SAMPLING OF CHEMICAL ATTRIBUTES IN FOREST SOILS

Bruno Oliveira Lafetá[1*], Tamires Mousslech Andrade Penido[2], Nivaldo de Souza Martins[3], Márcio Leles Romarco de Oliveira[2], Danielle Piuzana Mucida[2], Reynaldo Campos Santana[2]

[1]Federal Institute of Education, Science and Technology of Minas Gerais, Department of Forest Engineering, São João Evangelista, State of Minas Gerais, Brazil – *bruno.lafeta@ifmg.edu.br
[2]Federal University of the Jequitinhonha and Mucuri Valleys, Diamantina, State of Minas Gerais, Brazil – penidotma@gmail.com, marcioromarco@gmail.com, dpiuzana@yahoo.com.br, silviculturaufvjm@yahoo.com.br
[3]Aperam Bioenergia S/A, Itamarandiba, State of Minas Gerais, Brazil – nivaldo.martins@aperam.com

Abstract

Information about sample adequacy that represents soil chemical attributes distribution are fundamental for a better rationalization of the use of correctives and fertilizers. The objective was to evaluate the variability of these attributes and to size the minimum number of composite samples to represent the fertility of forest soils. The total area planted was 9,101ha, constituted of 265 commercial eucalypt stands. The 687 soil composite samples obtained were for chemical analysis. It was evaluated the performance of two exploratory analysis techniques and six sampling procedures. The attributes P, K, Ca, Mg and S presented higher coefficient of variation (>35%). In contrast, the distributions of Al, organic matter and, mainly, pH were the most homogeneous. The sample error was smaller as the amount of composite samples increased. The representative of all chemical attributes (sample error of 5%) was achieved with a minimum of 309 (one each 29ha, 1:29) and 295 (1:31) composite samples from sampling procedures simple casual and stratified by altitude class, respectively. Both procedures were promising for soil sampling, especially, when applying the boxplot for identification and removal of outliers.
Keywords: Boxplot; stratification; sample size.

Resumo

Amostragem de atributos químicos em solos florestais. Informações sobre a suficiência amostral que represente a distribuição de atributos químicos do solo são fundamentais para melhor racionalização do uso de corretivos e fertilizantes. O objetivo foi avaliar a variabilidade desses atributos e dimensionar o número mínimo de amostras compostas para representar a fertilidade de solos florestais. A área total de plantio foi de 9.101ha, constituída por 265 talhões comerciais de eucalipto. Foram obtidas 687 amostras compostas de solo para a análise química. Avaliou-se o desempenho de duas técnicas de análise exploratória e seis procedimentos de amostragem. Os atributos P, K, Ca, Mg e S apresentaram maior coeficiente de variação (>35%). Em contrapartida, as distribuições do Al, matéria orgânica e, principalmente, pH foram as mais homogêneas. O erro amostral foi menor à medida que aumentou a quantidade de amostras compostas. A representatividade de todos os atributos químicos (erro amostral de 5%) foi alcançada com o mínimo de 309 (uma a cada 29ha, 1:29) e 295 (1:31) amostras compostas para os procedimentos de amostragem casual simples e estratificada por classe de altitude, respectivamente. Ambos os procedimentos se mostraram promissores para a amostragem de solo, especialmente, quando se aplicou o boxplot para identificação e remoção de outliers.
Palavras-chave: Boxplot; estratificação; tamanho da amostra.

INTRODUCTION

Corrective and fertilizer recommendations are based on information from sampling procedures that, if properly carried out, contribute to desired forest productivity achievement. The samples should be collected in a planned manner and adequately represent the fertility of the site to be fertilized (HEIM *et al*., 2009; OLIVEIRA *et al*., 2014). The distribution of the sample units needs to be performed in sufficient quantity so that the estimates of nutrient averages are accurate and reliable (CANTARUTTI *et al*., 1999; BRUS, 2015).

Brazil has 7.84 million hectares in which eucalypt and pinus stands and, in general, soil sampling has been carried out per plot, usually with an area of less than 50ha. Although sampling is carried out per plot, the recommendations of fertilization occur in a generic way without considering stratification criteria, such as soil type, geological formation, textural class or topography. The reasons for the generalization are based on the operational ease or the technological limitations that the companies have in order to implement precision forestry. One one hand, it is not difficult to find companies that adopt between one and three fertilization recommendations per crop year. Often, these companies implant more than 1,000ha monthly. On the other hand, farmers who do not have adequate technological infrastructure rarely apply more than one fertilization recommendation, regardless of the size of the area to be implanted.

The chemical analysis is routinely done with composite soil sample, formed by the homogeneous mixture of several simple samples. Given the predominance of research aimed at the definition of the number of simple samples to form a composite (MACHADO *et al*., 2007; SANTOS *et al*., 2009; LIMA *et al*., 2010; SANTOS *et al*., 2013, GUARÇONI *et al*., 2017) it becomes relevant the dimensioning quantity of composite samples for the evaluation of soil fertility. The sample size is influenced by the purpose of sampling, precision required, method of selection and distribution of sample units, available resources and, above all, variation of the characteristic of interest (SHIVER; BORDERS, 1996). The difficulty in representing fertility results from the spatial unevenness of soil chemical composition, since variations can occur naturally due to pedogenetic processes and/or anthropic actions such as the use and management of the soil (GÓMEZ *et al*., 2009; WASTOWSKI *et al*., 2010; TIAN *et al*., 2017; JIMÉNEZ-AGUIRRE *et al*., 2018; LEOPIZZI *et al*., 2018).

There is a widespread thought which states that random sampling is not appropriate for determining fertility, due to spatial dependence of soil chemical attributes (GUARÇONI *et al*., 2017). However, it should be pointed out that the random sampling itself, based on "classical" statistics, can generate independence between sample units when made actually at random, which would justify their use (SHIVER; BORDERS, 1996; GUARÇONI *et al*., 2017).

The simplest way to do estimated calculation of soil attributes distribution is the use of simple random sampling. It is a probabilistic procedure in which each population unit has the same chance of being selected, independently of the other selected ones (SHIVER; BORDERS, 1996). It is a fundamental procedure to select sample units, from which all other random sampling procedures were derived in order to obtain greater economy and/or precision. The random or zigzag distribution of collection points is extensively used to obtain simple soil samples. In this perspective, several surveys have contemplated the analysis of chemical attributes representativeness applying simple random sampling or geostatistical analysis (HEIM *et al*., 2009; LIMA *et al*., 2010; OLIVEIRA *et al*., 2014; OLIVEIRA *et al*., 2015, GUARÇONI *et al*., 2017). It is emphasized that the strata definition is a laborious task compared to the unrestricted sample units distribution, but it has great potential for sample error control (SHIVER; BORDERS, 1996).

Another usual procedure is stratified casual sampling, which is based on dividing the population into more homogeneous subpopulations called strata. In this one, the sample units are selected randomly with a prior or post imposed restriction by the area of each stratum. The success in the stratification process depends on the knowledge of the area, adequately establishing the strata so they are as homogeneous as possible internally and distinct from each other. Traditionally, soil sampling is done by performing a prior division of the property into homogeneous lands for the distribution of collection points (CANTARUTTI *et al*., 1999; IAC, 2018).

The fertility spatial variability between distant points is mainly a consequence of the pedogenetic processes action (SANTOS *et al*., 2009). Therefore, the stratification of extensive areas regarding the factors that influence pedogenesis (climate, material of origin, relief, time and organisms) can be an alternative to reduce the amostral error, the one which incurs when evaluating only part of the population (SIQUEIRA *et al*., 2017, SHIVER; BORDERS, 1996). The stratus attainment that discriminates the chemical attributes distribution reduces fertilizer wastefulnesses in ground without deficiency of nutrients and, when deficient, of underdosage, it makes possible to achieve excellent levels for bigger productivity (MACHADO *et al*., 2007).

Negligence in collecting fewer sample units when there is a requirement for larger sample size can cause, in some cases, misinterpretations that do not represent actual fertility conditions (LIEß, 2015). Lack of precision can result in unbalanced application of fertilizers and compromise the yield of forest stands or other crops. On the other hand, it is well known that more selected sample units with no trend provide lower variability and error estimates (SHIVER; BORDERS, 1996). It is expected, from a given moment, that the increase in sample intensity will no longer correspond to an increase in precision and that, beyond this point, each unit measured becomes costly to the sampling performed.

The records on the application of simple and stratified casual sampling procedures for the determination of the optimum amount of composite samples for soil fertility evaluation, specific for forest plantations, are rare. The following hypotheses were tested: i) the sampling intensity of the soil chemical attributes decreases with the removal of outliers; and ii) the number of samples composed to represent chemical attributes varies among stratification criteria. Thus, the objective was to evaluate the variability of soil chemical attributes and to size the minimum number of composite samples to represent the fertility of forest soils using different sampling procedures.

MATERIAL AND METHODS

The present study was carried out in stands of *Eucalyptus* sp. located in the municipalities of Carbonita, Capelinha, Itamarandiba, Minas Novas, Tourmaline and Veredinha. The predominant climate in this region of the

Jequitinhonha Valley in the state of Minas Gerais is Cwa, temperate rainy (mesothermic) with dry winter and rainy summer, according to the international system of Köppen. According to the National Institute of Meteorology, the annual averages of precipitation and temperature are in the intervals between 850 and 1250mm and from 20 to 24ºC, respectively.

The total evaluated planted area was 9,101ha, constituted of 265 commercial stands. The plots were distributed in different soil types, geological formations, drainage densities (from 0.64 to 1.17km^{-2}), texture classes and altitudes (from 819 to 1127m). The plots were concentrated in the quadrant formed between latitude coordinates 17ºS and 18ºS and longitude 42ºW and 43ºW.

The soils were classified in: Red Latosol (RL), Red Argisol (RA) and Red Yellow Argisol (RYA); the geological formations in: metamorphic and inconsolidated sediments; and the textured classes in: clayey (from 35 to 60% clay) and very clayey (more than 60% clay). Based on the variability of the chemical attributes, two groups of greater homogeneity were defined by drainage densities (G1: 0 to 0,80km^{-2} and G2: > 0.80km^{-2}), and altitudes (G1: 0 to 1000m and G2: > 1000m). The administrative boundaries were derived from the Global Administrative Areas database, version 2.8 (Figure 1).

Each sample unit (s.u.) was characterized by a composite sample of soil. The number of composite samples, or pilot sample, was defined according to the area of each plot, a routine procedure adopted by the company between the years of 2013 and 2015. Two composite samples were obtained in the fields with less than 30ha, three for those with an area between 30 and 50ha and four when they had an area larger than 50ha. Each composite sample consisted of the homogeneous mixture of five simple ones, randomly collected between the lines of planting (a simple one per interlining). According to this routine procedure, 687 samples composed of soil within 0 to 20cm depth layer were obtained.

The soil samples were conditioned in plastic containers, identified and sent to the chemical analysis. The analytical determination of the chemical composition had been made according to Raij *et al*. (2001): phosphorus (p), potassium (k), calcium (Ca) and magnesium (Mg) using the method of the ions exchanging resi; sulphur (S, SO_4^{-2}) by turbidimetry extracted with calcium phosphate, aluminum (Al) in KCl, organic matter (MO) by colorimetry and potential hidrogen (pH) in H_2O.

For each chemical attribute, the performance of two exploratory analysis techniques was evaluated; the first without removal of outliers (P1), using all the observations, and the second with removal of outliers applying the boxplot technique (P2). This technique is based on graphical analysis to represent the data variation through quartiles. All observations beyond the boxplot critical limits were identified as outliers. The critical limits were defined from the interquartile dispersion ($Q_3 - Q_1$), with the upper limit represented by "$Q_3 + 1,5(Q_3 - Q_1)$" and the lower limit by "$Q_1 - 1,5.(Q_3 - Q_1)$", in which Q_1 and Q_3 are first and third quartile, respectively.

For each exploratory analysis technique, the following sampling procedures were evaluated: Simple Casual Sampling (ACS), Stratified Casual Sampling with strata according to soil type (ACE1), geological formation (ACE2), drainage density (ACE3), textural class (ACE4) and altitude (ACE5). Twelve combinations of the two exploratory analysis techniques and the six sampling procedures were performed. Descriptive statistics were performed for each combination, using coefficient of variation (CV), mean ($\bar{x}$) and, in percentage, sample error (E).

The sample size representing the mean value of each chemical attribute was calculated in order to meet the pre-established error of 5%. Assuming the population as infinite, the calculation of the sample size for the sampling procedures was performed according to Shiver and Borders (1996).

The exploratory analysis technique and the sampling procedure that resulted in a smaller sample size to meet the pre-established precision were selected for subsequent analysis (regression and calculation of the maximum size of the soil to obtain a s.u.). Also, the selection of simple random sampling was opted by virtue of its unrestricted random characteristic and wide use.

The minimum amount of composite samples was calculated to meet sample errors of 1, 5, 10, 15, 20, 30, 40, ... and 100%. These data were submitted to non-linear regression analysis using the Levenberg-Marquardt iterative method. The logistic model of three parameters, $n = \alpha(1 + \beta\, e^{-\gamma E})^{-1}$, was adjusted to estimate the sample size (n) as a function of the sample error; α, β and γ are template parameters.

The adherence test employed was Graybill's F (F (H_0)). The accuracy of the adjustments was evaluated through the Mean Error Square Root (RQEM) and Mean Absolute Deviation (MDA). Lower values of RQEM and MDA imply higher predictive quality. The data related to the number of outliers removed, coefficient of variation and asymptote of the equations were submitted to correlation analysis according to Pearson (r).
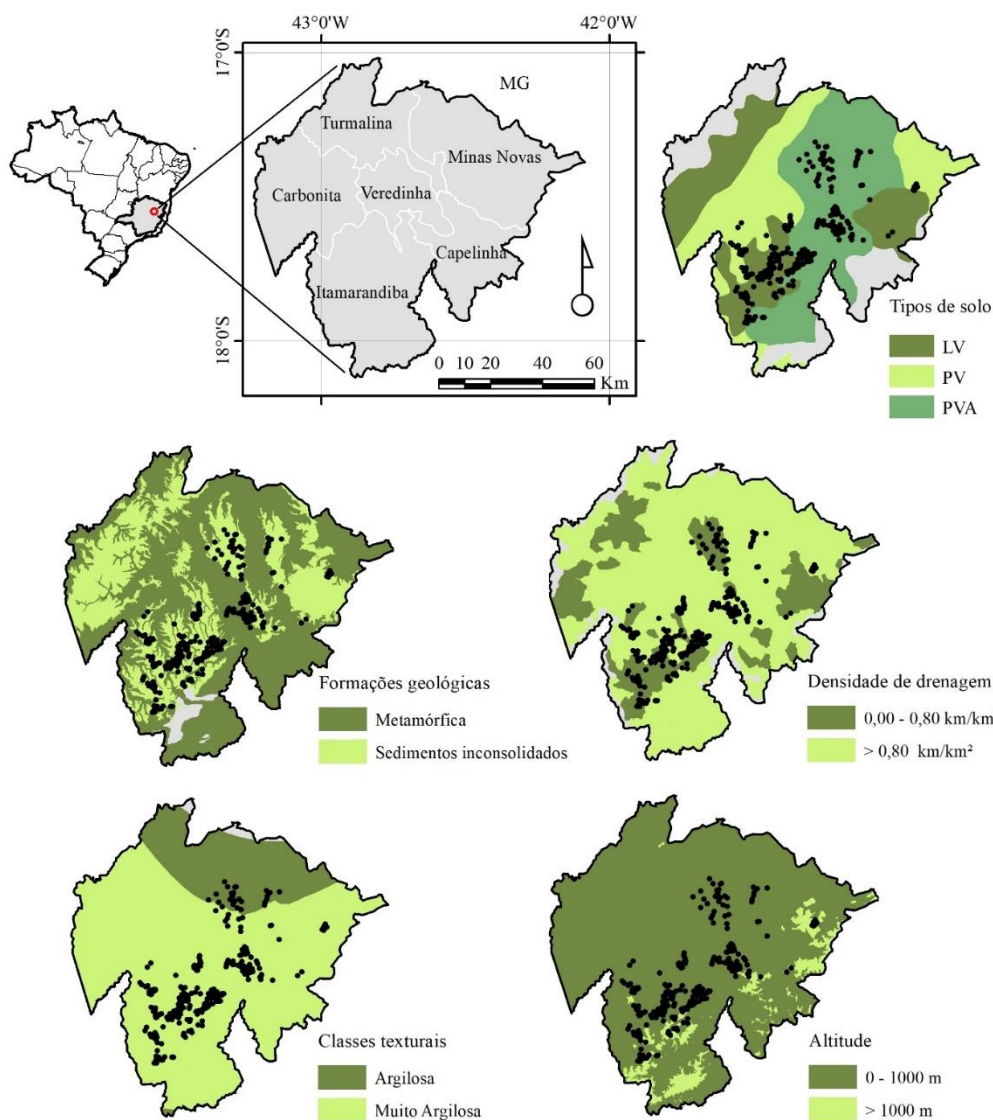
Figure 1. Location of eucalypt stands in different strata (Datum WGS84). In which: LV - Red Latosol, PV - Red Podzolic and PVA – Red-Yellow Podzolic.

Figura 1. Localização dos talhões de eucalipto em diferentes estratos (Datum WGS84). Em que: LV – Latossolo Vermelho, PV – Argissolo Vermelho e PVA – Argissolo Vermelho Amarelo.

The maximum size of the soil for obtaining a s.u. was obtained by the ratio between the planting total area (dividend) and the minimum representative quantity of composite samples.

For statistical effect diagnosis, 5% significance was used in all analysis. Those were performed with the help of the *softwares* ESRI ArcMap 10.3.1, Curve Expert 1.4 and R version 3.3 (R CORE TEAM, 2017).

RESULTS

The coefficient of variation of the evaluated chemical attributes decreased in the following order: P (66,60%) > Mg (64,63%) > Ca (54,45%) > S (48,79%) > K (43,81%) > Al (25,22%) > MO (17,26%) > pH (4,77%). The sampling error of the database was less than 5% in all procedures and attributes evaluated (Table 1). The largest error was observed for the P, ranging from 4.55 to 4.99%.

In the presence of outliers, the stratification criteria reduced the sampling error on average from 0.02 (pH) to 0.44 (P) percentage point in relation to simple random sampling. Stratification according to soil type decreased the error related to P, K, MO and pH. The definition of altitude classes increased the accuracy of Ca and Mg estimates. The stratifications made by geological formations and textural classes favored the representativeness (smaller sample error) of S and Al, respectively.

Table 1.   Abstracts of the sampling procedures statistics to represent the chemical attributes distribution in soil.

Tabela 1. Resumos das estatísticas dos procedimentos de amostragem para representar a distribuição de atributos químicos no solo.

Procedure	$\bar{x}$	E (%)	$\bar{x}^1$	E (%)^1	$\bar{x}$	E (%)	$\bar{x}^1$	E (%)^1
	\multicolumn{4}{c}{---------- P (mg dm^{-3}) ----------}	\multicolumn{4}{c}{------- K (cmol$_c$ dm^{-3}) -------}
ACS	5.713	4.99	4.807	3.37	0.041	3.28	0.040	2.80
ACE1	5.603	4.55	4.768	3.15	0.041	3.17	0.040	2.75
ACE2	5.745	4.89	4.852	3.26	0.041	3.29	0.040	2.82
ACE3	5.655	4.97	4.767	3.31	0.041	3.26	0.040	2.81
ACE4	5.799	4.86	4.852	3.29	0.041	3.29	0.040	2.81
ACE5	5.723	4.81	4.829	3.28	0.041	3.28	0.040	2.80
	\multicolumn{4}{c}{------- Ca (cmol$_c$ dm^{-3}) -------}	\multicolumn{4}{c}{------- Mg (cmol$_c$ dm^{-3}) -------}
ACS	0.520	4.08	0.471	3.20	0.198	4.84	0.163	3.55
ACE1	0.515	4.07	0.467	3.18	0.198	4.85	0.163	3.55
ACE2	0.525	3.98	0.475	3.09	0.199	4.79	0.164	3.51
ACE3	0.518	4.00	0.470	3.18	0.197	4.78	0.163	3.53
ACE4	0.522	4.00	0.473	3.12	0.197	4.79	0.163	3.55
ACE5	0.521	3.95	0.472	2.99	0.198	4.63	0.163	3.46
	\multicolumn{4}{c}{---------- S (mg dm^{-3}) ----------}	\multicolumn{4}{c}{---------- MO (g dm^{-3}) ----------}
ACS	10,227	3.65	9.656	3.29	43,262	1.29	43,224	1.28
ACE1	10.160	3.64	9.621	3.30	43,369	1.25	43,333	1.25
ACE2	10.298	3.61	9.714	3.26	43.261	1.29	43.222	1.28
ACE3	10.189	3.65	9.619	3.26	43.254	1.28	43.219	1.27
ACE4	10.294	3.62	9.716	3.26	43.084	1.26	43.048	1.25
ACE5	10.228	3.64	9.657	3.28	43.245	1.27	43.212	1.27
	\multicolumn{4}{c}{------- Al (cmol$_c$ dm^{-3}) -------}	\multicolumn{4}{c}{-------------- pH --------------}
ACS	0.658	1.89	0.658	1.89	4.759	0.36	4.756	0.34
ACE1	0.657	1.87	0.657	1.87	4.755	0.34	4.752	0.33
ACE2	0.655	1.88	0.655	1.88	4.767	0.34	4.763	0.32
ACE3	0.658	1.88	0.658	1.88	4.757	0.35	4.754	0.34
ACE4	0.660	1.86	0.660	1.86	4.755	0.35	4.752	0.33
ACE5	0.658	1.88	0.658	1.88	4.760	0.35	4.756	0.34

$\bar{x}$ = average; E = sampling error; [1]Identification and removal of outlier by applying boxplot; ACS = simple random sampling; ACE1, ACE2, ACE3, ACE4 and ACE5 = stratified random sampling with stratification according to soil type, geological formation, drainage density, textural class and altitude, respectively; Coefficient of variation before removal of outliers: P = 66,60%, K = 43,81%, Ca = 54,45%, Mg = 64,63%, S = 48,79%, Al = 25,22%, MO = 17,26%, and pH = 4,77%; The amount of observations left after identification and removal of outliers applying boxplot to P (CV = 43,02%), K (CV = 37,02%), Ca (CV = 41,51%), Mg (CV = 44,70%), S (CV = 43,01%), Al (CV = 25,22%), MO (CV = 17,14%) and pH (CV = 4,54%) was of 627, 673, 648, 613, 659, 686, 687 e 680 s.u., respectively.

After the identification and removal of outliers applying the boxplot, the order of variability (CV) was modified as follows: Mg (44,70%) > P (43,02%) > S (43,01%) > Ca (41,51%) > K (37,02%) > Al (25,22%) > MO (17,14%) > pH (4,54%). In this context, Mg presented the highest sampling error, from 3.46 to 3.55%. The predominance of outliers above the interquartile range was verified.

In the application of the boxplot technique, there was no composite sample whose total chemical attributes were identified as outlier. The maximum number of outliers identified for the same composite sample was 4 (pH, P, Ca and S), a condition found in only one s.u. In the sequence, 12 s.u. presented 3 outliers (multiple attributes) and 35 s.u. had 2 outliers (multiple attributes). Al did not show outliers.

The effect of each outlier removed was more pronounced in P; assigning the maximum allowed error of 5%, each elimination reduced around 6 to 7 composite samples to represent it. The combination of the procedures that included the removal of outliers and the casual sampling stratified by altitude classes presented better performance (Table 2).

Table 2.   Minimum representative quantity of composite samples to attend maximum error of 5%.
Tabela 2. Quantidade mínima representativa de amostras compostas para atender ao erro máximo de 5%.

Attribute	ACS	ACE1	ACE2	ACE3	ACE4	ACE5
P	685	569	658	679	650	636
P[1]	286	250	267	276	273	270
K	296	277	298	292	297	296
K[1]	212	205	215	213	214	212
Ca	458	456	435	441	439	430
Ca[1]	266	263	248	262	253	232
Mg	645	647	632	629	632	589
Mg[1]	309	309	302	307	310	295
S	368	364	359	366	360	364
S[1]	286	288	280	281	281	284
MO	46	44	46	45	44	45
MO[1]	46	43	46	45	43	44
Al	99	96	98	98	95	97
Al[1]	99	96	98	98	95	97
pH	4	4	4	4	4	4
pH[1]	4	3	3	4	3	4

[1]Identification and removal of outlier applying boxplot; ACS = simple random sampling; ACE1, ACE2, ACE3, ACE4 and ACE5 = stratified random sampling with stratification according to soil type, geological formation, drainage density, textural class and altitude, respectively.

For casual sampling stratified by altitude class, the removal of outliers decreased the minimum representative quantity of composite samples to be collected from 636 (one every 14ha, 1:14) to 295 s.u. (1:31) (Table 2). The difference of this sampling procedure relating to the simple casual was of 49 and 14 s.u. in the presence and absence of outliers, respectively. Both procedures, without outliers, were selected for subsequent analysis.

The absence of significant statistical effect by the Graybill's F test ($p > 0.05$) was observed in all equations generated to estimate the chemical attributes sample size (Table 3). The equations of those attributes of lower variability were more accurate, with lower values of RQEM and MDA. As the coefficient of variation increased (Table 1), the asymptote of the ACS equations ($r = 0.95$; $p \leq 0.05$) and ACE5 ($r = 0.92$; $p \leq 0.05$). Fewer composite samples implied larger sample errors (Figure 2). The maximum size of the soil to obtain 1 s.u. assuming different sample errors is found in Table 4.

Table 3.   Precision statistics of the equations obtained for estimation of the minimum representative quantity of composite samples (without outliers) as a function of the sample error.
Tabela 3. Estatísticas de precisão das equações obtidas para estimação da quantidade mínima representativa de amostras compostas (sem outliers) em função do erro amostral.

Attribute	α	β	γ	RQEM	MDA	p F(H$_0$)
--------------------------------- ACS ---------------------------------						
P	6.4576	-1.0059	0.0068	26.98	23.51	0.08
K	4.8048	-1.0059	0.0068	19.83	17.23	0.09
Ca	6.0454	-1.0059	0.0068	24.95	21.72	0.08
Mg	6.9730	-1.0059	0.0067	28.96	25.24	0.08
S	6.4398	-1.0059	0.0067	26.98	23.50	0.08
Al	2.2386	-1.0058	0.0067	9.15	7.89	0.09
MO	0.4876	-1.0026	0.0031	3.80	3.12	0.13
pH	0.8879	-1.0617	0.0710	0.31	0.21	0.97
--------------------------------- ACE5 ---------------------------------						
P	6.1408	-1.0059	0.0068	25.42	22.12	0.08
K	4.8048	-1.0059	0.0068	19.83	17.23	0.09
Ca	5.2827	-1.0059	0.0068	21.81	18.97	0.08
Mg	6.6535	-1.0059	0.0067	27.73	24.14	0.08
S	6.4377	-1.0059	0.0068	26.81	23.35	0.08
Al	0.0510	-1.0027	0.0032	8.50	7.27	0.10
MO	0.4537	-1.0026	0.0030	3.74	3.02	0.15
pH	0.8878	-1.0611	0.0709	0.31	0.21	0.97

$n = \alpha(1 + \beta\, e^{-\gamma E})^{-1}$, in which α, β and γ are parameters of the logistic model; RQEM = mean error square root; MDA = mean of absolute deviations; p F(H$_0$) = probability value of the Graybill's F test; ACS = Simple random sampling; and ACE5 = stratified random sampling, with two strata of altitude.
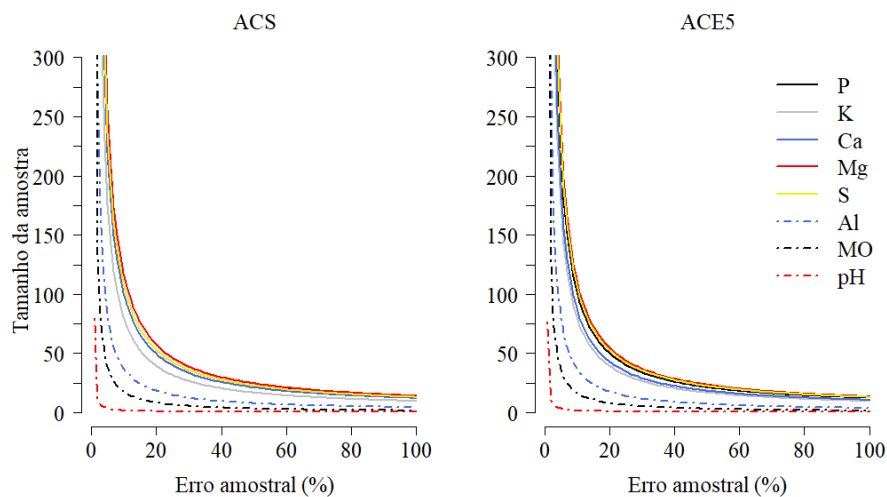

Figure 2. Graphic representation of the minimum representative quantity of composite samples (without outliers) as a function of sample error, using simple random sampling (ACS) and stratified random sampling, with two altitude strata (ACE5).

Figura 2. Representação gráfica da quantidade mínima representativa de amostras compostas (sem outliers) em função do erro amostral, empregando as amostragens casual simples (ACS) e estratificada, com dois estratos de altitude (ACE5).

Table 4.   Summary of the relation between maximum size of the plot (ha) to obtain a composite sample and sampling error.
Tabela 4. Resumo da relação entre o tamanho máximo da gleba (ha) para obtenção de uma amostra composta e o erro amostral.

Scenery	Sample error (%)									
	5	10	20	30	40	5	10	20	30	40
	-------------------- ACS --------------------					-------------------- ACE5 --------------------				
C1	29	117	455	1011	1820	31	123	479	1011	1820
C2	32	126	506	1138	1820	32	128	506	1138	1820
C3	32	126	506	1138	1820	34	134	535	1138	1820

C1 = scenario in which all chemical attributes are considered; C2 = scenario where only Ca and Mg are not direct objects of the recommendation; C3 = scenario where only Ca, Mg and S are not direct objects of the recommendation; ACS = Simple random sampling; and ACE5 = stratified random sampling, with two strata of altitude.

DISCUSSION

The coefficients of variation indicated that P, Mg, Ca, S and K presented greater heterogeneity (CV > 35%), requiring a greater sample effort to represent them. This effect is due to the wide variation of their contents in the sampled area. The variability of these attributes corroborates the one observed by Machado *et al*. (2007); Lima *et al*. (2010) and Oliveira *et al*. (2015).

The initial amount of composite samples represented the total planting area. Although this area was composed of genotypes with different ages and planting spaces, the OM and pH distributions were the most homogeneous (CV < 20%), with few outliers (0.15 to 1.02% of the observations) being identified. Lower variability of these chemical attributes (CV <12%) for a eucalyptus crop was also verified by Lima *et al*. (2017). The outliers were removed with boxplot application only in order to concentrate the analytical results closer to the respective central trends, reducing the sampling error. Although nutritional mean values were relatively close between the techniques of exploratory analysis for most soil chemical attributes, the probability of obtaining the same value from the mean in a new sampling was higher after the removal of outliers. It is emphasized that the elimination of outliers should be viewed with caution, if possible, considering statistical and biological aspects. The use of reference limits for the interpretation of chemical analysis makes it possible to evaluate the consistency of analytical results and of inconsistent values identification (CANTARUTTI *et al*., 1999).

As expected, the elimination of the outliers reduced the coefficient of variation of the chemical attributes; this reduction (in percentage points) was intensified with the increasing of removed outliers quantity (r = 0.95, *p* ≤ 0.05). There were more outliers related to Mg (10.77%), followed by P (8.73%) and Ca (5.68%) (Table 1). The other chemical attributes had less than 5% of their data identified as outliers. The sample error decreased by up to 1.66 percentage points (P, equivalent to a reduction of 33.40%).

It is important to emphasize the need to control sample and non-sample errors to minimize variation in soil properties. Non-sampling errors occur when samples are collected, recorded or analyzed erroneously, giving rise to values that deviate from the central tendency (SHIVER; BORDERS, 1996; HEIM *et al*., 2009; OLIVEIRA *et al*., 2014). Different instruments and collection professionals, nutrient mobility, irregularity of factors responsible for pedogenesis and inappropriate soil management contribute to the variability of the analytical results (ACQUA *et al*., 2013; NICOLITCH *et al*., 2016; TIAN *et al*., 2017). Even in the minimum cultivation system, haulage fertilizers, when ununiform or in rows, are activities that accentuate variability. The elimination of the entire composite sample should be considered when there is evidence of contamination in the collection or preparation of material. Whenever possible, it is recommended to compare analytical results with the history of chemical analysis and to opt for qualified laboratories. According to IAC (2018), this history allows observing trends, monitoring the evolution of fertility, detecting possible analytical problems and, if necessary, correcting soil management.

The largest sample error found in the collected composite samples was P (E = 4.99%, P1 + ACS) (Table 1). The minimum representative quantity of composite samples calculated to meet the maximum error of 5% varied between the sampling procedures. When simple random sampling without outliers was adopted, the representativeness of all chemical attributes was obtained with 309 s.u. (1:29) (Table 2). This sampling intensity was defined in agreement with the attribute of greater variability; in this case, Mg (CV of 44.70%). In established forest stands, it is suggested a division of the area into homogeneous areas smaller than 29ha for better distribution of collection points. Similarly, other ways of distributing collection points can be found in the literature, such as the division into areas not greater than 10 (CANTARUTTI *et al*., 1999) or 20ha (IAC, 2018). The minimum of 2

composite samples at sites with a total area of less than or equal to 29ha is indicated to avoid risks with the discard of some analytical result.

The most efficient stratification criterion used two altitude classes; minimum of 295 composite samples (1:31, equivalent to 95.47% in relation to ACS), calculated in accordance with Mg. In practice, this difference was relatively modest. This fact, in the 9,101ha scale, is an indication that the distribution of chemical attributes tended to randomness. It should be noted that the heterogeneity of soil properties depends on the scale analyzed (SIQUEIRA *et al.*, 2017). Therefore, it is probable that stratification becomes quite advantageous at other scales, including adopting other stratum definition criteria.

The definition of homogeneous strata allows the implementation of different fertilization recommendations, one recommendation per stratum. Since strata discriminate the distribution of chemical attributes, nutritional estimates are more accurate as stratification intensifies. Consequently, the quest for precision and accuracy requires investments for sampling strategy planning, field teams training and fertilization quality control. In this scenario, stratification by altitude class (ratio of 1:31) resulted in 248 s.u. for the stratum of up to 1000m and, when greater than 1000m, in 47 s.u.; assuming the maximum of two simultaneous recommendations in the area sampled. It is emphasized that, for stratified random sampling, the allocation of composite samples can be performed proportionally to the area of each stratum (SHIVER; BORDERS, 1996).

The sample intensity that exclusively represents Al, MO and pH varied relatively little or did not change when the boxplot was applied for the outliers removal. Therefore, the stratification and the elimination of its outliers are not indicated to meet the respective sampling adequacy. The economic viability of the unique chemical analysis for certain attributes should be investigated for the purpose of reducing analytical costs.

The equations for estimating the minimum representative quantity of composite samples showed adherence to the data (Table 3). Based on the shape of the curves generated (Figure 2), it was verified the existence of differences in the rate of decrease of s.u. quantity with the increase of the sample error. The curves of the lower variability attributes (pH, MO and Al) can be easily discriminated below the other curves, due to the greater homogeneity of the distribution of these chemical attributes in the soil of the sampled area. Sampling became progressively more costly as more composite samples were obtained to increase accuracy. Assuming larger sample errors for simple random sampling, such as 20 or 30%, the minimum representative quantity of all attributes was 20 (1: 455) and 9 s.u. (1: 1011), respectively. It is noted that the variation of 15 percentage points in the error (from 5 to 20%) reduced 289 s.u. and 10 percentage points (from 20 to 30%), only 11 s.u. This asymptotic trend followed up to 7 composite samples (sampling error beyond 100%).

In order to standardize the applied amount of limestone and/or other industrial waste to supply Ca and Mg, the representativeness (E=5%, Scenario C2) of the other attributes was obtained with a minimum of 284 (1:32) and 286 s.u. (1:32) for simple random sampling and stratified by altitude classes, respectively (Tables 2 and 4). A common operational practice in eucalyptus cultivation of large companies is the use of industrial waste from the pulp and charcoal/steel production chain to supply the demand for Ca and Mg. The application of these wastes, in general, is standardized in a single dose for all management units, considering the operational ease and reuse of waste. Similar to industrial waste, the application of standard doses of dolomitic limestone has become routine in forest areas due to the low levels of Ca and Mg in soils and to adopt the nutritional balance method. By this method, even if the Ca content in the soil is 0 cmolc dm^{-3}, 600kg ha^{-1} of CaO is recommended for high productivity, such as 50m^3 ha^{-1} year^{-1} (SANTANA *et al.*, 2014), amount equivalent to 2.2ton ha^{-1} of dolomitic limestone (30% CaO).

In cases where S is not a direct object of the recommendation, because it is an accompanying ion in fertilizers, and there is standardization of the applied amount of Ca and Mg, the sample intensity to represent the other attributes (E=5%, Scenario C3) reduces to 270 composite samples (1:34) for casual sampling stratified by altitude classes (Tables 2 and 4). This premise was based on the exclusion of three chemical attributes in the sample size designation to represent soil fertility.

The initially sampled intensity of 1:13 (687 composite samples: 9,101ha) can be reduced by more than 50%, as long there is outliers removal. This result has great practical importance because it demonstrates that soil sampling can be faster, less costly and laborious. The choice of the sampling procedure should be planned considering the exploratory analysis and sampling procedure, as well as the costs involved with geoprocessing, displacement, sampling and chemical analysis.

The balance between the amount of composite samples, sample errors, non-sample errors and resource availability is required to represent soil fertility. By neglecting the representativeness of essential chemical attributes for plant establishment and growth, silvicultural assessments and decisions become inaccurate and can be mistaken. Sampling adequacy analysis is crucial to minimize costs with soil sampling accurately.

CONCLUSIONS

- The simple and stratified random sampling procedures by altitude classes are promising for soil sampling, especially when applying the boxplot for the identification and removal of outliers.
- The minimum representative quantity of composite samples to meet the sampling error of 5% was 309 s.u. for simple random sampling, equivalent to 1 s.u. per 29 hectares (1:29). In the stratified sampling by altitude, the amount was 295 s.u., or 1:31.
- With regard to the practicality, representativeness and ease of distribution of soil samples, simple random sampling was more efficient than the stratified one.

REFERENCES

ACQUA, N. H. D.; SILVA, G. P.; BENITES, V. M.; ASSIS, R. L.; SIMON, G. A. Métodos de amostragem de solo em áreas sob plantio direto no Sudoeste Goiano. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 17, n. 2, p. 117-122, 2013.

BRUS, D. J. Balanced sampling: a versatile sampling approach for statistical soil surveys. **Geoderma**, Amsterdam, v. 253-254, p. 111-121, 2015.

CANTARUTTI, R. B.; ALVAREZ V. V. H.; RIBEIRO, A. C. Amostragem do solo. In: RIBEIRO, A. C.; GUIMARÃES, P. T. G.; ALVAREZ V., V. H., eds. **Recomendações para o uso de corretivos e fertilizantes em Minas Gerais (5ª Aproximação)**. Viçosa: Ed. UFV, 1999. p. 13-20.

GÓMEZ, J. A.; GUZMÁN, M. G.; GIRÁLDEZ, J. V.; FERERES, E. The influence of cover crops and tillage on water and sediment yield, and on nutrient, and organic matter losses in a olive orchard on a sandy loam soil. **Soil & Tillage Research**, Amsterdam, v. 106, p. 137-144, 2009.

GUARÇONI, A.; ALVAREZ, V. H.; SOBREIRA, F. M. Fundamentação teórica dos sistemas de amostragem de solo de acordo com a variabilidade de características químicas. **Terra Latinoamericana**, Chapingo, v. 35, n. 4, 2017.

HEIM, A.; WEHRLI, L.; EUGSTER, W.; SCHMIDT, N. W. I. Effects of sampling design on the probability to detect soil carbon stock changes at the Swiss CarboEurope site Lägeren. **Geoderma**, Amsterdam, v. 149, p. 347-354, 2009.

INSTITUTO AGRONÔMICO DE CAMPINAS (IAC). **Análise do solo, plantas, substratos resíduos e fertilizantes**. Disponível em: < http://www.iac.sp.gov.br/> Acesso em: 03 de jan. 2018.

JIMÉNEZ-AGUIRRE, M. T.; ISIDORO, D.; USÓN, A. Soil variability in la Violada Irrigation District (Spain): I delineating soil units for irrigation. **Geoderma**, Amsterdam, v. 311, p. 78-90, 2018.

LEOPIZZI, S.; GONDRET, K.; BOIVIN, P. Spatial variability and sampling requirements of the visual evaluation on soil structure in cropped fields. **Geoderma**, Amsterdam, v. 314, p. 58-62, 2018.

LIEß, M. Sampling for regression-based digital soil mapping: closing the gap between statistical desires and operational applicability. **Spatial Statistics**, Amsterdam, v. 13, p. 106-122, 2015.

LIMA, J. S. S.; SOUZA, G. S.; SILVA, S. A. Amostragem e variabilidade espacial de atributos químicos do solo em área de vegetação natural em regeneração. **Revista Árvore**, Viçosa, v. 34, n. 1, p. 127-136, 2010.

LIMA, E. S.; MONTANARI, R.; LOVERA, L. H.; TEIXEIRA FILHO, M. C. M.; GONZÁLEZ, A. P. Variabilidade espacial das propriedades dendrométricas do eucalipto e atributos químicos de um Neossolo Quartzarênico. **Revista de Agricultura Neotropical**, v. 4, n. 1, p. 1-11, 2017.

MACHADO, L. O.; LANA, A. M. Q.; LANA, R. M. Q.; GUIMARÃES, E. C.; FERREIRA, C. V. Variabilidade espacial de atributos químicos do solo em áreas sob sistema plantio convencional. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 31, p. 591-599, 2007.

NICOLITCH, O.; COLIN, Y.; TURPAULT, M.-P.; UROZ, S. Soil type determines the distribution of nutrient mobilizing bacterial communities in the rhizosphete of beech trees. **Soil Biology & Biochemistry**, Amsterdam, v. 103, p. 429-445, 2016.

OLIVEIRA, I. A.; MARQUES JUNIOR, J.; CAMPOS, M. C. C.; AQUINO, R. E.; FREITAS, L.; SIQUEIRA, D. S.; CUNHA, J. M. Variabilidade espacial e densidade amostral da suscetibilidade magnética e dos atributos de argissolos da região de Manicoré, AM. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 39, p. 668-681, 2015.

OLIVEIRA, I. A.; MARQUES JUNIOR, J.; CAMPOS, M. C. C.; AQUINO, R. E.; SIQUEIRA, D. S.; FREITAS, L. Spacial variability and sampling density of chemical attributes in archaeological black Earth and native forest soil in Manicoré, AM. **Revista Floresta**, Curitiba, v. 44, n. 4, p. 735-746, 2014.

R CORE TEAM. **R**: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2017.

RAIJ, B. Van; CANTARELLA, H.; QUAGGIO, J. A.; ANDRADE, J. C. (eds.) **Análise química para avaliação da fertilidade de solos tropicais**. Campinas: Instituto Agronômico, 2001. 285p.

SANTANA, R. C.; FONTAN, I. C. I.; OLIVEIRA, S. L. Implantação, manutenção e produtividade dos povoamentos. In: VALE, A. B. do; MACHADO, C. C.; PIRES, J. M. M.; VILAR, M. B.; COSTA, C. B.; NACIF, A. P. (Org.). **Eucaliptocultura no Brasil**: silvicultura, manejo e ambiência. Viçosa: Ed. SIF, 2014. p. 161-186.

SANTOS, E. O. J.; GONTIJO, I.; SILVA, M. B. Planejamento amostral de propriedades química do solo em lavoura de café conilon. **Coffee Science**, Lavras, v. 8, n. 4, p. 423-431, 2013.

SANTOS, H. C.; OLIVEIRA, F. H. T.; ARRUDA, J. A.; LOPES, A. R. S.; SOUZA JÚNIOR, R. F.; FARIAS, D. R. Amostragem para avaliação da fertilidade solo em função da variabilidade de suas características químicas. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 13, p. 849-854, 2009.

SHIVER, B. D.; BORDERS, B. E. **Sampling techniques for forest resource inventory**. Nova York: John Wiley & Sons, 1996, 356p.

SIQUEIRA, G. M.; SILVA, E. F. F.; VIDAL-VÁZQUEZ, E.; PAZ-GONZÁLES, A. Multifractal and joint multifractal analysis of general soil properties and altitude along transect. **Biosystems Engineering**, Amsterdam, p. 1-16, 2017.

TIAN, L.; ZHAO, L.; WU, X.; FANG, H.; ZHAO, Y.; YUE, G.; LIU, G.; CHEN, HAO. Vertical patterns and controls of soil nutrients in alpine grassland: implications for nutrient uptake. **Science of the Total Environment**, Amsterdam, v. 607-608, p. 855-864, 2017.

WASTOWSKI, A. D.; ROSA, G. M.; CHERUBIN, M. R, RIGON, J. P. G. Caracterização dos níveis de elementos químicos em solo, submetido a diferentes sistemas de uso e manejo, utilizando espectrometria de fluorescência de raios-x por energia dispersiva (EDXRF). **Química Nova**, São Paulo, v. 33, n. 7, p. 1449-1452, 2010.