Revista Brasileira de Ciência do Solo

Division - Soil in Space and Time | Commission – Pedometrics

Multivariate Analysis and Machine Learning in Properties of Ultisols (*Argissolos*) of Brazilian Amazon

Cristiano Marcelo Pereira de Souza[(1)*], André Thomazini[(2)], Carlos Ernesto Gonçalves Reynaud Schaefer[(3)], Gustavo Vieira Veloso[(4)], Guilherme Musse Moreira[(4)] and Elpídio Inácio Fernandes Filho[(3)]

[(1)] Universidade Estadual de Montes Claros, Programa de Pós-Graduação em Geografia, Montes Claros, Minas Gerais, Brasil.
[(2)] Universidade Federal de São João Del-Rei, Departamento de Ciências Agrárias, Sete Lagoas, Minas Gerais, Brasil.
[(3)] Universidade Federal de Viçosa, Departamento de Solos, Viçosa, Minas Gerais, Brasil.
[(4)] Universidade Federal de Viçosa, Departamento de Solos, Programa de Pós-Graduação em Solos e Nutrição de Plantas, Viçosa, Minas Gerais, Brasil.

*** Corresponding author:**
E-mail: cristiano.souza@ufv.br

Received: January 5, 2018

Approved: May 4, 2018

How to cite: Souza CMP, Thomazini A, Schaefer CEGR, Veloso GV, Moreira GM, Fernandes Filho EI. Multivariate analysis and machine learning in properties of Ultisols (*Argissolos*) of Brazilian Amazon. Rev Bras Cienc Solo. 2018;42:e0170419.
https://doi.org/10.1590/18069657rbcs20170419

ABSTRACT: Ultisols are the most common soil order in the Brazilian Amazon. The Legal Amazon (LA) has an area of 5 × 10^6 km^2, with few accessible areas, which restricts studies of soils at a detailed level. The pedological properties can be estimated more efficiently using statistical procedures and machine learning techniques, tools which are capable of recognizing patterns in a large soil database. We analyzed the main chemical and physical properties of the B horizons of the Ultisols of the Brazilian Amazon, as well as the spatial variability of the most explanatory properties of these horizons. Physical and chemical data of 1,068 profiles of the RadamBrasil Project were used. A principal component analysis (PCA) was applied and the most explanatory variables were separated by morphostructural units and climate zones. The technique of machine learning was used for spatialization of the explanatory variables based on predictive covariates. In general, the horizons are thick, clay, with a predominance of negative charges, and low levels of exchangeable cations. The variables retained in the PCA were: sum of bases (SB), Al^{3+}, degree of flocculation (Floc), ΔpH, and organic carbon content (C). Areas of greater precipitation have low SB, with higher values in the basement complex (BC) and in areas under the Andean influence. Higher levels of Al^{3+} and degrees of flocculation were also associated with greater precipitation. However, the soils are predominantly electronegative, showing a kaolinitic mineralogy. The C contents in general were low, with an increase in more humid zones due to the process of mineralization and illuviation (podzolization), and in the BC due to the protection of C by the aggregation of clay. The use of multivariate analysis allowed a better understanding of the Ultisols' main properties in different morphostructural and climatic domains, and its spatialization facilitated the interpretation of properties and their relationships with environmental characteristics in the Legal Amazon.

Keywords: Legal Amazon, principal component analysis, morphostructural domains, climatic zones.

INTRODUCTION

The Legal Amazon (LA) consists of an administrative region of 5.2×10^6 km², which includes the Lowland Amazon Forest and Cerrado biomes. Despite the overall dominance of rainforest ecosystems (Ab'saber, 2002), it is now recognized that the region presents much more varied systems (Almeida et al., 2010), ranging from sandy ecosystems with dwarf plants (Mendonça et al., 2015) to transitional ecotones between Cerrado and Forest (Marimon et al., 2006) and Dense Equatorial Forest (ombrophilus).

Due to the extensive territorial dimensions and the limited accessibility, studies of soils in the Amazon have mainly been based on local scales and on toposequences (Chauvel, 1981; Bernoux et al., 1998; Feitosa et al., 2016). However, in the 1970s an exploratory survey of the soils of this region was conducted, called the RADAM project (Brasil, 1973), and contributed significantly to increasing the knowledge on the soils and their environmental interpretations (Sanchez et al., 1982; Rodrigues, 1996; Nepstad et al., 2004).

Subjected to wet, hot climates, Amazon soils are highly developed, with a predominance of Oxisols, Ultisols, Plinthosols (Plintaquults) according to the system Soil Survey Staff (2014), mostly dystrophic, and with few primary minerals and nutrient reserves (Sanchez et al., 1982; Curi and Franzmeier, 1984; Vale Júnior et al., 2011). Unlike the Oxisols, the Ultisols, which match the order *Argissolos* in the Brazilian Soil Classification System (Santos et al., 2013), have characteristics such as less structural development at the subsurface and a textural gradient attributed to clay losses at the surface. Like the Amazon Oxisols, Ultisols have similar chemical and physical properties, and parent materials.

The study of soils in large areas can be effective if homogeneous zones can be recognized, allowing the extrapolation of properties influenced by the dominant characteristics (Trangmar et al., 1986; Corá et al., 2004), for example, the climate, which is key factor in the landscape evolution and pedogenesis of the Amazon (Yaalon, 1983; Stockmann et al., 2014; Delarmelinda et al., 2017). The local landform is also an excellent stratifier of environments, translating into similar hydrological and geological characteristics that have a direct influence on the pedological properties (Mulla and McBratney, 2001; Park and Burt, 2002).

Studies of large territorial dimensions and databases need robust methods of analysis. In this regard, statistical procedures can help with the identification of patterns and dataset groupings, as recently reported in soil studies (Gomes et al., 2004; Carvalho Junior et al., 2008). In addition, the use of techniques of machine learning algorithms (MLA) has recently gained strength. These are a set of models that are able to find patterns in data and perform predictions (Witten et al., 2016) and have already been applied in pedology (Bui and Moran, 2003; Behrens et al., 2005; Kovačević et al., 2010; Brungard et al., 2015; Forkuor et al., 2017).

Although the distribution of Amazon soils and their associations in relation to certain environmental characteristics are recognized on a generalized scale, no study has yet evaluated the relationship between the soil properties of a given order in relation to the different morphoclimatic domains, by using statistical tools and machine learning techniques in a regional approach. The objective of this study was to (i) determine and select the main chemical and physical properties that characterize the diagnostic Bt horizon of the Amazon Ultisols (*Argissolos*) through principal component analysis, and (ii) determine the spatial variability of the selected properties, in order to deepen the understanding of the spatial relationships with the geomorphology and landforms, helping to understand how Ultisols vary across the Amazon.

MATERIALS AND METHODS

Characterization of the study area

The study area comprises all the territory of the Legal Amazon Region (Figure 1). It is located between the coordinates 5° 28' N and – 18° 04' S latitude, and – 73° 10' to – 44° 00’ W longitude, with an area of approximately 5 × 10^6 km², corresponding to 61 % of the Brazilian territory. The climate is humid to sub-humid equatorial, with rainfall ranging from 1,400 to 3,500 mm yr^{-1}, with climatic types Af, Am until Aw (Vale Júnior et al., 2011). The average air temperature varies from 22 to 28 °C. In this research, the divisions of climatic zones considered were: semi-arid (SemiA); semi-humid (SemiU); humid (H), and super-humid (SU) (Figure 1a).

At a macroscale, the area has three geotectonic units (morphostructural domains) (Figure 1b): crystalline basement complex (BC) represented by granites and gneisses rocks related to the Brazilian crystalline basement, forming hills, mountains, crystalline massifs, plateaus and depressions; sedimentary deposits (SD), the unit mainly having Quaternary sediments, notably the Içá and Boa Vista Formations, which form low plateaus, tablelands, depressions and, at the edges, occasionally have sediment of the Alter do Chão Formation (Cretaceous); sedimentary Basins (SB), mainly related to the alluvial plain of the Amazon, having floodplain, mangroves, and terraces deposits (Sombroek, 2000).

The Ultisols occupy 1.79 million km^2 (Santos et al., 2011), representing 35 % of the area of the Legal Amazon and concentrated mainly in the state of Amazonas and Pará (Figure 1c).

Selection of the variables

Data from 1,068 profiles of Ultisols were used (considering only the horizon Bt), of the soil survey RADAM project (Brasil, 1973). We analyzed: depth (m); silt, clay, and sand (g kg^{-1}); degree of flocculation (Floc); pH(H_2O); total organic carbon (C, g kg^{-1}); exchangeable calcium (Ca^{2+}, $cmol_c$ dm^{-3}); magnesium (Mg^{2+}, $cmol_c$ dm^{-3}); potassium (K^+, $cmol_c$ dm^{-3}); sodium (Na^+, $cmol_c$ dm^{-3}); exchangeable aluminum (Al^{3+}, $cmol_c$ dm^{-3}); cation exchange capacity (CEC, $cmol_c$ dm^{-3}); sum of bases (SB, $cmol_c$ dm^{-3}); bases saturation (V, %), aluminum saturation (m, %), and the difference between pH(H_2O) and pH(KCl) (ΔpH).

The Ultisols order was selected as it presented great expressivity in the region, in addition to which the soil showed evolutionary characteristics that better preserve properties related to climate, geology, and landforms, especially when compared to the Oxisols, which have a much greater development stage.

A descriptive statistical analysis was performed to identify the main characteristics of the Ultisols. The chemical and physical properties dataset was also submitted to principal component analysis (PCA) (Abdi and Williams, 2010) in order to reduce the number of variables and select those that are more explanatory on the characterization of the Bt horizon.

The Kaiser criterion (Kaiser, 1960) was used to reduce the number of variables and main components, where components with eigenvalue ≤1.00 are discarded. In filtered components, the variable with the greatest percentage of explanation/importance was analyzed separately by morphostructural and climatic units.

Spatial variability

The spatialization of properties was based on the Extremely randomized trees algorithm (Extra-Trees ET), which is a variant of the Random Forest (RF) introduced by Geurts et al. (2006), with an added layer of randomness, because instead of seeking the ideal division, the extra-tree uses a limit of random division during the decision trees training (Maier et al., 2016). Extra-Trees shows good performance in data analysis, because it can improve accuracy and reduces the computational complexity (Geurts et al., 2006; Wehenkel et al., 2006; Devaney et al., 2015). The whole procedure was performed in software R (R Core Team, 2017).

A database was elaborated with predictive covariates, originated from the Digital Elevation Model (DEM) SRTM (30 m) (Shuttle Radar Topography Mission) and data on the geology, geomorphology, and pedology available in the Brazilian Institute of Geography and

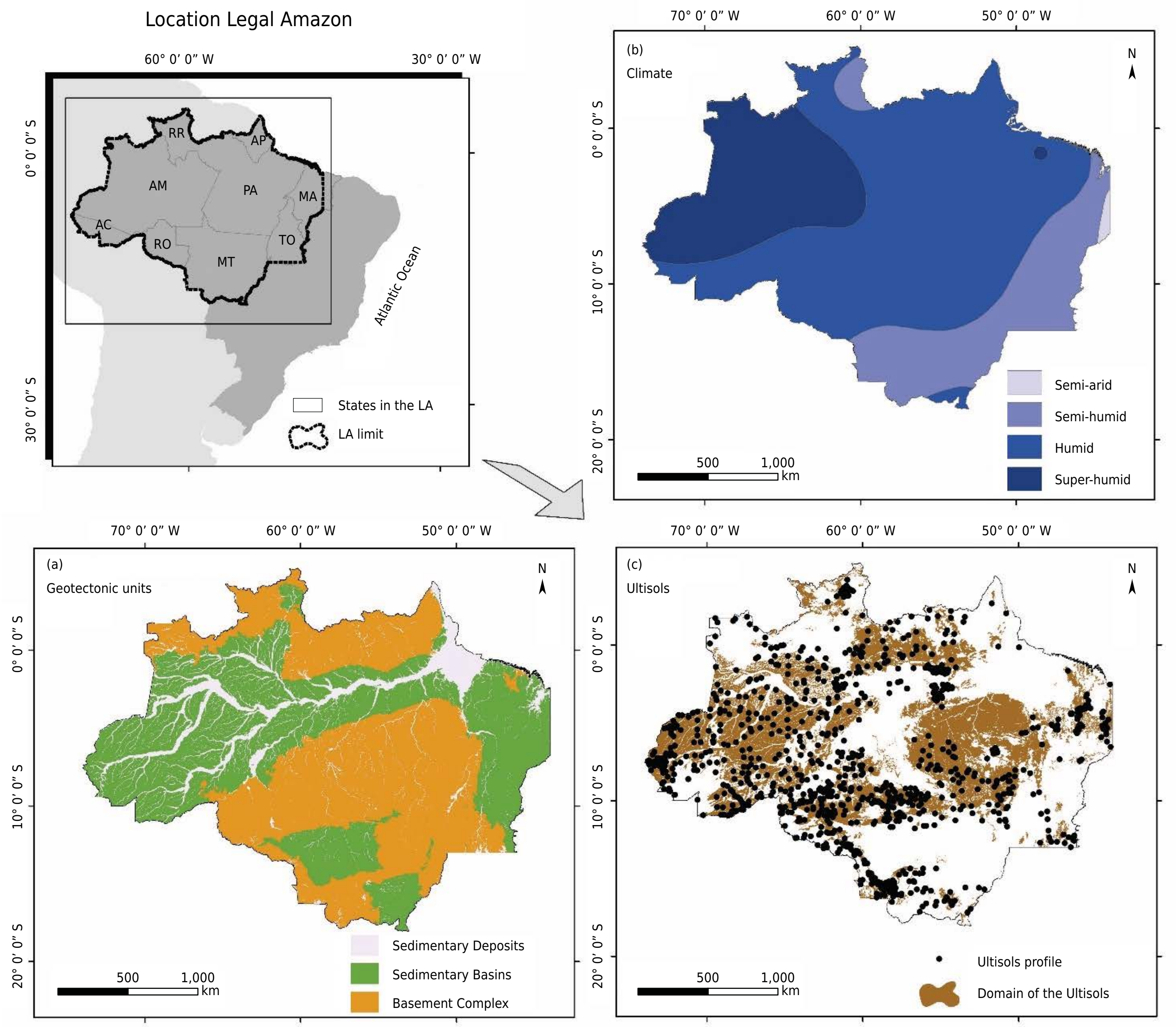

Figure 1. Location of the Legal Amazon with limits of the States, with an emphasis on morphostructural domains (a); climate zones (b), and profiles and domain of the Ultisols (c). Data (a) and (b) are available at IBGE (2017).

Statistics (IBGE, 2017), 19 items (Bioclimatic variables) of spatial information from the climatic data of WorldClim (Hijmans et al., 2005), the NDVI (Normalized Difference Vegetation Index) calculated from satellite images from the MODIS sensor, and also 41 geomorphometric data extracted from MDE using the software RSAGA (Brenning, 2008), associated with R, a method already applied by Hengl et al. (2008), Olaya and Conrad (2009), and Brungard et al. (2015). The individual definition of the variables can be found in Wilson and Gallant (2000) and MacMillan and Shary (2008). The raster files were sampled again for the cell size of 1 km^2.

From the predictors database, only 15 were selected, since with fewer predictors, the time and computational complexity are reduced (Kuhn and Johnson, 2013; Forkuor et al., 2017). The selection of predictors (numerical) is performed in two steps, the first based on the level of correlation (function findcorrelation- package caret) (Kuhn and Johnson, 2013), which analyzes the absolute values of the correlations between the predictors, which are analyzed in pairs, concomitantly; these are compared with all other predictors (numerical) of the database, and the one that exhibits the highest correlation with the database is thus eliminated.

The limit of correlation to eliminate the predictor was >95 % (a value defined by the operator). The second step employed the function Random Forest-Recursive Feature Elimination (RF-RFE) (Guyon et al., 2002), with the objective of selecting the smallest possible subset of predictors and with a certain spatialization capacity. To do this, we chose to select the subset that had lower performance than 3 % below the best R^2 found, because the best R^2 can be a reflection of the excess of predictors; furthermore, using fewer predictors ensures simplicity in the method and this factor can be essential so that the mapping is efficient, particularly for large territorial extensions (McKenzie and Gallant, 2006; Hartemink et al., 2008).

The training process/prediction was applied 100 times, where the classifier each time separated randomly a share of 75 % of the samples for training and 25 % for validation, also generating statistical data for each routine and a map by means of the results. The application of this procedure may provide a better estimate because it evaluates with different groups of data (Kohavi, 1995; Grimm et al., 2008; Kuhn and Johnson, 2013). Every procedure is represented briefly in figure 2.

RESULTS

General characterization of the Bt horizon

In relation to the general properties of the Ultisols - *Cambissolos* (Table 1), it was observed that the Bt horizon of the Ultisols of Legal Amazon has an average thickness of 1.34 m, with mean clay content of clay of 406.6 g kg^{-1}. However, the data show that these values can reach up to 2.90 m depth, and the clay content up to 870.0 g kg^{-1}. The pH values range from 1.6 to 7.5 (mean of 4.91). Organic carbon contents were low, with an average of 3.6 g kg^{-1} and maximum value of 40.0 g kg^{-1}. The levels of Ca^{2+} and Mg^{2+} can reach 54.27 and 9.35 $cmol_c$ dm^{-3}, respectively. In general, the degree of flocculation is 76.22 % and the Al^{3+} content is 2.33 $cmol_c$ dm^{-3}. The majority of Bt horizons are dystrophic (V = 23.32 %), with mean aluminum saturation (m) of 57.40 %. The ΔpH indicated a predominance of negative charges in most of the analyzed Bt horizons, despite the low activity clay.

Multivariate analysis

The results of the PCA (Table 2) showed that only the first five main components were retained for analysis, based on the Kaiser criterion (main component with eigenvalue ≥1.0) (Kaiser, 1960).

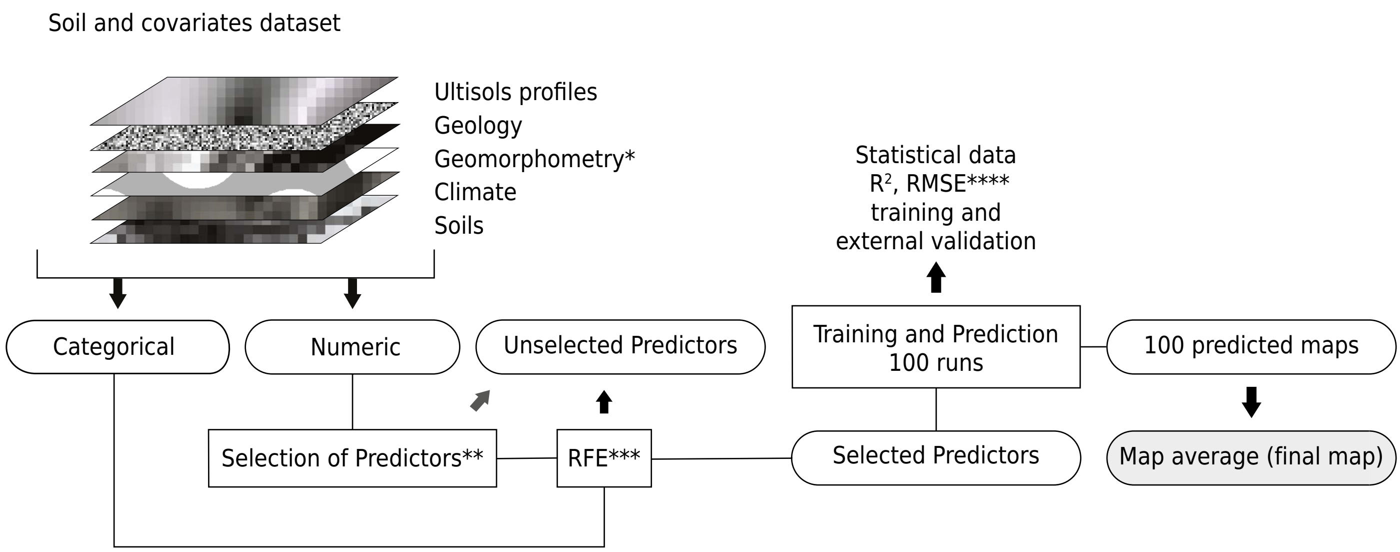

Figure 2. Flow chart representing the steps of selection of covariates and spatialization of properties of the Ultisols using the software R. Elements in the figure: * Data of the MDE-SRMT; ** findcorrelation function = refers to the data elimination that present collinearity (>97 %); *** Recursive Feature Elimination; **** Root-mean-square error.

Table 1. Descriptive statistics considering all profiles of Ultisols (Bt horizon) described by survey of RADAMBrazil.

Statistic	Depth	Silt	Clay	Sand	Floc	C	Ca^{2+}	Mg^{2+}	K^{+}	Na^{+}	Al^{3+}	CEC	SB	V	m	pH(H_2O)	ΔpH
	m	g kg^{-1}			%	g kg^{-1}	cmol$_c$ dm^{-3}							%			
Minimum	0.10	10.0	20.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.39	0.06	0.33	0.00	1.60	-5.70
Maximum	2.90	610	870	940	100.0	40.0	53.27	9.35	1.63	1.61	26.04	63.00	58.68	100.00	99.24	7.50	2.80
1st quartile	1.10	100	290	250	61.0	2.00	0.02	0.01	0.02	0.01	0.25	2.76	0.22	4.89	22.73	4.50	-1.00
Median	1.40	170	400	400	95.0	3.00	0.14	0.08	0.04	0.02	1.00	4.43	0.44	11.47	69.77	4.80	-0.70
3rd quartile	1.60	270	510	550	99.3	4.40	0.50	0.36	0.10	0.03	2.90	7.48	1.19	32.14	89.29	5.30	-0.50
Mean	1.35	196	407	397	76.2	3.60	0.94	0.47	0.10	0.05	2.33	6.44	1.62	23.32	57.40	4.91	-0.84
Variance	1261.7	147.0	226.0	406.6	1055.4	0.08	10.46	1.20	0.02	0.01	11.70	43.34	16.43	690.57	1271.20	0.43	0.49
Standard deviation	35.5	12.1	15.0	20.2	32.5	0.29	3.23	1.09	0.16	0.09	3.42	6.58	4.05	26.28	35.65	0.66	0.70
CV	0.30	0.60	0.4	0.5	0.4	0.79	3.42	2.31	1.63	2.06	1.47	1.02	2.50	1.13	0.62	0.13	-0.84
Asymmetry	-0.20	0.90	0.3	0.1	-1.2	5.29	8.58	4.26	4.41	9.08	2.55	3.50	7.01	1.45	-0.52	0.73	-3.41
Kurtosis	0.10	0.26	-0.41	-0.79	-0.31	52.91	99.13	21.49	26.12	121.17	8.05	17.39	67.99	0.95	-1.27	1.38	19.27
Standard error	1.10	0.36	0.46	0.61	1.04	0.01	0.10	0.03	0.00	0.00	0.10	0.20	0.12	0.80	1.09	0.02	0.02

Floc = flocculation; C = organic Carbon; Ca^{2+} = exchangeable calcium; Mg^{2+} = exchangeable magnesium; Na^{+} = exchangeable sodium; Al^{3+} = exchangeable aluminum; CEC = effective cation exchange capacity; SB = sum of base; V = base saturation; m = Al^{3+} saturation; ΔpH (H_2O - KCl) (Donagema et al., 2011). CV = coefficient of variation.

Table 2. Eigenvalue, variability, and percentage explained for the first five principal components selected according to the Kaiser criterion

Soil properties	Principal Components				
	CP1	CP2	CP3	CP4	CP5
	%				
Depth	0.53	0.01	0.78	15.48	16.62
Silt	2.78	10.56	0.87	1.52	1.27
Clay	0.86	7.02	19.94	12.71	6.18
Sand	2.84	15.39	7.51	11.54	1.44
Floc	1.48	0.54	22.51	4.57	13.50
pH	4.45	11.81	0.65	8.92	3.48
C	0.24	1.21	12.23	0.00	36.70
Ca^{2+}	15.49	0.00	0.17	6.37	1.47
Mg^{2+}	15.55	0.00	0.27	2.05	0.00
K^{+}	5.09	0.96	0.05	3.17	5.69
Na^{+}	3.96	0.20	0.11	6.43	0.89
Al^{3+}	0.71	19.44	4.28	0.06	0.03
CEC	10.21	9.51	2.05	1.69	0.47
SB	18.60	0.00	0.26	5.46	0.74
V	10.52	8.35	3.25	0.22	0.55
m	5.63	14.03	7.12	0.57	0.00
ΔpH	1.06	0.98	17.94	19.22	10.97
Eigenvalue	4.49	3.34	1.68	1.33	1.07
Variability (%)	26.41	19.67	9.87	7.83	6.27
Cumulative variability (%)	26.41	46.08	55.95	63.78	70.05

Floc = flocculation; C = organic carbon; CV = coefficient of variation; CEC = cation exchangeable capacity; SB = sum of bases; V = base saturation; m = aluminum saturation; and ΔpH = difference between H_2O and KCl pH.

The retained components explain about 70 % of the total variability in the data set. The variables with the greatest percentage of explanation/importance for each component were: SB (18.60 %) - CP1, Al (19.44 %) - CP2, Floc (22.51 %) - CP3, ΔpH (19.22 %) - CP4, and C (36.70 %) - CP5. These variables are considered the most explanatory ones among the data set and that effectively describe the chemical and physical properties of the Bt horizon Ultisols of the Legal Amazon.

Comparison between morphostructural units

With regard to morphostructural units (Figure 3), Amazon Ultisols tend to decrease the sum of bases (SB) from the crystalline basement complex (BC) for the sedimentary deposits (SD) that presented mean values of 1.33 and 0.82 $cmol_c$ dm^{-3}, respectively. The amplitudes of 50 % of the core values (1st quartile and 3rd quartile) of these properties were 0.18 and 0.68 $cmol_c$ dm^{-3} in the SD and 0.26 and 1.50 $cmol_c$ dm^{-3} in the BC. In the SD area, the values of the 1st quartile and 3rd quartile are in the range 0.19 and 0.94 $cmol_c$ dm^{-3}, i.e., lower than in the BC areas.

In relation to the Al^{3+}, the highest levels tended to occur in the SD, with an average value of 5.25 $cmol_c$ dm^{-3} and central values distributed between 1.41 and 8.48 $cmol_c$ dm^{-3}. The other morphostructural units had lower levels of Al^{3+} (3.33 and 0.91 $cmol_c$ dm^{-3} in the SB and BC, respectively). The flocculation decreased in the order BC>SD>SB, with mean values 83.43, 77.00, and 68.90 %. The most electronegative Bt horizons were observed in the SB, presenting ΔpH means of -0.88. The other morphostructural units showed similar values (0.78 and -0.79). Outstanding variations were not observed in organic carbon contents in morphostructural units, with higher mean values for the BC (4.0 g kg^{-1}).

Comparison among different climates

The analysis of Ultisols by climate units (Figure 3) indicated a decreasing sum of bases with increasing levels of rainfall, which corresponds on average to 2.92, 2.57, 1.43, and 1.17 $cmol_c$ dm^{-3}, respectively from the hydric regimes semi-arid (SemiA), semi-humid (SemiH), humid (H), and super-humid (SH). The opposite behavior was observed for the levels of Al^{3+}, which increased with greater rainfall, showing respectively the averages 0.55, 0.71, 1.77, and 4.96 $cmol_c$ dm^{-3} for the same regions respectively.

There is a tendency of increase in the degree of flocculation of Bt horizons with increased rainfall. The averages were 48.6, 64.7, 79.9, and 78.0 % from the driest index to the most humid. The data of ΔpH showed consistent decrease of electronegativity with increased rainfall, with mean values of -1.06, -1.00, -0.83, and -0.70 from the driest to the wettest. All Ultisols, however, had a predominance of negative charges in the Bt horizons. Similar behavior was also observed for the amounts of soil organic carbon, although with less difference. The average contents were 2.9, 3.3, 3.6, and 4.0 g kg^{-1} for the SemiA, SemiH, H, and SH regimes respectively. The central limiting values were 2.5 and 2.9 g kg^{-1}; 2.0 and 4.1 g kg^{-1}; 2.1 and 4.3 g kg^{-1}, and 2.4 and 5.0 g kg^{-1}, from the driest to the wettest regime, indicating that most Bt horizons have low organic carbon contents.

Spatial variability

The use of machine learning algorithms allowed us to consider various predictive covariates; however, the function (findcorrelation) excluded data with high correlation, since a common limitation of the regression models is the multicollinearity that occurs when there is a significant correlation among the predictors (Forkuor et al., 2017). Based on the selected predictors, values of R^2 were obtained, with a higher value for Al^{3+} ($R^2 = 0.47$) and lower for carbon ($R^2 = 0.06$). Among the predictors used, the most explanatory were those related to climatic factors, which influenced the mapping of the variables degree of flocculation, organic carbon contents, and sum of bases. On the other hand, the predictors that best explained the distribution of Al^{3+} were those related to landforms (Table 3).

The results of the training and validation process, run 100 times (Table 4), showed that the average of R^2 of the training and the validation did not exhibit large variations, which indicates a low effect of overfitting, since the technique of machine learning is able to reduce this statistical problem (Drake et al., 2006; Were et al., 2015). The values of the coefficient of variation in the training process did not exhibit large variations for Al^{3+}, Floc, and ΔpH, and were high for sum of bases (CV = 24 %) and carbon (CV = 29 %).

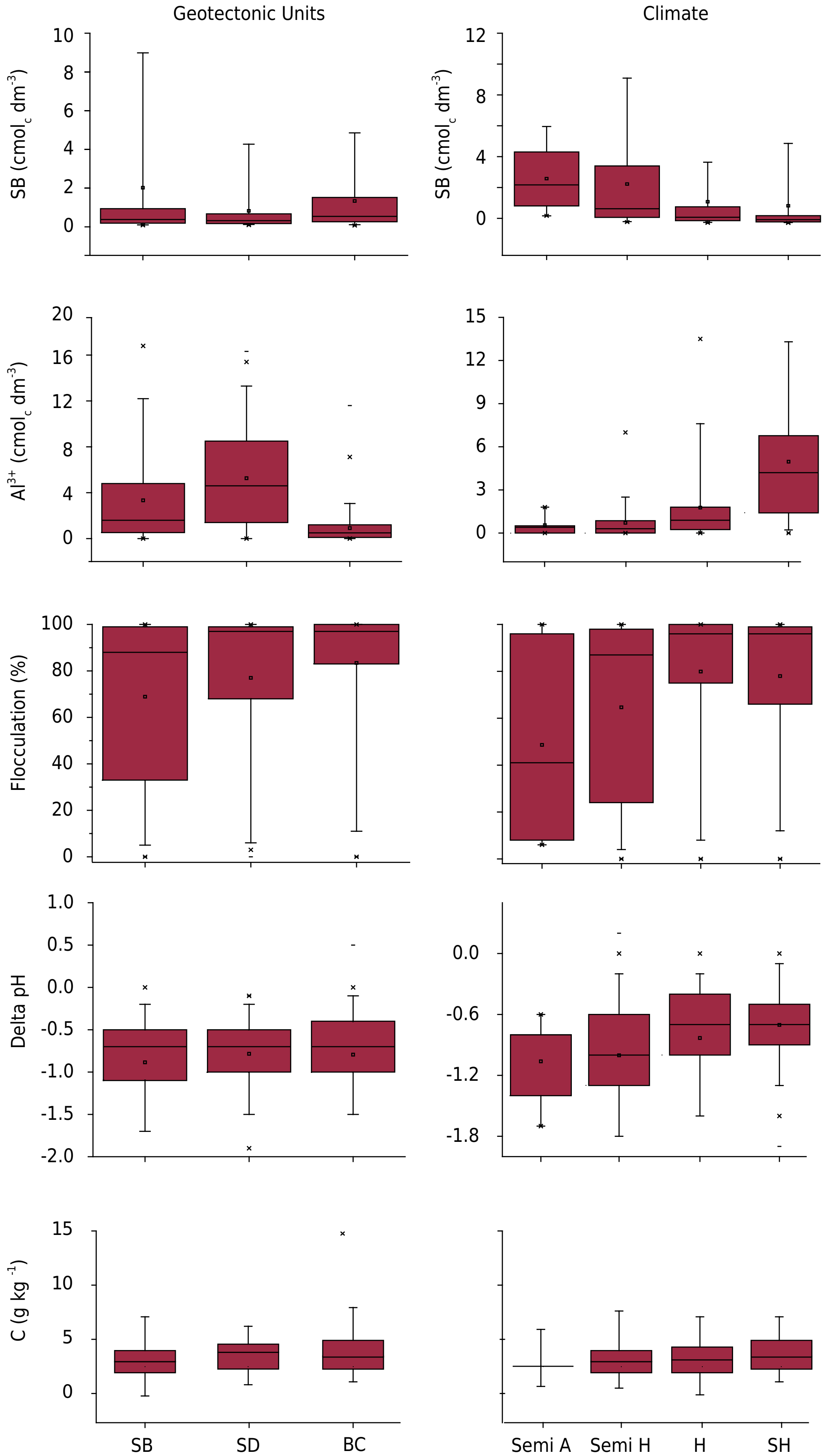

Figure 3. Mean, median, 1st, and 3rd quartile, near and distant outliers, maximum and minimum values of the variables among the groups of morphostructural domains: SB = sedimentary basins; SD = sedimentary deposits; BC = basement complex. Climate units: SemiA = semi-arid; SemiH = semi-humid; H = humid; SH = super-humid. SB = sum of bases ($cmol_c\ dm^{-3}$); Al^{3+} = exchangeable aluminum ($cmol_c\ dm^{-3}$); Flocculation = degree of flocculation (%); C = organic carbon ($g\ kg^{-1}$).

Table 3. Geostatistical results and covariates selection for predicting soil properties

Soil properties	Ncov	R^2	RMSE	Degree of importance						
				1	2	3	4	5	6	7
Floc	5	0.20	27.37	MRVBF	Bio 4	Bio 9	Bio 5	Bio 2	NA	NA
C	5	0.03	0.28	Wetness Index	Bio 14	Bio 15	Bio 12	DEM	NA	NA
Al^{3+}	5	0.47	2.54	MRVBF	MRRTF	Bio 15	Slope Height	Valley Depth	NA	NA
SB	7	0.23	3.31	DEM		Bio 4	Bio 11	bio_2	MRVBF	MRRTF
ΔpH	7	0.24	0.62	Bio 12	MRRTF	MRVBF	Bio 4	DEM	Terrain.S.C	Bio 15

Ncov = number of covariates; Floc = degree of flocculation; SB = sum of bases; R^2 = coefficient of determination; RMSE = root-mean-square error. Covariables derived from the RSAGA package: MRVBF = multiresolution index of valley bottom flatness; MRRTF = multiresolution index of ridge top flatness; Slope Heigth = vertical distance between base and slope ridge; DEM = digital elevation model; Terrain.S.C = Terrain surface convexity; Bioclimatic variables of WorldClim: Bio 4, Bio 5, Bio 9, Bio 11, Bio 12, Bio 14, Bio 15.

The spatial distribution (Figure 4) revealed that the values of sum of bases tend to be higher in the southwestern region, extending from the state of Acre to Mato Grosso State. These areas have either high elevations associated with the basement complex, or areas under the influence of Andean sediments, which are richer in primary minerals (Gama et al., 1992; Lips and Duivenvoorden, 1996; Lima et al., 2006; Guyot et al., 2007; Schaefer et al., 2017). On the other hand, in lowland areas of the sedimentary basins, the values of the sum of bases decrease, increasing the exchangeable Al^{3+}. It should be emphasized that, although the values of Al^{3+} could reach 26.0 $cmol_c\ dm^{-3}$, in the training/prediction process little importance is assigned to the extreme values, because the extra-trees algorithm is based on the distribution of higher frequency data, eliminating extremes.

The degree of flocculation means the level of non-dispersible clay, which is usually higher in areas of more weathered soils, where Al and Fe-oxides favor the flocculation of clay particles. In lowland areas, sedimentary basins and sedimentary deposits, the degree of flocculation was not high, and is unrelated to Al^{3+} levels, which is consistent with higher rainfall, as observed. The least flocculated Ultisols tend to occur in the peripheral region of the Legal Amazon, covering the southern regions of the state of Acre, Rondônia, and Mato Grosso State, and tend to rise in the easternmost area between Tocantins and Maranhão State. In relation to pH, the most electronegative soils are distributed in the eastern portion of the Legal Amazon, associated with areas of lesser rainfall. The highest amounts of organic carbon were recorded in the northeastern Legal Amazon, where crystalline rocks usually favor the formation of more clayey soils (Sombroek, 2000).

DISCUSSION

Statistical analysis indicated that the vast majority of the Bt horizon of the Legal Amazon (LA) Ultisols are generally deep, loamy, electronegative, and with low exchangeable cations contents. These properties are strongly influenced by climatic conditions. Although the LA region presents a wide variety of environments, it is also marked by a wide lithological variety (Quesada et al., 2011). Thus, the long process of chemical weathering under a humid to super-humid equatorial climate, associated with small temperature oscillations during all the Neogene period, has resulted in a very deep mantle of leached soils (Costa et al., 1997).

The predictors that best explain the distribution of the Ultisols properties were the bioclimatic variables, as also confirmed by multivariate analysis. For landform data, the predictors flatness, valley incision and concavities were more explanatory predictors; this result is consistent with the fact that the LA soils do not show large variations of altimetry, in a general way (Quesada et al., 2011).

Table 4. Descriptive statistics of estimated parameters and statistical cross-validation by machine learning algorithm (Extremely randomized trees) after 100-fold repetition

Statistics	Training			Validation		
	R^2	RMSE	MAE	R^2	RMSE	MAE
Observations number	100.00	100.00	100.00	100.00	100.00	100.00
			Sum of bases ($cmol_c$ dm^{-3})			
Minimum	0.10	2.41	0.84	0.03	1.84	1.10
1st quartile	0.21	3.06	1.22	0.12	2.84	1.39
Median	0.24	3.28	1.39	0.18	3.49	1.51
3rd quartile	0.30	3.58	1.61	0.27	4.08	1.62
Maximum	0.42	4.01	2.46	0.51	6.47	2.23
Mean	0.25	3.29	1.44	0.20	3.53	1.51
Coefficient of variation (%)	24.06	10.09	22.62	51.54	25.12	12.87
Standard deviation (n-1)	0.06	0.33	0.33	0.10	0.89	0.19
			Exchangeable aluminum ($cmol_c$ dm^{-3})			
Minimum	0.43	2.40	1.18	0.32	2.06	1.34
1st quartile	0.46	2.52	1.44	0.41	2.40	1.48
Median	0.48	2.57	1.51	0.47	2.53	1.55
3rd quartile	0.50	2.64	1.66	0.52	2.74	1.64
Maximum	0.54	2.71	2.02	0.64	3.22	2.00
Mean	0.48	2.58	1.56	0.46	2.56	1.57
Coefficient of variation (%)	5.42	3.00	11.26	15.65	9.74	8.25
Standard deviation (n-1)	0.03	0.08	0.18	0.07	0.25	0.13
			Flocculation (%)			
Minimum	0.15	26.01	16.76	0.06	23.97	17.78
1st quartile	0.19	26.93	18.83	0.15	26.76	19.31
Median	0.20	27.41	19.83	0.19	27.56	20.33
3rd quartile	0.22	27.80	20.97	0.23	28.71	21.01
Maximum	0.27	28.55	24.32	0.33	31.26	22.37
Mean	0.20	27.37	20.01	0.19	27.72	20.22
Coefficient of variation (%)	11.90	2.08	8.43	29.65	4.99	5.03
Standard deviation (n-1)	0.02	0.57	1.69	0.06	1.38	1.02
			ΔpH			
Minimum	0.16	0.49	0.29	0.03	0.41	0.31
1st quartile	0.25	0.58	0.34	0.14	0.57	0.36
Median	0.28	0.59	0.37	0.22	0.61	0.38
3rd quartile	0.31	0.62	0.40	0.32	0.68	0.40
Maximum	0.38	0.67	0.48	0.46	0.84	0.47
Mean	0.28	0.59	0.37	0.23	0.62	0.38
Coefficient of variation (%)	15.00	6.00	11.74	48.01	14.15	8.13
Standard deviation (n-1)	0.04	0.04	0.04	0.11	0.09	0.03
			Organic carbon (g kg^{-1})			
Minimum	0.02	0.20	0.12	0.00	0.18	0.13
1st quartile	0.05	0.25	0.15	0.01	0.22	0.16
Median	0.06	0.26	0.16	0.03	0.27	0.16
3rd quartile	0.07	0.28	0.19	0.05	0.33	0.17
Maximum	0.12	0.30	0.23	0.14	0.43	0.20
Mean	0.06	0.26	0.16	0.04	0.28	0.16
Coefficient of variation (%)	29.77	7.46	15.57	79.02	23.92	8.53
Standard deviation (n-1)	0.02	0.02	0.03	0.03	0.07	0.01

R^2 = coefficient of determination; RMSE = root-mean-square Error; MAE = mean absolute error.

The statistical data of the maps (Table 4) demonstrated a higher R^2 (0.48) for Al^{3+}, which can also be explained by climatic variables, while the lowest value of R^2 (0.06) was observed for organic carbon (C), which can be explained by the low variation of the values of C on the Bt horizon, which presents low values in relation to the upper horizons.

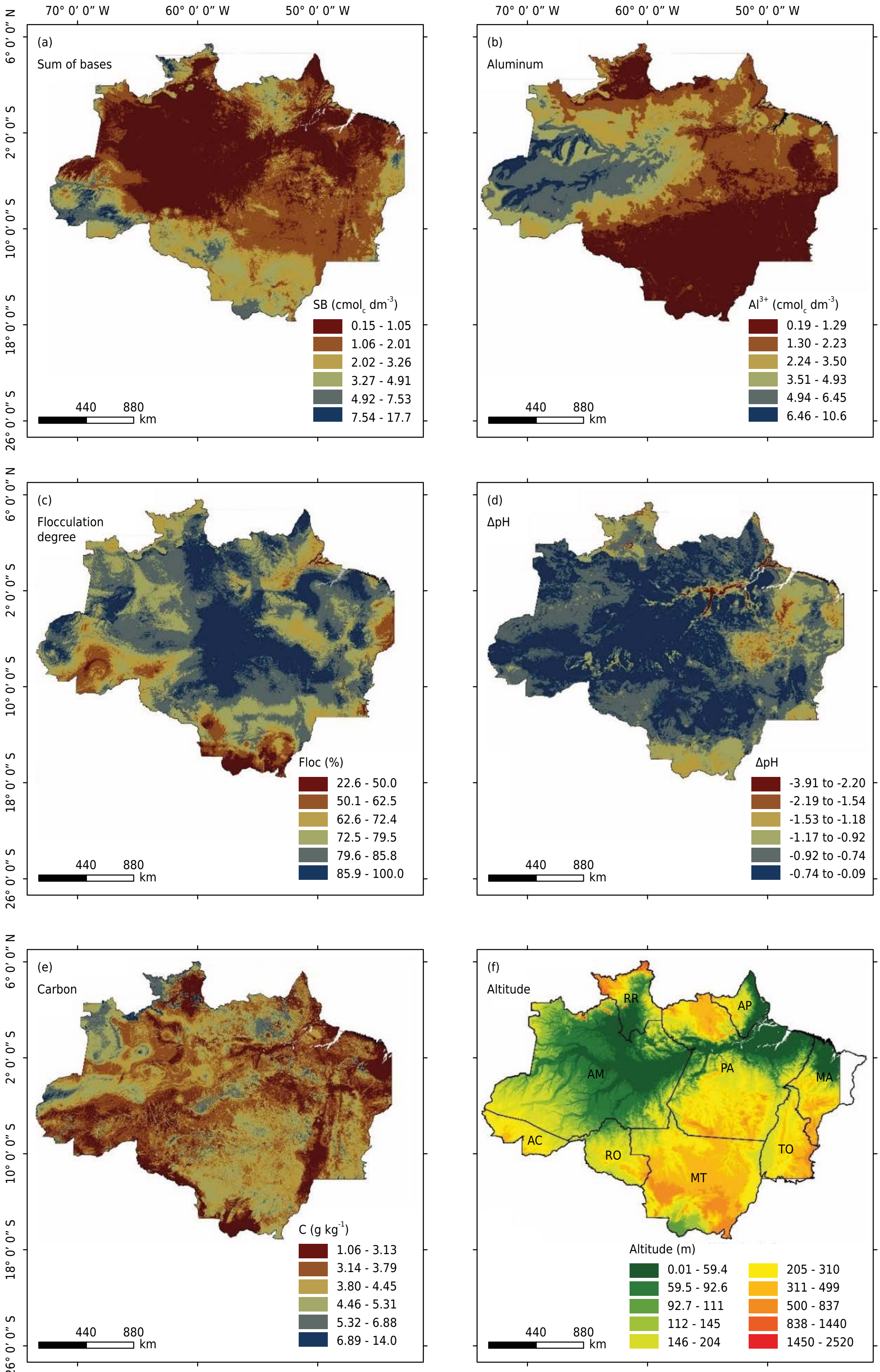

Figure 4. Thematic maps obtained by machine learning of soil attributes: sum of bases (a), aluminum (b), degree of flocculation (c), ΔpH (d), carbon (e). Map of altitude and location of the states for correlation of the attributes of soils mapped with geographical location and altitude (f).

The predictors used are not suitable to predict variable C in the subsurface. According to Grimm et al. (2008), topographic data are better for the prediction of organic carbon in layers from 0.00-0.10 m, considering also that the data must have better resolution.

In general, the values of R^2 were satisfactory, since for spatial quantitative models of soil, R^2 values higher than 0.70 are uncommon and values of 0.50 or less are common (Malone et al., 2009); in addition, the statistical results are equivalent to applied studies with different spatialization methods (Tarr et al., 2005; Kumar et al., 2012; Vaysse and Lagacherie, 2015) and considering that the scale adopted was 1 km.

The statistical results of machine learning can be interpreted as the result of a deficit of predictive information related to soils (Vaysse and Lagacherie, 2015). The choice of the analyzed information, which is only the Bt horizons of Ultisols for the entire extension of Legal Amazon, is a factor that can negatively interfere in the model.

When considering the explanatory variables retained in each component in the PCA, the results indicate that SB tends to be higher in the BC areas. This domain exhibits crystalline, metamorphic rocks or metasediments, which are characterized by gneisses-granite lands, with a predominance of gneisses with granodioritic to granitic composition; subordinately, there are amphibolites, trondhjemites, and granulites (Oliveira Junior and Corrêa, 2002), richer parent materials compared to other domains. However, the climatic factor in the BC region, in general, should contribute to the lower SB, while on the other hand, the rainfall regime can also cause erosion of surface soil layers, rejuvenating the soil profile.

In the Amazon, 78 % of soils are acid and have low natural fertility (Sanchez, 1977; Delarmelinda et al., 2017), characteristics well represented in the areas in the central part of the SB and SD domains. This chemical poverty is directly influenced by climatic factors, since the highest rainfall in the central portion of the state of Amazonas, promotes the greatest leaching. According to Schaefer et al. (2000) and Melo et al. (2006), the exceptions concerning higher fertility in the Amazon region occur in the floodplain Aquents (*Gleissolos*) of the Amazonas river alluvial plain, in the soils of low plateaus, and terraces of the basins of Acre and Solimões/ Upper Amazon, both influenced by carbonate-rich or Andean sediments, or scattered areas of outcrops of mafic rocks.

The levels of sum of bases are higher mainly in the western part of the states from Acre to Mato Grosso, a fact that is related to the influence of the Andean region. The Solimões Formation (Miocene-Pliocene) is a late Cenozoic sedimentation associated with the Andes orogenesis, and general subsidence of the basin (Westaway, 2006). The soils of this region are characterized by the presence of 2:1 minerals (smectites, vermiculites, and chlorites), with high amounts of exchangeable Al^{3+}, Ca^{2+}, and Mg^{2+} (Volkoff et al., 1989; Lima et al., 2006).

Another aspect that denotes an influence of the Andes on Amazon soils is the chemical features of floodplains in rivers coming from Andean sources, often having higher levels of nutrients (Ca, P, and Mg) compared to other tributaries (Gaillardet et al., 1997; Guyot et al., 2007; McClain and Naiman, 2008). The SB values tend to rise again in the easternmost part of the LA, associated with the semi-humid and semi-arid climate and less weathered soils.

The levels of Al^{3+} were higher in the areas of higher precipitation (super-humid) and also in the SD domain. These results are closely associated among themselves, since the origin material is subjected to intense removal of basis by the rainfall regime. However, the Amazon soils usually have a clay mineralogy dominated by kaolinite, with minor chloritized vermiculite (Demattê and Demattê, 1993; Cornu et al., 1998; Marques et al., 2002). Silica leaching is reduced by cycling of Si promoted by the forest vegetation, which contributes to the maintenance of kaolinite at the surface soils (Lucas et al., 1993; Rose et al., 1993; Lucas, 2001; Kleber et al., 2007).

Soils with a low degree of flocculation in areas of lower rainfall (semi-humid and semi-arid) indicate younger Ultisols. Despite the role of Al^{3+} in flocculation phenomena (Goldberg et al., 1990; Itami and Kyuma, 1995; Pedrotti et al., 2003), soils of the basement complex (BC) showed a higher flocculation with low levels of Al^{3+}, indicating an oxidic mineralogy and high weathering. This result is related more with the amphoteric character of Al and Fe-oxides than with the Al^{3+} content (Andrade et al., 1997). In addition, the association between Fe/Al oxides and kaolinite generates soil with a neutral reaction between the high pH of zero-point charge of oxides (Fontes et al., 2001). Moreover, the reduction in the Al^{3+} activity is related to organic matter complexation, which is not evaluated here (Lima and Anderson, 1997; Moreira and Costa, 2004).

In sedimentary domains, Al^{3+} contents were high, but with a lower flocculation, which can be explained by the kaolinitic mineralogy, which provides less stable aggregates (Ferreira et al., 1999; Schaefer, 2001). In addition, clayey and less weathered soils, for example, under Andean influence and in regions of lower rainfall, have higher levels of water-dispersible clay, and a lower degree of flocculation in the Bt horizon (Muller et al., 2001).

The variable ΔpH [pH(H_2O) - pH(KCl)] demonstrated the predominance of surfaces of negative charge in the Bt horizons of the Ultisols of the Legal Amazon. The low CEC and the negative ΔpH are characteristic of soils with kaolinitic mineralogy (Mekaru and Uehara, 1972; Benedetti et al. 2011). However, when compared by climatic zones, it has been observed that the maintenance of negative charge in soils has a close relationship with the lower rainfall; there is a direct association among more electronegative Bt horizons in arid areas and semi-humid rainfall zones

The areas with ΔpH with positive trends are also associated with the regions to the south of the Legal Amazon, which coincide with areas with a predominance of the Cerrado (Brazilian savanna), and this fact can be associated with the predominant mineralogy of the soils of this region, which have oxides of Al and Fe in crystalline form (gibbsite, hematite, and goethite) that contribute to low values of Ki index and characterize the soils as electropositive (Demattê and Demattê, 1993; Motta et al., 2002). In addition, although the pH of some soils from the Cerrado may be around 6.0, the levels of SB (Ca^{2+} and Mg^{2+}) mainly in the lower layers are low, and thus the values of sum of bases for the same region were low.

In relation to the levels of carbon in the Ultisols, these were low in all areas and the highest levels occur in the BC domain, which features rocks related mainly to the crystalline basement, which develop soils with higher clay contents. Studies in the Amazon region corroborate these results and describe positive correlations among the content of clay, organic matter, and biomass (Laurance et al., 1999; McGrath et al., 2001; Telles et al., 2003; Novaes Filho et al., 2007).

Although Amazon soils can have considerable amounts of organic carbon (Neill et al., 1997), it is often concentrated at the superficial layer, and the authors Batjes and Dijkshoorn (1999) argue that approximately 50 % of the pool of organic carbon is maintained within 0.3 m depth. Besides, in the Ultisols the incorporation of organic carbon into the subsurface is limited by the Bt horizon. However, some areas identified with higher levels of carbon coincide with the mapping performed by Schaefer et al. (2008), who observed stock C that was higher for the region of the state of Acre and Rio Negro, with lower values in the central part of the Amazon valley.

In relation to the climatic zones, areas with higher rainfall showed greater amounts of carbon, because they have a greater intensity of organic matter mineralization, a process that facilitates the illuviation of C and incorporation into sub-horizons (Tognon et al., 1998). In zones considered to be semi-arid, the amounts of carbon are even lower, since this transitional zone (Ab'Sáber, 1970) has high seasonality in litter levels deposited in the soil, with a peak during the dry season (Silva et al., 2007; Sanches et al., 2009). These results highlight that the litter production and amounts of soil carbon are closely associated with temperature and rainfall, and consequently this trend is manifested in the macro scale, where low rainfall is accompanied by low levels of carbon in the soil.

CONCLUSIONS

A large database of soils was effectively treated by multivariate analysis, which identified the variables that best explain the properties of the Bt horizons of the Ultisols of the Legal Amazon (SB, Al^{3+}, Floc, ΔpH, and organic carbon).

The machine learning algorithm (Extremely randomized trees) led to the spatial prediction of properties of the Ultisols, which was based on 70 covariables (predictors). The prediction was consistently more accurate for the property Al^{3+} and less efficient for organic carbon. The predictors that best explain the distribution were associated with climate and landforms. We assume that the predictions can be improved by the inclusion of covariates linked more strongly to soils, and the results are improved when working with higher resolution data. However, the results of this research were considered satisfactory, due to the high variability of the environments of the Legal Amazon.

The climatic factor is most influential on the Ultisols properties. Greater rainfall leads to lower SB, higher Al^{3+}, and higher degree of flocculation. The separation by morphostructural domains indicates that the soils developed on crystalline rocks of the basement complex have a frequency of Bt horizons with higher SB, organic carbon accumulation, electronegativity, and lower content of Al^{3+}.

Although there is a domain of low fertility Ultisols, those under the influence of Andean sediments in the western part have a higher sum of bases, despite the elevated levels of Al^{3+} and high rainfall.

The results of the machine learning technique presented here help to clarify the relationships between soils, climate and landforms at the macro scale. Future work with other classes of soils may be conducted regionally and will certainly contribute to the mapping at more detailed scales, meeting a growing need for environmental, urban, and agricultural planning.

REFERENCES

Ab'Saber AN. Bases para o estudo dos ecossistemas da Amazônia brasileira. Estud Av. 2002;16:7-30. https://doi.org/10.1590/S0103-40142002000200002

Ab'Sáber AN. Províncias geológicas e domínios morfoclimáticos no Brasil. São Paulo: Instituto de Geografia da USP; 1970.

Abdi H, Williams LJ. Principal component analysis. Computation Stat. 2010;2:433-59. https://doi.org/10.1002/wics.101

Almeida CA, Valeriano DM, Escada MIS, Rennó CD. Estimativa de área de vegetação secundária na Amazônia Legal Brasileira. Acta Amaz. 2010;40:289-302. https://doi.org/10.1590/S0044-59672010000200007

Andrade H, Schaefer CEGR, Demattê JLI, Andrade FV. Pedogeomorfologia e micropedologia de uma sequência Latossolo-Areia Quartzosa hidromórfica sobre rochas cristalinas do Estado do Amazonas. Revista Geonomos. 1997;5:55-66. https://doi.org/10.18285/geonomos.v5i1.189

Batjes NH, Dijkshoorn JA. Carbon and nitrogen stocks in the soils of the Amazon Region. Geoderma. 1999;89:273-86. https://doi.org/10.1016/S0016-7061(98)00086-X

Behrens T, Förster H, Scholten T, Steinrücken U, Spies E-D, Goldschmitt M. Digital soil mapping using artificial neural networks. J Plant Nutr Soil Sc. 2005;168:21-33. https://doi.org/10.1002/jpln.200421414

Benedetti UG, Vale Júnior JF, Schaefer CEGR, Melo VF, Uchôa SCP. Gênese, química e mineralogia de solos derivados de sedimentos Pliopleistocênicos e de rochas vulcânicas básicas em Roraima, Norte Amazônico. Rev Bras Cienc Solo. 2011;35:299-312. https://doi.org/10.1590/S0100-06832011000200002

Bernoux M, Cerri C, Arrouays D, Jolivet C, Volkoff B. Bulk densities of Brazilian Amazon soils related to other soil properties. Soil Sci Soc Am J. 1998;62:743-9. https://doi.org/10.2136/sssaj1998.03615995006200030029x

Brasil. Departamento Nacional de Produção Mineral. Projeto Radambrasil. Folhas da Amazônia Legal: Geologia, geomorfologia, solos, vegetação e uso potencial da terra. Rio de Janeiro; 1973 (Levantamento dos recursos naturais).

Brenning A. Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. In: Böhner J, Blaschke T, Montanarella L, editors. SAGA–Seconds out. Hamburger: Beiträge zur Physischen Geographie und Landschaftsökologie; 2008. p. 23-32.

Brungard CW, Boettinger JL, Duniway MC, Wills SA, Edwards Jr TC. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma. 2015;239-40:68-83. https://doi.org/10.1016/j.geoderma.2014.09.019

Bui EN, Moran CJ. A strategy to fill gaps in soil survey over large spatial extents: an example from the Murray–Darling basin of Australia. Geoderma. 2003;111:21-44. https://doi.org/10.1016/S0016-7061(02)00238-0

Carvalho Junior WD, Schaefer CEGR, Chagas CS, Fernandes Filho EI. Análise multivariada de Argissolos da faixa atlântica brasileira. Rev Bras Cienc Solo. 2008;32:2081-90. https://doi.org/10.1590/S0100-06832008000500029

Chauvel A. Contribuição para o estudo da evolução dos Latossolos amarelos, distróficos, argilosos na borda do platô, na região de Manaus: mecanismos da gibbsitização. Acta Amazon. 1981;11:227-45. https://doi.org/10.1590/1809-43921981112227

Corá JE, Araujo AV, Pereira GT, Beraldo JMG. Variabilidade espacial de atributos do solo para adoção do sistema de agricultura de precisão na cultura de cana-de-açúcar. Rev Bras Cienc Solo. 2004;28:1013-21. https://doi.org/10.1590/S0100-06832004000600010

Cornu S, Lucas Y, Ambrosi JP, Desjardins T. Transfer of dissolved Al, Fe and Si in two Amazonian forest environments in Brazil. Eur J Soil Sci. 1998;49:377-84. https://doi.org/10.1046/j.1365-2389.1998.4930377.x

Costa J, Hasui Y, Costa ML, Angélica RS. Evolução geológica da Amazônia. In: Costa ML. Angélica RS, editors. Contribuições à geologia da Amazônia. Belém: FINEP/SBG-Núcleo Norte; 1997. p. 16-90.

Curi N, Franzmeier DP. Toposequence of Oxisols from the Central Plateau of Brazil. Soil Sci Soc Am J. 1984;48:341-6. https://doi.org/10.2136/sssaj1984.03615995004800020024x

Delarmelinda EA, Souza Júnior VS, Wadt PGS, Deng Y, Campos MCC, Câmara ERG. Soil-landscape relationship in a chronosequence of the middle Madeira River in southwestern Amazon, Brazil. Catena. 2017;149:199-208. https://doi.org/10.1016/j.catena.2016.09.021

Demattê JLI, Demattê JAM. Comparações entre as propriedades químicas de solos das regiões da Floresta Amazônica e do cerrado do Brasil Central. Sci Agric. 1993;50:272-86. https://doi.org/10.1590/S0103-90161993000200016

Devaney J, Barrett B, Barrett F, Redmond J, O`Halloran J. Forest cover estimation in Ireland using radar remote sensing: a comparative analysis of forest cover assessment methodologies. PLoS One. 2015;10:e0133583. https://doi.org/10.1371/journal.pone.0133583

Donagema GK, Campos DVB, Calderano SB, Teixeira WG, Viana JHM. Manual de métodos de análises de solos. 2a ed. Rio de Janeiro: Embrapa Solos; 2011.

Drake JM, Randin C, Guisan A. Modelling ecological niches with support vector machines. J Appl Ecol. 2006;43:424-32. https://doi.org/10.1111/j.1365-2664.2006.01141.x

Feitosa KKA, Vale Júnior JF, Schaefer CEGR, Sousa MIL, Nascimento PPRR. Relações solo-vegetação em "ilhas" florestais e savanas adjacentes, no nordeste de Roraima. Cienc Florest. 2016;26:135-46. https://doi.org/10.5902/1980509821098

Ferreira MM, Fernandes B, Curi N. Influência da mineralogia da fração argila nas propriedades físicas de Latossolos da região sudeste do Brasil. Rev Bras Cienc Solo. 1999;23:515-24. https://doi.org/10.1590/S0100-06831999000300004

Fontes MPF, Camargo OD, Sposito G. Eletroquímica das partículas coloidais e sua relação com a mineralogia de solos altamente intemperizados. Sci Agric. 2001;58:627-46. https://doi.org/10.1590/S0103-90162001000300029

Forkuor G, Hounkpatin OKL, Welp G, Thiel M. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. PLoS One. 2017;12:e0170478. https://doi.org/10.1371/journal.pone.0170478

Gaillardet J, Dupré B, Allègre CJ, Négrel P. Chemical and physical denudation in the Amazon River Basin. Chem Geol. 1997;142:141-73. https://doi.org/10.1016/S0009-2541(97)00074-0

Gama J, Kusaba T, Ota T, Amano Y. Influência de material vulcânico em alguns solos do Estado do Acre. Rev Bras Cienc Solo. 1992;16:103-6.

Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn. 2006;63:3-42. https://doi.org/10.1007/s10994-006-6226-1

Goldberg S, Kapoor BS, Rhoades JD. Effect of aluminum and iron oxides and organic matter on flocculation and dispersion of arid zone soils. Soil Sci. 1990;150:588-93. https://doi.org/10.1097/00010694-199009000-00004

Gomes JBV, Curi N, Motta PEF, Ker JC, Marques JJGSM, Schulze DG. Análise de componentes principais de atributos físicos, químicos e mineralógicos de solos do bioma cerrado. Rev Bras Cienc Solo. 2004;28:137-53. https://doi.org/10.1590/S0100-06832004000100014

Grimm R, Behrens T, Märker M, Elsenbeer H. Soil organic carbon concentrations and stocks on Barro Colorado Island - Digital soil mapping using Random Forests analysis. Geoderma. 2008;146:102-13. https://doi.org/10.1016/j.geoderma.2008.05.008

Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn. 2002;46:389-422. https://doi.org/10.1023/a:1012487302797

Guyot JL, Jouanneau JM, Soares L, Boaventura GR, Maillet N, Lagane C. Clay mineral composition of river sediments in the Amazon Basin. Catena. 2007;71:340-56. https://doi.org/10.1016/j.catena.2007.02.002

Hartemink AE, McBratney A, Mendonça-Santos ML. Digital soil mapping with limited data. Dordrecht: Springer Science; 2008.

Hengl T, Bajat B, Blagojević D, Reuter HI. Geostatistical modeling of topography using auxiliary maps. Comput Geosci. 2008;34:1886-99. https://doi.org/10.1016/j.cageo.2008.01.005

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol. 2005;25:1965-78. https://doi.org/10.1002/joc.1276

Instituto Brasileiro de Geografia e Estatística - IBGE. Bases cartográficas e referenciais do território; 2017 [cited 2017 Nov 15]. Available from: https://mapas.ibge.gov.br/bases-e-referenciais.

Itami K, Kyuma K. Dispersion behavior of soils from reclaimed lands with poor soil physical properties and their characteristics with special reference to clay mineralogy. Soil Sci Plant Nutr. 1995;41:45-54. https://doi.org/10.1080/00380768.1995.10419557

Kaiser HF. The application of electronic computers to factor analysis. Educ Psychol Meas. 1960;20:141-51. https://doi.org/10.1177/001316446002000116

Kleber M, Schwendenmann L, Veldkamp E, Rößner J, Jahn R. Halloysite versus gibbsite: silicon cycling as a pedogenetic process in two lowland neotropical rain forest soils of La Selva, Costa Rica. Geoderma. 2007;138:1-11. https://doi.org/10.1016/j.geoderma.2006.10.004

Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th international joint conference on Artificial intelligence; August 1995; Montreal. Montreal: Morgan Kaufmann Publishers; 1995. p. 1137-43.

Kovačević M, Bajat B, Gajić B. Soil type classification and estimation of soil properties using support vector machines. Geoderma. 2010;154:340-7. https://doi.org/10.1016/j.geoderma.2009.11.005

Kuhn M, Johnson K. Applied predictive modeling. New York: Springer; 2013.

Kumar S, Lal R, Liu D. A geographically weighted regression kriging approach for mapping soil organic carbon stock. Geoderma. 2012;189-90:627-34. https://doi.org/10.1016/j.geoderma.2012.05.022

Laurance WF, Fearnside PM, Laurance SG, Delamonica P, Lovejoy TE, Merona JMR, Chambers JQ, Gascon C. Relationship between soils and Amazon forest biomass: a landscape-scale study. Forest Ecol Manag. 1999;118:127-38. https://doi.org/10.1016/S0378-1127(98)00494-0

Lima HN, Mello JWV, Schaefer CEGR, Ker JC, Lima AMN. Mineralogia e química de três solos de uma topossequência da bacia sedimentar do Alto Solimões, Amazônia Ocidental. Rev Bras Cienc Solo. 2006;30:59-68. https://doi.org/10.1590/S0100-06832006000100007

Lima JM, Anderson SJ. Aggregation and aggregate size effects on extractable iron and aluminum in two Hapludoxs. Soil Sci Soc Am J. 1997;61:965-70. https://doi.org/10.2136/sssaj1997.03615995006100030036x

Lips JM, Duivenvoorden JF. Regional patterns of well drained upland soil differentiation in the middle Caquetá basin of Colombian Amazonia. Geoderma. 1996;72:219-57. https://doi.org/10.1016/0016-7061(96)00027-4

Lucas Y. The role of plants in controlling rates and products of weathering: importance of biological pumping. Annu Rev Earth Pl Sc. 2001;29:135-63. https://doi.org/10.1146/annurev.earth.29.1.135

Lucas Y, Luizao FJ, Chauvel A, Rouiller J, Nahon D. The relation between biological activity of the rain forest and mineral composition of soils. Science. 1993;260:521-3. https://doi.org/10.1126/science.260.5107.521

MacMillan RA, Shary PA. Landforms and landforms elements in geomorphometry. In: Hengl T, Reuter HI, editors. Geomorphometry: concepts, software, applications. Amsterdan: Elsevier; 2008. p. 227-54. https://doi.org/10.1016/S0166-2481(08)00009-3

Maier O, Schröder C, Forkert ND, Martinetz T, Handels H. Classifiers for ischemic stroke lesion segmentation: a comparison study. PLoS One. 2016;11:e0149828. https://doi.org/10.1371/journal.pone.0149828

Malone BP, McBratney AB, Minasny B, Laslett GM. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma. 2009;154:138-52. https://doi.org/10.1016/j.geoderma.2009.10.007

Marimon BS, Lima ES, Duarte TG, Chieregatto LC, Ratter JA. Observations on the vegetation of northeastern Mato Grosso, Brazil. IV. An analysis of the Cerrado - Amazonian Forest ecotone. Edinb J Bot. 2006;63:323-41. https://doi.org/10.1017/S0960428606000576

Marques JJ, Teixeira WG, Schulze DG, Curi N. Mineralogy of soils with unusually high exchangeable Al from the western Amazon Region. Clay Miner. 2002;37:651-61. https://doi.org/10.1180/0009855023740067

McClain ME, Naiman RJ. Andean influences on the biogeochemistry and ecology of the Amazon River. BioScience. 2008;58:325-38. https://doi.org/10.1641/B580408

McGrath DA, Smith CK, Gholz HL, Oliveira FDA. Effects of land-use change on soil nutrient dynamics in Amazônia. Ecosystems. 2001;4:625-45. https://doi.org/10.1007/s10021-001-0033-0

McKenzie NJ, Gallant JC. Digital soil mapping with improved environmental predictors and models of pedogenesis. In: Lagacherie P, McBratney AB, Voltz M, editors. Developments in soil science. Amsterdam: Elsevier; 2006. v. 31. p. 327-49.

Mekaru T, Uehara G. Anion adsorption in ferruginous tropical soils. Soil Sci Soc Am J. 1972;36:296-300. https://doi.org/10.2136/sssaj1972.03615995003600020027x

Melo VF, Schaefer CEGR, Fontes LEF, Chagas AC, Lemos Júnior JB, Andrade RP. Caracterização física, química e mineralógica de solos da colônia agrícola do Apiaú (Roraima, Amazônia), sob diferentes usos e após queima. Rev Bras Cienc Solo. 2006;30:1039-50. https://doi.org/10.1590/S0100-06832006000600013

Mendonça BF, Fernandes Filho EI, Schaefer CEGR, Simas FNB, Paula MD. Os solos das campinaranas na Amazônia brasileira: ecossistemas arenícolas oligotróficos. Cienc Florest. 2015;25:827-39. https://doi.org/10.5902/1980509820581

Moreira A, Costa DG. Dinâmica da matéria orgânica na recuperação de clareiras da floresta Amazônica. Pesq Agropec Bras. 2004;39:1013-9. https://doi.org/10.1590/S0100-204X2004001000009

Motta PEF, Carvalho Filho A, Ker JC, Pereira NR, Carvalho Junior W, Blancaneaux P. Relações solo-superfície geomórfica e evolução da paisagem em uma área do Planalto Central Brasileiro. Pesq Agropec Bras. 2002;37:869-78. https://doi.org/10.1590/S0100-204X2002000600017

Mulla D, McBratney AB. Soil spatial variability. In: Warrick AW, editor. Soil physics companion. Boca Raton: CRC Press; 2001. p. 343-77.

Muller MML, Guimarães MF, Desjardins T, Martins PFS. Degradação de pastagens na Região Amazônica: propriedades físicas do solo e crescimento de raízes. Pesq Agropec Bras. 2001;36:1409-18. https://doi.org/10.1590/S0100-204X2001001100012

Neill C, Melillo JM, Steudler PA, Cerri CC, De Moraes JFL, Piccolo MC, Brito M. Soil carbon and nitrogen stocks following forest clearing for pasture in the southwestern Brazilian Amazon. Ecol Appl. 1997;7:1216-25. https://doi.org/10.1890/1051-0761(1997)007[1216:SCANSF]2.0.CO;2

Nepstad D, Lefebvre P, Silva UL, Tomasella J, Schlesinger P, Solórzano L, Moutinho P, Ray D, Benito JG. Amazon drought and its implications for forest flammability and tree growth: a basin-wide analysis. Glob Change Biol. 2004;10:704-17. https://doi.org/10.1111/j.1529-8817.2003.00772.x

Novaes Filho JP, Selva EC, Couto EG, Lehmann J, Johnson MS, Riha SJ. Distribuição espacial de carbono em solo sob floresta primária na Amazônia meridional. Rev Arvore. 2007;31:83-92. https://doi.org/10.1590/S0100-67622007000100010

Olaya V, Conrad O. Geomorphometry in SAGA. In: Hengl T, Reuter HI, editors. Developments in soil science. Amsterdan: Elsevier; 2009. p. 293-308.

Oliveira Junior RC, Corrêa JRV. Solos do município de Alenquer, Estado do Pará. Belém: Embrapa Amazônia Oriental; 2002.

Park SJ, Burt TP. Identification and characterization of pedogeomorphological processes on a hillslope. Soil Sci Soc Am J. 2002;66:1897-910. https://doi.org/10.2136/sssaj2002.1897

Pedrotti A, Ferreira MM, Curi N, Silva MLN, Lima JM, Carvalho R. Relação entre atributos físicos, mineralogia da fração argila e formas de alumínio no solo. Rev Bras Cienc Solo. 2003;27:1-9. https://doi.org/10.1590/S0100-06832003000100001

Quesada CA, Lloyd J, Anderson LO, Fyllas NM, Schwarz M, Czimczik CI. Soils of Amazonia with particular reference to the RAINFOR sites. Biogeosciences. 2011;8:1415-40. https://doi.org/10.5194/bg-8-1415-2011

R Core Team. R language definition. Vienna, Austria: R Foundation for Statistical Computing; 2000.

Rodrigues T. Solos da Amazônia. In: Alvarez V VH, Fontes LEF, Fontes MPF, editores. O solo nos grandes dominios morfoclimaticos do Brasil e o desenvolvimento sustentado. Viçosa, MG: Sociedade Brasileira de Ciência do Solo/Universidade Federal de Viçosa; 1996. p. 19-60.

Rose AW, Kato T, Machesky ML. The significance of biogenic element cycling in ancient tropical soils. Chem Geol. 1993;107:401-3. https://doi.org/10.1016/0009-2541(93)90219-9

Sanches L, Valentini CMA, Biudes MS, Nogueira JS. Dinâmica sazonal da produção e decomposição de serrapilheira em floresta tropical de transição. Rev Bras Eng Agr Amb. 2009;13:183-9. https://doi.org/10.1590/S1415-43662009000200012

Sanchez PA. Properties and management of soils in the tropics. New York: Wiley; 1977.

Sanchez PA, Bandy DE, Villachica JH, Nicholaides JJ. Amazon basin soils: management for continuous crop production. Science. 1982;216:821-7. https://doi.org/10.1126/science.216.4548.821

Santos HG, Carvalho Junior W, Dart RO, Aglio MLD, Sousa JS, Pares JG, Fontana A, Martins ALS, Oliveira AP. O novo mapa de solos do Brasil: legenda atualizada. Rio de janeiro: Embrapa Solos; 2011 [cited 2018 Mar 18]. Available from: https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/920267.

Santos HG, Jacomine PKT, Anjos LHC, Oliveira VA, Oliveira JB, Coelho MR, Lumbreras JF, Cunha TJF. Sistema brasileiro de classificação de solos. 3. ed. rev. ampl. Rio de Janeiro: Embrapa Solos; 2013.

Schaefer CERG. Brazilian Latosols and their B horizon microstructure as long-term biotic constructs. Aust J Soil Res. 2001;39:909-26. https://doi.org/10.1071/SR00093

Schaefer CEGR, Amaral EF, Mendonça BAF, Oliveira H, Lani JL, Costa LM, Fernandes Filho EI. Soil and vegetation carbon stocks in Brazilian Western Amazonia: relationships and ecological implications for natural landscapes. Environ Monit Assess. 2008;140:279-89. https://doi.org/10.1007/s10661-007-9866-0

Schaefer CEGR, Lima HN, Teixeira WG, Vale Junior JF, Souza KW, Corrêia GR, Mendonça BAF, Amaral EF, Campos MCC, Ruivo MLP. Solos da região Amazônica. In: Curi N, Ker JC, Novais RF, Vidal-Torrado P, Schaefer CERG, editores. Pedologia - Solos dos biomas brasileiros. Viçosa, MG: Sociedade Brasileira de Ciência do Solo; 2017. p. 111-75.

Schaefer CERG, Lima HN, Vale Júnior JF, Mello JWV. Uso dos solos e alterações da paisagem na Amazônia: cenários e reflexões. Bol Mus Para Emílio Goeld, Cienc Terra. 2000;12:63-104.

Silva CJ, Sanches L, Bleich ME, Lobo FA, Nogueira JS. Produção de serrapilheira no Cerrado e Floresta de transição Amazônia-Cerrado do centro-oeste brasileiro. Acta Amazon. 2007;37:543-8. https://doi.org/10.1590/S0044-59672007000400009

Soil Survey Staff. Keys to soil taxonomy. 12th ed. Washington, DC: United States Department of Agriculture, Natural Resources Conservation Service; 2014.

Sombroek W. Amazon landforms and soils in relation to biological diversity. Acta Amazon. 2000;30:81-100. https://doi.org/10.1590/1809-43922000301100

Stockmann U, Minasny B, McBratney AB. How fast does soil grow? Geoderma. 2014;216:48-61. https://doi.org/10.1016/j.geoderma.2013.10.007

Tarr AB, Moore KJ, Burras CL, Bullock DG, Dixon PM. Improving map accuracy of soil variables using soil electrical conductivity as a covariate. Precis Agric. 2005;6:255-70. https://doi.org/10.1007/s11119-005-1385-9

Telles ECC, Camargo PB, Martinelli LA, Trumbore SE, Costa ES, Santos J, Higuchi N, Oliveira Junior RC. Influence of soil texture on carbon dynamics and storage potential in tropical forest soils of Amazonia. Global Biogeochem Cy. 2003;17:1040. https://doi.org/10.1029/2002GB001953

Tognon AA, Demattê JLI, Demattê JAM. Teor e distribuição da matéria orgânica em Latossolos das regiões da floresta amazônica e dos cerrados do Brasil central. Sci Agric. 1998;55:343-54. https://doi.org/10.1590/S0103-90161998000300001

Trangmar BB, Yost RS, Uehara G. Application of geostatistics to spatial studies of soil properties. In: Brady NC, editors. Advances in agronomy. Cambridge: Academic Press; 1986. p. 45-94.

Vale Júnior JF, Souza MIL, Nascimento PPR, Souza DLS. Solos da Amazônia: etnopedologia e desenvolvimento sustentável. Agro@mbiente On-line. 2011;5:158-65. https://doi.org/10.18227/1982-8470ragro.v5i2.562

Vaysse K, Lagacherie P. Evaluating Digital Soil Mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). Geoderma Regional. 2015;4:20-30. https://doi.org/10.1016/j.geodrs.2014.11.003

Volkoff B, Melfi AJ, Cerri CC. Solos Podzólicos e Cambissolos eutróficos do alto rio Purus (Estado do Acre). Rev Bras Cienc Solo. 1989;13:363-72.

Wehenkel L, Ernst D, Geurts P. Ensembles of extremely randomized trees and some generic applications. In: Proceedings of Robust Methods for Power System State Estimation and Load Forecasting; Maio 2006; Paris. Paris: RTE-VT workshop; 2006. p. 1-10.

Were K, Bui DT, Dick ØB, Singh BR. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecol Indic. 2015;52:394-403. https://doi.org/10.1016/j.ecolind.2014.12.028

Westaway R. Late Cenozoic sedimentary sequences in Acre state, southwestern Amazonia: Fluvial or tidal? Deductions from the IGCP 449 fieldtrip. J S Am Earth Sci. 2006;21:120-34. https://doi.org/10.1016/j.jsames.2005.08.004

Wilson JP, Gallant JC. Digital terrain analysis in terrain analysis. In: Wilson JP, Gallant JC, editors. Terrain analysis: principles and applications. Nova Jersey: John Wiley & Sons; 2000. p. 1-28.

Witten IH, Frank E, Hall MA, Pal CJ. Data Mining: Practical machine learning tools and techniques. 2nd ed. Amsterdam: Elsevier; 2016.

Yaalon DH. Climate, time and soil development. In: Wilding LP, Smeck NE, Hall GF, editors. Developments in soil science. Part A. Elsevier; 1983. v. 11. p. 223-51. https://doi.org/10.1016/S0166-2481(08)70603-2