ISSN 0104-7760
v.24 n.3 2018

Diego dos Santos Vieira[1a+], Marcio Leles Romarco de Oliveira[1b], João Ricardo Vasconcellos Gama[2a], Bruno Lafetá Oliveira[3a], Anna Karyne Costa Rego[4a], Talita Godinho Bezerra[4b]

SAMPLING PROCESSES FOR *Carapa guianensis* AUBL. IN THE AMAZON

VIEIRA, D. S.; OLIVEIRA, M. L. R.; GAMA, J. R. V; OLIVEIRA, B. L.; REGO, A. K. C.; BEZERRA, T. G. Sampling processes for *Carapa guianensis* Aubl. in the Amazon. **CERNE**, v. 24, n. 3, p. 169-179, 2018.

Keywords:
Adaptive cluster sampling
Simple random sampling
Systematic samplingn

HIGHLIGHTS

The process of adaptive cluster sampling (ACS) is not legally viable.

The processes of systematic sampling and simple random are legally viable.

The use of larger parcels in the ACS overestimates the variable of interest.

The higher the share in the ACS, the greater the sampling intensity.

ABSTRACT

The objective of this study was to analyze the adaptive cluster sampling (ACS), simple random sampling (SRS) and systematic sampling (SS) processes to obtain the number of ha-1 trees of *Carapa guianensis* Aubl. in the Amazon. The data were obtained through 100% inventory and sampling simulations, considering a DBH ≥ 25 cm, a sampling intensity of 4%, a maximum error of 10% and plots of 0.09, 0.16 and 0.25 ha. The last two sizes were only used to analyze their effect on the ACS estimators. The processes were evaluated for accuracy, precision ($E_{\%}$) and confidence interval (CI), while the mean ha^{-1} of the processes were compared with that of the 100% inventory by the Z test. The ACS process showed no significant difference between its average ha^{-1} trees and the 100% inventory, and it was also the most accurate and the only one whose CI was true. However, it presented a final sample intensity 3.6 times greater than the simple and systematic random samplings, in addition to $E_{\%}$ above 10%, which makes it unacceptable, legally, and economically unfeasible. The other processes had densities significantly higher than the 100% inventory, with sample intensities lower than ACS and $E_{\%}$ lower than 10%, making them legally viable. The use of larger plots in the ACS implies larger clusters and a greater tendency to underestimate the number of trees, resulting in larger sample errors and less accuracy.

Historic:
Received 31/01/2018
Accepted 25/06/2018

+Correspondence:
diegovieir4@gmail.com

[1] Federal University of the Jequitinhonha and Mucuri Valleys, Diamantina, Minas Gerais, Brazil - ORCID: 0000-0003-3780-1189[a], 0000-0002-8097-1135[b]
[2] Federal University of Western Pará, Cidade, Estado, Brazil - ORCID: 0000-0002-3629-3437[a]
[3] Federal Institute of Education, Science and Technology of Minas Gerais, Cidade, Estado, Brazil - ORCID: 0000-0003-2913-6617[a]
[4] Federal Rural University of Amazônia, Cidade, Estado, Brazil - ORCID: 0000-0002-1514-7801[a], 0000-0001-9968-7874[b]

DOI:
10.1590/01047760201824032514

INTRODUCTION

The *Carapa guianensis* Aubl. (andiroba) is one of the most promising non-timber forest species due to its versatility and variety of uses. It is a native species of the Amazon, with medium to large trees that can reach 2 m in diameter and 30 m in height, with a bitter bark that easily detaches in large plates (Ferraz et al., 2002). Its fruits are globose and subglobosa type capsules with four to six indiscriminate leaflets, which are rapidly dispersed and consumed by collared and white-lipped peccaries, agoutis and pacas, or attacked by insects of the genus *Hypsipyla* sp. (Dionisio et al., 2016). It occurs both in flooded areas and in firm land forests. It is more abundant in the Amazon river banks and streams (Klimas et al., 2007).

The main use of andiroba is timber, due to the excellent quality of the wood, which is resistant to insect attacks (Klimas et al., 2007; Tonini et al., 2009). However, the extraction of the oil from its fruits has also gained recognition in the national and international market in recent years. The major consumers are the cosmetic and homeopathic industries, which use the oil of their fruits to produce soaps, creams, shampoos and repellent, or as a healer, anti-inflammatory, anthelmintic and insecticide (Ferraz et al., 2002). In addition, traditional populations usually use leaf and bark tea for the same therapeutic purposes as oil (Silva; Almeida, 2014). With all these functions, andiroba became part of the National List of Medicinal Plants of the Brazilian National Health System, defined as the list of plant species with the potential to advance in the stages of the production chain and generate products of interest for the Ministry of Health of Brazil, thus reinforcing its importance in the country.

Considering the potential of generating multiproduct, the low number of trees and the intensive extraction of the andiroba in the Amazon, it is necessary to use techniques that mitigate the environmental costs of logging and non-timber extraction and that result in sustainable long-term production (Klimas et al., 2007; Coelho, 2013). Therefore, studies become essential on the ecology, structure, spatial distribution and especially on the sampling processes that evaluate the number of trees and subsidize the forest managers in the decision making during the execution and evaluation of the forest management activities, with the objective of conserving not only the trees, but also of the animals that feed on the fruits of these trees (Brown et al., 2013; Abreu et al., 2014). In addition, from a production point of view, studies on sampling processes are fundamental to characterize the available wood stock for the forest harvest, aiding forest planning (Abreu et al., 2004; Vieira, 2015).

The basic tool to obtain information for the accomplishment of these studies is the forest inventory by sampling, based on the simple (SRS), systematic (SS) or stratified (STS) sampling processes. However, with regard quantifying trees of a single species, which usually occurs when the interest is non-timber forest products, these processes are usually inefficient due to the numerical rarity and aggregation characteristics of non-timber species (Acharya et al., 2000; Bruzinga et al., 2014). Therefore, the use of SRS, SS and STS would result in low probability of obtaining plots with the occurrence of trees of the selected species and high number of zero plots, which would make SRS, SS and STS inefficient (Acharya et al., 2000).

Thompson (1990) and Thompson e Seber (1996) developed a sampling process for populations with numerical rarity and aggregation, called adaptive cluster sampling (ACS). The emphasis on rare and clustered populations is that they are often harder to inventory. There is no one definition for rare and clustered populations that all statisticians and ecologists agree on, but generally these are populations that are difficult to detect (so in fact may not be rare, but sightings are rare), or the population that is sparse in some sense. (Brown et al., 2013). In the ACS, an initial set of units is selected by some probabilistic process (e.g. SRS or SS), and whenever the variable of interest is observed in these initial units, neighboring units are investigated, thus increasing the efficiency of the estimator. This way, it is more likely to find a rare element in the neighborhood of another element when the population is aggregated (Thompson; Seber, 1996).

The variable of interest refers to the inclusion condition (C), determined as the presence of one or more trees of the species in the sample unit, but can be any variable, as long as it is relative to the sample unit. Given the condition C, the neighboring plots are investigated and so on, until no more trees of the species are found in neighboring units, forming clusters composed of a set of network units (those that met condition C) and margin units (which did not satisfy condition C, but were in the vicinity of the network units) (Thompson, 1990). The final sampling intensity is thex collection of clusters that were detected in the initial sample, including any of the sample units that were in the initial sample but did not meet condition C (Brown et al., 2013).

In the state of Pará (BR), there are few scientific records of the use of ACS for quantification of non-timber species, limited to the research carried out by Vieira et al. (2017) in the Tapajós National Forest. Thus, the objective of this study was to evaluate the use of ACS to estimate the number of trees ha^{-1} and the diametric and spatial distributions of andiroba, comparing the ACS, SS and 100% inventory processes. Hypotheses: H_0 = there are no differences between the number of trees, the spatial pattern and the diametric structure estimated by the sampling processes and by the 100% inventory; H_1 = rejection of H_0 and ACS is the best sampling process. The effect of sample unit size on ACS estimators was also evaluated.

MATERIALS AND METHODS

Area of study

The study was carried out in Work Unit No. 16 (WU-16), with 100 ha located in the Annual Production Unit No. 09 and administered by the National Cooperative of Tapajós (Coomflona), located in the municipality of Belterra, state of Pará, Brazil. The region's climate, according to the classification of Köppen, is Ami type, with annual temperature and precipitation of 25.5 °C and 1820 mm, respectively (Alvares et al., 2013). The land relief is slightly uneven and presents a topography varying from mild to undulating, with a predominance of Yellow Latosol (Alvares et al., 2013). The vegetation is classified as Dense Ombrophylous Forest of Terra Firme, characterized by the dominance of large trees and by the abundance of woody lianas, palms and epiphytes (IBGE, 2012).

Data collection

Data collection was performed by means of an inventory, with cartesian (X, Y) coordinates of all andiroba trees with diameter at 1.30 m above ground (DBH) equal to or larger than 25 cm. It is defined based on commercial seed productivity criteria of reproductively mature trees. WU-16 was divided into 20 parallel and contiguous blocks of 5 ha each (50 x 1,000 m) delimited by marks in the north-south direction (Vieira et al., 2017). Beacons were placed every 25 m in every mark, with the respective margins in relation to the origin, for later registration of the Y coordinates of each tree. The X coordinate of each individual was obtained by means of the number of the mark and the distance of each tree in relation to it. Then, the next mark was attended to, executing the same work routine, and so on, until the last WU-16 mark was completed (Cavalcanti et al., 2011).

Sampling processes

With the coordinates of the andiroba trees, maps were elaborated where grids represented contiguous sample units (plots) of 30 x 30 m (0.09 ha), 40 x 40 m (0.16 ha) and 50 x 50 m (0.25 ha). The 0.09 ha units were used in all the sampling processes evaluated, namely: simple random sampling (SRS), without replacement of sample units; systematic sampling (SS), with randomization only of the first plot and the others distributed equally; and adaptive cluster sampling (ACS), with inclusion condition (C) determined by the presence of at least one tree ($y_i \geq 1$) of the species, while the units of 0.16 and 0.25 ha were used only in the ACS to evaluate the effect of plot size on the ACS estimators.

The sample intensity used was 4% of the WU area. Thus, for SRS and SS, 45 sample units (30 x 30 m) were distributed throughout the area. The identification of the presence of trees in the samples was performed by crossing the X and Y coordinates of the trees with the limits of each unit. In the ACS, the initial samples used were those obtained by SRS, so that if there were one or more trees in one of the initial units, the adjacent units (north, south, east and west) would be attached to the sample, forming the networks (Thompson; Seber, 1996). In addition, for each sampling process, 30 random simulations were performed.

The maximum sample error ($E_{\%}$) allowed for all sampling processes was 10% of the mean trees.ha^{-1}, at 95% probability. The same $E_{\%}$ accepted by the Brazilian Institute of Environment and Natural Renewable Resources (IBAMA) was used for the variables abundance, basal area and community volume, because there are no specific normative instructions for $E_{\%}$ of population inventories (e.g. non-timber species), so this same value is adopted for other purposes, as observed in Soares et al. (2009), Bruzinga et al. (2014) and Vieira et al. (2017), taking into account local laws.

Sampling analysis

Five criteria were used to evaluate the quality of the sampling processes in obtaining basic information for the planning and execution of the forest management of andiroba: three are statistical (sample mean ha^{-1}, sampling error and accuracy) and two technical (diameter distribution and spatial pattern). The statistical criteria were used to evaluate the quality of the forest inventory, while the technical to evaluate if the information calculated from the sampling processes were the same as those observed in the inventory.

Statistical criteria

For each of the simulations, the estimate of the average number of trees per hectare and the mean variance were calculated. The mean (1) and the mean variance of the SRS (2) were calculated using the estimators commonly used and described by Shiver e Borders (1996). Where: n = number of units in the sample; a = sample unit area; N = number of units available in the population; y_i = number of trees in the i^{th} sample unit; and S^2 = estimator of the sample variance.

$$\bar{Y} = \frac{\sum_{i=1}^{n_i} y_i}{n} \cdot \left(\frac{10{,}000}{a}\right) \quad [1]$$

$$S_{\bar{Y}}^2 = \frac{S_2}{n_i} \cdot \left(1 - \frac{n_i}{N}\right) \quad [2]$$

In SS, the mean number of trees estimator per hectare was the same as for SRS, while the mean variance (3) was calculated by successive differences, as described by Loetsch e Haller (1964). In ACS, mean (4) and mean variance (5) estimates were obtained by the modified Hansen-Hurwitz estimator (Hansen; Hurwitz, 1943; Thompson, 1990), where: w_i = weight of the networks formed in each cluster, i.e., number of trees in relation to the number of network units i; and m_i = number of units in network i.

$$S_{\bar{Y}}^2 = \frac{\sum_{i=1}^{n_i} (y_i - y_{i+1})^2}{2n_i(n_i - 1)} \cdot \frac{(N - n_i)}{N} \quad [3]$$

$$\bar{Y}_{HH^*} = \left(\frac{1}{n_1} \sum_{i=1}^{n_i} w_i\right) \quad [4]$$

$$w_i = \frac{\sum_{i=1}^{m_i} y_i}{m_i} \quad [5]$$

$$S_{\bar{Y}_{HH^*}}^2 = \frac{N - n_1}{Nn_i(n_i - 1)} \sum_{i=1}^{n_1} \left(w_i - \bar{Y}_{HH^*}\right)^2 \quad [6]$$

The values of precision ($E_{\%}$) (7), accuracy ($ACR_{\%}$) (8) and confidence interval (CI) (9) were also calculated. In these cases, the lower the $E_{\%}$ and $ACR_{\%}$ values, the higher the accuracy, while in the CI, if the observed population mean was within the confidence interval estimated by the sampling processes, the largest reliability of the estimated number of trees per hectare, where t = tabulated value of t for a level of significance of 0.05%, with n - 1 degrees of freedom; S_x = standard error of the mean; T_{est} = total estimated; T_{obs} = total observed; and r = number of simulations (30).

$$E_{\%} = \pm \frac{t \cdot S_x}{\bar{Y}} \cdot 100 \quad [7]$$

$$ACR_{\%} = \frac{\sum_{i=1}^{r} \frac{T_{est} - T_{obs}}{T_{obs}} \cdot 100}{r} \quad [8]$$

$$CI = \bar{Y} - t \cdot S_x \leq \mu \leq \bar{Y} + t \cdot S_x \quad [9]$$

The statistical consistency of the estimated number of trees per hectare was assessed using the Z test, at 95% probability, using the mean ha^{-1} of the 100% inventory (μ) as the reference value (Péllico Netto et al., 2017). The hypotheses tested were: null (H_0), when the equality between the means of the sampling process and the inventory 100% is not rejected; and alternative (H_1), when this equality is rejected. In addition, the Graybill F test, at 95% probability, was used to evaluate significant differences between tree·ha^{-1} averages, accuracy rates and precision obtained by the sampling processes, considering as standard SRS and SS processes. In this case, the hypotheses tested were of similarity (H_0) and dissimilarity (H_1) between the average, the accuracy and the precision of the processes.

Technical criteria

Based on the diameter information, the diametric structures were calculated for the SRS, SS, ACS (0.09 ha) and 100% inventory, obtained by calculating the trees in diametric classes with a 10 cm amplitude, after the inclusion of the tree diameters. In this case, the number of trees per diameter class was obtained by means of the average of 30 simulations. The Kolmogorov-Smirnov test (K-S) was used to evaluate significant differences between the 100% inventory diameter structure and those obtained from SRS, SS and ACS processes (0.09 ha), at 95% of probability.

The spatial pattern of the trees in the SRS, SS and ACS processes (0.09 ha) was determined using the Morisita index (IM_i). For each of the 30 simulations a value of IM_i was calculated, with the final spatial pattern being the one with the highest frequency. If the value of IM_i equals 1.0, the spatial pattern is random, otherwise the randomness hypothesis is rejected and an aggregate standard is assumed when the IM_i value is greater than 1.0, and regular, when it is less than 1.0. The level of significance of the Morisita index was tested using the F statistic, at 95% probability. The value of F was compared with the tabulated value of F, with n^{-1} degrees of freedom for the numerator and infinity for the denominator.

The spatial pattern with the 100% inventory data was defined by the Ripley function K(s). The calculation of K(s) was performed as a function of a circle with a radius of 5 m centered in each tree, in which the number of events present in the area of that circle was counted (Araújo et al., 2014). By varying the radius at a maximum distance of 500 m, the spatial pattern of the species was detected at different distance scales. This way, the Ripley function K(s) evaluated the relationship between pairs of events every 5 m up to the maximum distance of 500 m, considered the half of the largest longitudinal axis of the area (Vieira et al., 2017).

Particular cases occurred when the trees were near the edges of the area, as K(s) is cumulative and computes all distances between all events. Trees near the edge of a radius greater than the boundary of the map could not be interpreted as if there were no neighbors (Capretz et al., 2012). The trees existed, but because they were outside the limits of the study area they were not computed. Therefore, the number of trees near events close to the boundaries of the map would be lower than the others, causing a bias in the calculation of the K(s) estimator. In view of this, the estimator of the function K(s) with isotropic edge correction proposed by Ripley (1979) (10) was used, where: g = number of trees in the study region; X_i and X_j = coordinates of the map points; = Euclidean distance between the location X_i and X_j; s = arbitrary distance vector; W_i $I(X_i, X_j)$ = correction function for edge effect, which represents the ratio of the circumference with center in X_i and with radius that is outside the region of study; $\hat{\lambda} = n/|A|$ = number of trees divided by the area of the study region, being an unbiased estimator of the intensity of the process; and I(U) = indicator function that assumes value 1 whenever the condition U is true, and zero when it is false.

$$K(s)=\frac{1}{\hat{\lambda} g}\sum_{i=1}^{g}\sum_{i=1}^{g} W_i^{-1}(X_i,X_j)\,I(\|X_i - X_j\| \le S) \qquad [10]$$

for $i \neq j$ and $s > 0$ [10.1]

Then, to analyze the data graphically, a reliable envelope was constructed by means of 1,000 Monte Carlo simulations, carried out through the model of complete spatial randomness (CSR) (Meira Junior et al., 2017). Afterwards, the function K(s) was calculated for the simulation results, storing the minimum and maximum estimates of K(s) used to generate confidence intervals at 99% probability (Capretz et al., 2012). In order to better visualize the characteristics of CSR, the values of the function K(s) were transformed to L(s), according to the expression (11), and distributed graphically, in which the abscissa and ordinate axes represent, respectively, the accumulated distances and the transformed values of the function K(s) (Ripley, 1979).

$$L(s)=\sqrt{\frac{\hat{K}(s)}{\pi}} \qquad [11]$$

If the values of L(s) remain within the reliable envelopes identified by two dotted boundary lines, one positive and one negative, the CSR hypothesis is accepted and the spatial pattern is completely random, otherwise the hypothesis is rejected and it is assumed that the spatial pattern of individuals is aggregated when it passes the upper boundary of the envelope and is regular when it passes the lower boundary (Ripley, 1979). The calculation of the Ripley K function and the Morisita index was done using the software R version 3.3.2, using the "splancs" and "vegan", respectively.

RESULTS AND DISCUSSION

In the 100% inventory, 2.45 trees.ha^{-1} were recorded, while the simple random sampling (SRS), systematic (SS) and adaptive clustering (ACS) presented mean values of 2.73, 2.79 and 2.54 trees.ha^{-1}, respectively (Table 1). These values were lower than those described by Klimas et al. (2007), Tonini et al. (2009) and Guarino et al. (2014), who found, on average, 15.4 trees.ha^{-1}, but was within the range specified by Ferraz et al. (2002) for terra firme forests (0.3 to 9.0 trees·ha^{-1}) in the Amazon. It was also observed that the SRS and SS processes presented sample errors ($E_{\%}$) lower than 10%, at 95% probability, but were statistically less accurate than the ACS (0.09 ha) and with significant tendencies to overestimate the mean trees·ha^{-1}, when compared to the 100% inventory (Table 2). In addition, according to Graybill's F test, SRS and SS presented statistical differences in relation to the mean of trees·ha^{-1}, accuracy and precision.

The simulation results demonstrated that in 27 out of 30 carried out in SRS, the true value of the number of trees·ha^{-1} (2.45 trees·ha^{-1}) was outside the confidence interval (CI), while 28 cases were recorded in SS (Figure

TABLE 1 Mean estimates of the 30 simulations and Z test for the sampling processes..

Categories	μ	$\bar{Y}$	$S^2_{\bar{Y}}$	$E_{\%}$	$\|ACR_{\%}\|$	Confidence interval	
						Upper	Lower
SRS	2.45	2.73*	0.007	5.11	29.9	2.87	2.60
SS	2.45	2.79*	0.007	5.15	31.6	2.93	2.65
ACS (0.09 ha)	2.45	2.54^{ns}	0.131	23.15	23.8	3.13	1.95

Where: SRS = simple random sampling; SS = systematic sampling; ACS = adaptive clusters sampling; μ = average ha^{-1} trees obtained by the inventory 100%; $\bar{Y}$ = the average trees·ha^{-1} estimated by the sampling processes; $S^2_{\bar{Y}}$ = mean variance; $E_{\%}$ = sampling error, in percentage; $|ACR_{\%}|$ = accuracy, in module; * = significant differences between the average of trees ha^{-1} of the 100% inventory and that of the sampling processes; and ns = equality between the means by the Z test, at 95% probability.

TABLE 2 Graybill F test.

Sampling Alternatives			$E_{\%}$	$\|ACR_{\%}\|$	$\bar{Y}$
SRS	x	SS	8.30*	11.85*	7.99*
SRS	x	ACS (0.09 ha)	84.10*	13.17*	11.37*
SS	x	ACS (0.09 ha)	92.75*	18.96*	15.38*

Where: SRS = simple random sampling; SS = systematic sampling; ACS = adaptive cluster sampling Fcrítico = 3.3404; $\bar{Y}$ =mean number of trees·ha^{-1} estimated by the sampling processes; $E_{\%}$ = sampling error, in percentage; $|ACR_{\%}|$ = accuracy, in module; * = significant differences between the averages of trees ha-1 by the F-test, at 95% probability.

1). However, considering the $E_{\%}$, which varied from 5.1 to 7.2% and 3.8 to 5.1% for SRS and SS, respectively, and the impossibility of obtaining the true mean (μ) by means of sample inventories and, consequently, the accuracy and certainty that the CI encompasses μ, it can be affirmed that all simulations of sample inventories were satisfactory according to IBAMA Normative Instruction No. 01/2007, although cases were recorded in which it was underestimated down to 1.45 tree·ha^{-1}, that is, more than half of the true average. Thus, all the information calculated from the dendrometric data obtained from these simulations are valid and can serve as a basis for the planning of conservation activities or forest management.

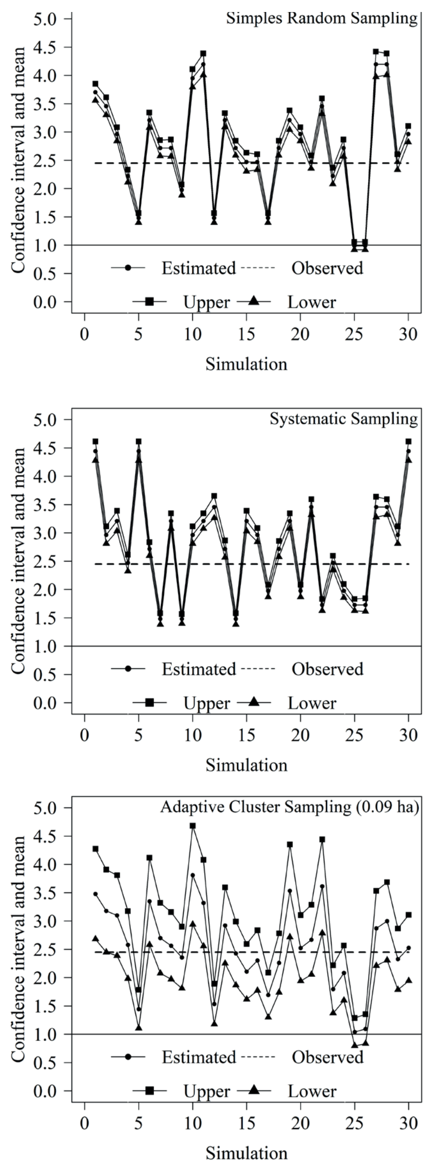

FIGURE 1 Estimation of the total number of individuals and the lower and upper limits of the confidence interval for the sampling processes.

The ACS process (0.09 ha) showed a sampling error ($E_{\%}$) greater than 10%, at 95% probability, but it was significantly more accurate than the other processes and the only one, on average, whose CI covered the mean (Table 1). In addition, according to the Z test, the averages of trees·ha^{-1} of the 100% inventory and the ACS (0.09 ha) were significantly similar at 95% probability. Results also showed that, in 17 out of 30 simulations performed to ACS (0.09 ha), the true value of the number of trees·ha^{-1} was within the confidence interval (CI), because the sampling errors were high ($\bar{E}_{\%}$ = 23.8, CV% = 0.96), which increased the amplitude of CIs and facilitated the inclusion of the true mean.

It was also found that the mean sample intensity for the ACS (0.09 ha) process was 3.6 times higher than for the SRS and SS alternatives, which leads to the belief in higher sample costs. In this case, the classical theory that the higher the sample intensity, the lower the $E_{\%}$ cannot be attributed to ACS (0.09 ha), because, according to Brown et al. (2013) and Bruzinga et al. (2014), the estimator of the variance of this process uses the net weight, not only the number of final sample units. Therefore, despite the greater sample size, the ACS (0.09 ha) remained with a high $E_{\%}$. In addition, the average size of the networks sampled in the ACS (0.09 ha) was 7 sample units, of which the largest network had 23 units, while the average number of trees per network unit was 2.0.

The low tree occurrence produced low net weight ($\bar{W}$ = 1,2), which influenced the modified Hansen-Hurwitz estimators. The low network weights can be results of factors such as the degree of occupation of the tree under the soil, the degree of aggregation and the area of the sample unit, which are directly related to each other and affect the accuracy and precision of the estimates obtained in the ACS, because trees with large diameter and that are reproductively mature can hardly make up aggregate and spaced groups, so that the network units present 10 to 20 trees. Treetop diameter and other abiotic and biotic factors limit the occurrence of nearby and reproductively mature andiroba trees, reducing the probability of occurrence of other trees of the species in the plot area. Thus, the probability of obtaining a final sample slightly larger than the initial sample and with clusters made up of few network units and high number of trees, which according to Brown et al. (2013) is the arrangement for efficiency in the ACS, is low.

From these results, it can also be inferred that the sampling error ($E_{\%}$) is not a good criterion for evaluating the quality of sample inventories of monospecific populations when the interest is to obtain the average of trees·ha^{-1}, because, despite the high average sampling error observed in ACS (0.09 ha) ($\bar{E}_{\%} = 23{,}8$, $CV_{\%} = 0{,}96$), the mean tree estimation was significantly similar to that described by the inventory 100%. Thus, the ACS (0.09 ha) process can be considered an acceptable option for andiroba sample inventories. However, it would be costlier due to the high final sample intensity, and imprecise, according to IBAMA Normative Instruction No. 01/2007, which is not the case with SRS and SS. However, although the $E_{\%}$ in the SRS and the SS are much lower than the maximum permissible by IBAMA, no statistical similarities were found between the averages of tree.ha^{-1} of these processes and those of the 100% inventory, in addition to the CI that does not cover the average.

Observing only the adaptive processes, it was verified that the plot size was inversely proportional to the average estimate and directly proportional to the mean variance and the sampling error (Table 3). In all three cases, the mean confidence interval (CI) covered the true mean (μ), with the ACS (0.09 ha) alternative being the least amplitude for the CI and, consequently, lower sampling error and greater precision. The estimates of the number of trees.ha^{-1}, precision and accuracy were, according to Graybill's F test, significantly different from each other: they had a p-value of less than 0.05. In contrast, the Z test captured significant differences only between the averages of trees.ha^{-1} of the inventory 100% and those of the ACS (0.25 ha) process. Thus, it can be stated that the size of the sample unit influenced the modified Hansen-Hurwitz estimators.

In all three adaptive processes, the cluster design and layout demonstrated that larger sample units form clusters that cover almost all trees (Figure 2), as found when 0.25 ha initial units were used, at which time only one cluster accounted for 81.6% of trees and 41.5% of the WU-16 area, a situation that, according to Thompson e Seber (1996), is caused by the low inclusion condition

TABLE 3 Mean estimates of 30 simulations and z test for adaptive cluster sampling processes, on different sample unit sizes.

Sampling alternatives	μ	$\bar{Y}$	$s^2_{\bar{Y}}$	$E_{\%}$	$\lvert ACR_{\%} \rvert$	Confidence interval	
							Lower
ACS (0.09 ha)	2.45	2.54 ns	0.131	23.15	23.8	3.13	1.95
ACS (0.16 ha)	2.45	2.39 ns	0.202	30.85	30.9	3.11	1.66
ACS (0.25 ha)	2.45	2.27 *	0.246	37.27	28.3	3.10	1.44

where: ACS, μ, $\bar{Y}$, $s^2_{\bar{Y}}$, $E_{\%}$, $\lvert ACR_{\%} \rvert$, * and ns = already mentioned.

(C = 1) adopted. Thus, the probability of a unit be selected in this cluster in the initial sample is high. This would increase the sampling effort due to cluster size and reduce process efficiency with respect to sampling estimates and costs (Bruzinga et al., 2014). In this case, the solution would be to determine values for C > 1.0, since 15.25% of the 400 units available in the area (N) presented at least two individuals of andiroba. However, nothing will happen if the units added adaptively do not attend a new inclusion condition (C).

With the reduction of the initial unit for 0.09 and 0.16 ha, the same cluster formed smaller networks, resulting in a much smaller final sample area (Figure 2). In these cases, the adoption of values for C > 1.0 would lead to non-clustering, because only 3.24 and 8.96% of the N units available in the area to 0.09 and 0.16 ha,

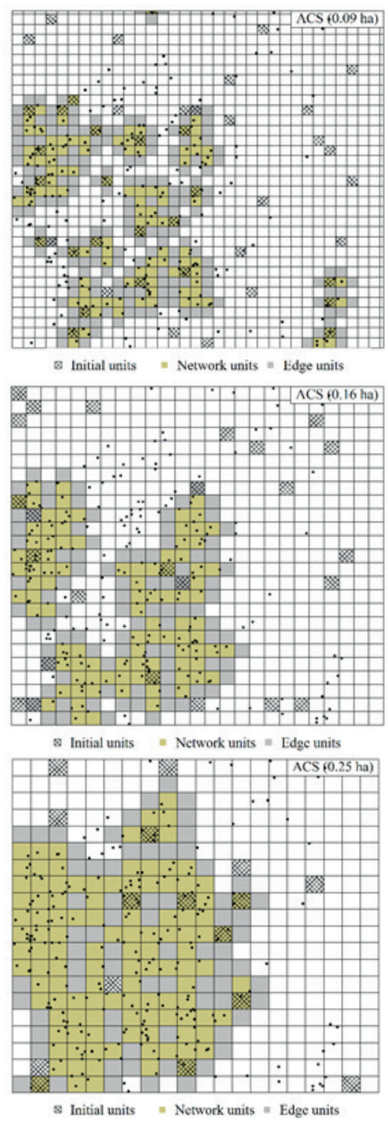

FIGURE 2 Illustration of the adaptive cluster sampling process by different sample unit sizes. In which ACS = adaptive cluster sampling.

respectively, presented more than one tree. It was also observed that for the same inclusion condition there was little variation in the number of trees and weights of the networks among the three sample unit sizes, which caused little discrepancy in the mean, because the networks that presented many individuals also had many networks. Thus, in the calculation of the mean estimator, the weight will be used once for each unit of the initial sample (Bruzinga et al., 2014), so the probability of larger networks having more than one of their selected plots in the initial sample is greater than in the smaller network plots.

The value of C for andiroba populations should be defined based on the abundance of the species and the size of the plot to be used. Thus, for populations with high abundance and larger plots, high C values are indicated, because when a low C is adopted, the size of the final sample will be large since the networks that will be formed may encompass many units (Brown, 2003) resulting in the degeneration of the sampling process, since almost 100% inventory is being carried out. On the other hand, in populations with low abundance and smaller plots, C is recommended since the value of C is too high, and few units will be adhered to the sampling process (Brown, 2003). Other important factors to be considered are the degree of aggregation and the distance between trees of the species. As a general principle, the more grouped the population, the more efficient the ACS will be compared to SRS and SS (Brown et al., 2013), so populations with low aggregation intensity will have many size-1 networks with no variance, hence the adaptive clustering would imply many edge units, which leaves an advantage in this context for SRS and SS.

Diameter distribution

The diameter distribution graph obtained from the 100% inventory and the sampling procedures showed that the species presented a low number of trees in the lower diameter classes and a high frequency in the intermediate classes, with a marked reduction in the larger classes (Figure 3). In this way, it can be inferred that the diameter distributions do not follow the negative (J-inverted) exponential trend described by Klimas et al. (2007), Tonini et al. (2009), Guarino et al. (2014) and Londres et al. (2017) in the Amazon, which indicates a probable imbalance between mortality and recruitment, caused by possible past or natural anthropogenic disturbances. The Kolmogorov-Smirnov adhesion test did not show significant differences between the diametric structures obtained by the sampling procedures and the 100% inventory (SRS D_{calc} = 0.254, SS D_{calc} = 0.125, ACS (0.09 ha) D_{calc} = 0.125, D_{tab} = 0.869). In spite of this, the ACS alternative (0.09 ha) was the most efficient, since the error associated with each class was lower when compared to the SRS and SS processes.

It is believed that if the same area were evaluated with a DBH ≥ 10 cm, which is used in most studies with andiroba, the mentioned imbalance would continue to occur, since the new inventoried trees would not compensate for the low number of trees.ha^{-1} observed in the initial classes of this study. In the Tapajós National Forest, with data from a prospecting inventory of 200 ha, 7 km away from the study area, Vieira et al. (2013) verified

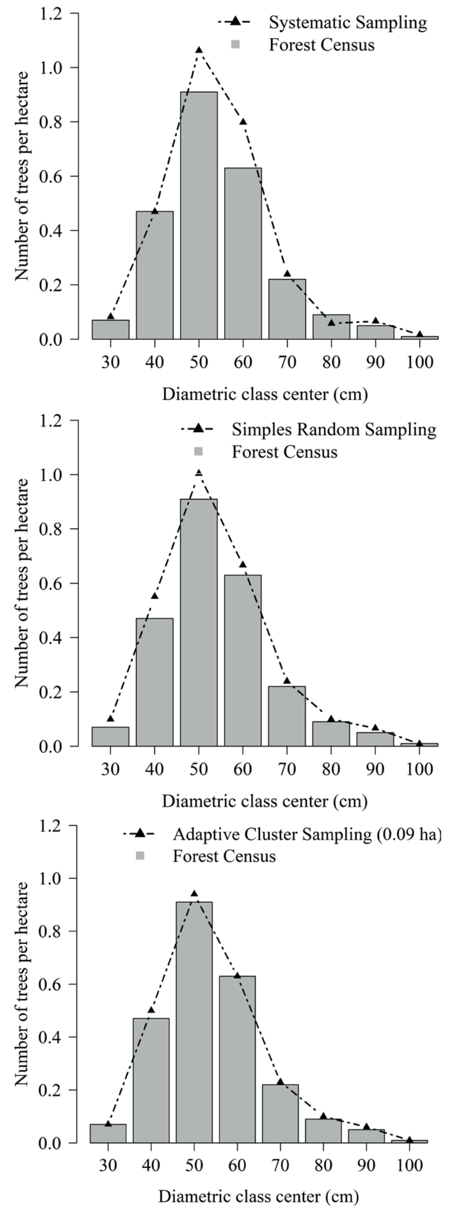

FIGURE 3 Observed (100% inventory) and estimated (sampling alternatives) diameter distribution for a population of andiroba in the Tapajós National Forest, state of Pará, Brazil.

absence of andiroba trees in diameter classes of 15 to 35 cm. The same trend was found by Coelho (2013), who carried out a 100% inventory of andiroba with DBH ≥ 20 cm in an area of 963.5 ha in the Santo Antônio community, far more than 80 km from the study area.

It is also believed that the low frequency in the initial classes may be a result of the high intra and interannual rates of mortality (20 to 50%) of the species in the terra firme area (Guarino et al., 2014), which, associated with strong temporal variations in precipitation, causes oscillations in the diameter structure in the long term, directly influencing the entry of trees into larger diameter classes. In addition, the seeds are also easily predated by insects, mainly of the genus *Hypsipyla* sp., whose larvae attack the fruits, making galleries, damaging more than 39% of the seed production, which reduces the germination rate and, consequently, the establishment of natural regeneration (Pinto et al., 2013). A second, but less probable, hypothesis is related to the current levels of andiroba seed collection in the area, which, according to the Annual Operational Plan (OPA), is made in 40% of the total number of trees in the UPA, where the study area is inserted.

Spatial pattern

The K(s) function of Ripley demonstrated that the andiroba population presents a predominantly aggregated spatial pattern (Figure 4). That is, the number of trees observed in the vicinity of any andiroba tree was higher than expected for a random pattern. This corroborates the results found by Klimas et al. (2007) and Guarino et al. (2014) but differs from the random pattern described by Tonini et al. (2009), in the south of the state of Roraima. In addition, it was observed that the confidence envelope indicates increasing values of the function L(s) up to ± 450 m, from which the curve assumed a decreasing behavior, indicating that the cluster size within the WU varies up to approximately 450 m, and which after this scale is a little less clear.

Using the Morisita index (IM_i), it was verified that 27 of the 30 simulations of SRS and SS presented values of the index statistically higher than 1.0, configuring an aggregated spatial pattern. Of the three cases where there was no aggregate distribution for SRS and SS, two were unable to estimate a value for the index, which occurred when the sampling process recorded only one tree in all samples, and one resulted in IM_i equal to zero, caused by the same abundance of trees in the sample units in which andiroba was present (Amaral et al., 2015). On the other hand, in the ACS all the simulations resulted in an aggregate spatial distribution. Thus, it can be stated

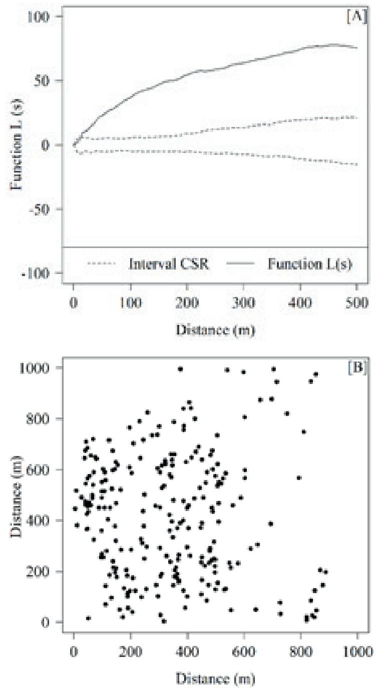

FIGURE 4 Point pattern, Ripley's K(s) function for a population of andiroba in the Tapajós National Forest, state of Pará, Brazil. where: CSR = complete spatial randomness.

that the pattern obtained by the 100% inventory is similar to that described by the sampling procedures.

It is believed that the observed spatial pattern is strongly determined by the zoocoric and barocoric dispersion syndromes of the species. This occurs because, normally, these syndromes limit the distribution of propagules, causing high rates of germination and recruitment in areas close to the source trees, while in other areas these processes are smaller and random (Negrini et al., 2012). According to Guarino et al. (2014), edaphic factors can provide the occurrence of microsomes favorable to the occurrence of aggregates of the species, since the establishment of the trees occurs according to the soil characteristics conducive to their development. In general, more clayey soils, with higher base saturation and high levels of manganese, exchangeable calcium and total magnesium, are more favorable to the establishment and growth of andirobas trees (Magalhães et al., 1986).

Selection of the sampling process

In this study, all alternatives are considered valid, however the cost of sampling and the accuracy of the estimates are factors to be considered, mainly due to the maximum admissible error of 10%, at 95% probability, required by IBAMA. ACS (0.09 ha) was the most accurate and the only one that did not show significant differences between its average ha^{-1} and 100% of the trees but presented a final sampling intensity of 3.6 times greater than the other processes and $E_{\%}$ above the maximum admissible, which made it unacceptable to IBAMA. The other processes, although showing significantly higher mean values than those from the 100% inventory and lower accuracy, presented an $E_{\%}$ lower than 10% and a lower sample intensity than the ACS (0.09 ha), which results in lower costs and compliance with IBAMA standards. In addition, in the three processes, the diameter structure and the spatial patterns were similar to those of the 100% inventory, not influencing the decision.

Thus, among the evaluated processes, the traditional ones are recommended, that is, SRS and SS. Among these alternatives, although the SS is performed faster and at a lower cost, provided that the choice of plots is regular and uniform, the statistical differences observed among their averages of trees.ha^{-1}, accuracy and precision, by the Graybill F test, make SRS the best choice. However, it should be noted that, according to the Z test, these methods presented statistically higher averages than the 100% inventory, in addition to the true mean outside the confidence interval (CI). These inconsistencies prove that, although the sampling procedures meet the legal norms, the estimation of the variable of interest can be statistically invalid.

CONCLUSIONS

The adaptive cluster sampling process was the only one that presented the average trees.ha^{-1}, the diameter structure and the spatial pattern similar to those of the 100% inventory, however the high final sample intensity and the high sampling errors turn it inappropriate for the purposes of this study.

Although they have significantly higher mean trees.ha^{-1} than the 100% inventory, the SRS and SS processes provided a sampling error less than 10% of the mean value and a spatial pattern and diametric structure similar to those of the 100% inventory, making them suitable for inventory of monospecific populations. Among these, SRS is the most indicated.

The use of larger sampling units in the ACS, considering an inclusion condition equal to 1.0, is less efficient and underestimates the variable of interest (trees.ha^{-1}). In addition, larger units result in the degeneration of the process, since almost 100% inventory is being performed.

ACKNOWLEDGMENTS

To the Mixed Cooperative of the Tapajós National Forest for the supply of 100% inventory data.

REFERENCES

ABREU, J.C.; GUEDES, M.C.; GUEDES, A.C.L.; BATISTA, E.M. Estrutura e distribuição espacial de andirobeiras (*Carapa* spp.) em floresta de várzea do estuário amazônico. **Ciência Florestal**, v. 24, n. 4, p. 1009-1019, 2014.

ACHARYA, B.; BHATTARAI, G.; GIER, A.; STEIN, A. Systematic adaptive cluster sampling for the assessment of rare tree species in Nepal. **Forest Ecology and Management**, v. 137, n.1-3, p. 65-73, 2000.

ALVARES, C.A.; STAPE, J.L.; SENTELHAS, P.C.; MORAES, G.; LEONARDO, J.; SPAROVEK, G. Koppen's climate classification map for Brasil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711-728, 2013.

AMARAL, MK; NETTO, S. P.; LINGNAU, C; FILHO, A.F. Evaluation of the Morisita Index por determination of the spatial distribution of species in a fragment of araucaria forest. **Applied Ecology and Environmental Research**, v. 13, n. 2, p. 361-372, 2015.

ARAÚJO, E.J.G.; DAVID, H.C.; NETTO, S. P.; MORAIS, V.A.; SCOLFORO, J.R.S. Padrão espacial de espécies árboreas em fragmento de floresta estacional semidecidual. **Revista de Ciências Agrárias**, v. 57, n. 2, p. 166-171, 2014.

BROWN, J. A. Designing an efficient adaptive cluster sample. **Environmental and Ecological Statistics**, v. 10, n. 1, p. 95-105, 2003.

BROWN, J.A.; SALEHI, S.M.; MORADI, M.; PANAHBEHAGH, B.; SMITH, D.R. Adaptive survey designs for sampling rare and clustered populations. **Mathematics and Computers in Simulation**, v. 93, p. 108-116, 2013.

BRUZINGA, J.S.; OLIVEIRA, M.L.R.; NOGUEIRA, G.S.; PEREIRA, I.M.; LEITE, H.G.; MACHADO, E.L.M. Métodos de amostragem para quantificar indivíduos adultos de pequi (*Caryocar brasiliense* Cambess). **Ciência Rural**, v. 44, n. 8, p. 1341-1347, 2014.

CAPRETZ, R.L.; BATISTA, J.L.F.; SOTOMAYOR, J.F.M.; CUNHA, C.R.; NICOLETTI, M.F.; RODRIGUES, R.R. Padrão espacial de quatro formações florestais do estado de São Paulo, através da função K de Ripley. **Ciência Florestal**, v. 22, n. 3, p. 551-565, 2012.

CAVALCANTI, F.J.B.; MACHADO, S.A.; OSOKAWA, R.T.; CUNHA, U.S. Comparação dos valores estimados por amostragem na caracterização da estrutura de uma área de floresta na Amazônia com as informações registradas no censo florestal. **Revista Árvore**, v. 35, n. 5, p. 1061-1068, 2011.

COELHO, A. A. **Estrutura populacional e comercialização do óleo de andiroba (***Carapa guianensis* Aubl.**) em área de assentamento no município de Santarém, estado do Pará**. 2013. 132 p. Dissertaçao de mestrado. Universidade Federal Rural da Amazônia, Belém, 2013.

FERRAZ, I.D.K.; CAMARGO, J.L.C.; SAMPAIO, P.T.B. Sementes e plântulas de andiroba (*Carapa guianensis* Aubl. e *Carapa procera* DC): aspectos botânicos, ecológicos e tecnológicos. **Acta amazônica**, v. 32, n.4, p. 647-661, 2002.

DIONISIO, L.F.S.; LIMA, A.C.S.; CASTRO, T.M.M.G.; CORREIA, R.G.; MARTINS, W.B.R.; ABREU, V.S. Ocorrência de Hypsipyla grandella Zeller (Lepidoptera, Pyralidae) no Sul do estado de Roraima. **EntomoBrasilis**, v. 9, n. 2, p. 97-100, 2016.

GUARINO, E.D.S.G.; GESSNER, C.M.; WADT, L.H.O.; FONSECA, F.L.; RAPOSO, A. Estrutura etária e espacial de uma população natural de *Carapa guianensis* Aubl. (Meliaceae) na Amazônia Sul Ocidental. **Scientia Forestalis**, v. 42, n. 101, p. 91-99, 2014.

IBGE, INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Manual técnico da vegetação brasileira**. IBGE, 2012. 275p.

HANSEN, M.H.; HURWITZ, W.N. On the theory of sampling from finite populations. **The Annals of Mathematical Statistics**, v. 14, n.4, p. 333-362, 1943.

KLIMAS, C.A.; KAINER, K.A.; WADT, L.H.O. Population structure of *Carapa guianensis* in two forest types in the southwestern Brazilian Amazon. **Forest Ecology and Management**, v. 250, n. 3, p. 256-265, 2007.

LOETSCH, F.H.; HALLER, K.E. **Forest inventory**. BLV Verlagsgesellschaft, 1964. 436p.

LONDRES, M.; SCHULZE, M.; STAUDHAMMER, C.L.; KAINER, K.A. Population Structure and Fruit Production of Carapa guianensis (Andiroba) in Amazonian Floodplain Forests: Implications for Community-Based Management. **Tropical Conservation Science**, v. 10, p. 1-13, 2017.

MAGALHÃES, L.M.S.; BLUM, W.E.H.; FERNANDES, N.P. Características edáfico-nutricionais de plantios florestais na região de Manaus. II. Crescimento de *Carapa guianensis* Aubl. em solos de diferentes texturas. **Acta Amazonica**, v. 16/17, p 523-548, 1986/87.

MEIRA JUNIOR, M.S.; MOTA, S.L.L.; MACHADO, E.L.M.; PEREIRA, I.M. Distrubuição espacial de *Eremanthus incanus* (Less.) Less. (Asteraceae) em duas áreas com diferentes níveis de conservação. **Revista Brasileira de Biociência**, v. 15, n. 1, p. 27-31, 2017.

NEGRINI, M.; AGUIAR, M.D.; VIEIRA, C.T.; SILVA, A.C.; HIGUCHI, P. Dispersão, distribuição espacial e estratificação vertical da comunidade arbórea em um fragmento florestal no Planalto Catarinense. **Revista Árvore**, v. 36, n. 5, p. 919-930, 2012.

PÉLLICO NETTO, S.; PELISSARI, A.L.; CYSNEIROS, V.C.; BONAZZA, M.; SANQUETTA, C.R. Sampling procedures for inventory of commercial volume tree species in Amazon Forest. **Anais da Academia Brasileira de Ciências**, v. 89, n. 3, p. 1829-1840, 2017.

PINTO, A.A.; TELES, B.R.; ANJOS, N.; COUCEIRO, S.R.M. Predação de sementes de Andiroba [*Carapa guianensis* Aubl. e *Carapa procera* DC.(Meliaceae)] por insetos na Amazônia. **Revista Árvore**, v. 37, n. 6, p. 1115-1123, 2013.

RIPLEY, B.D. Tests ofrandomness' for spatial point patterns. **Journal of the Royal Statist ical Society. Series B (Methodological)**, v. 41, n. 3, p. 368-374, 1979.

SHIVER, B.D.; BORDERS, B.E. **Sampling techniques for forest resource inventory**. John Wiley and Sons, 1996. 368p.

SILVA, F.R.P.; ALMEIDA, S.S.M.S. Análise fitoquímica e microbiólogica das atividades do extrato bruto etanólico da Andiroba, *Carapa guianensis* Aubl. **Biota Amazônia**, v. 4, n. 4, p. 10-14, 2014.

SOARES, C.P.B.; RODELLO, C.M.; SOUZA, A.L.; LEITE, H.G.; SOARES, V.P.; SILVA, G.F. Comparação entre procedimentos de amostragem para espécies florestais raras e padrão de distribuição espacial agregado. **Revista Árvore**, v. 33, n. 3, p. 545-553, 2009.

THOMPSON, S. K.; SEBER, G. A. F. **Adaptive sampling**. John Wiley & Sons, 1996. 288p.

THOMPSON, S.K. Adaptive cluster sampling. **Journal of the American Statistical Association**, v. 85, n. 412, p. 1050-1059, 1990.

TONINI, H.; COSTA, P.; KAMISKI, P.E. Estrutura, distribuição espacial e produção de sementes de andiroba (*Carapa guianensis* Aubl.) no sul do estado de Roraima. **Ciência Florestal**, v. 19, n. 3, p. 247-255, 2009.

VIEIRA, D.S. **Padrão espacial de espécies arbóreas no Baixo Rio Tapajós**. 2015. 100 p. Dissertação mestrado. Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina.

VIEIRA, D.S.; GAMA, J.R.V.; RIBEIRO, R.B.S.; XIMENES, L.C. Estrutura, distribuição espacial e volumetria de *Carapa guianensis* Aubl. na Floresta Nacional do Tapajós. **Nature and Conservation**, v. 6, n. 2, p. 18-25, 2013.

VIEIRA, D.S.; OLIVEIRA, M.L.R.; GAMA, J.R.V.; MACHADO, E.L.M.; GORGENS, E.B. Patrón espacial y métodos de muestreo para Bertholletia excelsa (castaña de Pará) en la Mesorregión del Bajo Amazonas, estado de Pará, Brasil. **Bosque**, v. 38, n.1, p. 97-107, 2017.