinary kriging that uses the position and values of dichotomous data to produce a local distribution of the spatial probability rather than a global distribution of the analyzed properties. IK is not influenced by the effects of discrepant values (MOTOMIYA et al., 2006; PAZ-FERREIRO et al., 2010; and PELISSARI et al., 2014).

Forest planning incorporated georeferenced information system techniques to turn the decision-making process more conscious regarding productivity, economic gains, and ecological aspects. Machado & Lopes (2014) claim that the detailing of forest micro-planning is extremely important in the analyzes of the stand within the forest. A stand is a production unit, therefore, observing its size, shape, and distribution in the forest to obtain accurate estimates reduces time and costs and contributes to business planning.

Thus, this study aimed to test the performance of geostatistical techniques as a tool to reduce the sampling error and, consequently, the sampling intensity of forest inventories.

## MATERIAL AND METHODS

This study was carried out in a 932-ha area of *Pinus taeda* L. planted forest in central southern Paraná, Brazil. The stands were between 17 and 42 years old. The sampling process used was systematic with circular plots (500 m<sup>2</sup> with a 12.62 m radius). In the sampling units, the diameter at breast height (DBH) of all individuals was measured with a measuring tape, and the total height of 20% of the trees and five dominant trees was measured with a Vertex IV hypsometer. Also, to generate the site index curves, 283 pairs of dominant height and age values were used. The Schumacher (Equation 1) and Chapman-Richards 2 and 3 coefficients (Equations 2 and 3) models were tested in their original form.

$$\ln(\text{DH}) = \ln(S) + \beta_1 * \left[ \left( \frac{1}{A_{ref}} \right) * \left( \frac{1}{A} \right) \right] \quad (\text{Equation 1})$$

$$\text{DH} = \frac{S * [1 - e^{(-\beta_1 * I)}]}{1 - e^{(-\beta_1 * A_{ref})}} \quad (\text{Equation 2})$$

$$\text{DH} = \frac{S * [1 - e^{(-\beta_1 * A)}]^{\beta_2}}{[1 - e^{(-\beta_1 * A_{ref})}]^{\beta_2}} \quad (\text{Equation 3})$$

In which:

DH = Assmann's dominant height,

$A_{ref}$  = reference age,

A = age,

$e$  = exponential,

$\ln$  = Napier's log,

S = site/site index and,

$\beta_1 e \beta_2$  = parameters.

To assess the quality of the adjustments of the dominant height models, the values of the adjusted coefficient of determination ( $R^2_{aj}$ ), relative estimate standard error (Syx%), and the graphic analysis of residuals were observed. After adjusting and selecting the models, site index curves were built from the most suitable model. The indexed age considered was 37 years, as it has a greater range of data and a greater number of data pairs. The amplitude of the site class was 1.83 m, grouped into four classes (21.7-25.4 m; 25.4-29.0 m; 29.0-32.70m; 32.7-36.4m). To improve the accuracy of the forest inventory, geostatistical analysis was used to verify the spatial dependence of the data by calculating the semi-variance (Equation 4).

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{ [Z(x_i + h) - Z(x_i)]^2 \} \quad (\text{Equation 4})$$

In which:

$\gamma(h)$  = semi-variance of the variable  $Z(x_i)$ ,

$h$  = distance (m) and,

$N(h)$  = number of pairs of measured points  $Z(x_i)$  and  $Z(x_i+h)$  separated by a  $h$  distance.

The regularization of the sampling grid through the angular tolerance is performed with the semi-variance determined between the equidistant sampling points. This method is performed in the four directions of the spatial plane, 0° (S-N); 45° (SW-NE); 90° (W-E), and 135° (NW-SE) to obtain the matrix of mean semi-variances between the equivalent distances and the number of pairs of selected sampling units (PELISSARI et al., 2014). Three stochastic semivariogram models (spherical (Eq. 5), exponential (Eq. 6), and Gaussian (Eq. 7)) were tested and the best model was chosen based on the estimated parameters: average standardized error (ASE), coefficient of determination ( $R^2$ ), nugget effect (C0), and contribution (C). The structure of the semivariograms was composed by the nugget effect (C0), which corresponded to the semi-variance at the zero distance and indicated random variation; the plateau (C0+C), which represents the stabilization of the semivariogram values; the contribution (C), which was given by the

difference between the plateau ( $C_0 + C$ ), the nugget effect ( $C_0$ ), and the range ( $A$ ), which was defined by the distance where the semivariogram reaches the plateau and indicated the limit of the correlation between the sampling units (YAMAMOTO; LANDIM, 2013).

$$\left\{ \begin{array}{l} C_0 + C \left[ \left( \frac{3}{2} \right) \left( \frac{h}{A} \right) - \left( \frac{1}{2} \right) \left( \frac{h}{A} \right)^3 \right] \\ C_0 + C \left[ \left( \frac{3}{2} \right) \left( \frac{h}{A} \right) - \left( \frac{1}{2} \right) \left( \frac{h}{A} \right)^3 \right] \end{array} \right. \quad \text{(Equation 5)}$$

$$\gamma(h) = C_0 + C (1 - e^{-h/A}) \quad \text{(Equation 6)}$$

$$\gamma(h) = C_0 + C (1 - e^{-h^2/A^2}) \quad \text{(Equation 7)}$$

In which:

$\gamma(h)$  = semi-variance of the variable  $Z(x_i)$ ,

$h$  = distance (m),

$C_0$  = nugget effect,

$C$  = variance "a priori",

$C_0 + C$  = plateau and,

Yamamoto and Landim's (2013) classification was used to analyze the degree of dependence. Semivariograms with nugget effects less than or equal to 25% of the threshold were considered to have a strong spatial dependence, followed by moderate dependence between 25% and 75%, and weak dependence when higher than 75%.

For the mapping of site classes and construction of thematic maps, a two-stage approach was adopted after selecting the best model. In the first stage, ordinary kriging was applied by using the limits of the site index classes generated for the spatialization of the sites, analyzing the spatial dependence and minimal to non-bias variances. In the second stage, indicatrix kriging was applied to diagnose the probabilities of occurrence of productive sites. The

ArcGis 10.3 software was used to generate the thematic maps and for data processing.

Finally, the total forest inventory was carried out with systematic sampling. For its processing, stratified systematic sampling was considered with weights of different strata resulting from the classes generated via geostatistics. The criterion used to evaluate the procedures was the absolute and relative sampling error. The equations used for both procedures can be found in SANQUETTA et al. (2014).

## RESULTS AND DISCUSSION

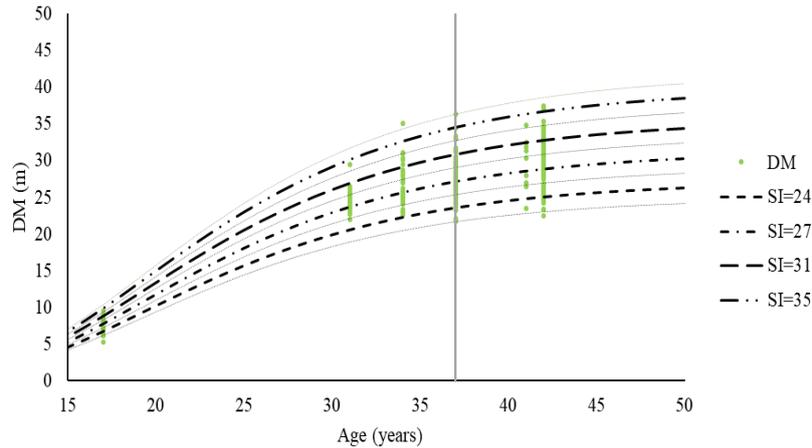
Table 1 shows the results for the adjusted site models. The Chapman-Richards (3 coefficients) model was the best as it presented a higher coefficient of determination and a lower estimate of the relative standard error. Therefore, the site curves were constructed by using this coefficient. The data showed an amplitude of 14.64 m at the indexed age. This amplitude was defined as a function of the distribution of the observed values of DH x Age at the indexed age. Thus, four site classes were defined for each model and the site indexes were SI=24, SI=27, SI=31, and SI=35 at the indexed age of 37 years.

In Figure 1, it is possible to observe the curves built for each adjusted model. The curves generated by the tested models did not adequately cover the observed data, and some data at younger ages are not covered by the lower and upper limits of the respective curves.

Furthermore, Table 2 presents the statistical parameters of the adjusted models with emphasis on the exponential ones, which had the lowest average standard errors (ASE), the lowest nugget effect values ( $C_0$ ), and the highest coefficient of determination ( $R^2$ ) in both the ordinary and indicative kriging. The exponential model also showed a smaller nugget effect in the adjustment of dominant heights in teak trees (PELLISSARI et al., 2014). However, the lowest errors found by these authors were obtained with the application of the spherical model.

**TABLE 1** - Coefficients and statistics of site models adjusted for *Pinus taeda* L.

Model	Coefficients			$S_{yx}$ (cm)	$S_{yx}$ %	$R^2$	$R^2_{aj}$
	S	$\beta_1$	$\beta_2$				
Schumacher	4.31960	-37.6346	-	2.621	9.93	0.788	0.787
Chapman-Richards 2 Coeff.	1800.032	-0.00041	-	2.912	11.04	0.749	0.748
Chapman-Richards 3 Coeff.	32.15753	0.11044	8.405074	2.141	8.12	0.818	0.817



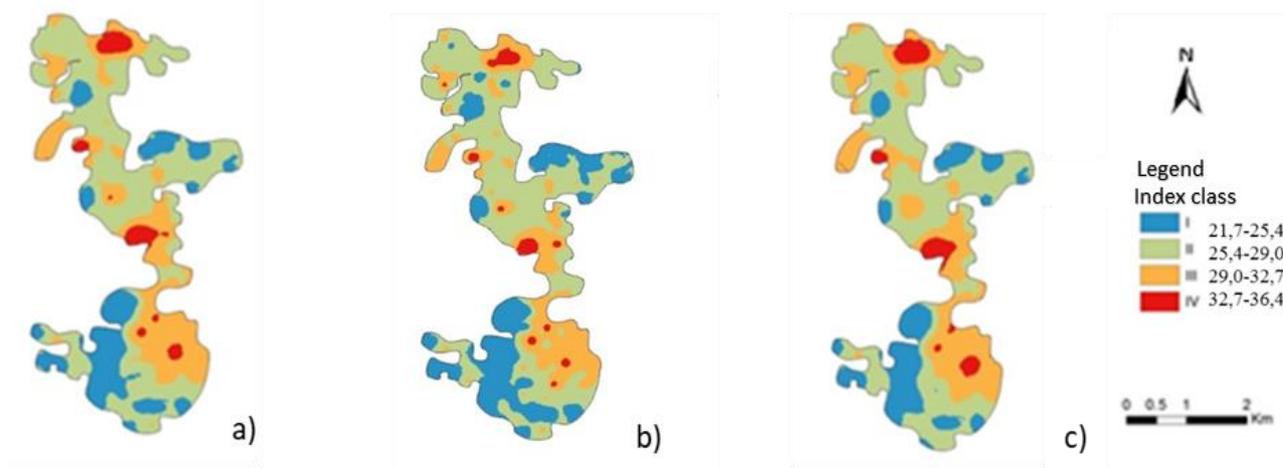
**FIGURE 1** - Mean curves of four site classes estimated by the Chapman-Richards model (3Ccoeff) for *Pinus* sp.

The strong spatial dependence was obtained with the exponential model and the moderate one was obtained

with the other models, both for ordinary and indicatrix kriging (Table 2) (YAMAMOTO; LANDIM, 2013).

**TABLE 2** - Semivariographic parameters and adjustment statistics of the fitted models.

Ordinary kriging – site classes									
Model	C <sub>0</sub>	C	RMS	RMSS	ASE	GD	R <sup>2</sup>	a	B
Spherical	3,068	8,773	2,5510	1,0034	2,5712	34.97	0.898	0,3540	17,9838
Exponential	0,701	8,867	2,5657	1,0274	2,5544	7.91	0.916	0,3494	18,0812
Gaussian	3,993	8,862	2,5459	1,0369	2,5774	45.06	0.851	0,3595	17,8141
Indicatrix kriging – site classes									
Model	C <sub>0</sub>	C	RMS	RMSS	ASE	GD	R <sup>2</sup>	a	B
Spherical	0,097	0,2261	0,4415	1,0220	0,4339	43.14	0.745	-	-
Exponential	0,000	0,2259	0,4520	1,1538	0,4013	0.00	0.788	-	-
Gaussian	0,124	0,2263	0,4393	1,0377	0,4248	55.10	0.694	-	-
Ordinary kriging – productivity classes									
Model	C <sub>0</sub>	C	RMS	RMSS	ASE	GD	R <sup>2</sup>	a	B
Spherical	9287.37	49430.87	156.4424	1.063866	147.261	18.7886	0.520	-0.240	1.070
Exponential	9531.22	61951.67	156.486	1.032936	151.383	15.3449	0.460	-0.680	1.018
Gaussian	14826.71	35993.35	156.0198	1.022439	152.414	41.1929	0.510	0.880	0.960



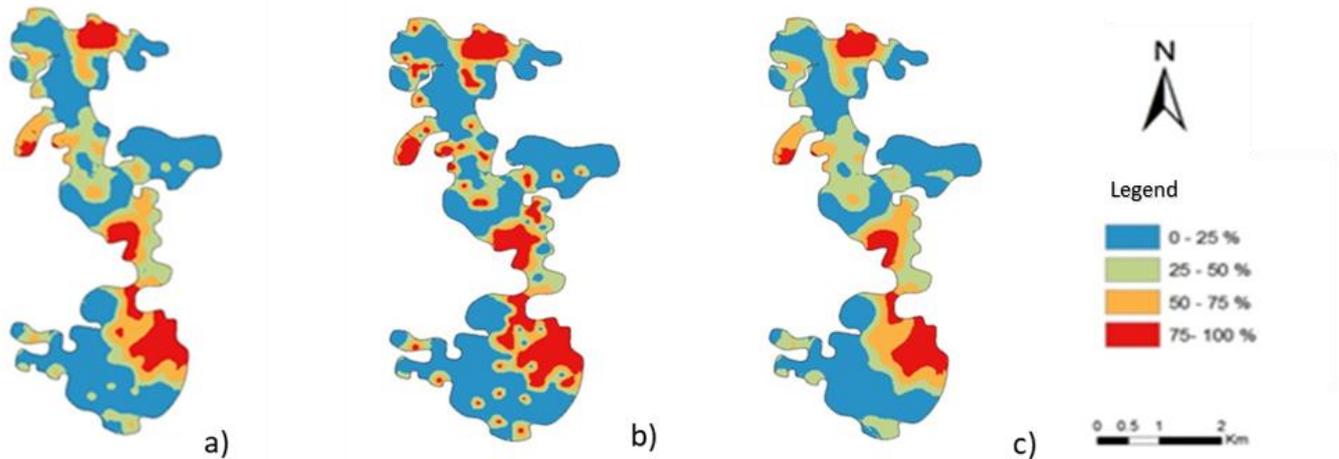
**FIGURE 2** - Maps of site index classes generated by ordinary kriging for the spherical (A), exponential (B), and Gaussian (C) models.

Thematic maps generated by the same procedure (Figure 2) showed similar spatial patterns for the spherical

and Gaussian models. The exponential model stood out and it was selected as the most suitable one due to the

delimitation of a smaller area for the sites with higher productivity (Class IV) with smaller and more dispersed cores. Indicative kriging allowed to spatialize the probability of occurrence of high productivity sites. The thematic maps generated for the exponential model

estimated larger areas with a probability of occurrence of high productivity sites (Figure 3). The spherical and Gaussian models showed similar patterns and “pure nugget effect” values.



**FIGURE 3** - Maps of site index classes generated by indicatrix kriging for the spherical (A), exponential (B), and Gaussian (C) models.

According to Santos et al. (2017), the information generated by the thematic maps allows visualizing the spatial characteristics of the population, making it possible to carry out the zoning of growth and volumetric stock of trees. Thus, the authors emphasize kriging as an alternative tool for estimating the wood stock in *Eucalyptus grandis* stands, highlighting the method's advantages for sampling planning and stand management. The effectiveness of geostatistical modeling is also mentioned by Pelissari et al. (2015) for *Tectona grandis*. The authors claim that the technique allows mapping the productive capacity of the place and inferring on environmental characteristics that limit the development of the species.

The stand was divided into five volumetric strata with the lowest value being  $<300$  to  $>700$   $\text{m}^3 \text{ha}^{-1}$  with intervals of  $150 \text{m}^3 \text{ha}^{-1}$ . The stratum with the highest productivity was the fifth, it presented a total volume of  $799.47 \text{m}^3$  in  $383.18 \text{ha}$ . On the other hand, the least productive stratum was the first with a total volume of  $101.25 \text{m}^3$  in  $147.3 \text{ha}$ . The intermediate stratum, the third, presented productivity of  $499.72 \text{m}^3$  in  $264.0 \text{ha}$ . The delimitation of the areas with the probability of the sites allows creating strategies for logistics and harvest, and also to plan the conduction of the plantations by verifying interventions to minimize costs and maximize productivity and profits.

Therefore, the productivity classes by ordinary kriging (Table 2) presented the best results for  $R^2$  with the spherical and Gaussian models, the best ASE for the spherical model, and the best value for RMSS. In Figure 4, it was observed that the model with the best standard for the productivity classes is the Gaussian. The applicability of thematic maps helps to carry out the zoning of the productive capacity of the populations to reduce costs and decision-making time in possible adversities in the economy and the climatic conditions.

The statistics of the sampling processes, traditional inventory (simple random), and inventory supported by geostatistical techniques (stratified), are presented in Table 3. There was a difference in the average variance between the two sampling processes due to the large amplitude of the data (minimum = 4.45; maximum = 2,276.96) of volume per hectare.

It was observed that the relative sampling error for the simple random process was high (6.55%), which can be explained by the heterogeneity of the stands due to the wide age difference of the stands. For the stratified sampling using the geostatistical techniques, there was a reduction in the relative sampling error (1.93%), which indicates that the application of these sampling processes in heterogeneous stands could cause great inaccuracies in population estimates.

**TABLE 3** - Sampling statistics by the simple random and the stratified random process using geostatistic techniques in *Pinus taeda* L.stands.

Statistics	Traditional inventory	Inventory supported by geostatistics
Average	1.086.57	579.10
Variance	36.1256.20	6.847.43
Variance of the mean	1.298.04	32.15
Standard error	36.02	5.67
Absolute sampling error	71.2301	11.1763
Relative sampling error	6.55	1.93

Therefore, geostatistics contributes to forestry planning and harvesting by indicating the best areas of productivity for the decision-making process in planted forest stands. The tool has shown advances in studies with *Tectona grandis*, *Araucaria angustifolia*, *Pinus* sp., and *Eucalyptus* sp. with different purposes such as carbon and biomass quantification, growth modeling, and forest harvest planning.

### CONCLUSIONS

The Chapman-Richards (3 coefficients) was the best model for the construction of the site curves.

The exponential model was the best one because of its statistical parameters to identify the best areas of the probability of occurrence of high productivity sites.

Indicative kriging generated thematic maps to delimit the likely locations of high-quality sites in smaller areas, reducing costs in the planning of forest inventories and the decision-making process on management plans.

The sampling process supported by geostatistics proved to be promising with a sampling error of 1.93%, a reduction of approximately 30% when compared to simple random sampling (6.55%).

### REFERENCES

- IBÁ. INDÚSTRIA BRASILEIRA DE ÁRVORES. **Relatório Ibá 2019**. Brasília, DF: Pöyry, 2019. 79p.
- MACHADO, C.C.; LOPES, E.S. Planejamento. Colheita Florestal. 3.ed. Viçosa: UFV, 2014.
- MATHERON, G. Principles of geostatistics. **Economic Geology**, v.58, n.8, p.1246-1266, 1963.
- MOTOMIYA, A.V.A.; CORÁ, J.E.; PEREIRA, G.T. Uso da krigagem indicatriz na avaliação de indicadores de fertilidade do solo. **Revista Brasileira de Ciência do Solo**, v.30, n.3, p.485-496, 2006.
- PELISSARI, A.L.; CALDEIRA, S.F.; FIGUEIREDO FILHO, A.; MACHADO, S.A. Propostas de mapeamento da capacidade produtiva de sítios florestais por meio de análises geoestatísticas. **Scientia Forestalis**, v.43, n.107, p.601-608, 2015.
- PELISSARI, A.L.; FIGUEREDO, A.F.; CALDEIRA, S.F.; MACHADO, S.A. Geoestatística aplicada ao manejo de povoamentos florestais de teca, em períodos pré-desbaste seletivo, no estado do Mato Grosso. **Revista Brasileira de Biometria**, v.32, n.3, p.430-444, 2014.
- SANQUETTA, C.R.; CORTE, A.P.D.; RODRIGUES, A.L.; WATZLAWICK, L.F. **Inventários Florestais: Planejamento e Execução**. 3a. edição. 1. ed. Curitiba: Dos Autores, 2014. v.1. 406p.

SANTOS, M.C.; ROVEDA, M.; ZANON, M.L.B.; FIGUEIREDO FILHO, A.; ROIK, M.; PACHECO, J.M.; SCAVINSKI, V. Inventário florestal utilizando técnicas de silvicultura de precisão em povoamentos de *Eucalyptus grandis* Hill ex Maiden. **Floresta e Ambiente**, v.24, n.1, p. 1-9, 2016.

YAMAMOTO, J.K.; LANDIM, P.M.B. **Geoestatística: conceitos e aplicações**. São Paulo: Oficina de Textos, 2013. 215p.