

Multiscalar Geomorphometric Generalization for Soil-Landscape Modeling by Random Forest: A Case Study in the Eastern Amazon

Cauan Ferreira Araújo¹, Raimundo Cosme de Oliveira Junior², Troy Patrick Beldini¹

¹Federal University of the West of Pará, Santarém, Brazil

²Eastern Amazon Agroforestry Research Center, Belém, Brazil

Email: cauan.araujo@ufopa.edu.br

How to cite this paper: Araújo, C.F., de Oliveira Junior, R.C. and Beldini, T.P. (2021) Multiscalar Geomorphometric Generalization for Soil-Landscape Modeling by Random Forest: A Case Study in the Eastern Amazon. *Journal of Geographic Information System*, 13, 434-451.
<https://doi.org/10.4236/jgis.2021.134024>

Received: June 10, 2021

Accepted: July 27, 2021

Published: July 30, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Multiscalar topography influence on soil distribution has a complex pattern that is related to overlay of pedological processes which occurred at different times, and these driving forces are correlated with many geomorphologic scales. In this sense, the present study tested the hypothesis whether multiscale geomorphometric generalized covariables can improve pedometric modeling. To achieve this goal, this case study applied the Random Forest algorithm to a multiscale geomorphometric database to predict soil surface attributes. The study area is in Phanerozoic sedimentary basins, in the Alter do Chão geological formation, Eastern Amazon, Brazil. The multiscale geomorphometric generalization was applied at general and specific geomorphometric covariables, producing groups for each scale combination. The modeling was run using Random Forest for A-horizon thickness, pH, silt and sand content. For model evaluation, visual analysis of digital maps, metrics of forest structures and effect of variables on prediction were used. For evaluation of soil textural classifications, the confusion matrix with a Kappa index, and the user's and producer's accuracies were employed. The geomorphometry generalization tends to smooth curvatures and produces identifiable geomorphic representations at sub-watershed and watershed levels. The forest structures and effect of variables on prediction are in agreement with pedological knowledge. The multiscale geomorphometric generalized covariables improved accuracy metrics of soil surface texture classification, with the Kappa Index going from 43% to 62%. Therefore, it can be argued that topography influences soil distribution at combined coarser spatial scales and is able to predict soil particle size contents in the studied watershed. Future development of the multiscale geomorphometric generalization framework could include generalization methods concerning preservation of features,

landform classification adaptable at multiple scales.

Keywords

Digital Soil Mapping, Upscaling, Machine Learning, Random Forest Algorithm, Multiscale Geomorphometric Generalization

1. Introduction

Elements of the landscape control the processes acting on soils; therefore, the soil-landscape approach constitutes one of the most powerful conceptual tools in mapping activities, especially at scales with an intermediate or greater level of detail. The soil-landscape relationship is related to the concept of the catena, coined by Milne [1]. In a catena, variations in soils along a slope are attributed to the translocation of soluble elements and to erosive and depositional processes, not excluding different source materials. Subsequently, the analyses of soil-landscape relations proposed by Hugget [2], contemplated three-dimensional models of the slopes. In the context of digital soil mapping, in the scorpan model paradigm [3], soil-landscape process modeling can be described as an interdisciplinary object of the interface between pedometry-geomorphometry [4].

The scale issues of soil-landscape relationships are related to the complex interactions of both elements and how these processes occur and are perceived. Several geomorphology studies report a time-space coupling between landform size-scales and lifetime [5] [6]. From a soil perspective, different pedological process will manifest influence at short, mid or long time-scales [7] [8]. In contrast, topography studies in soil physics demonstrate a complex water dynamic related to the nested geometry of slopes, considering relief and micro-relief patterns, resulting in trends in the movement of particles and solutes and changes in texture and chemical parameters of soils [9] [10]. Therefore, multiscalar topography influences a particular soil distribution in two general aspects, overlay of pedological processes that occurred at different times, and driving forces in the present time, determined by the sum of forces better correlated with one, several, or many geomorphologic scales.

Some aspects of spatial scaling in digital soil mapping have been summarized in non-exhaustive reviews [11] [12]. The hierarchical definition of scale can be used to understand the soil phenomena, from the soil region, passing through watersheds, catena, pedon, horizons and finally molecular interactions. The characteristics of measurement affect the results of analysis, and in this sense, the modifiable areal unit problem (MAUP) represents a key issue. Information transfer across scales could be classified in upscaling, in less detail, or downscaling, with greater detail, but both of these require must consider bias.

With respect to the scale of covariables for predictive soil mapping, the highest DEM resolutions do not necessarily produce the highest accuracy [13] [14]. Despite the potential of machine learning to produce complex and non-linear

predictions, few studies have investigated multiscale perspective in covariables to account for physical process that are not predictable by finer scale environmental information [15]. Some studies propose data-driven techniques for selection of pixel size or neighborhood size for a particular landscape [16] [17], but this approach produces a variety of results in different geomorphic units, which complicate the interpretation of scalar components in the soil-landscape relationship. For a friendly interpretation of scale relationships on soil-landscape models, this study proposed a cartographic-based criterion to formalize the scale correspondence to pixel size for geomorphometric covariables.

The present study tested the hypothesis whether multiscale geomorphic representation, obtained from cartographic generalization of a digital elevation model, can improve pedometric modeling. To achieve this goal, this case study applied the Random Forest algorithm to a multiscale geomorphometric database to predict soil surface attributes.

2. Material and Methods

The procedures described in this section were performed using the open-source software QGIS 3.10; SAGA GIS 2.3; GRASS GIS 7.8; and R Programming 3.5 [18] [19] [20] [21].

2.1. Study Area

The study was conducted in Iripixi Lake (ILW) and Caipuru Lake (CLW) watersheds, with an area of 27.137 and 28.315 ha respectively, located in the Trombetas basin in Oriximiná-Pará in the Eastern Amazon. Pilot areas, consisting of small farms, are distributed in the upper and lower courses in both basins and adjacent basin boundaries, as shown in **Figure 1**, totaling 697 ha, approximately 1% of the extent of the watersheds.

The study area is in Phanerozoic sedimentary basins, in the Alter do Chão geological formation, whose local characteristics were evaluated from a faciological analysis near Óbidos [22], a neighboring municipality. This region is classified as humid equatorial climate with 3 dry months [23]. The geomorphic units of these watersheds are classified as a homogeneous dissection with coarse drainage density and weak incision depth [24]. The pedoenvironment of this study area are in the upper lands of the lower Amazon basin and the most abundant soil classes in the area are *Latossolo Amarelo* (Ferrasols), *Latossolo Vermelho-Amarelo* (Ferrasols), *Argissolo Vermelho-Amarelo* (Acrisols) and *Gleissolos Haplicos* (Gleysols) [25].

2.2. Environmental Covariates

In the proposed mapping scale, vegetation and topography factors are the main sources of soil variation. In this study, to represent such factors, the source multispectral images Landsat 8 and Shuttle Radar Topography Mission Digital Elevation Model 30 m (SRTM DEM 30 m), were used, respectively.

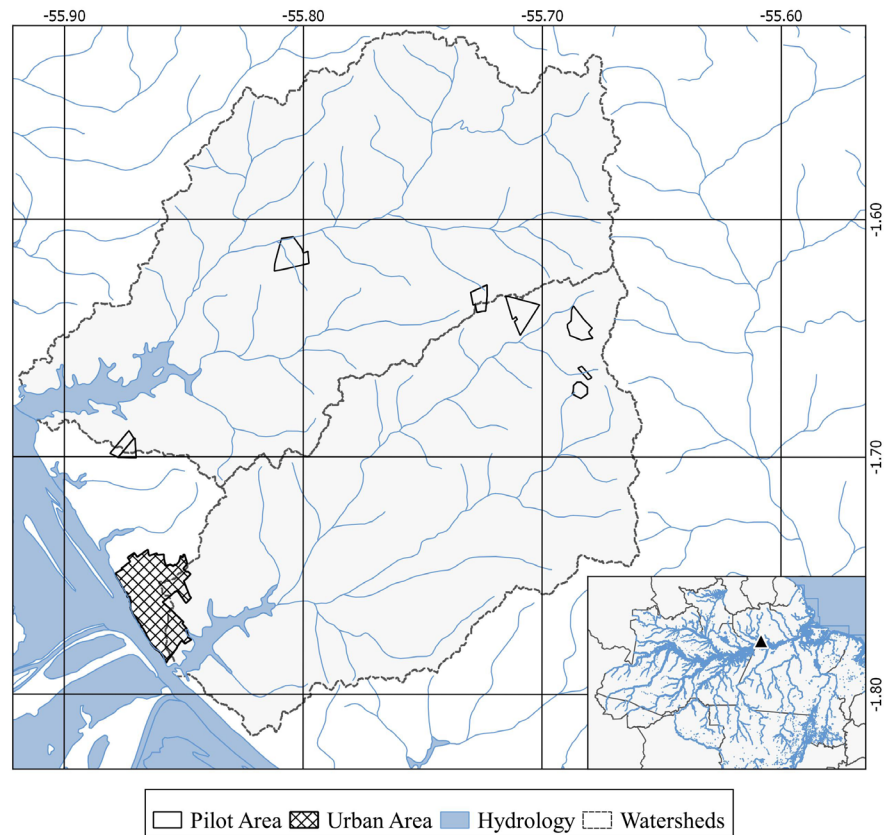


Figure 1. Study area location in the Eastern Amazon.

The Landsat 8 images are from September 11, 2017, corrected for surface reflectance with the LaSRC algorithm by USGS [26]. The SRTM DEM is a digital elevation model based on stereoscopic radar survey, and has 30m pixels [27]. The corrections made in SRTM DEM were filling in sinks, and reduction of deforestation effect by the estimated canopy addition method [28].

The topography information was upscaled and organized into generalized multiscale geomorphometric variable groups, as detailed in the next section.

2.3. Multiscale Geomorphometric Generalization

The multiscale geomorphometric generalization (MGG) is an upscaling operation, based on cartographic concepts of generalization of digital maps [29] [30]. This approach can be applied for any geomorphic variable, including elevation models, landforms units, and primary and secondary derivatives.

This operation results in variables at different scales, arranged in groups according to criteria required for the analysis. The framework of these upscaling methods brings to the pedometric perspective the understanding that the soil-landscape relationship occurs through complex and multiscale interactions. In this sense, the formalization of the desirable scales of analysis and modeling occurs both in their definition and in the group arrangements.

For the MGG operation, vector and raster representations demand different

approaches for upscaling, because each of them has specific scale transformation problems due to their mathematical structures. Furthermore, it is necessary to have a unique reference for the scales for compatible representation of geomorphic features in both types of variables, thus allowing for joint interpretation. In this study, the concept of minimum mappable area for soil surveys [31] was considered to define pixel sizes in relation to cartographic scale. The detailed descriptions for each of the four scales used are in **Table 1**. The area equivalence between raster and vector is calculated as a function of a 5×5 pixel grid, considered a conservative parameter to determine a geomorphic feature.

The MGG was applied to the following geomorphometric covariables: elevation (Elev), slope, relative slope position (RSP), topographic wetness index (TWI), plan curvature (PlanCurv), profile curvature (ProfCurv), topographic factor of water erosion (LS) and geomorphons.

These geomorphic variables, at different scales, were obtained from SRTM DEM from two upscaling methods, as illustrated in **Figure 2**. Using local averages on covariable elevation, in 2×2 , 3×3 , 4×4 windows, for resolutions 60 m, 90 m and 120 m, respectively, and subsequent derivatives covariable calculation. Classification of geomorphons [32] was followed by the exclusion of polygons smaller than the minimum mappable area for each scale. Such methods correspond

Table 1. Correspondence between scale and pixel size for Multiscale Geomorphometric Generalization (MGG), using the concept of minimum mappable area.

Scale	Minimal mappable area (m ²)	Pixel size (m)	Pixel area ^a (m ²)	pa/mma ^b (%)
1:25,000	25,000	30	22,500	90
1:50,000	100,000	60	90,000	90
1:75,000	225,000	90	202,500	90
1:100,000	400,000	120	360,000	90

^aFor a 5×5 window. ^bRatio pixel area (pa) by minimal mappable area (mma), in percentage.

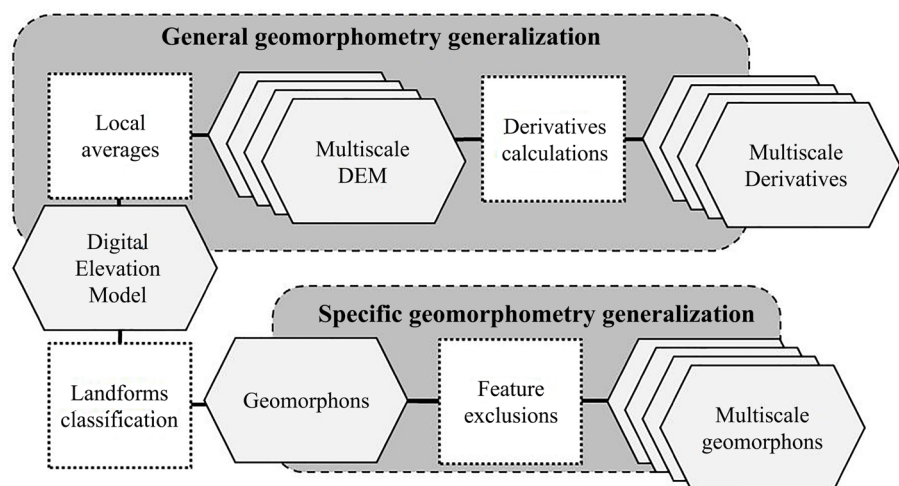


Figure 2. Methodology flowchart of MGG for the topography covariables.

to cartographic generalization applied to general geomorphometry and specific geomorphometry, respectively [33].

In this case study was used a machine learning approach to identify and select the optimum scales of variables for modeling. It was therefore necessary to provide a multiple database for training and evaluation each scale. In this sense, the variables were organized from the combination of the set of topography variables, arranged in all possible combinations.

2.4. Soil Sampling and Analysis

Soil sampling was performed in 9 pilot areas, considering covariables to evaluate the effect of topography related to variation in soil distributions. Each pilot area was sampled at 10 points, a sufficient density for semi-detailed soil surveys, compatible with the 1:25,000 scale soil maps [31]. The sample points were distributed according to a stratified random arrangement by the conditioned latin hypercube method [34] [35], with raster topography covariates, described in the previous section, at a 1:25,000 scale. At the total of 90 sample points, the morphological description of the A horizon was performed [36] and soil samples were collected in the 0 - 30 cm depth for physical and chemical analyses [37] [38]. To evaluate variances and patterns in the sample dataset, principal component analysis [39] was conducted.

2.5. Modeling by Random Forest

The modeling of the soil-landscape was done using Random Forest [40], a machine learning algorithm frequently used to produce digital soil maps [41]. Some characteristics of this algorithm that are worth mentioning are that it can handle categorical and continuous variables, it can do regression and classifications, it is robust for overfitting problems, and is feasible for interpretation of variable relationships, including linear and nonlinear systems [42].

First, the set of training cases and those intended for validation were defined. The selection was made randomly, with proportions of 70% and 30%, respectively. The training was done for every group of geomorphometric covariables, one group at a time, for modeling of some soil attributes, namely, A horizon thickness, pH, silt and sand content. Next, for the prediction of the multiscale generalized geomorphometric groups, the groups with the best adjustment were evaluated and selected based on the highest values of % variation explained by the model. Finally, the modeling structure and results of original scale geomorphometric and generalized geomorphometrics groups were compared. For evaluation of these predictions, visual analysis of digital maps and multi-way plots of forest structures and effect of variables on prediction were used, calculated with Random Forest Explainer in the R package [43].

The prediction of soil texture was made by considering the silt and sand raster layers using the Brazilian soil classification system [44]. For evaluation of this prediction, the confusion matrix was calculated with the Kappa index, and the user's and producer's accuracy [45].

3. Results and Discussion

3.1. Generalized Multiscale Geomorphometrics

The multiscale geomorphometric generalization produces a dataset of covariables at each scale. For evaluation of that operation, in this section, the tendency of geomorphometry distributions and a geomorphology interpretation of this generalized data will be described.

The application of MGG to the original DEM database resulted in 28 continuous geomorphometric variables, the distributions of which are illustrated in **Figure 3**. Some distributions had smaller changes, like TWI and LS, with a reduction of occurrence of the extreme values and some scatter reduction in the upscaling direction. This could be explained by the lowest representation of smaller geomorphic features that contributed to counting of upper and lower limits on dynamics of the water on the slope. Some variable distributions had highly significant changes, such as slope and geomorphic curvatures, which had a sharp and progressive reduction of scatter related to the smoothest generalized surface. Considering the PlanCurv, it is observable that the upscaled derivative better describes the features of the valleys and spikes in the study area, as illustrated in **Figure 4(b)**.

Some studies test upscaling effects on digital topographic information, with comparable results. The contextual spatial modeling, using gaussian space scale rates to produce a set of coarse resolution DEMs, had a similar result of smooth

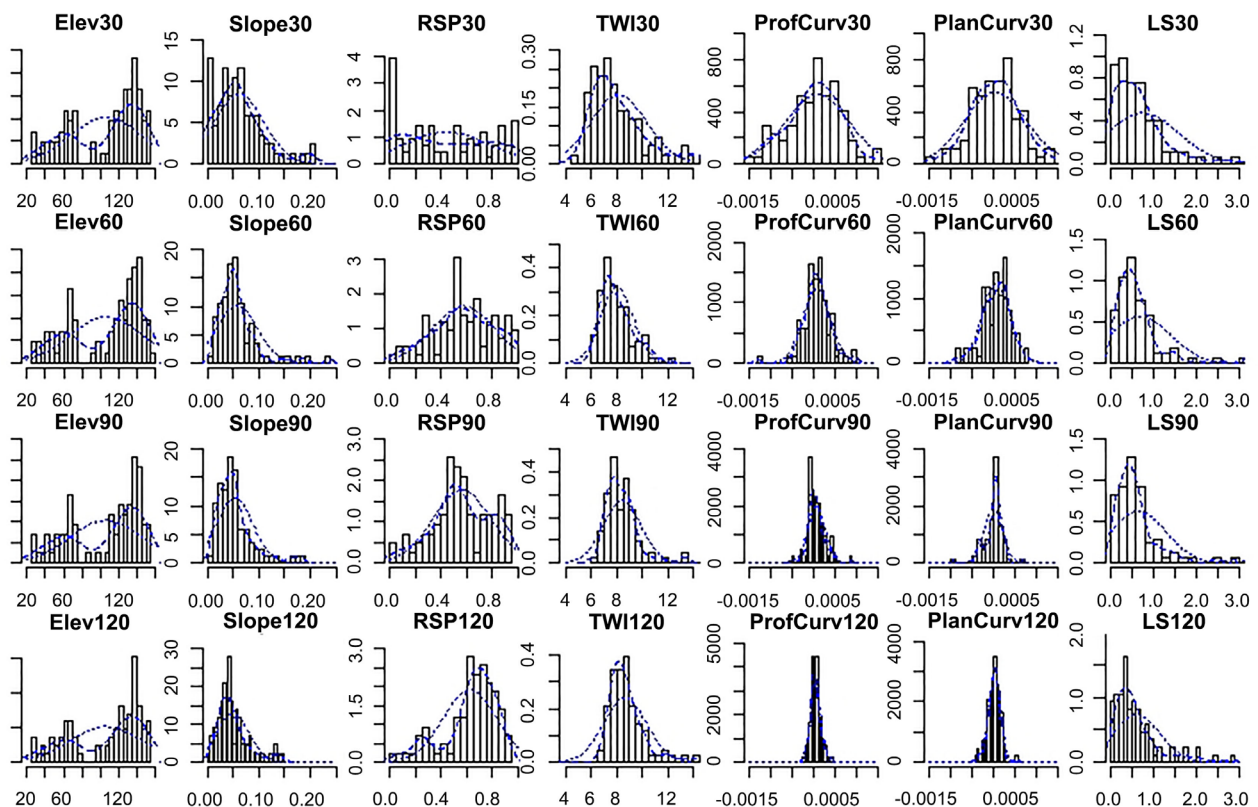


Figure 3. Distributions of geomorphometry raster variables at original and generalized scales.

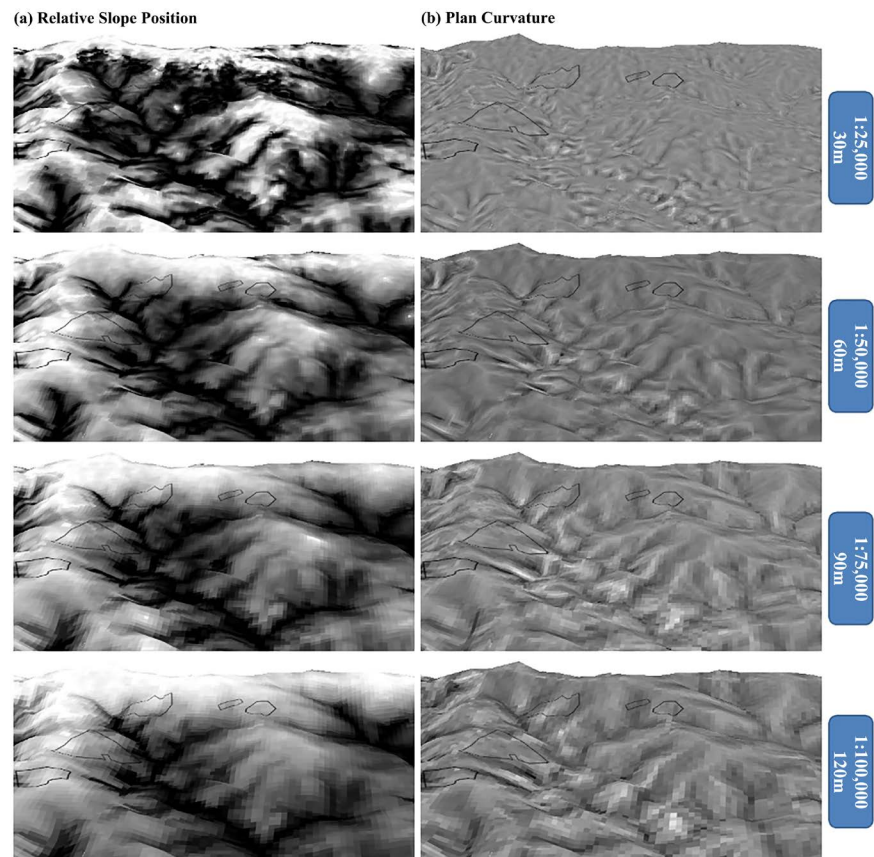


Figure 4. Comparative 3D view of original and generalized geomorphometric covariables: (a) Relative slope position; (b) Plan curvature.

slope and geomorphic curvatures [46]. In contrast, other studies have proposed sophisticated calculations for generalization of DEM considering questions of feature preservation, and could be tested with the MGG framework. The Feature Preserving DEM Smoothing (FPDEMS) method reduces the complexity of the surface at the detailed spatial scales at which roughness dominates, while not significantly altering the topographic complexity at larger spatial scales [47]. Other approaches use a multi-point algorithm to rapidly and accurately retrieve the critical points, using drainage-constrained TIN, to produce coarser-resolution DEMs [48] [49].

The distribution of relative slope positions had the most complex variation, and was related to dissection process acting on a flattened surface of phanerozoic sedimentary basins, which results in organization of this type of landform, a homogeneous dissection with tabular top, coarse drainage density and weak incision depth [24] [50]. A comparative 3D view of original and generalized relative slope positions is illustrated in **Figure 4(a)**. This variable at its original scale identifies the variation at a sub-watershed level, with 0 - 1 interval values that occur at each drainage division. In contrast, at 1:100,000, this variable shows variation at the watershed level, with high values only at the most elevated ridge-line section. In this perspective, the upscaled variable can identify upper and

lower courses of basins, a relevant information for soil mapping, because it allows analysis of base level, exposition time to weathering and stratigraphic differentiation on sedimentary geological formation [5].

The geomorphon scale transformations resulted in incorporation to the valley portions of some local landform segments which originally were classified in topo and slope, as illustrated in **Figures 5(a)-(c)**. Therefore, at a 1:100,000 scale, some geomorphic features were disconnected from the main patterns and were omitted and substituted by the local prevalent geomorphon type. Regarding the original scale geomorphons, the results show consistent valley-slope-top landforms sequences in the most of mapping site. However, at same scale of soil mapping in South Africa, the geomorphons were compared with expert manual delineation with a high degree of mismatch (57% of the area), despite reasonable prediction results [51]. Alternatively, considering the low importance of geomorphons in decision trees, as we will see in the next section, it is possible to test other approaches to generalization, including the variation of search parameters in the definition of relief units.

Other approaches to classification of landforms, like the topographic position index [52], and k-median clustering [53], also had parameters that could be adjusted for the proposed correspondence of scales. Therefore, these techniques can be included and tested with the MGG framework in future research.

3.2. Evaluation of Soil-Landscape Models

This section presents the relationship of soil variables to geomorphometric covariables, in the context of MGG and will focus on whether the decision tree forest structures have an expected pedological meaning, and if a greater number of topographic variables could mask other factors.

The main soil variation is related to particle size distribution, as shown by the contribution of this variable to the first component of PCA, as illustrated in **Figure 6**. The A horizon thickness has an opposite tendency with respect to silt,

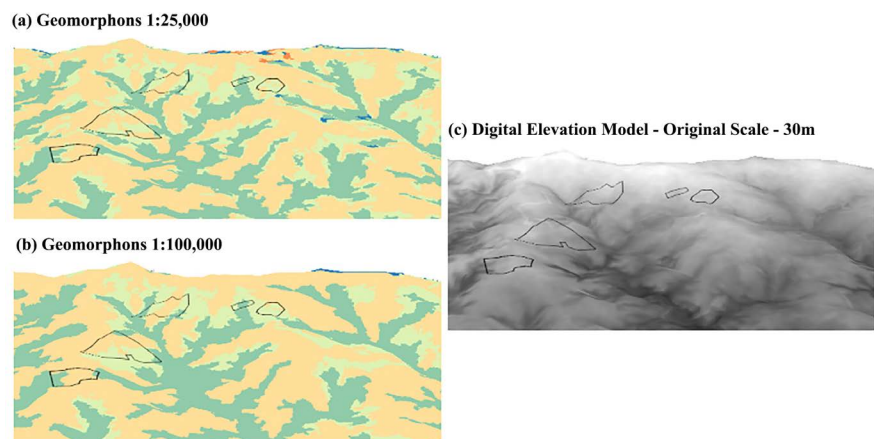


Figure 5. Comparative 3D view of original and generalized geomorphometric covariables: (a) Geomorphons 1:25,000; (b) Geomorphons 1:100,000; (c) Digital elevation model at original scale.

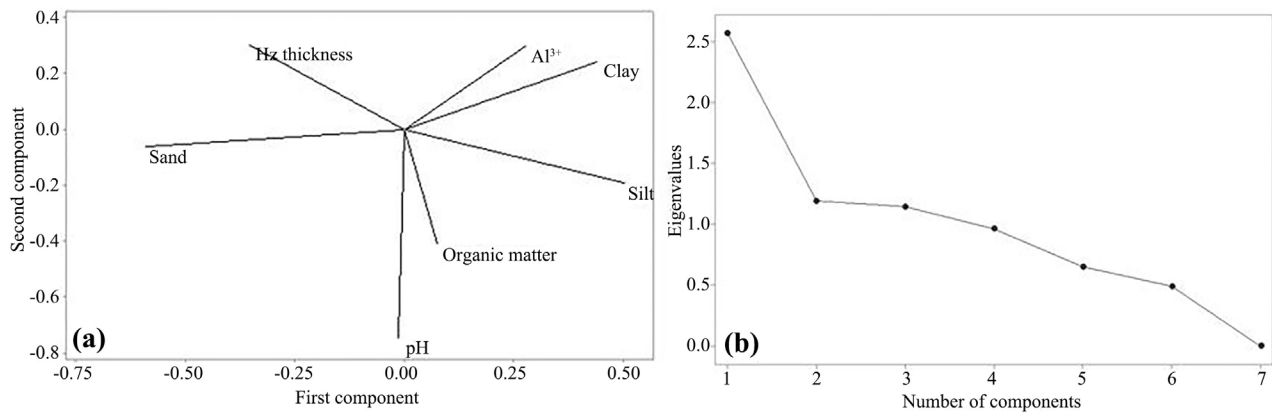


Figure 6. (a) Vectors and (b) variable loadings of principal components for soil attributes.

and is only slightly in agreement with sand content, revealing the weathering dynamics related to translocation of organic matter in sandy rainforest soils. Organic matter has a relevant influence on soil differentiation and is directly related to pH, as expected in this pedoenvironment, a upper land of the lower Amazon basin [25].

Also, exchangeable aluminum has only a slight agreement with clay content, and this is probably associated with clay mineralogy dominated by kaolinite, common in that fraction in acid soils [54]. Studies indicate that predominance of kaolinite at the surface soil is explained by reduced Silica leaching by cycling of Si promoted by the forest vegetation, and the resulting soil properties patterns was found on Amazon soils located on sedimentary domains with high exchangeable aluminum content, like this study area [55] [56].

The variation in the A horizon is strongly related to landcover and topographic factors [57]. For modeling the A horizon, the MGG group 15 had the best adjustment, with 40.5% of variation explained, in contrast to 30.2% for the original scale DEM, indicating the influence of coarse variables. This could mean that there is prevalence of older pedological processes associated with larger and older geomorphic features.

The decision tree forest structures, as illustrated in **Figure 7**, clearly show the main controllers of tendencies, represented by Elev and RSP, such as the high portion of watersheds that have conserved and currently present long-term pedological processes of soil development that have resulted in an overall greater thickness, in contrast with the dissected portion. The water content and dynamics, represented by TWI, PlanCurv, ProfCurv, influence the variation in the A horizon thickness, and landcover also has a significant importance in the model, represented by surface reflectance variables.

These results show similarity with another study that reported increased accuracy of soil prediction by Random Forest on four datasets across the globe with addition of a coarser scale DEM [58]. These authors also proposed an approach with a variogram of soil properties to a priori approximations of the effective scale for modelling. In the context of MGG, this can be used for scale definition and group arrangement for soil attributes on specific pedoenvironment.

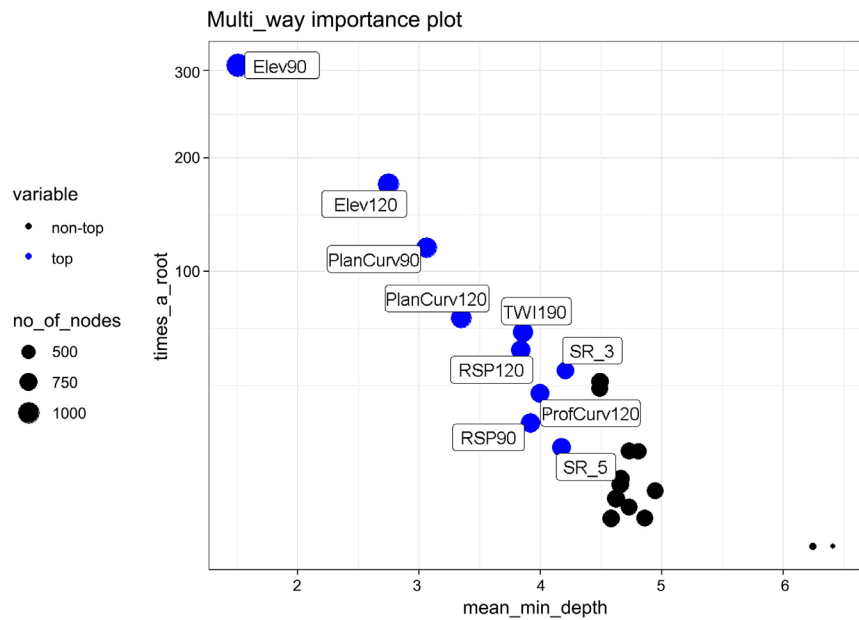


Figure 7. MGG decision tree forest structures for A horizon thickness, geomorphometric group 1:75,000 plus 1:100,000 scales.

Soil pH influences microbiota, nutrient uptake, root growth and therefore plant development in general [59]. Soil pH is strongly related to landuse, or in scorpan, the organism factor. As expected, the most important variable for modeling of this soil attribute, at the original scale, is the surface reflectance, as shown in **Figure 8(a)**. When considering the importance plot of the MGG group 22, illustrated in **Figure 8(b)**, the same five variables have with highest importance index. In this sense, both models show topographic factors with less importance than vegetation factors, in the same proportion, despite twice the number of geomorphometric variables in the MGG.

3.3. Prediction of Superficial Layer Soil Texture

This section discusses the Random Forest prediction of particle size, which is able to produce soil textural classifications for land users and stakeholders and will focus on the question of whether model adjustment can be improved by the MGG, and if the soil particle size maps result in a more accurate soil textural classification.

For prediction of particle size of surface layer, sand and silt fractions were used because of the low contribution of clays in the Alter do Chão lithology. This could be explained by local characteristics of that geological formation, with an overall fine to medium grained sandstone content in the upper portion, and medium and coarse sandstones with small contribution of red claystones in the lower portion [22].

The Random Forest models for silt and sand, at original scale geomorphometrics, had a poor adjustment, as shown in **Table 2**. When considering the best adjusted MGG groups, despite the considerable portion of randomness related

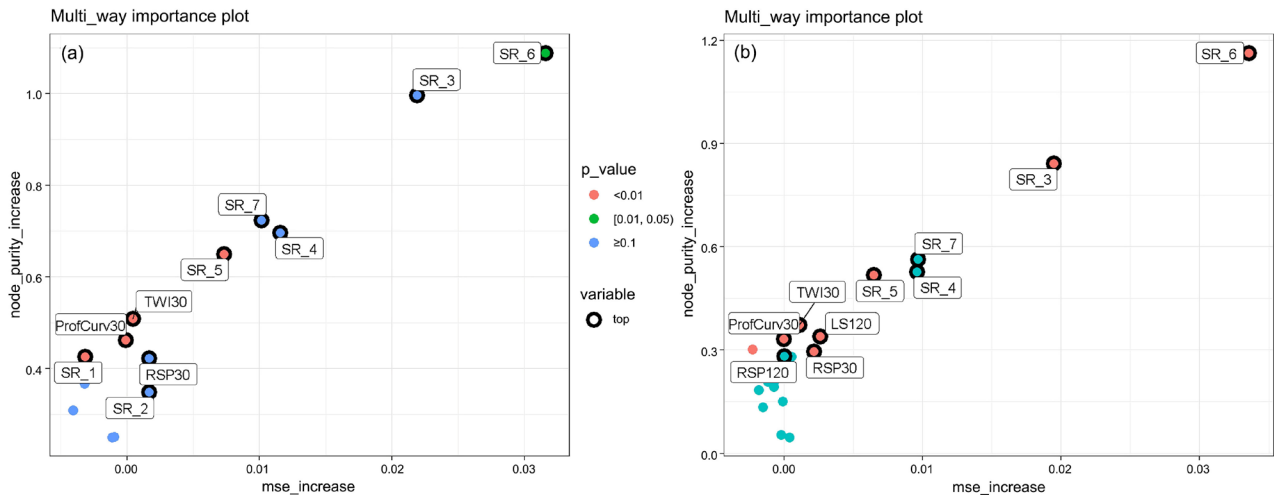


Figure 8. Variable prediction importance for pH at (a) original geomorphometric scale and (b) multiscale generalized geomorphometrics.

Table 2. Model adjustment for silt and sand.

Particle Size	Predictors	% Var explained	Mean of squared residuals
Silt	Original scale	3.82	0.0007345
	MGG 1:75,000 + 1:100,000	31.73	0.0000674
Sand	Original Scale	6.9	0.0023651
	MGG 1:25,000 + 1:75,000	32.43	0.0002224

to the heterogeneity of soils, the % variation explained is reasonably higher and mean of squared residuals is reasonably lower. The model's predictors have a principal contribution from variables at 1:75,000 and 1:100,000 and can identify tendencies at the watershed scale.

The most significant covariables of particle size prediction are shown in **Figure 9(a)**, **Figure 9(b)**: Elev and RSP at coarser scales, associated with stratigraphy and long-term hillslope transportation; ProfCurv at coarser scales, and PlanCurv at original and coarser scales, related with accumulation, transit and dissipation zones [60].

The prediction of sand and silt content, at original and multiscale generalized geomorphometrics, is illustrated in **Figure 10**. In both variable groups, the MGG has produced maps with less noise and more recognizable patterns related to geomorphic features. These results corroborate the hypothesis that the topography has an influence, in a larger spatial context, and has prevalence on prediction of soil particle size contents in the tested basin. In contrast, a case study with Random Forest with 30 m and 90 m DEM did not achieve significant differences in prediction [61]. Despite some similarity with covariables importance, like Elev and RSP, the modeling is done on single scale datasets. In this sense, we can argue the importance of observing soil-landscape phenomena from a multiscale perspective.

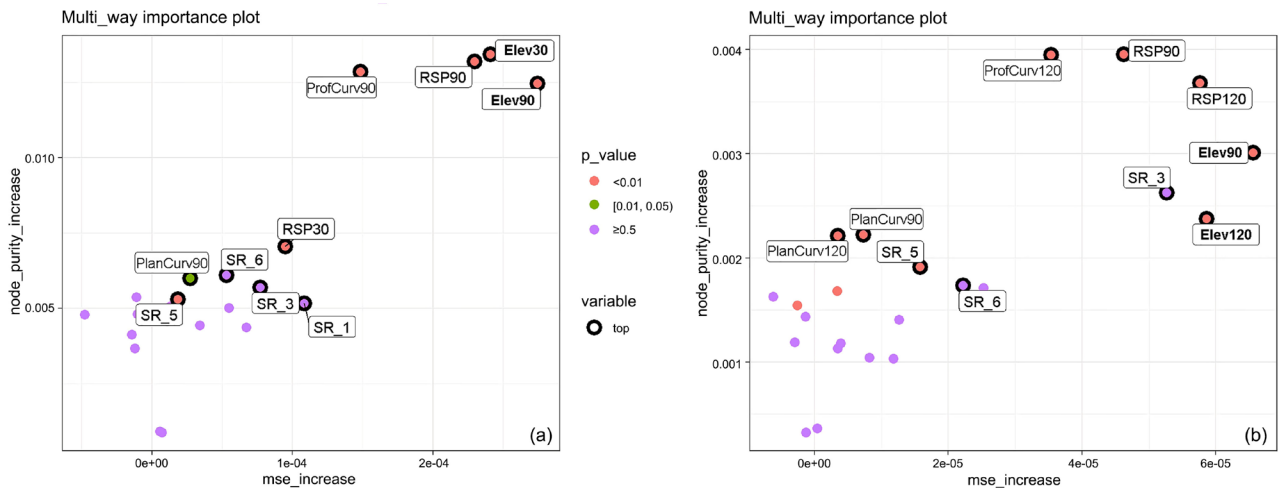


Figure 9. Variable prediction importance: (a) model of sand by geomorphometric group 1:25,000 plus 1:75,000; (b) model of silt by geomorphometric group 1:75,000 plus 1:100,000.

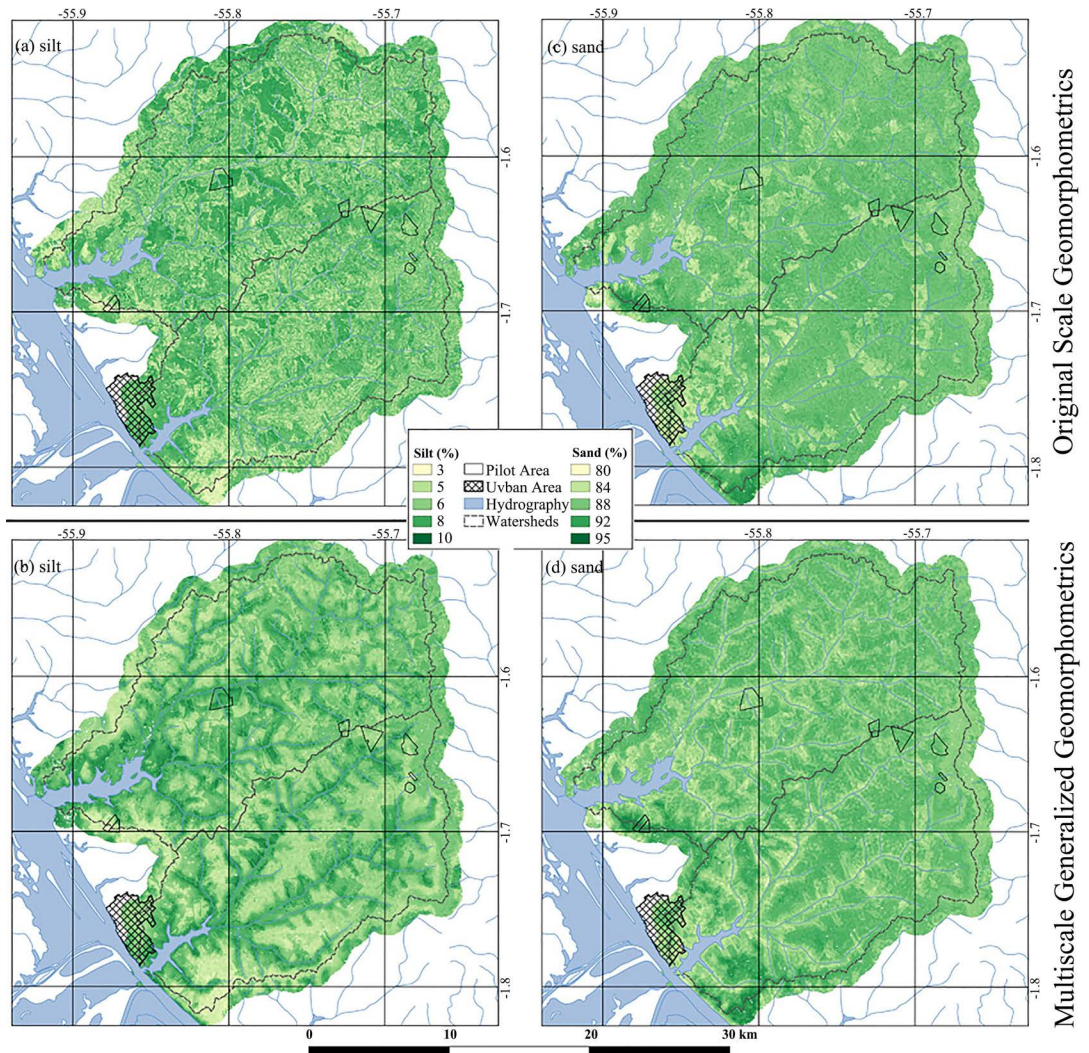


Figure 10. Predictive maps of silt (a, b) and sand (c, d) at original scale geomorphometrics and multiscale generalized geomorphometrics, respectively.

Table 3. Accuracy evaluation for soil texture classification.

Geomorphometric variables	User's/Producer's accuracy			Kappa Index
	MAr	ArMe	MeAr	All Classes
Original Scale	75%/84%	72%/58%	20%/20%	43%
MGG	81%/88%	76%/71%	100%/67%	62%

The MGG was able to increase the accuracy of superficial layer soil texture classifications, as shown in **Table 3**. The most significant improvements occur in the franco-arenosa (MeAr) and areia franca (ArMe) soil texture classes, both with a smaller contribution area at the mapping site in relation to the more relevant areia (MAr) class. Also, the user's accuracy has a considerably higher result, so the MGG increased the reliability of each mapped class. In the same way, the Kappa index also has higher values for MGG geomorphometric variables.

A case study done on a farm in China, with machine learning on single and multiple scales variables [62], also found better results with a range of appropriate scales, even using only local derivatives. In this sense, the MGG framework has greater potential because DEM transformation before derivate calculation allows for use of both local and regional derivatives in a multiscale group arrangement.

4. Conclusions

This study evaluates the multiscale geomorphometric generalization with the purpose of modeling the soil-landscape relationship. The comparison of original scale with multiscale generalized variables was based on Random Forest prediction of soil attributes. Regarding the proposed methodological framework, the following results and issues stand out from this case study:

- The general geomorphometry generalization tends to smooth slopes and curvatures and produce identifiable representations of relative slope position at sub-watershed and watershed level. The specific geomorphometry generalization results in incorporation into valley portions some local landform segments which originally were classified in topo and slope.
- Random Forest modeling with MGG variables did not mask greater importance of the vegetation factor for prediction of soil attributes related to land-cover. The forest structures and effect of variables on prediction agree with pedological knowledge. Comparing these results with those from other studies, it could be argued that the soil-landscape phenomena from a multiscale perspective are highly relevant.
- The MGG improved model adjustment for silt and sand particles and also improved the accuracy of metrics of soil texture classification of surface layer, especially for the most unusual classes, with the Kappa Index going from 43% to 62%. Topography influences at a coarser spatial scale and has prevalence on prediction of soil particle size contents in the studied watershed.

- Future development of the MGG framework should address generalization of DEM concerning feature preservation and comparison of landform classification adaptable at multiple scales.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Milne, G. (1935) Some Suggested Units of Classification and Mapping Particularly for East African Soils. *Soil Research*, **4**, 183-198.
- [2] Hugget, R.J. (1975) Soil Landscape Systems: A Model of Soil Genesis. *Geoderma*, **13**, 1-22. [https://doi.org/10.1016/0016-7061\(75\)90035-X](https://doi.org/10.1016/0016-7061(75)90035-X)
- [3] McBratney, A.B., Mendonça Santos, M.L. and Minasny, B. (2003) On Digital Soil Mapping. *Geoderma*, **117**, 3-52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- [4] Ma, Y., Minasny, B., Malone, B.P. and Mcbratney, A.B. (2019) Pedology and Digital Soil Mapping (DSM). *European Journal of Soil Science*, **70**, 216-235. <https://doi.org/10.1111/ejss.12790>
- [5] Wysocki, D.A. and Schoeneberger, P.J. (2011) Geomorphology of Soil Landscapes. In: Huang, P.M., Li, Y. and Sumner, M.E., Eds., *Handbook of Soil Science. Properties and Processes*, Chemical Rubber Company Press, Florida, 1-26.
- [6] Schmidt, J. and Andrew, R. (2005) Multi-Scale Landform Characterization. *Area*, **37**, 341-350. <https://doi.org/10.1111/j.1475-4762.2005.00638.x>
- [7] Targulian, V.O. and Krasilnikov, P.V. (2007) Soil System and Pedogenic Processes: Self-Organization, Time Scales, and Environmental Significance. *Catena*, **71**, 373-381. <https://doi.org/10.1016/j.catena.2007.03.007>
- [8] Kämpf, N. and Curi, N. (2012) Formação e evolução do solo (pedogênese). In: Ker, J.C., Curi, N., Schaefer, C.E.G.R. and Vidal-Torrado, P., Eds., *Pedologia: Fundamentos*, SBCS (School of Business and Computer Science), Viçosa, 207-302.
- [9] Hu, G.R., Li, X.Y. and Yang, X.F. (2020) The Impact of Micro-Topography on the Interplay of Critical Zone Architecture and Hydrological Processes at the Hillslope Scale: Integrated Geophysical and Hydrological Experiments on the Qinghai-Tibet Plateau. *Journal of Hydrology*, **583**, 12-46. <https://doi.org/10.1016/j.jhydrol.2020.124618>
- [10] Florinsky, I.V. (2016) Influence of Topography on Soil Properties. In: Florinsky, I., Ed., *Digital Terrain Analysis in Soil Science and Geology*, 2nd Edition, Elsevier, London, 482.
- [11] Malone, B.P., McBratney, A.B. and Minasny, B. (2013) Spatial Scaling for Digital Soil Mapping. *Soil Science Society of America Journal*, **77**, 890-902. <https://doi.org/10.2136/sssaj2012.0419>
- [12] Pachepsky, Y. and Hill, R.L. (2017) Scale and Scaling in Soils. *Geoderma*, **287**, 4-30. <https://doi.org/10.1016/j.geoderma.2016.08.017>
- [13] Samuel-Rosa, A., Heuvelink, G.B.M., Vasques, G.M. and Anjos, L.H.C. (2015) Do More Detailed Environmental Covariates Deliver More Accurate Soil Maps? *Geoderma*, **243**, 214-227. <https://doi.org/10.1016/j.geoderma.2014.12.017>
- [14] Cavazzi, S., Corstanje, R., Mayr, T., Hannam, J. and Fealy, R. (2013) Are Fine Resolution Digital Elevation Models always the Best Choice in Digital Soil Mapping?

- Geoderma*, **195-196**, 111-121. <https://doi.org/10.1016/j.geoderma.2012.11.020>
- [15] Wadoux, A.M.J.C. Minasny, B. and McBratney, A.B. (2020) Machine Learning for Digital Soil Mapping: Applications, Challenges and Suggested Solutions. *Earth-Science Reviews*, **210**, Article ID: 103359. <https://doi.org/10.1016/j.earscirev.2020.103359>
- [16] Hengl, T. (2006) Finding the Right Pixel Size. *Computers & Geosciences*, **32**, 1283-1298. <https://doi.org/10.1016/j.cageo.2005.11.008>
- [17] Smith, M.P., Zhu, A.X., Burt, J.E. and Stiles, C. (2006) The Effects of DEM Resolution and Neighborhood Size on Digital Soil Survey. *Geoderma*, **137**, 58-69. <https://doi.org/10.1016/j.geoderma.2006.07.002>
- [18] R Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <http://www.r-project.org/>
- [19] Grass, D.T. (2019) Geographic Resources Analysis Support System (GRASS) Software, Version 7.8. Open Source Geospatial Foundation, Beaverton.
- [20] QGIS (2019) QGIS Geographic Information System. Open Source Geospatial Foundation Project, Beaverton. <http://qgis.osgeo.org>
- [21] Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., *et al.* (2015) System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, **8**, 1991-2007. <https://doi.org/10.5194/gmd-8-1991-2015>
- [22] Mendes, A.C., Truckenbrod, W. and Rodrigues, A.C.R.N. (2012) Análise faciológica da Formação Alter do Chão (Cretáceo, Bacia do Amazonas), próximo à cidade de Óbidos, Pará, Brasil. *Revista Brasileira de Geociências*, **42**, 39-57.
- [23] IBGE (Brazilian Institute of Geography and Statistics) (2002) Mapa de Clima do Brasil. Diretoria de Geociências, Brazilian Institute of Geography and Statistics, Rio de Janeiro, 1.
- [24] IBGE (Brazilian Institute of Geography and Statistics) (2008) Estado do Pará: Mapa de Geomorfologia. Brazilian Institute of Geography and Statistics, Rio de Janeiro, 1.
- [25] Schaefer, C.E.G.R., *et al.* (2017) Solos da região amazônica. In: Curi, N., Ker, J.C., Novais, R.F., Vidal-Torrado, P. and Schaefer, C.E.G.R., Eds., *Pedologia-Solos dos Biomas Brasileiros*, School of Business and Computer Science, Trinidad, 75-111.
- [26] Geological Survey, U.S. (2019) Landsat 8 Surface Reflectance Code (LASRC) Product Guide. No. LSDS-1368 Version 2.0, Geological Survey, U.S., Reston, 40. <https://www.usgs.gov/media/files/landsat-8-surface-reflectance-code-lasrc-product-guide>
- [27] Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., *et al.* (2007) The Shuttle Radar Topography Mission. *Reviews of Geophysics*, **45**, Article No. RG2004. <https://doi.org/10.1029/2005RG000183>
- [28] Brochado, G.T. (2015) Atenuação do efeito do desflorestamento em dados SRTM por meio de diferentes técnicas de interpolação. Instituto Nacional de Pesquisas Espaciais, São José dos Campos.
- [29] Li, Z. and Openshaw, S. (1993) A Natural Principle for the Objective Generalization of Digital Maps. *Cartography and Geographic Information Systems*, **20**, 19-29. <https://doi.org/10.1559/152304093782616779>
- [30] Guilbert, E., Boguslawski, P. and Isikdag, U. (2019) Multidimensional and Multiscale GIS. *ISPRS International Journal of Geo-Information*, **8**, Article No. 523. <https://doi.org/10.3390/ijgi8120523>
- [31] IBGE (Brazilian Institute of Geography and Statistics) (2015) Manual Técnico de Pedologia, 3a. Brazilian Institute of Geography and Statistics, Rio de Janeiro.
- [32] Jasiewicz, J. and Stepinski, T.F. (2013) Geomorphons—A Pattern Recognition Ap-

- proach to Classification and Mapping of Landforms. *Geomorphology*, **182**, 147-156. <https://doi.org/10.1016/j.geomorph.2012.11.005>
- [33] Zinck, J.A. (2016) The Geomorphic Landscape: Criteria for Classifying Geoforms. In: Zinck, J.A., Metternicht, G., Bocco, G. and Del Valle, H.F., Eds., *Geopedology*, Springer International Publishing, Cham, 77-99. https://doi.org/10.1007/978-3-319-19159-1_6
- [34] Minasny, B. and McBratney, A.B. (2006) A Conditioned Latin Hypercube Method for Sampling in the Presence of Ancillary Information. *Computers & Geosciences*, **32**, 1378-1388. <https://doi.org/10.1016/j.cageo.2005.12.009>
- [35] Biswas, A. and Zhang, Y. (2018) Sampling Designs for Validating Digital Soil Maps: A Review. *Pedosphere*, **28**, 1-15. [https://doi.org/10.1016/S1002-0160\(18\)60001-3](https://doi.org/10.1016/S1002-0160(18)60001-3)
- [36] dos Santos, R.D., dos Santos, H.G., Ker, J.C., dos Anjos, L.H.C. and Shimizu, S.H. (2015) *Manual de descrição e coleta de solo no campo*, 7a edição. Sociedade Brasileira de Ciência do Solo, Viçosa.
- [37] Brazilian Agricultural Research Corporation, National Soil Research Center (2017) *Manual de Métodos de Análise de Solo*, 3a Edição. Brazilian Agricultural Research Corporation, National Soil Research Center, Rio de Janeiro.
- [38] Kettler, T.A., Doran, J.W. and Gilbert, T.L. (2001) Simplified Method for Soil Particle-Size Determination to Accompany Soil-Quality Analyses. *Soil Science Society of America Journal*, **65**, 849-852. <https://doi.org/10.2136/sssaj2001.653849x>
- [39] Abdi, H. and Williams, L.J. (2010) Principal Component Analysis. *WIREs Computational Statistics*, **2**, 433-459. <https://doi.org/10.1002/wics.101>
- [40] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/A:1010933404324>
- [41] Lamichhane, S., Kumar, L. and Wilson, B. (2019) Digital Soil Mapping Algorithms and Covariates for Soil Organic Carbon Mapping and Their Implications: A Review. *Geoderma*, **352**, 395-413. <https://doi.org/10.1016/j.geoderma.2019.05.031>
- [42] Malone, B.P., Minasny, B. and McBratney, A.B. (2017) *Use R for Digital Soil Mapping*. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-319-44327-0>
- [43] Paluszynska, A., Biecek, P. and Jiang, Y. (2019) randomForestExplainer: Explaining and Visualizing. *R Package*.
- [44] Santos, H.G., Jacomine, P., dos Anjos, L.H.C., Oliveira, V. and Lumberreras, J.F. (2018) Sistema brasileiro de classificação de Solos, 5a. Sociedade Brasileira de Ciência do Solo, Viçosa.
- [45] Liu, C., Frazier, P. and Kumar, L. (2007) Comparative Assessment of the Measures of Thematic Classification Accuracy. *Remote Sensing of Environment*, **107**, 606-616. <https://doi.org/10.1016/j.rse.2006.10.010>
- [46] Behrens, T., Schmidt, K., MacMillan, R.A. and Viscarra Rossel, R.A. (2018) Multiscale Contextual Spatial Modelling with the Gaussian Scale Space. *Geoderma*, **310**, 128-137. <https://doi.org/10.1016/j.geoderma.2017.09.015>
- [47] Lindsay, J.B., Francioni, A. and Cockburn, J.M.H. (2019) LiDAR DEM Smoothing and the Preservation of Drainage Features. *Remote Sensing*, **11**, Article No. 1926. <https://doi.org/10.3390/rs11161926>
- [48] Wu, Q., Chen, Y., Wilson, J.P., Liu, X. and Li, H. (2018) An Effective Parallelization Algorithm for DEM Generalization Based on CUDA. *Environmental Modelling & Software*, **114**, 64-74. <https://doi.org/10.1016/j.envsoft.2019.01.002>
- [49] Zhou, Q. and Chen, Y. (2011) Generalization of DEM for Terrain Analysis Using a

- Compound Method. *ISPRS Journal of Photogrammetry and Remote Sensing*, **66**, 38-45. <https://doi.org/10.1016/j.isprsjprs.2010.08.005>
- [50] IBGE (Brazilian Institute of Geography and Statistics) (2009) Manual técnico de geomorfologia, 2a. Brazilian Institute of Geography and Statistics, Rio de Janeiro.
- [51] Flynn, T., Rozanov, A., Ellis, F., de Clercq, W. and Clarke, C. (2019) Farm-Scale Soil Patterns Derived from Automated Terrain Classification. *Catena*, **185**, Article ID: 104311. <https://doi.org/10.1016/j.catena.2019.104311>
- [52] De Reu, J., Bourgeois, J., Bats, M., Zwertvaegher, A., Gelorini, V., De Smedt, P., et al. (2013) Application of the Topographic Position Index to Heterogeneous Landscapes. *Geomorphology*, **186**, 39-49. <https://doi.org/10.1016/j.geomorph.2012.12.015>
- [53] Szypuła, B. and Wieczorek, M. (2020) Geomorphometric Relief Classification with the K-Median Method in the Silesian Upland, Southern Poland. *Frontiers of Earth Science*, **14**, 152-170. <https://doi.org/10.1007/s11707-019-0765-9>
- [54] de Mello Cunha, G.O., de Almeida, J.A. and Barboza, B.B. (2014) Relação entre o alumínio extraível com KCL e oxalato de amônio e a mineralogia da fração argila, em solos ácidos Brasileiros. *Revista Brasileira de Ciência do Solo*, **38**, 1387-1401. <https://doi.org/10.1590/S0100-06832014000500004>
- [55] de Souza, C.M.P., Thomazini, A., Schaefer, C.E.G.R., Veloso, G.V., Moreira, G.M. and Fernandes Filho, E.I. (2018) Multivariate Analysis and Machine Learning in Properties of Ultisols (Argissolos) of Brazilian Amazon. *Revista Brasileira de Ciência do Solo*, **42**, e0170419. <https://doi.org/10.1590/18069657rbc20170419>
- [56] Lucas, Y., Luizão, F.J., Chauvel, A., Rouiller, J. and Nahon, D. (1993) The Relation between Biological Activity of the Rain Forest and Mineral Composition of Soils. *Science*, **260**, 521-523. <https://doi.org/10.1126/science.260.5107.521>
- [57] Hartemink, A.E., Zhang, Y., Bockheim, J.G., Curi, N., Silva, S.H.G., Grauer-Gray, J., et al. (2020) Soil Horizon Variation: A Review. *Advances in Agronomy*, **160**, 125-185. <https://doi.org/10.1016/bs.agron.2019.10.003>
- [58] Behrens, T., Viscarra Rossel, R.A., Kerry, R., MacMillan, R., Schmidt, K., Lee, J., et al. (2019) The Relevant Range of Scales for Multi-Scale Contextual Spatial Modelling. *Scientific Reports*, **9**, Article No. 14800. <https://doi.org/10.1038/s41598-019-51395-3>
- [59] Novais, J., Roberto and Mello, F. (2007) Relação Solo-Planta. In: Novais, R.F., de Barros, N.F., Fontes, R.L.F., Cantarutti, R.B. and Lima, J.C, Eds., *Fertilidade do Solo*, School of Business and Computer Science, Viçosa, 1-10.
- [60] Florinsky, I., Eilers, R., Manning, G. and Fuller, L. (2002) Prediction of Soil Properties by Digital Terrain Modelling. *Environmental Modelling & Software*, **17**, 295-311. [https://doi.org/10.1016/S1364-8152\(01\)00067-6](https://doi.org/10.1016/S1364-8152(01)00067-6)
- [61] Bhering, S.B., Chagas, C.S., de Carvalho Jr., W., Pereira, N.R., Filho, B.C. and 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**, 1359-1370. <https://doi.org/10.1590/S0100-204X2016000900035>
- [62] Shi, J., Yang, L., Zhu, A., Qin, C., Liang, P., Zeng, C., et al. (2018) Machine-Learning Variables at Different Scales vs. Knowledge-based Variables for Mapping Multiple Soil Properties. *Soil Science Society of America Journal*, **82**, 645-656. <https://doi.org/10.2136/sssaj2017.11.0392>