

ESTIMATING PRECISION OF SYSTEMATIC SAMPLING IN FOREST INVENTORIES

Estimativa da precisão na amostragem sistemática em inventários florestais

José Marcio de Mello¹, Henrique Ferrazzo Scolforo², Marcel Régis Raimundo²,
José Roberto Soares Scolforo², Antônio Donizette de Oliveira², Antônio Carlos Ferraz Filho²

ABSTRACT

The sampling technique commonly used in forest inventories is the systematic sampling. This study aimed to evaluate the estimator of the variance of the mean proposed by Cochran for a systematic sampling technique in forests with high and low percentages of the sampled area. The study areas comprised native vegetation in Minas Gerais. To assess the efficiency of the estimators in situations involving high sampling rates (determined as the percentage of the area sampled), a fragment where a census was conducted was used. The remaining fragments comprised situations involving low sampling rates, and for these fragments, inventory accuracy was determined using the Cochran estimator. As a result it was observed, in the fragment where the census was conducted, that the structure of the correlation coefficient proposed by Cochran remained approximately constant for the area, and to the extent that sampling rate reduced, the impact of the Cochran estimator on the inventory accuracy decreased. For the fragments with a low sampling rate, it could be inferred that the sampling rate was a key factor for the correlation proposed by Cochran to have an impact on the forest inventory accuracy. The use of this estimator is indicated for fragments with a sampling rate greater than 10% of the area.

Index terms: Sampling, forest inventory, Minas Gerais.

RESUMO

A técnica de amostragem comumente usada em inventários florestais é a amostragem sistemática. Objetivou-se avaliar o estimador da variância da média proposto por Cochran, no procedimento de amostragem sistemática, em florestas com alto e baixo percentual da área amostrado. As áreas contemplaram a vegetação nativa de Minas Gerais. Para verificar o comportamento dos estimadores numa situação com alto percentual de área amostrado, utilizou-se o fragmento onde foi efetuado o censo. Os demais se adequaram às situações de baixo percentual amostrado, observando-se o comportamento da precisão do inventário, a partir do uso do estimador de Cochran. Como resultado para o fragmento onde se fez o censo, a estrutura do coeficiente de correlação proposto por Cochran se manteve, aproximadamente, constante para a área e, reduzindo a intensidade de amostragem, o impacto do estimador de Cochran diminuiu. Para os fragmentos com baixo percentual amostrado, a intensidade amostral é a chave para a correlação proposta por Cochran: impactar na precisão do inventário florestal, sendo indicado o uso desse estimador para fragmentos onde há uma intensidade amostral superior a 10% da área inventariada.

Termos para indexação: Amostragem, inventário florestal, Minas Gerais.

INTRODUCTION

The growing demand for products derived from native or planted forest resources requires accurate assessment of forest growth dynamics through forest inventory (Ubialli et al., 2009). For a precise assessment of forest growth dynamics, it is necessary to apply sampling techniques that can faithfully capture the forest reality (Reis et al., 2007; Assis et al., 2009; Druszcz et al., 2012; Guedes et al., 2012).

Forest inventory is an activity that uses sampling techniques to obtain information on quantitative and qualitative characteristics of existing forest resources in a predetermined area (Mello et al., 2009; Vibrans, 2010). These sampling techniques can estimate population

parameters based on the characteristics that are being evaluated. The most common sampling techniques used in forest inventories are: simple random, stratified random, and systematic sampling (Mello; Scolforo, 2000). Estimators based on randomization were developed by means of probabilities generated by the randomization of the sampling units (Brus; Gruijter, 1997).

The sampling technique commonly used in forest inventories is systematic sampling. The most of ecological studies use techniques that incorporate systematization principles. In forests, randomization is restricted to achieve a better spatial coverage of the plots. Cochran (1965) designated this type of plot distribution in planted forests as “mismatched systematic sampling”. This type of distribution does not “negatively affect” the principle of statistical

¹Universidade Federal de Lavras/UFLA – Departamento de Ciências Florestais/DCF – Cx. P. 3037 – 37200-000 – Lavras – MG – Brasil – josemarcio@dcf.ufla.br

²Universidade Federal de Lavras/UFLA – Departamento de Ciências Florestais/DCF – Lavras – MG – Brasil

Received in July 26, 2014 and approved September 29, 2014

randomness. However, this technique produces good spatial coverage of the area and therefore can generate accurate and reliable estimations (Soares et al., 2009).

The systematic sampling techniques do not possess their own estimators of the desired population parameters. The simple random sampling (SRS) estimators are frequently used, as reported by authors associated with the sampling field such as Scolforo and Mello (2006). The use of the estimator of the variance of the mean of the SRS for systematic sampling increases statistical problems, because systematic sampling can better explain the dependencies between the sampling units (spatial correlation). On average, there is no bias between two given sampling units. However, in terms of accuracy, there may be substantial bias according to the spatial correlation between the sampling units.

Such correlations are directly associated with the proportionality between the installed sampling units and the statistical population in question; in that, for larger distances between the sampling units, it is expected that the correlation between them tends to be null, thereby causing minimal impact on the inventory accuracy. Conversely, shorter

distances between the sampling units result in correlation values that may cause bias in the inventory accuracy if these are not taken into consideration (Scolforo; Mello, 2006).

Cochran (1965) proposed an estimator to evaluate the effects of the correlation between the sampling units on the estimation of the variance of the mean using the systematic sampling technique.

Therefore, the aim of this study was to evaluate the estimator of the variance of the mean proposed by Cochran (1965) for the systematic sampling technique in the following two distinct situations: forests with high and low percentages of the sampled area.

MATERIAL AND METHODS

Characterization of the areas and data collection

The study areas comprised various phytophysionomies of native arboreous vegetation in the state of Minas Gerais (Table 1). These areas were surveyed during the Project – “Forest Inventory of Minas Gerais” according to the methodology described in Scolforo, Mello and Oliveira (2008a).

Table 1 – Location, phytophysionomy, identification (ID), longitude, latitude, area (ha), mean altitude (m), and soil type of the studied fragments.

County	Phytophysionomy	ID	Longitude	Latitude	Area (ha)	Mean altitude (m)	Soil
Mateus Leme	Montane Seasonal Semideciduous Forest	33	-44.3585	-20.0118	4649.9	812	Cambisol
Brasilândia	Cerrado Senu Stricto	58	-45.8191	-16.9414	476.7	963	Cambisol
Rio Pardo de Minas	Campo Cerrado	95	-42.4968	-15.5969	364.7	572	Cambisol
Camanducaia	Montane High Ombrophilous Forest	97	-46.0583	-22.8871	181.8	1890	Latosol
Araguari	Cerradão	105	-48.2839	-18.5270	49.8	895	Latosol
Araçuaí	Submontane Deciduous Seasonal Forest	110	-44.1101	-16.9108	209.7	354	Argisol
Lavras	Cerrado Senu Stricto	126	-45.9839	-21.2262	3.9	890	Latosol
Mato Verde	Montane Deciduous Seasonal Forest	144	-43.9849	-15.4400	154.6	551	Latosol
Muriae	Submontane Seasonal Semideciduous Forest	145	-42.9646	-21.0838	185.8	266	Latosol
Baependi	Montane Ombrophilous Forest	152	-44.7472	-21.9824	18.9	1647	Latosol

ID is the identification number of the fragment according to the methodology by Scolforo, Mello and Oliveira (2008 a).

The selected fragments were intended to faithfully represent the plant biodiversity found in the state of Minas Gerais (Figure 1), thus providing a demonstration of the impact of adopting the correlation coefficient proposed by Cochran (1965).

The cerrado fragment located in Lavras (ID 126), with an area of 3.89 hectares, was completely inventoried, enabling the use of the high sample rate (determined as the percentage of the area sampled) in the performed simulations.

For the present study, data were collected from plots of 400 m² (for fragments with IDs 126 and 152), and 1000 m² (for the remaining fragments). In all evaluated fragments, the plots were standardized, with the distance between them being defined based on the total area and the number of the sampling units to be evaluated in each situation.

In each plot, the total height was measured with the aid of a height pole, in addition to the circumference at 1.30 meter above ground or at breast height (CBH) of all specimens with a minimum circumference of 15.7 cm.

The volume for each sample unit was determined from the sum of the individual volumes, which in turn were obtained from adjusted equations from the Forest Inventory of Minas Gerais project (Scolforo, Oliveira; Acerbi Junior, 2008b; Rufini et al., 2010). Finally, the volume of each sample unit evaluated was converted to hectare.

Data Processing

The response variable evaluated in the present study was the volume per hectare for each fragment studied. Data processing consisted of calculating the mean, variance of the mean, and confidence interval using the estimator of SRS and the estimator proposed by Cochran (1965). These processes were performed in the two situations proposed in the study.

Fragment with a high sampling rate (Case 1)

To determine the efficiency of the estimators in a situation with a high sampling rate, we used the fragment with ID 126 where a census was conducted. In this situation, we evaluated the precision and the accuracy

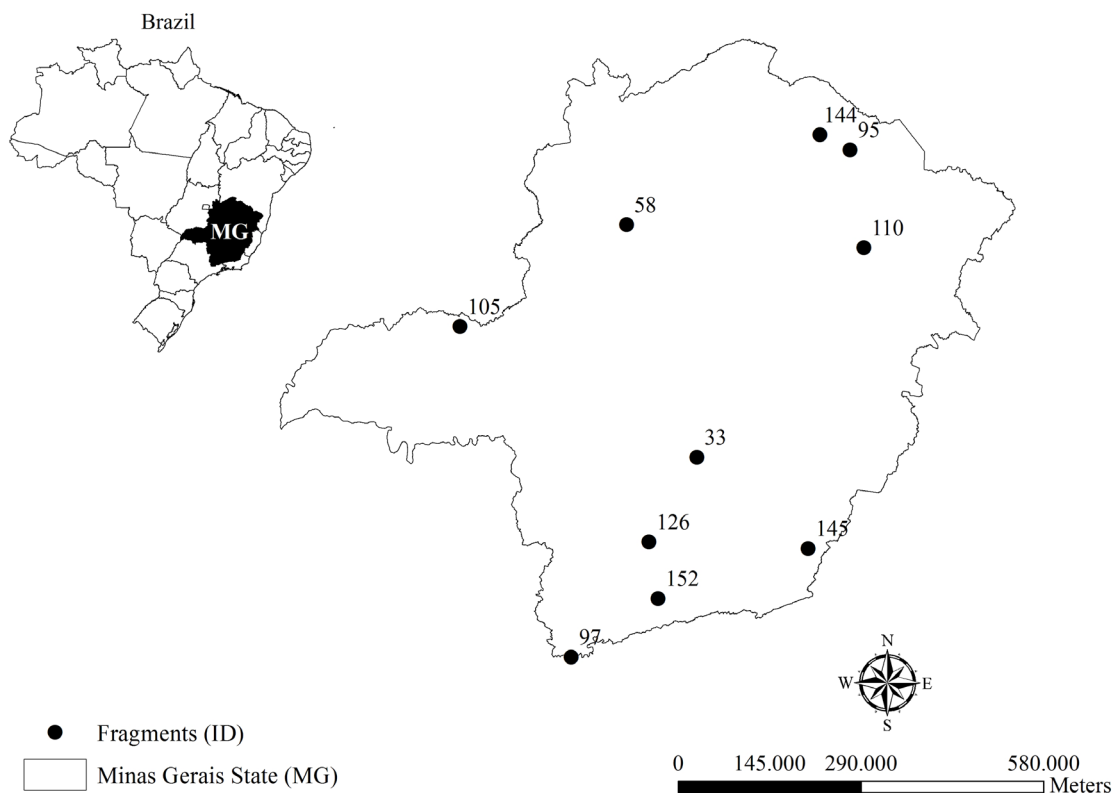


Figure 1 – Fragment distribution in the state of Minas Gerais.

of the simulations performed, considering that the true mean of the fragment (μ) was known. Furthermore, it was possible to evaluate two high sampling rates (33% and 50%) in the total area, with the plots being systematically distributed throughout the study area.

Fragment with a low sampling rate (Case 2)

This is the most common situation in forest inventories. These surveys generally involve large areas and suffer from budget constraints for inventories. This implies large distances between the plots and results in a low degree of correlation between them.

Inventory processing

For each fragment, the forest inventory was processed according to the estimator of the mean, variance of the mean, and confidence interval (this interval only being generated for the fragment where the census was conducted to determine the sampling accuracy) of the SRS using the traditional estimators (Scolforo; Mello 2006). The estimator of the mean is the same for the situations involving randomization or standardization of the plots. In the present study, there is a basic distinction between the variance of the SRS mean (Equation 1) and the estimator proposed by Cochran (1965) (Equation 2), according to the following equations:

$$S^2_{\bar{y}} = \frac{S^2}{n} \left(\frac{N-n}{N} \right) \quad (1)$$

$$S^2_{\bar{y}} = \frac{S^2}{n} \left(\frac{N-n}{N} \right) [1 + (n-1)\hat{\rho}] \quad (2)$$

Where: S^2 is the sampling variance; n is the sampling size; N is the number of possible plots present in each area; $\hat{\rho}$ is the correlation coefficient proposed by Cochran (1965) between pairs of sampling units.

The correlation coefficient proposed by Cochran (1965) between the sampling units (Equation 3) is represented in the following equation:

$$\hat{\rho} = \frac{2}{(n-1)(N-1)S^2} \sum_{i=j}^k \sum_{j<u} (y_{ij} - \bar{y})(y_{iu} - \bar{y}) \quad (3)$$

Where: n , N and S^2 were defined above; k is equal to N/n ; y_{ij} denotes the j th member of the i th systematic sample so that $j = 1$ to n , $i = 1$ to k ; y_{iu} denotes the j th member at an distance u from y_{ij} ; \bar{y} represents the sample mean.

Substituting (3) from (2), we obtain the variance of the mean (Equation 4):

$$S^2_{\bar{y}} = \frac{S^2}{n} \left(\frac{N-n}{N} \right) \{1 + (n-1) \left[\frac{2}{(n-1)(N-1)S^2} \sum_{i=j}^k \sum_{j<u} (y_{ij} - \bar{y})(y_{iu} - \bar{y}) \right]\} \quad (4)$$

After obtaining the variance of the mean, it was possible to determine the effect of adding the correlation coefficient proposed by Cochran to the inventory accuracy for both situations.

After obtaining the estimates for the completely inventoried fragment, it is possible to observe beyond the effect of precision, the accuracy provided by the performed inventory.

RESULTS AND DISCUSSION

Effect of correlation coefficient proposed by Cochran - Case 1

In this first case, we evaluated the correlation coefficient in a study area of 3.89 ha, which revealed the parametrical statistical value of the forest inventory processing. The aim was to evaluate the effect of the distances between the evaluated plots.

In this fragment, it was possible to generate 5 systematic samples with sampling rates of 33% and 50% of the total area. The maximum amount of plots that fit in this area is 34 plots. Therefore, two databases with 17 plots representing 50% of the area were generated. For a sampling rate of 33%, 2 databases with 12 plots and a third database with 10 plots were generated. In all these situations, the plots were distributed systematically. The descriptive statistics for each sample present in the fragment are shown in table 2 and it is important to note the maximum distance separating the sampling units (D).

The mean parameter (μ) in Case 1 was 116.78 m^3/ha . Considering the estimated means for each sampling rate, it was observed that the mean of sample 5 was the most accurate, because it was the closest to the parameter.

Notably, even with a reduction in the sampling rate, the structure of the correlation coefficient proposed by Cochran (1965) tends to remain approximately constant for the area. A study conducted by Sé et al. (2013) corroborates this fact, as it clearly explains how the structure of the correlation coefficient remains approximately constant for the same area with different sampling rates.

Table 2 – Descriptive statistics for the 5 databases generated for simulation purposes.

Sample	n	D (m)	Mean (m ³ /ha)	Standard deviation (m ³ /ha)	CV (%)	ρ
1	17	48	124.82	43.40	34.77	-0,0303
2	17	48	108.73	42.02	38.65	-0,0303
3	12	57	119.38	49.85	41.76	-0.0303
4	12	57	113.90	42.30	37.14	-0.0303
5	10	62	117.11	38.58	32.94	-0.0303

n: number of plots; D: distance between sample plots; CV: coefficient of variation; ρ : correlation coefficient proposed by Cochran (1965).

This can be explained by the fact that the variations in sampling size (n) present in the denominator of the formula of the correlation of Cochran (1965) also impacts the estimated variance. Therefore, variations in the numerator (differences between each plot and the mean) result in a proportional change between the two components of this ratio (numerator and denominator), generating values for the correlation of Cochran that are very similar to each other for the same fragment. Therefore, different sampling rates cause a tendency of continuity for the correlation structure, for the population, considering systematic sampling, to cover the spatiality of the whole area.

As the correlation structure for the fragment remains constant, it could be verified that the sampling size (n) will determine how the addition of the correlation coefficient will impact the estimator of the variance of the mean, i.e., the greater the sampling rate, the larger the impact of this factor on the inventory error. Table 3 clearly emphasizes this point by demonstrating that as the sampling rate decreases, the impact of correlation on the inventory accuracy decreases [difference in the error estimations between SRS and Cochran (1965)]; however, this still had a good advantage over the use of the classical estimator for the evaluated fragment. It was verified that there was a reduction in the percentage error between the SRS and the Cochran estimator, ranging from 14% to 28% approximately. It was observed that even when 50% of the area was sampled with the SRS estimator, there was no error lower than 10%. Using the correlation coefficient for samples 1 and 2, the error was approximately 10%.

As the precision of the error estimation of the inventory for this fragment increases, the use of the estimator proposed by Cochran reflects a more accurate confidence interval (Table 4), because the range of variation of the interval was small. Notably, the true mean (116.78 m³/ha) was found between the generated intervals.

Table 3 – Error estimations (%) of the inventory considering the classical estimator (SRS) and the estimator proposed by Cochran (1965).

Sample	SRS (%)	Cochran (%)	Difference (%)
1	12.64	09.07	28.24
2	14.05	10.08	28.26
3	21.34	17.43	18.32
4	18.98	15.50	18.34
5	19.80	16.88	14.75

SRS: error estimation using the estimator of simple random sampling; Cochran: error estimation using the estimator coefficient proposed by Cochran (1965).

Table 4 – Confidence intervals (probability level of 5%) generated for wood volume.

Sample	SRS (m ³ /ha)	Cochran (m ³ /ha)
1	109.05 ≤ μ ≤ 140.60	113.50 ≤ μ ≤ 136.15
2	93.46 ≤ μ ≤ 124.01	97.77 ≤ μ ≤ 119.70
3	93.90 ≤ μ ≤ 144.86	98.58 ≤ μ ≤ 140.18
4	92.28 ≤ μ ≤ 135.52	96.25 ≤ μ ≤ 131.55
5	93.92 ≤ μ ≤ 140.30	97.34 ≤ μ ≤ 136.88

SRS: confidence interval using the estimator of simple random sampling; Cochran: confidence interval using the estimator coefficient proposed by Cochran (1965).

Effect of the correlation coefficient proposed by Cochran - Case 2

In this case, we used fragments of larger areas, where the sampling rate was low, and consequently, the distance between the plots (D) was higher.

The first interesting point to be noted is the values of the coefficient of variation (CV) that represent the spatial variability of the vegetation in the state of Minas

Gerais. In this regard, because the presented CVs vary widely, starting from an intermediate range, e.g., for ID 145 that is a Submontane Seasonal Semideciduous Forest, to very high amplitudes, e.g., ID 152 that is a Montane Ombrophilous Forest, indicating dispersion from the average.

Therefore, table 5 contains the information related to fragments, distance values between the plots and descriptive statistics, and also highlights the behavior of the correlation structure. One can clearly see that this correlation increasingly approaches zero as the fragment size increases.

Increasing the area leads to a decrease in the correlation value, because according to the formulation of the correlation proposed by Cochran (1965), N (number of plots that fit the area, i.e., as the area increases, the number of potential plots to be measured increases) represents a strong impact on the denominator of the correlation. This shows that as N increases, the correlation value (ρ) will approach zero.

This indicates that the impact on the correlation is dependent on the percentage of the area sampled (sample rate), i.e., both for large and small areas. The size of n (sampling size) will be essential for improving the estimation of the variance of the mean when using the estimator proposed by Cochran (1965). For large areas, where ρ will certainly approach zero, a high sampling rate will compensate this effect to make this correlation relevant for the inventory accuracy.

Despite the fact that the correlations presented in this study were negative, according to Cochran (1965), this correlation value may be presented as positive for certain areas. This indicates that the use of the SRS underestimated the inventory error (keeping in mind that a negative correlation indicates that SRS overestimates the error).

Thus, table 6 clearly shows that the effect of using the estimator proposed by Cochran worked better as the sample rate increased. The two smaller areas, ID 105 and ID 152, were the most highly sampled areas, representing 3% and 4%, respectively, of their total area. Yet, the effect of using the correlation was low.

The use of the estimator proposed by Cochran (1965) will be effective for sampling rates of least above 10%. Sé et al. (2013) observed that a sampling rate of 16% of the total area was sufficient for the correlation effect to have a good impact on the precision of the inventory, and this impact was approximately 7%.

Naturally, it can be inferred that the estimator proposed by Cochran (1965) is a more feasible alternative to be applied to small areas, considering the high probability that at least 10% of these areas will be sampled. Regarding very large areas, even with high sampling size, the sampling rate will be probably low on account of the total area. This indicates an almost null effect for the use of the estimator proposed by Cochran (1965).

Table 5 – Descriptive statistics for the fragments.

ID	Area (ha)	n	D (m)	Mean (m ³ /ha)	Standard Deviation (m ³ /ha)	CV (%)	ρ
33	4649.89	35	1153	118.62	84.75	71.44	-0.000022
58	476.73	30	399	42.47	16.61	39.10	-0.000210
95	364.73	20	427	30.17	24.42	80.96	-0.000274
97	181.83	20	302	347.81	142.08	40.85	-0.000550
105	49.77	16	176	180.94	80.96	44.75	-0.002016
110	209.73	16	362	166.60	96.71	58.04	-0.000477
144	154.63	15	321	85.52	32.60	38.12	-0.000647
145	185.80	15	352	216.24	61.26	28.33	-0.000539
152	18.90	19	100	72.52	65.36	90.11	-0.002132

n: number of plots; D: distance between sample plots; CV: coefficient of variation; ρ : correlation coefficient proposed by Cochran (1965).

Table 6 – Inventory error in % and m³/ha from the use of the SRS and Cochran (1965) estimators.

ID	SRS (%)	SRS (m ³ /ha)	Cochran (%)	Cochran (m ³ /ha)	Difference (%)
33	24.53	14.32	24.52	14.31	0.04
58	14.55	3.02	14.50	3.01	0.34
95	37.79	5.45	37.69	5.43	0.26
97	19.01	31.59	18.91	31.43	0.53
105	23.46	19.91	23.10	19.61	1.53
110	30.81	24.08	30.70	24.00	0.36
144	21.01	8.38	20.91	8.34	0.48
145	15.63	15.75	15.56	15.69	0.45
152	43.95	15.11	43.14	14.83	1.84

SRS: inventory error using the estimator of simple random sampling; Cochran: inventory error using the estimator coefficient proposed by Cochran (1965).

CONCLUSIONS

In the area completely inventoried, the structure of the correlation coefficient proposed by Cochran (1965) remains approximately constant.

As the sampling rate of an area decreases, the impact on the inventory accuracy from the use of the estimator proposed by Cochran (1965) decreases. Therefore, the Cochran estimator and the SRS estimator will present similar results.

The sampling rate is a key factor for the correlation proposed by Cochran (1965) to have an impact on the forest inventory accuracy. The use of this estimator is particularly indicated for fragments where the sampling rate is greater than 10%. This suggests that its use is more adequate for small areas.

REFERENCES

ASSIS, A. L. et al. Development of a sampling strategy for young stands of *Eucalyptus* sp. using geostatistics. **Cerne**. 15(2):166-173, 2009.

BRUS, D. J. D.; GRUIJTER, J. J. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). **Geoderma**. 80(1/2):1-59, 1997.

COCHRAN, W. G. **Sampling techniques**. 2.ed. New York: Wiley, 1965. 555p.

DRUSZCZ, J. P. et al. Custos de inventário florestal com amostragem de Bitterlich (PNA) e conglomerado

em cruz (CC) em plantação de *Pinus taeda* L. **Scientia Forestalis**. 40(94):231-239, 2012.

GUEDES, I. C. de L. et al. Técnicas geoestatísticas e interpoladores espaciais na estratificação de povoamentos de *Eucalyptus* sp. **Ciência Florestal**. 22(3):541-550, 2012.

MELLO, J. M.; SCOLFORO, J. R. S. Análise comparativa de procedimentos de amostragem em um remanescente de Floresta Estacional Semidecídua Montana. **Revista Árvore**. 24(1):55-62, 2000.

MELLO, J. M. de. et al. Continuidade espacial para características dendrométricas (número de fustes e volume) em plantios de *Eucalyptus grandis*. **Revista Árvore**. 33(1):185-194, 2009.

REIS, H. et al. Análise da composição florística, diversidade e similaridade de fragmentos de mata atlântica em Minas Gerais. **Cerne**. 13(3):280-290, 2007.

RUFINI, A. L. et al. Equações volumétricas para o cerrado *sensu stricto*, em Minas Gerais. **Cerne**. 16(1):1-11, 2010.

SCOLFORO, J. R. S.; MELLO, J. M. de. **Inventário florestal**. Lavras: UFLA/FAEPE, 2006. 561p.

SCOLFORO, J. R. S.; MELLO, J. M.; OLIVEIRA, A. D. de. Amostragem e caracterização dos fragmentos inventariados. In: SCOLFORO, J. R. S. et al.

Inventário Florestal de Minas Gerais: Cerrado - Florística, Estrutura, Diversidade, Similaridade, Distribuição Diamétrica e de Altura, Volumetria, Tendências de Crescimento e Áreas aptas para Manejo Florestal. 1.ed. Lavras: Editora UFLA, 2008a. v. 1, p. 1-78.

SCOLFORO, J. R. S.; OLIVEIRA, A. D. de ; ACERBI JUNIOR, F. W. **Inventário Florestal de Minas Gerais: Equações de Volume, Peso de Matéria Seca e Carbono para diferentes Fitofisionomias da Flora Nativa.** 1.ed. Lavras: Editora UFLA, v. 1, 2008b. 216p.

SÉ, D. C. et al. O uso do coeficiente de correlação entre parcelas visando aumento de precisão em

inventários florestais. **Cerne**. 19(4):575-580, 2013.

SOARES, C.P. B. et al. Comparação entre procedimentos de amostragem para espécies florestais raras e padrão de distribuição espacial agregado. **Árvore**. 33(3):545-553. 2009.

UBIALLI, J. A. et al. Comparação de métodos e processos de amostragem para estimar a área basal para grupos de espécies em uma floresta ecotonal da região norte mato-grossense. **Acta Amazonica**. 39(2):305-314. 2009.

VIBRANS, A. C. et al. Inventário florístico florestal de Santa Catarina (IFFSC): aspectos metodológicos e operacionais. **Pesquisa Florestal Brasileira**. 30(64):291-302, 2010.