REVISTA FLORESTA
Engenharia Florestal
ISSN 1982-4688
Edição 56
Universidade Federal do Paraná
Setor de Ciências Agrárias

ENVIRONMENTAL SELECTIVITY OF THE BRAZIL NUT TREE: GEOSTATISTICAL MODELING AND RANDOM FOREST

Luciana Jefferson de Moraes[1*], Jandresson Dias Pires[2], Kátia Emídio da Silva[3], Santiago Linorio Ferreyra Ramos[4], Gérson Rodrigues dos Santos[5], Fernando de Souza Bastos[6]

[1*]Universidade Federal do Amazonas, Programa de Pós-Graduação em Ciências Florestais e Ambientais – Manaus – AM - Brasil - lucyjeferson22@gmail.com
[2]Instituto Federal da Bahia – Brumado – BA - Brasil - jandresson.pires@ifba.edu.br
[3]Embrapa Amazônia Ocidental, Manaus – AM - Brasil - katia.emidio@embrapa.br
[4]Universidade Federal do Amazonas, Instituto de Ciências Exatas e Tecnologia, Itacoatiara – AM - Brasil - slfr@ufam.edu.br
[5]Universidade Federal de Viçosa - Instituto de Ciências Exatas e Tecnológicas – Florestal – MG - Brasil - gerson.santos@ufv.br
[6]Universidade Federal de Viçosa - Departamento de Estatística – Viçosa – MG - Brasil - fernando.bastos@ufv.br

Received for publication: 31/07/2025 – Accepted for publication: 05/10/2025

Resumo

Seletividade ambiental da castanheira-da-amazônia: modelagem geoestatística e random forest. A castanheira-da-Amazônia é uma espécie florestal de grande relevância para a conservação das florestas tropicais e como fonte de renda e alimentação para milhares de famílias agroextrativistas. Este estudo analisou a relação entre os atributos físico-químicos do solo e a distribuição espacial das castanheiras em um castanhal nativo no Amazonas. As amostras foram coletadas na Propriedade do Jutica – Tefé/AM, na camada superficial do solo (0-20 cm) de uma parcela permanente (300 x 300 m), seguindo uma amostragem sistemática com 60 pontos distribuídos a cada 50 m entre linhas e 30 m entre pontos. Todas as castanheiras com DAP ≥ 10 cm foram inventariadas, e os dados do solo foram avaliados por estatística descritiva e geoestatística. A Krigagem Indicativa foi aplicada para interpolação do DAP, separando árvores com DAP ≥ 50 cm e DAP < 50 cm. A regressão espacial foi realizada com o algoritmo *Random Forest*, e todas as análises foram conduzidas no software R. Os resultados indicaram um grau moderado de heterogeneidade nos atributos do solo. O modelo *Random Forest* gerou mapas de probabilidade que caracterizaram a distribuição espacial das castanheiras e identificaram as regiões mais propícias à ocorrência de indivíduos com diferentes classes de DAP. Os atributos mais seletivos foram cálcio, sódio, matéria orgânica, cobre, areia total e nitrogênio, evidenciando sua influência na seletividade ambiental da espécie.

Palavras-chave: DAP, Krigagem Indicativa, Regressão Espacial, Mapas de Probabilidade, Heterogeneidade.

Abstract

The Brazil nut tree is a forest species of great relevance for the conservation of tropical forests and as a source of income and food for extractive families. This study analyzed the relationship between soil physico-chemical attributes and the spatial distribution of Brazil nut trees in a native stand in the Amazon. Soil samples were collected at the Jutica Property – Tefé/AM, from the surface layer (0–20 cm) of a permanent plot (300 x 300 m), following a systematic sampling approach with 60 points distributed every 50 m between lines and 30 m between points. All Brazil nut trees with DBH ≥ 10 cm were inventoried, and soil data were evaluated using descriptive statistics and geostatistics. Indicator Kriging was applied to interpolate DBH, separating trees with DBH ≥ 50 cm and DBH < 50 cm. Spatial regression was performed using the Random Forest algorithm, and all analyses were conducted in R software. The results indicated a moderate degree of heterogeneity in soil attributes. The Random Forest model generated probability maps that characterized the spatial distribution of Brazil nut trees and identified the most favorable regions for the occurrence of individuals with different DBH classes. The most selective attributes were calcium, sodium, organic matter, copper, total sand, and nitrogen, highlighting their influence on the environmental selectivity of the species.

Keywords: DBH, Indicator Kriging, Spatial Regression, Probability Maps, Heterogeneity.

INTRODUCTION

The sustainable use of tropical forests has been widely discussed as a crucial strategy for the conservation of natural environments in the Amazon (Wadt *et al.*, 2017). In this context, Non-Timber Forest Products (NTFPs) play an essential role, as they contribute to the preservation of these forests, in addition to providing food and serving as a source of income for numerous families (Serviço Florestal Brasileiro - SFB, 2014).

Among the main NTFPs of the Amazon, the almonds of the Brazil Nut Tree (Bertholletia excelsa, Bonpl.), also known as Brazil nut, Amazon nut or Pará nut, stand out. This product of great economic and nutritional relevance is widely appreciated and represents one of the main extractive activities in the region (Tonini & Baldoni, 2019). The Brazil Nut Tree, belonging to the Lecythidaceae family and the only representative of the genus Bertholletia, is considered a symbol of Amazonian tropical forests due to its majestic presence and its significant social, economic, and environmental importance (Wadt *et al.*, 2019).

The species occurs predominantly in upland forests, forming Brazil nut groves in specific environments, often associated with other forest species (Brasil. Ministério do Meio Ambiente – MMA. Secretaria de Extrativismo e Desenvolvimento Rural Sustentável. Departamento de Extrativismo, 2017). Although it thrives in clayey or clay-sandy, well-structured and well-drained soils, the Brazil nut tree also adapts to low-fertility soils, characteristic of the Amazon forest, where acidity and low nutrient levels predominate (Spera *et al.*, 2019). This contrast between the exuberance of the forest and the challenging soil conditions reinforces the need to understand the edaphic conditions that favor the occurrence and productivity of the Brazil nut tree (Wadt *et al.*, 2017).

Considering its ecological, social, and economic importance, this study is based on the hypothesis that the spatial distribution of Brazil nut trees is directly associated with specific physical and chemical soil attributes. Variables such as calcium, sodium, organic matter, copper, total sand, and nitrogen are assumed to be determinants for the occurrence and development of individuals with different DBH classes (≥ 50 cm and < 50 cm). To test this hypothesis, the study proposes the use of geostatistical techniques and machine learning, which make it possible to identify spatial patterns and establish relationships between the environmental selectivity of the species and soil attributes.

Understanding these relationships is essential for the sustainable management of the Brazil nut tree. In this context, geostatistics, grounded in the Theory of Regionalized Variables, plays a crucial role by enabling the identification of spatial correlations and the prediction of unsampled values based on distance and observed data (Oliveira *et al.*, 2015; Yamamoto & Landim, 2015). This approach not only contributes to understanding the environmental selectivity of the species but also provides support for more efficient management practices.

Additionally, Random Forest (RF), a machine learning algorithm, has stood out as a robust and effective tool in prediction studies (Hengl *et al.*, 2018; Heung *et al.*, 2022; Sharma *et al.*, 2025; Suleymanov *et al.*, 2023). RF works by aggregating multiple decision trees, generated from random subsets of the training data. For regression, it combines independent trees and provides, as the final result, the overall mean of the predictions. Its flexibility and ability to select relevant predictors make it particularly suitable for analyses involving complex environmental variables.

Thus, the present study aimed to analyze the relationship between the physical–chemical attributes of the soil and the occurrence of Brazil nut trees in a native grove in the state of Amazonas. Using geostatistical tools and the RF algorithm, the study sought to understand the spatial patterns of the species and identify the main associated environmental factors, providing a scientific basis for sustainable management strategies for the Brazil nut tree.

MATERIALS AND METHODS

Characterization of the study area

The study was conducted in a 9-hectare area located on the Jutica Property, in the municipality of Tefé, state of Amazonas, at the coordinates 3° 38' 8.323" S and 64° 18' 42.648" W (Figure 1). The region is situated in Central Amazonia, characterized by predominantly Dense Ombrophilous Forest, which also includes areas of secondary vegetation and agricultural activities, according to the classification of the Brazilian Institute of Geography and Statistics - (Instituto Brasileiro de Geografia e Estatística - IBGE, 2012).

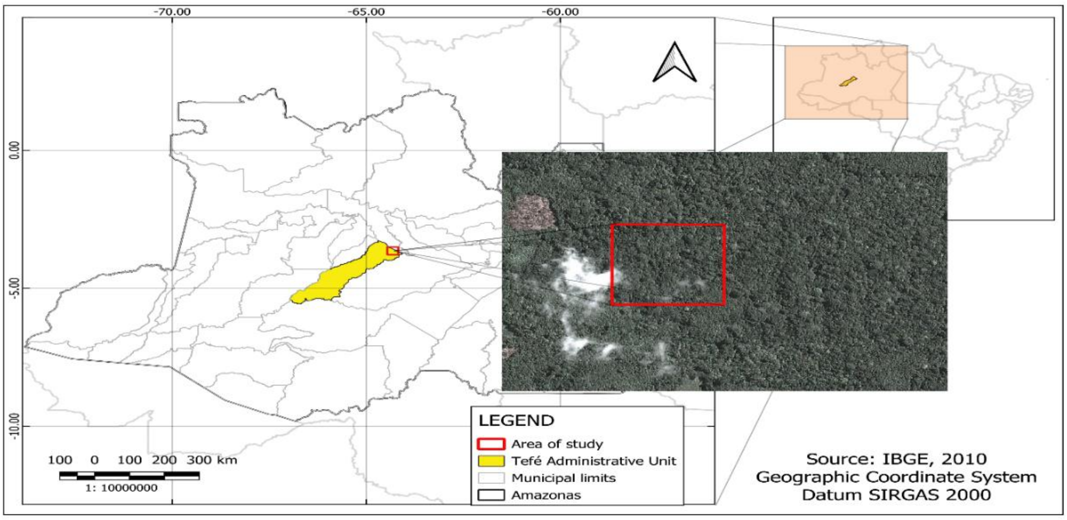


Figure 1. Location map of the study area: Brazil nut stand in the Jutica property, Tefé municipality, Amazonas.
Figura 1. Mapa de localização da área de estudo: Castanhal na propriedade Jutica, município de Tefé, Amazonas.

The climate of the region is classified as Afi, according to the Köppen classification, characteristic of humid tropical areas, with rainfall well distributed throughout the year. The average annual precipitation ranges between 2,400 and 2,700 mm, and the annual maximum temperature reaches 32.8°C (Aleixo & Silva Neto, 2018; Alvares *et al.*, 2013). These climatic conditions, typical of the Amazon, are decisive for the development of the local vegetation and for the dynamics of the soils in the study area.

Data collection and analysis

The soil samples were collected in 2017 as part of the MapCast project (Mapping of native Brazil nut groves and socio-environmental and economic characterization of Brazil nut production systems in the Amazon), funded by the Brazilian Agricultural Research Corporation (Embrapa). The study area consisted of a permanent 9-hectare plot (300 × 300 meters), where a systematic sampling strategy was adopted to ensure adequate representation of the data. A regularly spaced grid was established, totaling 60 sampling points, with distances of 50 meters between the lines (six lines in total) and 30 meters between points along the lines (Figure 2). This configuration allowed for uniform coverage of the area, ensuring the collection of representative data for subsequent analyses.

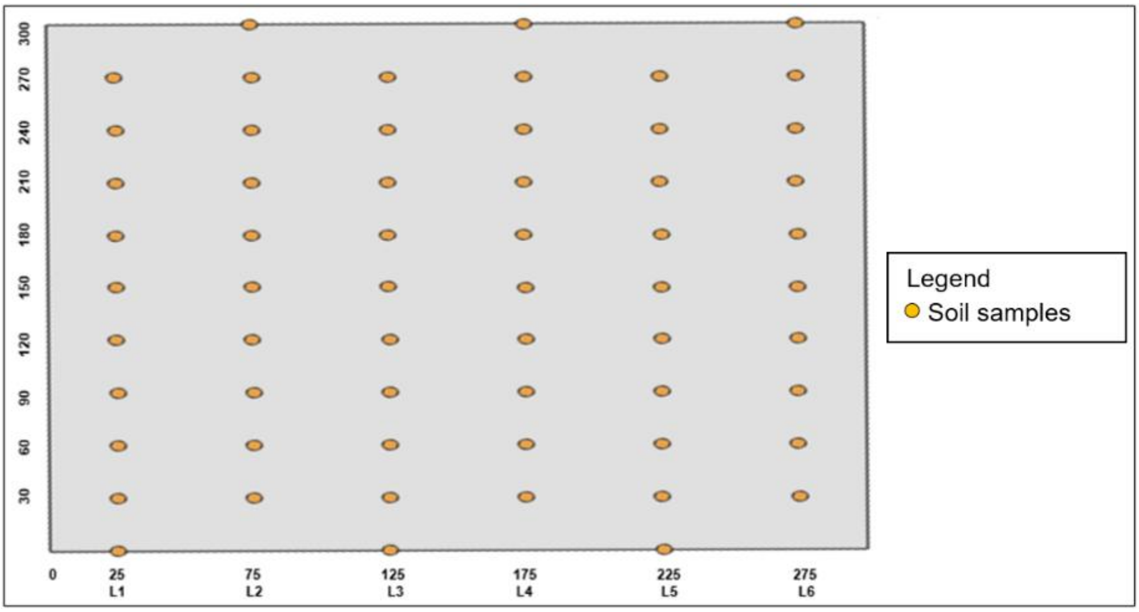


Figure 2. Systematic soil sampling scheme in the study area (9 ha), with a regular grid of 60 points, spaced 50 meters between lines and 30 meters between points along the lines.

Figura 2. Esquema de amostragem sistemática do solo na área de estudo (9 ha), com um grid regular de 60 pontos, espaçados a cada 50 metros entre linhas e 30 metros entre pontos ao longo das linhas.

The soil samples were collected using a Dutch auger at a depth of 0–20 cm. The sampling locations were georeferenced in UTM (Universal Transverse Mercator) coordinates, using the WGS84 (World Geodetic System 84) Datum, a system widely employed in cartographic projections. After collection, the soil samples were stored in labeled plastic bags and transported to the laboratory of Embrapa Western Amazon, in Manaus/AM. In this laboratory, the physical and chemical soil analyses were performed following the methods described by the Brazilian Agricultural Research Corporation (EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária, 2017). The soil variables analyzed included:

- Physical analyses: Total Sand, Silt, Clay, Textural Classification, Soil Density, Total Pore Volume (TPV, %), Microporosity (%), and Macroporosity (%).
- Chemical analyses: pH, organic matter (OM), nitrogen (N), phosphorus (P), potassium (K), sodium (Na), calcium (Ca), magnesium (Mg), aluminum (Al), potential acidity (H+Al), iron (Fe), zinc (Zn), manganese (Mn), and copper (Cu).

In the study area, a forest inventory was conducted for all Brazil nut trees with diameter at breast height (DBH) ≥ 10 cm, totaling 42 individuals distributed across the 9 ha (300 × 300 m). For the statistical analysis, descriptive statistics were applied to the soil data in order to characterize its variability and provide support for the spatial analyses. The modeling of spatial dependence of the variables was performed through geostatistical analyses, using theoretical models (spherical and exponential) fitted to the experimental variograms. The selection criterion was based on the weighted least squares method, in which the model with the lowest sum of squared residuals between the empirical and theoretical values was selected for each variable. These models allowed the estimation of structural semivariogram parameters, such as nugget effect, sill, and range (Vieira, 2000). These models were subsequently used in spatial interpolation by kriging, enabling the generation of continuous surfaces for the soil variables. The model that exhibited the best fit to explain spatial dependence was selected, and based on it, ordinary kriging interpolation was carried out (Yamamoto & Landim, 2015). These procedures allowed the development of maps for all soil variables.

To assess the relationship between the distribution of Brazil nut trees and soil attributes, indicator kriging was used, allowing the generation of continuous surfaces based on the DBH of the Brazil nut trees. The response variable DBH was categorized into two groups: DBH ≥ 50 cm and DBH < 50 cm. This procedure resulted in probability maps of occurrence for these DBH categories, which supported the subsequent analyses.

With the continuous surfaces of the soil variables and DBH obtained, machine learning was applied through the RF algorithm to build a prediction model. The main model was configured with 500 decision trees, using sampling with replacement. At each split of the tree nodes, 8 predictor variables were tested. In parallel, a second model with repeated cross-validation was implemented, in which the data were divided into five partitions of similar size, using four partitions for training and one for testing. This approach tested different numbers of variables per split (2, 3, 4, and 5), with the optimal value of 2 variables selected based on the lowest mean squared error. This multivariate spatial regression procedure used the Gini Index to identify the most important variables associated with the DBH of the Brazil nut trees, efficiently splitting the data and selecting the best variable to compose the root node (Carvalho Junior *et al.*, 2016; Cutler *et al.*, 2007). The RF regression trees were generated using the randomForest package, available in the R Software (R Core Team, 2024). The final result consisted of probability maps indicating the regions most favorable for the occurrence of Brazil nut trees, associated with the soil characteristics identified as the most relevant.

RESULTS

The results of the descriptive analyses of the physical–chemical soil attributes are presented in Tables 1 and 2. The descriptive statistics revealed that most of the physical variables of the samples exhibited moderate variability (12% < CV < 60%), with the exception of microporosity (7.21%) and pH (3.41%), which showed low variability.

Table 1. Descriptive statistics of soil physical variables (clay, silt, total sand, soil density, total pore volume – TPV, microporosity, and macroporosity) at a depth of 0–20 cm in the Jutica property, Tefé, Amazonas.

Tabela 1. Estatística descritiva das variáveis físicas do solo (argila, silte, areia total, densidade do solo, volume total de poros – VTP, microporosidade e macroporosidade) na profundidade de 0-20 cm, na propriedade Jutica, Tefé, Amazonas.

Variable	**Mean**	**Median**	**Min. Value**	**Max. Value**	**Coefficients**			**Standard Dev.**
					Variation (%)	**Asymmetry**	**Kurtosis**	
Total Sand (g kg^{-1})	417,74	424,42	274,92	594,41	16,77	0,14	-0,58	70,03
Silt (g kg^{-1})	299,86	289,41	197,85	482,48	21,40	0,96	0,58	64,17
Clay (g kg^{-1})	282,40	284,75	195,00	383,50	14,48	0,07	-0,46	40,89
Soil dens. (g cm^{-3})	0,87	0,87	0,60	1,11	14,37	-0,18	-0,59	0,12
VTP (%)	43,76	42,38	35,20	62,89	13,31	1,12	0,86	5,82
Microporosity (%)	66,57	66,44	57,45	76,85	7,21	0,18	-0,59	4,80
Macroporosity (%)	22,81	23,29	6,79	32,53	22,30	-0,61	0,57	5,09

Subsequently, the geostatistical analysis allowed the assessment of spatial dependence among these variables, enabling predictive modeling through the RF algorithm. This approach allowed the generation of probability maps for the occurrence of Brazil nut trees with different DBH classes (≥ 50 cm and < 50 cm) and the identification of the most influential edaphic attributes in the species' distribution. The results obtained provide a detailed view of the spatial patterns of the Brazil nut trees and reinforce the potential of the applied methodologies for understanding environmental selectivity and the sustainable management of the species.

Among the physical attributes, the highest coefficients of variation were observed for macroporosity (22.30%) and silt (21.40%) (Table 1), indicating greater spatial heterogeneity of these variables in the soil of the study area. The elevated CV of silt suggests a less uniform distribution, which may be associated with natural deposition and erosion processes, directly affecting moisture retention and soil structuring. Macroporosity, a variable that influences soil aeration and water infiltration, showed the greatest variation among the physical attributes, reflecting structural differences that may impact root development in Brazil nut trees. Other physical attributes, such as total sand (16.77%) and clay (14.48%), exhibited moderate variability, indicating relative homogeneity in the soil's particle-size composition in the study area. These results suggest that the spatial heterogeneity of these attributes may directly influence the environmental selectivity of the Brazil nut tree, determining locations more conducive to its growth and development.

For the chemical attributes, the highest coefficients of variation (CV) were recorded for copper (Cu), with 45.43%, calcium (Ca), with 42.51%, and manganese (Mn), with 40.25% (Table 2), indicating high spatial heterogeneity of these variables in the study area. This variation may be related to natural processes of leaching and redistribution of nutrients within the soil profile, as well as to different rates of biogeochemical cycling associated with the existing vegetation. Additionally, organic matter (OM) presented a CV of 26.38%, while nitrogen (N) varied by 21.14%, demonstrating moderate variability, which may reflect differences in plant material decomposition and nutrient availability throughout the sampled area.

Table 2. Descriptive statistics of soil chemical variables (pH, organic matter - OM, nitrogen - N, phosphorus - P, potassium - K, sodium - Na, magnesium - Mg, hydrogen + aluminum - H+Al, iron - Fe, zinc - Zn, manganese - Mn, and copper - Cu) at a depth of 0-20 cm in the Jutica property, Tefé, Amazonas.

Tabela 2. Estatística descritiva das variáveis químicas do solo (pH, matéria orgânica - MO, nitrogênio - N, fósforo - P, potássio - K, sódio - Na, magnésio - Mg, hidrogênio + alumínio - H+Al, ferro - Fe, zinco - Zn, manganês - Mn e cobre - Cu) na profundidade de 0-20 cm, na propriedade Jutica, Tefé, Amazonas.

Variable	**Mean**	**Median**	**Mín. Value**	**Max. Value**	**Coefficients**			**Standard Dev.**
					Variation (%)	**Asymmetry**	**Kurtosis**	
pH (H2O)	4,07	4,07	3,75	4,45	3,41	0,40	0,77	0,14
MO (g kg^{-1})	35,02	33,63	15,75	66,64	26,38	0,76	1,25	9,24
N (g kg^{-1})	1,51	1,46	0,80	2,48	21,14	0,44	0,63	0,32
P (mg dm^{-3})	3,00	2,87	1,72	4,59	19,30	0,36	-0,43	0,58
K (mg dm^{-3})	28,97	27,00	15,00	68,00	33,72	1,60	3,35	9,77
Na (mg dm^{-3})	2,08	2,00	1,00	9,00	32,22	0,90	1,63	0,67
Ca (cmol$_c$ dm^{-3})	0,03	0,02	0,01	0,25	42,51	0,74	0,03	0,01
Mg (cmol$_c$ dm^{-3})	0,11	0,10	0,07	0,20	27,97	0,97	-0,14	0,03
H+Al (cmol$_c$ dm^{-3})	10,55	10,65	6,72	14,29	13,95	-0,03	-0,15	1,47
Fe (mg dm^{-3})	313,92	299,00	180,00	604,00	28,26	0,86	0,57	88,72
Zn (mg dm^{-3})	0,45	0,41	0,24	1,19	39,43	1,42	2,94	0,18
Mn (mg dm^{-3})	0,68	0,63	0,31	1,87	40,25	2,49	8,09	0,27
Cu (mg dm^{-3})	0,56	0,49	0,23	1,28	45,42	0,86	0,24	0,25

Additionally, the pH values ranged from 3.75 to 4.45, with an average of 4.07, confirming the acidity of the soil in the region, as shown in Table 2. This acidic condition is typical of highly weathered tropical soils and may influence the availability of essential nutrients, consequently affecting the environmental selectivity of the Brazil nut tree. Soils with pH below 5.0 often exhibit low availability of calcium and magnesium, as well as greater aluminum solubility, which can impact plant growth. These factors highlight the need for appropriate management to ensure favorable conditions for the development of the Brazil nut tree, particularly regarding acidity correction and the replenishment of essential nutrients.

The values of the physical–chemical variables were subjected to geostatistical analysis to verify the presence of spatial dependence. The interpolation results indicated that the physical–chemical attributes of the soil displayed moderate to strong spatial dependence, a pattern also observed for the DBH data (Table 3). For each variable, theoretical models were fitted to the experimental semivariograms, with the spherical model being predominant in fitting the main physical and chemical variables, such as organic matter and total sand, which presented ranges of 39.05 m. However, some variables, such as calcium, sodium, and copper, showed the best fit with the exponential model, with ranges varying between 23.37 m and 29.97 m. The nugget effect close to zero observed for calcium, sodium, and copper indicates that these variables exhibit a strong spatial structure, meaning that most of the variability can be explained by spatial dependence rather than random factors. In contrast, organic matter and total sand showed considerably higher nugget effects compared with the other variables, suggesting the need for a denser sampling structure in future studies to reduce this undesirable effect. These results reinforce the importance of geostatistical analysis for characterizing the spatial distribution of soil attributes, enabling a better understanding of the environmental selectivity of the Brazil nut tree and supporting management strategies for the species.

After the univariate analyses of the soil attributes and DBH, a multivariate spatial regression process was performed using machine learning through the Random Forest algorithm. This method enabled the generation of probability maps for the occurrence of Brazil nut trees in two DBH categories (≥ 50 cm and < 50 cm), as well as maps for the most important soil variables associated with the species. The predictive modeling results using Random Forest indicated the influence of different environmental zones on the occurrence of Brazil nut trees with DBH ≥ 50 cm (Figure 3). Green areas represent zones of high probability of occurrence of these individuals, suggesting more favorable edaphic conditions for their growth. Yellow zones indicate medium probability, while salmon-colored areas represent relative probability of occurrence. White zones are classified as low-probability areas, possibly due to limiting factors such as lower soil fertility or unfavorable structural conditions. The black points in the figure represent Brazil nut trees with DBH ≥ 50 cm, whereas white points indicate individuals with DBH < 50 cm.

Table 3. Results of the geostatistical analysis for the main physical and chemical soil attributes and the diameter at breast height (DBH) in the Jutica property, Tefé, Amazonas.
Tabela 3. Resultados da análise geoestatística para os principais atributos físicos e químicos do solo e do diâmetro à altura do peito (DAP) na propriedade Jutica, Tefé, Amazonas.

Variable	**Model**	**Nugget Effect**	**Level**	**Reach**
Ca (cmol$_c$ dm^{-3})	Exponential	0	0,00015	23,3784
Na (mg dm^{-3})	Exponential	0	0,3841	29,9732
MO (g kg^{-1})	Spherical	72,7534	83,7375	39,0512
Cu (mg dm^{-3})	Exponential	0	0,054	29,7518
Areia Total (g kg^{-1})	Spherical	1605,27	4704,13	39,0512
N (g kg^{-1})	Spherical	0,0824	0,1001	39,0512
DAP (cm)	Spherical	0,09816	0,1934	59,1762

The results demonstrated that the Random Forest algorithm is efficient for estimating the probability of occurrence of Brazil nut trees in the analyzed categories (DBH < 50 cm and DBH ≥ 50 cm). Most individuals with DBH ≥ 50 cm (black points) were found in zones of medium to high probability of occurrence, as estimated by the model. Furthermore, the mean squared error (MSE) of 0.0049 and the coefficient of determination (R^2) of 77.16% indicate good explanatory capacity, reinforcing the applicability of the model for understanding the spatial patterns of the species in the Jutica–Tefé locality (Figure 3). These findings highlight the potential of machine learning for environmental analysis, allowing the identification of priority areas for the sustainable management of the Brazil nut tree.

In this study, the maps of the soil variables that most influenced the model were generated using the Random Forest machine learning algorithm. The selection of variables was based on the Gini impurity index, which evaluates the relevance of each attribute in predicting the occurrence of the Brazil nut tree. Each variable was analyzed individually, allowing the identification of which edaphic factors exerted the greatest influence on the spatial distribution of individuals with DBH ≥ 50 cm.

The absence of black points in white zones, as well as the absence of white points in green zones, reinforces the accuracy of the model in predicting the spatial distribution of the Brazil nut tree. Furthermore, the presence of individuals with DBH ≥ 50 cm in yellow zones may indicate transitional areas, where environmental conditions still favor the species' growth, but less expressively than in high-probability zones. These findings highlight the model's ability to identify relevant spatial patterns and contribute to understanding the environmental selectivity of the Brazil nut tree.

The maps generated for the most important variables, as identified by the Random Forest model in the Jutica–Tefé region, show a heterogeneous distribution of soil attributes (Figure 4). Green-colored areas represent zones with higher levels of the analyzed attributes, while white areas indicate regions with lower expression of these variables. Organic matter and nitrogen display similar spatial patterns, with higher concentrations in the western and southern portions of the study area, suggesting a relationship between these attributes. Total sand also shows elevated levels in these regions, whereas calcium and sodium exhibit a more fragmented distribution, with scattered points of higher concentration. Copper, in turn, displays an irregular pattern, with localized areas of higher content. These spatial patterns underscore the differentiated influence of these attributes in the study area and reinforce the importance of predictive modeling for identifying the environmental selectivity of the Brazil nut tree in relation to edaphic characteristics.

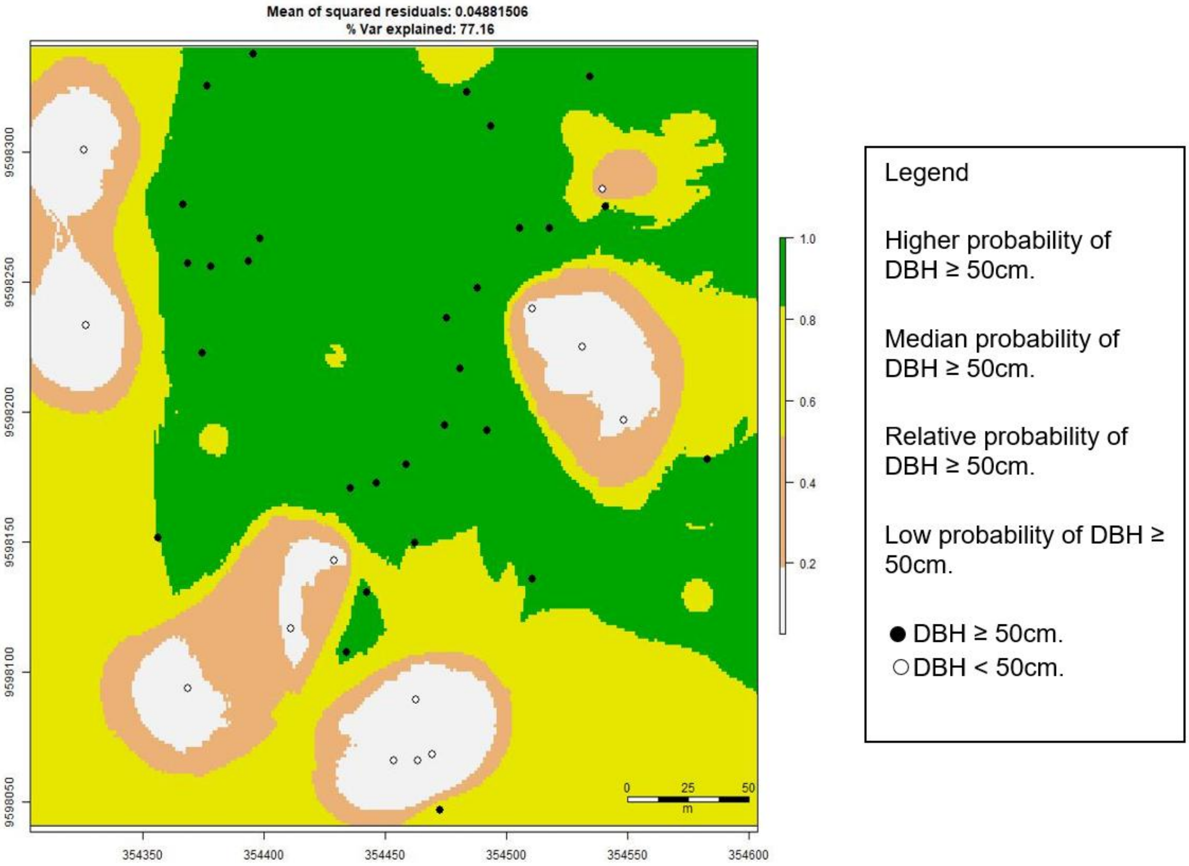
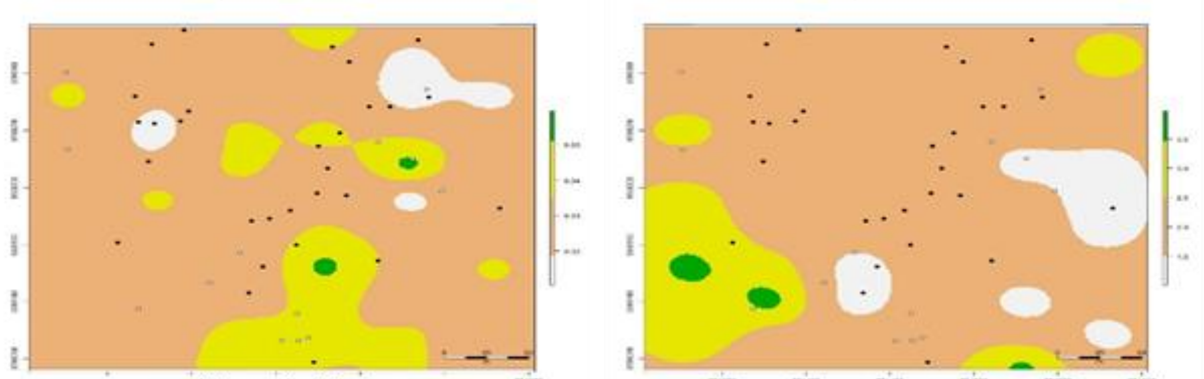


Figure 3. Probability map of the occurrence of individuals with DBH < 50 cm and DBH ≥ 50 cm in the Jutica property, Tefé, Amazonas, generated through prediction using the Machine Learning algorithm Random Forest.

Figura 3. Mapa de probabilidade de ocorrência de indivíduos com DAP < 50 cm e DAP ≥ 50 cm na propriedade Jutica, Tefé, Amazonas, gerado por predição utilizando o algoritmo de Aprendizagem de Máquina Random Forest.

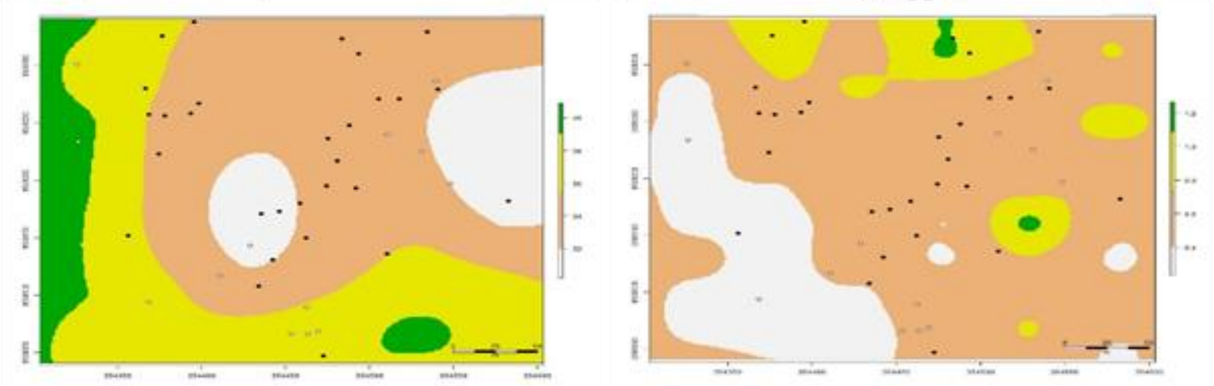
a) Calcium

b) Sodium



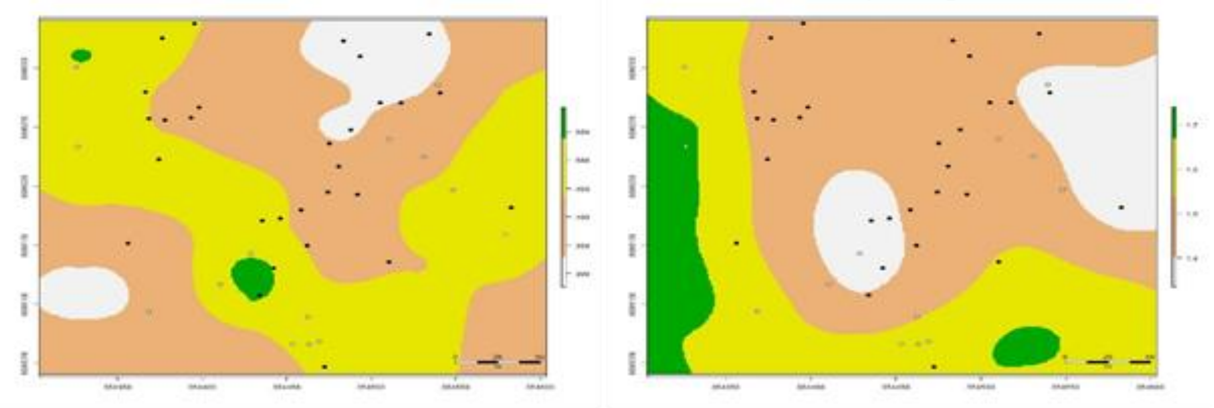
c) Organic Matter

d) Copper



e) Total Area

f) Nitrogen



Figures 4 (a, b, c, d, e, f). Maps of the most important variables for the data from the Jutica property, Tefé,Amazonas, identified through prediction using the Machine Learning algorithm Random Forest in conjunction with DBH.

Figuras 4 (a, b, c, d, e, f). Mapas das variáveis mais importantes para os dados da propriedade Jutica, Tefé, Amazonas, identificadas por predição utilizando o algoritmo de Aprendizagem de Máquina Random Forest em conjunto com o DAP.

DISCUSSION

Among the physical variables analyzed, macroporosity showed the highest coefficient of variation (CV = 22.30%). Among the chemical attributes, copper (CV = 45.42%) and calcium (CV = 42.51%) stood out, both classified as having moderate variability (Tables 1 and 2). In previous studies, Guerreiro *et al.* (2017) identified greater variability for microporosity, potassium, and iodine in soils of native Brazil nut groves in Pará, while zinc, copper, and pH showed lower variability. Costa *et al.* (2017), when analyzing soils in Roraima, evidenced the Brazil nut tree's requirement for calcium, emphasizing the importance of proper management to replenish this nutrient. Similarly, Aquino *et al.* (2014) observed predominance of moderate variability in the physical attributes of soils under native forest and pasture in the region of Manicoré, Amazonas.

In the present study, pH values ranged from 3.75 to 4.45, characterizing acidic soils. Guerreiro *et al.* (2017) also reported heterogeneous behavior of chemical variables in soils of native Brazil nut groves in Pará, observing low variation for pH. Additionally, the geostatistical interpolations indicated moderate to strong spatial dependence for the physical–chemical soil attributes and for diameter at breast height (DBH). Similar results were obtained by Guerreiro *et al.* (2017), who found strong spatial relationships between the distribution of Brazil nut trees and variables such as silt, clay, macroporosity, pH, phosphorus, zinc, and copper.

The modeling performed with the Random Forest (RF) algorithm, based on soil and DBH data, identified zones with high probability of occurrence of individuals with DBH ≥ 50 cm, represented by green areas (Figure 3). These zones indicate edaphic conditions favorable to the growth of the Brazil nut tree. Previous studies, such as Bhering *et al.* (2016), highlighted the potential of RF in predicting soil attributes such as sand, clay, and organic carbon, even with reduced datasets. In the present study, the maps generated by RF evidenced the model's selectivity for physical–chemical attributes such as calcium, sodium, organic matter, copper, total sand, and nitrogen (Figure 4). Previous works also emphasize the effectiveness of RF in modeling tropical soils. (Carvalho Junior *et al.*, 2016) identified that the most relevant covariates for estimating soil density in mountainous tropical regions included organic carbon, fine and coarse sand, base saturation, and cation exchange capacity (CEC). Bhering *et al.* (2016), in turn, indicated that variables such as drainage channel elevation, elevation, and lithology were determinant for explaining soil variability. These findings corroborate the results of the present study, reinforcing the efficiency of RF in identifying the most relevant factors for the distribution of Brazil nut trees.

It is concluded that the RF algorithm proved to be a robust tool for estimating the probability of occurrence of Brazil nut trees with DBH < 50 cm and DBH ≥ 50 cm. Most individuals with DBH ≥ 50 cm were recorded in zones of medium to high probability, demonstrating the effectiveness of the model in identifying spatial patterns of occurrence. Furthermore, the results allowed a better understanding of the species' spatial distribution and the identification of areas more conducive to its growth. RF also proved useful for highlighting the most relevant edaphic attributes in the studied environment, contributing to sustainable management strategies for the Brazil nut tree in the Jutica–Tefé region.

CONCLUSIONS

With the completion of this study, it was possible to conclude that:

- Most of the physical and chemical soil attributes showed a moderate degree of heterogeneity, with the exception of microporosity and pH, which demonstrated low variability. Among the attributes with the highest variability, silt and macroporosity stood out among the physical aspects, while copper (Cu), calcium (Ca), and manganese (Mn) were the most heterogeneous chemical attributes in the study area.
- The geostatistical analyses showed that the fitted semivariograms were effective in reproducing the behavior and spatial distribution of the physical–chemical soil attributes and the Brazil nut trees (DBH < 50 cm and DBH ≥ 50 cm), allowing the generation of univariate maps that satisfactorily represented these spatial relationships.
- The model created using the Random Forest algorithm generated probability maps of the occurrence of Brazil nut trees (DBH < 50 cm and DBH ≥ 50 cm), contributing to characterize and understand the spatial distribution of the species and to identify the regions or zones with greater propensity for the occurrence of Brazil nut trees in the two DBH categories.
- RF also allowed the identification of the main soil attributes related to the distribution of Brazil nut trees in the study area, generating maps that revealed selectivity for calcium, sodium, organic matter, copper, total sand, and nitrogen as the most relevant attributes.
- The results presented provide a valuable information base that can be applied to studies related to other productive areas of Brazil nut trees in the Amazon region, especially in the context of interactions with environmental variables and soil attributes. Furthermore, they confirm the effectiveness of the tools used in analyzing and predicting the distribution of Brazil nut trees in relation to soil attributes. Finally, the hypotheses proposed were confirmed, demonstrating that it is possible to understand the spatial patterns of the species and to identify regions more favorable to its development through techniques such as geostatistics and machine learning.

Despite the relevance of the results, this study presents limitations, such as the analysis being restricted to a single area of a native Brazil nut grove and the soil sampling depth being limited to 0-20 cm, which may not encompass all the edaphic variability that influences the species' development. Additionally, the number of inventoried individuals was small, which may limit the extrapolation of the results to other regions of the Amazon.

Future research may expand the spatial and temporal scope of sampling, include different soil depths, and evaluate the interaction between biotic and abiotic factors.

In summary, the results presented here not only confirm the environmental selectivity of the Brazil nut tree but also provide practical support for guiding strategies of sustainable management and conservation of the species. Furthermore, they offer a methodological basis that can be applied in future studies, enabling advances in understanding the relationships between soil attributes and the ecology of Amazonian tree species.

ACKNOWLEDGMENTS

To my advisors, Dr. Santiago Linorio and Dr. Kátia Emidio, as well as Dr. Gerson Rodrigues, Dr. Jandresson Pires and Dr. Fernando Bastos. I express my profound gratitude for their valuable guidance, dedication, patience, and trust throughout this work. I also thank the Coordination for the Improvement of Higher Education Personnel (Capes) for the financial support through the granting of a scholarship, which was essential for the development of this research. My appreciation extends to the Brazilian Agricultural Research Corporation (Embrapa) for the technical support and for providing its facilities and equipment for data analyses. I am also grateful to Mr. Márcio Cunha, owner of the study area in Jutica, for his generosity in allowing the use of the site for the execution of this work.

REFERENCES

Aleixo, N. C. R., & Silva Neto, J. C. A. (2018). Vegetation Index and Air Temperature Behavior in Tefé-Amazonas, Brazil. *Revista Brasileira De Geografia Física*, *11*(3), 864–876.

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. L. M., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, *22*, 711–728.

Aquino, R. E., Campos, M. C. C., Marques Jr, J., Oliveira, I. A., Mantovaneli, B. C., & Soares, M. D. R. (2014). Geoestatística na avaliação dos atributos físicos em Latossolo sob floresta nativa e pastagem na região de Manicoré, Amazonas. *Revista Brasileira de Ciência Do Solo*, *38*, 397–406.

Bhering, S. B., Chagas, C. S., Carvalho Junior, W., Pereira, N. R., Filho, B. C., & Pinheiro, H. S. K. (2016). Mapeamento digital de areia, argila e carbono orgânico por modelos Random Forest sob diferentes resoluções espaciais. *Pesquisa Agropecuária Brasileira*, *51*(9), 1359–1370.

Brasil. Ministério do Meio Ambiente – MMA. Secretaria de Extrativismo e Desenvolvimento Rural Sustentável. Departamento de Extrativismo. (2017). *Castanha-do-Brasil: boas práticas para o extrativismo sustentável orgânico* (1st ed.).

Carvalho Junior, W., Calderano Filho, B., Chagas, C. S., & Bhering, B. S. (2016). Regressão linear múltipla e modelo Random Forest para estimar a densidade do solo em áreas montanhosas. *Pesquisa Agropecuária Brasileira*, *51*(9), 1428–1437.

Costa, M. G. C., Tonini, H., & Mendes Filho, P. M. (2017). Atributos do Solo Relacionados com a Produção da Castanheira-do-Brasil (Bertholletia excelsa). *Revista Floresta e Ambiente*, 1–10.

Cutler, D. R., Edwards Jr., T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*, 2783–2792.

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária. (2017). *Manual de métodos de análise de solo* (3rd ed.).

Guerreiro, Q. L. de M., Oliveira Junior, R. C., Santos, G. R., Ruivo, M. L. P., Beldini, T. P., Carvalho, E. J. M., Silva, K. E., Guedes, M. C., & Santos, R. B. (2017). Spatial variability of soil physical and chemical aspects in a Brazil nut tree stand in the Brazilian Amazon. *African Journal of Agricultural Research*, *12*, 237–250.

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, *6*, e5518. https://doi.org/10.7717/peerj.5518

Heung, B., Bulmer, C. E., Schmidt, M. G., & Zhang, J. (2022). Provincial-scale digital soil mapping using a random forest approach for British Columbia. *Canadian Journal of Soil Science*, *102*(03), 597–620.

Instituto Brasileiro de Geografia e Estatística - IBGE. (2012). *Manual Técnico da Vegetação Brasileira*.

Oliveira, R. P., Grego, C. R., & Brandão, Z. N. (2015). *Geoestatística aplicada na agricultura de precisão utilizando o vesper* (1st ed.).

R Core Team. (2024). *R: A language and environment for statistical computing*. https://www.R-project.org/

Serviço Florestal Brasileiro - SFB. (2014). *Manejo da Castanha-do-Brasil (Bertholletia Excelsa) Orientações para as boas práticas de manejo, coleta e pós coleta das castanha-do-brasil, Bioma Brasileiro*.

Sharma, M., Goel, S., & Elias, A. A. (2025). Predictive modeling of soil profiles for precision agriculture: a case study in safflower cultivation environments. *Scientific Reports*, *15*(1), 44. https://doi.org/10.1038/s41598-024-83551-9

Spera, S. T., Magalhães, C. A. S., Baldoni, A. B., & Calderano, S. B. (2019). Caracterização pedológica de locais de estudo de populações naturais de castanheira-do-brasil no estado de Mato Grosso. *Pesquisas Agrárias e Ambientais. Nativa*, *7*(2), 145–161.

Suleymanov, A., Gabbasova, I., Komissarov, M., Suleymanov, R., Garipov, T., Tuktarova, I., & Belan, L. (2023). Random Forest Modeling of Soil Properties in Saline Semi-Arid Areas. *Agriculture*, *13*(5). https://www.mdpi.com/2077-0472/13/5/976

Tonini, H., & Baldoni, A. B. (2019). Estrutura e regeneração de Bertholletia excelsa Bonpl. em castanhais nativos da Amazônia. *Ciência Florestal de Santa Maria*, *29*(2), 607–621.

Vieira, S. R. (2000). Geoestatística em estudos de variabilidade espacial do solo. In R. F. Novais, V. H. Alvarez V., & C. E. G. R. Schaefer (Eds.), *Tópicos especiais em ciências do solo* (pp. 1–54). Sociedade Brasileira de Ciência do Solo.

Wadt, L. H. O., Santos, L. M. H., Bentes, M. P. M., & Oliveira, V. B. V. (2017). *Avaliação edáfica e nutricional em espécies arbóreas. Produtos florestais não madeireiros: guia metodológico da Rede Kamukaia* (1st ed.). EMBRAPA.

Wadt, L. H. O., Santos, L. M. H., Marocolo, F. J., Rego, D. S. G., & Emidio, K. (2019). *Panorama geral da produção extrativista de castanha-da-amazônia no Estado de Rondônia*. Embrapa.

Yamamoto, J. K., & Landim, P. M. B. (2015). *Geoestatística: conceitos e aplicações* (1st ed.). Oficina de texto.