

MULTICRITERIA ANALYSIS AND MACHINE LEARNING ALGORITHM FOR DEFINITION OF AREAS FOR MICRO-DAM, SOUTHEASTERN BRAZIL

Cristiano Marcelo Pereira de Souza

Universidade Estadual de Montes Claros, Pós-graduação em Geografia, Montes Claros, MG, Brasil
cristiano.souza@ufv.br

Gustavo Vieira Veloso

Universidade Federal de Viçosa, Pós-graduação em Solos e Nutrição de Plantas, Viçosa, MG, Brasil
cristiano.souza@ufv.br

André Luiz Lopes de Faria

Universidade Federal de Viçosa, Departamento de Geografia, Viçosa, MG, Brasil
andre@ufv.br

Marcos Esdras Leite

Universidade Estadual de Montes Claros, Departamento de Geociências, Montes Claros, MG, Brasil
marcos.leite@unimontes.br

Elpídio Inácio Fernandes Filho

Universidade Federal de Viçosa, Departamento de Solos, Viçosa, MG, Brasil
elpidio@ufv.br

ABSTRACT

Micro-dams are efficient in controlling erosion and increases the rate of water infiltration into the soil. The study aims to predict potential areas for micro-dam allocation. We prepared the map of potential areas for micro-dam from the multi-criteria analysis (AHP) using topographic data and land use. Furthermore, we developed a methodological framework with Random Forest (RF) algorithm to predict potential micro-dam areas from points extracted from the map by AHP and aided by the insertion of covariates; and we also apply the overlap of the two maps (AHP + RF). The AHP method overestimates high potential areas in zones without potential. The RF model used seven topographic covariates, and they are related to hydrological flows. The performance of R^2 in RF was 0.43, statically satisfactory, but spatially underestimated the very low and very high potential classes. The superposition of the low potential class of the AHP method on the RF map favored a 99% reduction of very high potential areas in zones not suitable for micro-dam allocation. Therefore, the combination of the AHP and ML method generates more consistent spatial results

Keywords: Soil erosion control. Spatial prediction. Random Forest.

ANÁLISE MULTICRITÉRIO E ALGORITMO DE APRENDIZAGEM DE MÁQUINA PARA DEFINIÇÃO DE ÁREAS DE BARRAGINHAS, SUDESTE DO BRASIL

RESUMO

As barraginhas são eficientes no controle da erosão e aumentam a taxa de infiltração de água no solo. Este estudo visa prever áreas potenciais para alocação de barraginhas. Preparamos o mapa de áreas potenciais para alocação de barraginhas a partir da análise multicritério (AHP) usando dados topográficos e uso do solo. Além disso, desenvolvemos uma estrutura metodológica com algoritmo Random Forest (RF) para prever áreas potenciais de barraginhas a partir de pontos extraídos do mapa gerado por AHP e auxiliados pela inserção de covariáveis; e também aplicamos a sobreposição dos dois mapas (AHP + RF). O método AHP superestima as áreas de alto potencial em zonas sem potencialidade. O modelo de RF usou sete covariáveis topográficas, e elas estão relacionadas aos fluxos hidrológicos. O desempenho de R^2 no RF foi de 0,43, estaticamente satisfatório, mas espacialmente subestimou as classes de potencial muito baixo e muito alto. A superposição da classe de baixo potencial do método AHP no mapa de RF favoreceu uma redução de 99% das áreas de muito alto potencial em zonas não adequadas para alocação de micro-barragens. Portanto, a combinação do método AHP e ML gera resultados espaciais mais consistentes

Palavras-chave: Controle de erosão do solo. Predição espacial. Random Forest.

INTRODUCTION

Soil erosion is a significant cause of soil degradation, affects soil organic matter and fertility levels, contaminates water bodies with transported sediments and agrochemicals, and increases the risk of flooding (GARCÍA-RUIZ et al., 2015; LABRIÈRE et al., 2015). In the world, only soil erosion is responsible for a loss of 10^6 ha/year of agricultural land (PIMENTEL, 2006). The main factors that promote erosion are the inappropriate use of the land and wrong soil management practices (ASSESSMENT, 2005; WUEPPER et al., 2020). Recent studies show that developing countries, especially in tropical regions, historically coexist with higher rates of soil loss due to erosion (WUEPPER et al., 2020). In addition to the history of land use, the natural characteristics of tropical regions have a predisposition to increase soil loss. In Brazil, the combination of high rainfall and sloping relief increases the kinetic energy of surface waters (OLIVEIRA et al., 2013). Notably, the relationship of these factors is very typical in the morphoclimatic domain known as “Mares de Morro”, which are a sequence of wavy landforms with a high slope, occurring in a vast area of the Brazilian territory (AB'SÁBER, 1970). This predisposition to erosion affects even soils with higher physical stability, such as oxisols, with erosion rates above 34% of the erosion tolerance limit (AYER et al., 2015; SAKUNO et al., 2020). Another characteristic in sloping areas is the formation of Inceptisols, with a higher silt content, making the area more susceptible to erosion (FU et al., 2011). Therefore, knowledge of environmental characteristics and adequacy of soil erosion control practices are essential in land use planning for these areas.

In the literature, soil erosion studies have direction towards two focuses in general: (i) measure and identify erosion factors, for example, soil loss studies (DIDONÉ et al., 2015; ANACHE et al., 2017). (ii) erosion control techniques and practices, for example, slope stabilization (HOLANDA et al., 2008), no-till system (DEUSCHLE et al., 2019), gullies control (VALENTIN et al., 2005), and construction of micro-dam (BARROS E RIBEIRO, 2009). Specifically, the micro-dam are small basins excavated in the ground, with an average diameter of 16 m and an average depth of 1.8 m, with the function of capturing water from the runoff, increasing soil infiltration and erosion control (BARROS E RIBEIRO, 2009; ARAGÃO et al., 2019; HIPÓLITO et al., 2019; XU et al., 2019; MESHARAM et al., 2020). The micro-dam also plays a socio-environmental role, as it favors infiltration to groundwater that feeds rivers, increasing the availability of water for urban and agricultural supply. For example, in Tunisia, especially during the dry years, micro-dam provides an additional amount of water for crops (SCHIETTECATTE et al., 2005). In the case of Brazil, the construction of micro-dam becomes relevant because of the possibility of mitigating the effects of soil erosion, so common in rural areas, especially small rural properties, where lower-income inhibits more expensive erosion control techniques (DIDONÉ et al., 2015). In addition to the factor, it contributes to attenuating the water crisis effects that routinely affect regions in southeastern Brazil, either due to changes in the precipitation regime and/or related to the increase in unsustainable rural and urban land-use practices (NOBRE et al., 2016).

The allocation of areas with potential for the construction of micro-dam must consider several restrictions and potential factors. Among some restrictive factors, there are perennial watercourses, permanent protection areas (PPA), inside gullies, narrow valleys, and slope greater than 12% (BARROS E RIBEIRO, 2009). The support for analyzing these factors is to combine data analysis of the Digital Elevation Model (DEM), land use with geoprocessing tools (ARAGÃO et al., 2019; HIPÓLITO et al., 2019). Notably, DEM provides several new data (topographic covariates) (GALLANT E AUSTIN, 2015), and some are highly correlated to hydrological flow, which is a factor closely linked to micro-dam dynamics. However, analyzing together the various topographic factors and land use requires more robust statistical and geostatistical methods. The effort in this regard is to consider the multicriteria analysis and hydrological data (ARAGÃO et al., 2019; HIPÓLITO et al., 2019). The advantage of multicriteria analysis is due to the process that transforms and combines geographic data with defined weights to obtain new information (SICAT et al., 2005; CHEN et al., 2010).

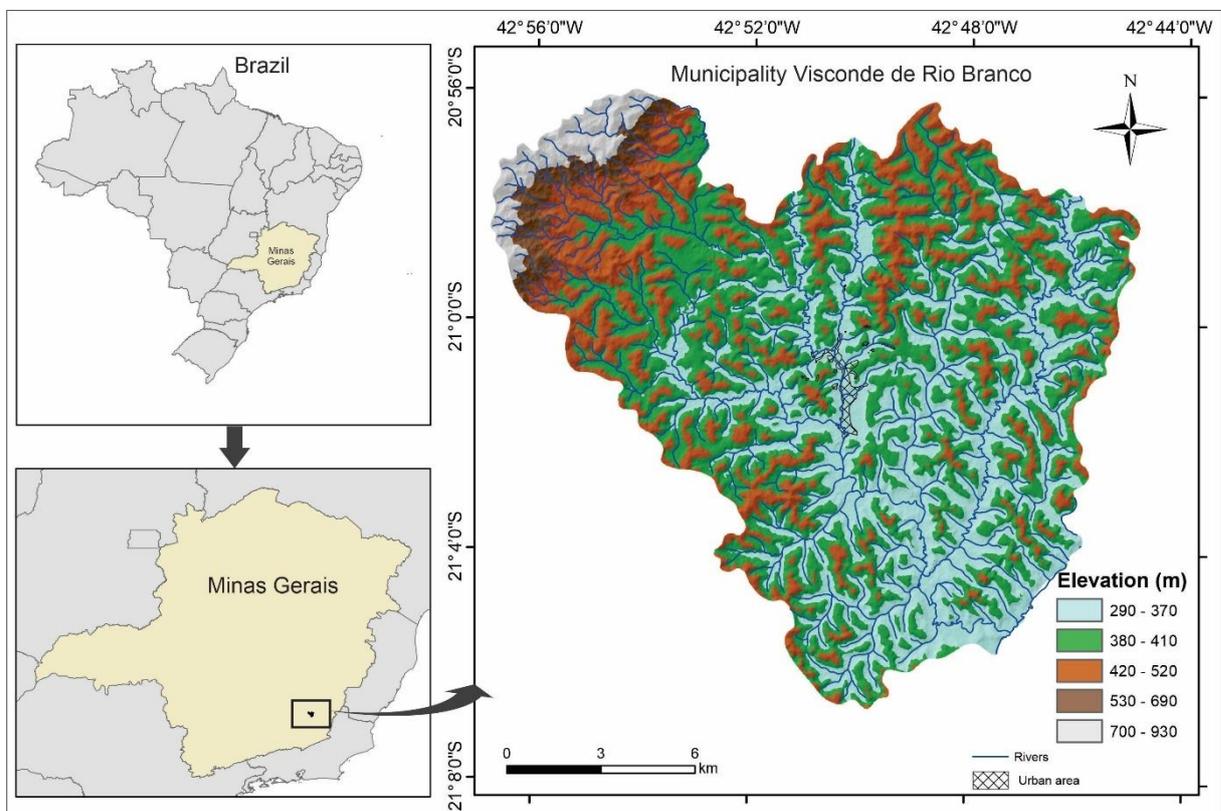
Also, in the field of robust statistical analysis applied to environmental studies, there are Machine Learning Algorithms (ML), which involve a wide range of models used to discover patterns in data and make predictions (WITTEN et al., 2016). Several studies confirm the efficiency of ML in a vast field of application in spatial prediction, involving research areas of climate, geology, soil, vegetation, among others (AMIRI et al., 2019; GOMES et al., 2019; MOHAMMADY et al., 2019; SOUZA et al., 2020). In general, the ML allows selecting significant covariates from a database and provides statistical data with error levels and prediction accuracy (KUHN E JOHNSON, 2013). Although ML is efficient, however, there are still no studies predicting micro-dam areas. This study aimed to determine potential areas for micro-dam and demonstrate that the association of the multi-criteria analysis, ML, and covariate dataset provides a more accurate estimate of potential sites for micro-dam allocation.

MATERIALS E METHODS

Study area

The study area is the municipality of Visconde do Rio Branco (Southeastern Brazil), between the coordinates $-20^{\circ} 55' 58''$ to $-21^{\circ} 6' 43''$ W and $-42^{\circ} 56' 51''$ S to $42^{\circ} 44' 53''$ S (Figure 1). It presents the Cwa climate in the Köppen classification, with annual precipitation of 1100 mm, concentrated in October to March. The region is part of the Mantiqueira Complex, with orthognathic amphibole-biotite, and pegmatites rocks (PINTO et al., 2003). The relief is part of the Mares de Morros domain (AB'SÁBER, 1970), with a sequence of hills forming landscapes with wavy to mountainous landforms, whose altitudes vary between 467 and 897 meters. The predominant soils are Ultisols and Oxisols, and on most slopes, there is a predominance of Inceptisols. The main land use is pasture, generally degraded, representing 80.69%, followed by exposed soil (10.98%), forest/eucalyptus (5.84%), and civil constructions (2.48%).

Figure 1. Visconde de Rio Branco (MG): Location of the study area with a digital elevation model.



Methodological procedures

Multicriteria analysis

To determine potential areas for micro-dam, we used the data of slope and accumulated flow model derived from DEM Alos-Palsar (12.5 m of spatial resolution). We also use the land use map prepared by the Sentinel-2 satellite image (10 m resolution) by the Maxver algorithm.

In the next step, we classify the intervals of the covariates according to the degree of potentiality for micro-dam construction. We assign zero value to classes without potential, for example, forest areas and perennial watercourses, areas with an accumulated flow higher than 500 pixels of contribution, and slopes greater than 15% (BARROS E RIBEIRO, 2009; ARAGÃO et al., 2019). Weights vary in four classes (0, 1, 5, 10), indicating less for higher potential. We use the Analytic Hierarchy Process (AHP) method to grant variable weights. The technique analyzes the variables by a correlation matrix (decision

matrix) (SAATY, 2013; SAATY E VARGAS, 2013) and provides a total contribution to the set of variables involved. In defining the weights, it is necessary to divide the value of each variable by the total contribution amount. As a final weight, we used the average between the columns of the variables (Table 1).

We evaluated the consistency of the weights by the consistency ratio (RC), which indicates whether the weights are satisfactory by determining the RC must be less than 0.10 (SAATY, 1986). Subsequently, we applied equation 1 to determine the potential of the micro-dam.

$$MDP = 0,63 \times AF + 0,26 \times S + 0,11 \times US \quad \text{Equation 1}$$

Where: MDP- Micro-dams potential, AF- Accumulated flow, S- slope, LS- land use.

Table 1. Top: Decision matrix of variables to analyze the total contribution of variables. Bottom: individual variable weights and average weight.

Contribution matrix of variables				
	Accumulated flow	Slope	Land use	
Accumulated flow	1	3	5	
Slope	1/3	1	3	
Land use	1/5	1/3	1	
Total	1.53	4.33	9	
Variable weights				
	Accumulated flow	Slope	Land use	Average
Accumulated flow	0.65	0.69	0.56	0.63
Slope	0.22	0.23	0.33	0.26
Land use	0.13	0.08	0.11	0.11

Machine Learning Algorithm

The stage of prediction of potential areas for micro-dam allocation involved: (i) definition of sampling points (variable); (ii) creation of a bank of covariates; (iii) selection of significant covariates; (iv) training, validation with Random Forest algorithm for map prediction, with steps described below.

From the micro-dam map by the AHP method, we determined a grid of 2000 points and extracted values to compose the input data for prediction (variable). As for the covariates, we used DEM Alos Palsar 12.5 m radar images in the RSAGA software to extract covariates (BRENNING, 2008; SENA et al., 2020; SOUZA et al., 2020), totaling 34 topographic covariates (Table 2). In the following steps, we apply covariates selection criteria to assist in constructing the model, since the excessive number does not provide fluidity to computational processing (KUHN E JOHNSON, 2013). We analyzed the correlation factor between the covariates by the find correlation function, discarding those with high correlation (Pearson >95%). The purpose is to discard covariates that contribute similarly to the identification of potential areas for micro-dam.

Subsequently, we used the Random Forest-Recursive Feature Elimination method (RF-RFE), to select the subsets of covariates that best explain prediction (RF-RFE) (GRANITTO et al., 2006; KUHN E JOHNSON, 2013). The RF-RFE successively removes the least relevant covariates to find the best performing set, using the value of R² as a reference. However, high values of R² are almost always the result of using a more significant number of covariates; this makes the prediction less fluid and can cause difficulties in applying it in other areas (MALCZEWSKI, 2006). Therefore, we configure the RF-RFE to select a set of covariates allowing an R-squared (R²) 3% less than the largest R² found. This step of the RF-RFE was based on a cross-validation with 10 folds and 5 repetitions (KUHN E JOHNSON, 2013; GOMES et al., 2019).

Using the ideal subset of covariates defined by the RF-RFE from the previous step, we configured the separation of these data in 75% for training and 25% for validation (holdout-test). We used the Random Forest (RF) algorithm, which is capable of performing classification and regression (BREIMAN, 2001), with wide use in spatial prediction, generating good results (PAL, 2005; JUNIOR et al., 2016; GOMES et al., 2019; MOHAMMADY et al., 2019; SOUZA et al., 2020). This entire process was repeated 100 times, randomly changing the samples present in the training and validation at each repetition. The advantage of this excessive repetition is to avoid potentially biased predictions, as it evaluates with different sample groups (GRANITTO et al., 2006; KUHN E JOHNSON, 2013; SOUZA et al., 2018; GOMES et al., 2019). Besides, at each step (100 times), this methodological framework provides statistical data indicating the accuracy and error of the prediction: R-square (R^2), root-mean-square error (RMSE), Mean Absolute Error (MAE) (KUHN E JOHNSON, 2013; GOMES et al., 2019).

The difference between R^2 values in the validation training steps was analyzed to understand if the model trains and validates satisfactorily. Discrepant values of R^2 denote low modeling efficiency (overfitting). This step assumes that the training dataset is used to build (or train) the model, while the validation dataset is used to estimate the performance of the trained predictive model. ML algorithms do not know the test dataset (validation) in the training stage, so the model trained on the test dataset is considered the model's performance estimates (KUHN E JOHNSON, 2013; CHUNG E LEE, 2019; SOUZA et al., 2020). Thus, when the predictive model well predicts the outcome in the training stage but not in the validation stage, this predictive model is considered overestimated in the training dataset.

Table 2. Group of predictive covariates from the Alos-Palsar image.

Terrain attributes	Description	Terrain attributes	Description
Aspect	Slope orientation	Plan curvature	Described as the curvature of the hypothetical contour line passing through a specific cell
Convergence index	Convergence/divergence index concerning runoff	Profile curvature	Describes surface curvature in the direction of the steepest incline
Cross sectional curvature	Measures the curvature perpendicular to the downslope direction	Real surface área	The actual calculation of cell área
Digital elevation model	Represents the elevation in each model cell	Slope	Represents local angular slope
Flow line curvature	Represents the projection of a gradient line to a horizontal plane	Slope height	The vertical distance between base and ridge of slope
General curvature	The combination of both plan and profile curvatures	Standardized height	The vertical distance between base and the standardized slope index
Gradient	Corresponds to the hydrological gradient	Surface specific points	Indicates differences between specific surface shift points
Hill	Demonstrates the hills	Tangential curvature	Measured in the normal plane in a direction perpendicular to the gradient
Hill Index	Simulation of diffusive hillslope evolution using an Alternating-Direction-Implicit (ADI) method.	Terrain ruggedness index	Quantitative index of topography heterogeneity
Longitudinal curvature	Measures the curvature in the downslope direction	Terrain surface convexity	The ratio of the number of cells that have positive curvature to the number of all valid cells within a specified search radius
Mass balance index	Balance index between erosion and deposition	Terrain surface texture	Splits surface texture into 8, 12, or 16 classes
Maximal curvature	Maximum curvature in local normal section	Topographic position index	Difference between a point elevation with surrounding elevation
Mid-slope position	Represents the distance from the top to the valley, ranging from 0 to 1	Topographic wetness index	Describes the tendency of each cell to accumulate water as a function of relief
Minimal curvature	Minimum curvature for local normal section	Total curvature	General measure of surface curvature
Multiresolution index of ridge top flatness	Indicates flat positions in high altitude áreas	Valey Index	

Multiresolution index of valley bottom flatness	Indicates flat surfaces at bottom of valley	Valley depth	Calculation of vertical distance at drainage base level
Normalized height	Vertical distance between base and ridge of normalized slope	Vector ruggedness measure	Measures the variation in terrain roughness

Overlay maps (AHP+RF) and analysis of micro-dam areas

This step included overlaying maps of potential areas for micro-dam generated by AHP and RF. We selected the low potentiality class for micro-dam allocation and applied the overlay on the RF map. This criterion is an alternative for the normalization of values practiced by the RF algorithm, which tends to eliminate extreme values. In addition, this procedure makes it possible to generate a more restrictive map to construct a micro-dam, making the map more environmentally compatible, this new map called AHP+RF.

In evaluating the efficiency of the three models (AHP, RF, and AHP+RF), we analyzed the distribution of potentiality classes over-restrictive zones to APP allocation. The input data were areas of slope greater than 12% extracted from the DEM. In addition, we consider Permanent Preservation Areas as restrictive zones, including watercourses, hilltops, and humid areas. These data are made available by the Geomorphology laboratory at UFV (GEOMORFOLOGIA-UFV, 2019).

RESULTS AND DISCUSSION

Covariates selection

Analyzing the statistical performance of prediction with Random Forest algorithm associated with Recursive Feature Elimination (RF-RFE), as it involves a more extensive set of steps, the results indicated that the accuracy and error metrics are negatively affected with an excess of covariates (Figure 2). In addition, the findcorrelation function previously discarded several high-correlation covariates. Therefore, using a smaller set of covariates generates more fluid modeling without overestimated predictions and meets the principle of parsimony (GRANITTO et al., 2006; KUHN E JOHNSON, 2013; SOUZA et al., 2018; GOMES et al., 2019). Therefore, the Random Forest (RF) selected seven covariates to determine potential areas for micro-dam and generated an R^2 of 0.43 in training and validation (Table 3; Table 4). Although there are no specific studies with the prediction of micro-dam by ML, R^2 values are consistent with other prediction studies (YESILNACAR E TOPAL, 2005; SOUZA et al., 2018; GOMES et al., 2019; ASSIS et al., 2021). Generally, the lower performance is due to the lack of covariates that can better explain the distribution of a variable (KUHN E JOHNSON, 2013). Therefore, for many spatial prediction studies by ML, R^2 values between 0.4 to 0.6 are common and considered a satisfactory performance.

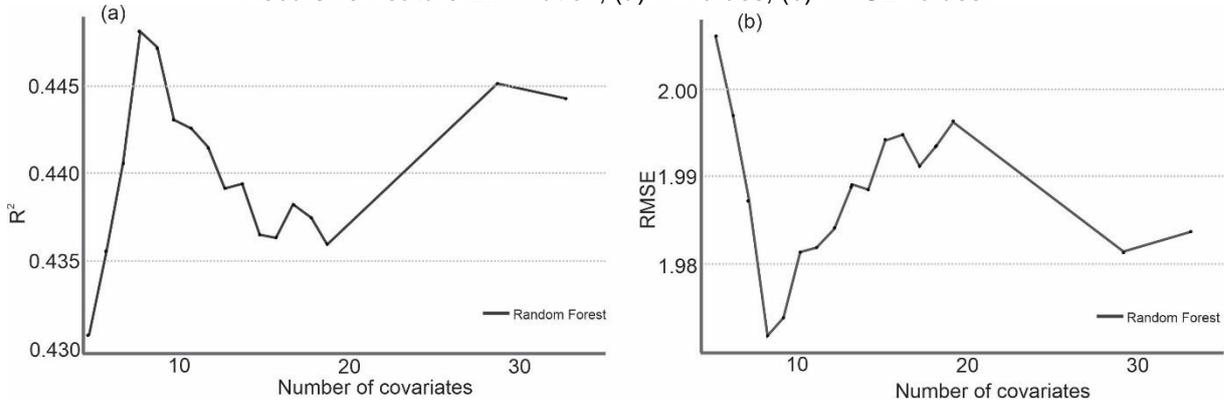
Although the value of R^2 is not high, in prediction studies comparing the training and validation accuracy metrics is a way to assess the modeling performance (KUHN E JOHNSON, 2013). Considering the behavior of precision and error metrics (R^2 , RMSE, and MAE), there are similarities in the values in training (cross-validation) and validation (holdout-test), showing a low overfitting statistical effect (Table 3). In general, ML predictions are subject to overfitting, deteriorating prediction accuracy, and this problem occurs when data are insufficient concerning the number of model parameters, generating differences between training and validation accuracy values (CHUNG E LEE, 2019). However, low overfitting is an advantage of the RF model, because as the number of trees increases, the generalization error reduces (BREIMAN, 2001; PAL, 2005).

Table 3. Performance of the Random Forest model during training and validation (100 times).

Value	R^2		RMSE			MAE		
	S. dev	CV	Value	S. dev	CV	Value	S. dev	CV
Training (cross-validation)								
0.43	0.02	3.74	2.01	0.03	1.41	1.53	0.03	1.84
Validation (holdout-test)								
0.43	0.04	9.05	2.00	0.07	3.44	1.51	0.07	3.46

R² – R square, RMSE - Root-mean-square error; MAE - Mean absolute error; CV – Coefficient of variation, S. dev – Standard deviation.

Figure 2. Performance of the Random Forest model in the covariate selection process, using Recursive Feature Elimination, (a) R² values, (b) RMSE values.



Among the covariates selected after the findcorrelation and RF-RFE function, there is a predominance of low correlation covariates (Pearson below 0.69) those significantly important according to the ranking established by the RF-RFE (Table 4). In general, covariates are relief information indicating flat areas and valleys and can explain the areas of high potential and low potential for micro-dam allocation. The most significant covariate was the slope length factor (Ls-factor), and this covariate indicates areas with high and low potential for micro-dam. Concerning the length of the slope, because the higher the distance, there is potential for increased kinetic energy, and the consequence is an increase in the erosion process (VAN REMORTEL et al., 2001). Therefore, the construction of micro-dam after steep slopes and long ramps can mitigate the erosion factor.

Table 4. Correlation table between selected covariates (Pearson) and the hierarchy of importance between the covariates by RF-RFE.

	Relative importance level						
	Low						High
M. ruggedness	WTI	Conv. Index	MRRTF	Valley depth	MRVBF	Ls factor	
M. ruggedness	1	-0.3	-0.22	-0.24	0.45	-0.33	0.58
WTI		1	-0.17	0.3	0.17	0.69	-0.5
Conv. Index			1	0.41	-0.3	-0.34	-0.04
MRRTF				1	-0.18	0.19	-0.29
Valley depth					1	0.06	0.29
MRVBF						1	-0.65
Ls factor							1

M. ruggedness: Melton ruggedness index; WTI: Topographic wetness index; Conv. Index: Convergence index; MRRTF: Multiresolution index of the ridge top flatness; MRVBF: Module multiresolution index of valley bottom flatness; Ls factor: Slope length factor.

The covariate Melton ruggedness index is an index that is related to accumulation areas, as it discriminates areas of deposition of alluvial sediment flows (debris-flow fans) (JACKSON JR. et al., 1987) was also selected; and one of the functions of micro-dams is to control the erosion process in lower areas (BARROS E RIBEIRO, 2009). Also, flattened areas that receive water flow are potentially suitable for construction micro-dam (SCHIETTECATTE et al., 2005; BARROS E RIBEIRO, 2009). Thus, the model selected the topographic wetness index, and this covariate is related to hydrological

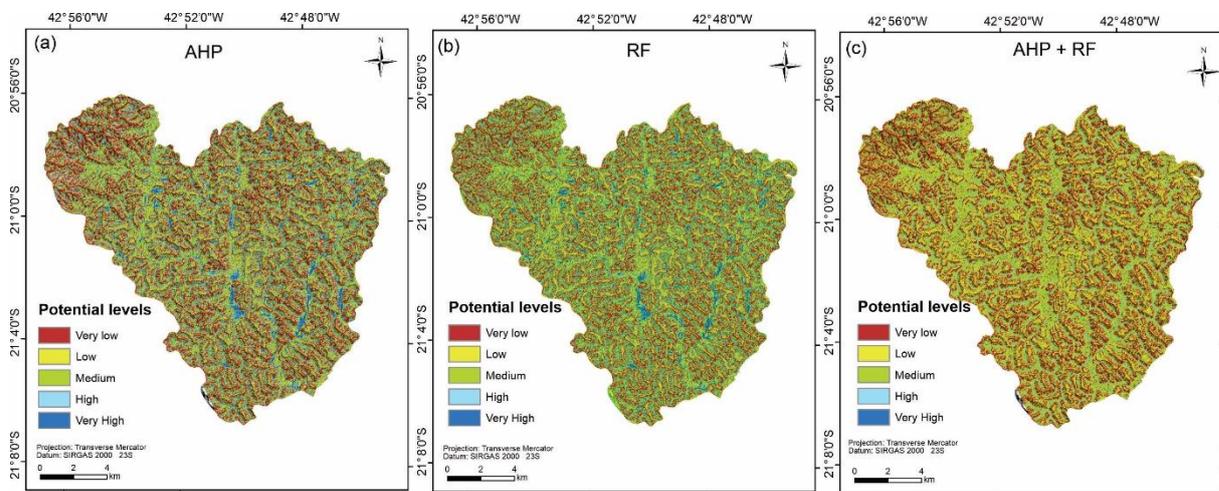
processes, describes the local trend of receiving water from the slope area (GRUBER E PECKHAM, 2009; KOPECKÝ E ČÍŽKOVÁ, 2010). In the same way, the convergence index indicates the fields of convergence of flow (PROSSER E ABERNETHY, 1996).

Finally, three covariates indicate areas of valleys (Table 4); this information is related to areas without the potential for micro-dam because the sites for construction of micro-dam should not be in slope >12%. Furthermore, the valleys are closely related to hydrological processes (GALLANT E DOWLING, 2003), can indicate areas of perennial rivers, and are unsuitable for micro-dam.

Prediction of potential areas for dams (AHP and RF)

The micro-dam potential maps derived from the multicriteria analysis method (AHP) and by Random Forest (RF), have different configurations (Figure 3). In the AHP method, there were much larger areas of the very low and very high potential areas, while in the RF, the same classes fell ~ 98% (Figure 4). In evaluating the efficiency of these two maps, we considered the prediction of classes for areas limiting the allocation of micro-dam. Therefore, certain factors do not favor the construction of micro-dam, for example, Permanent Preservation Area (PPA), slopes >12%, narrow valleys (valleys in "V"), and there must be a spatial dispersion in the allocation of micro-dams (BARROS E RIBEIRO, 2009). Considering the slope factor is fundamental in the allocation of micro-dam, several aspects linked to the slope reduce the efficiency of the micro-dam in controlling erosion. On high slopes, there is an increase in the kinetic energy of rainfall, a lower rate of water infiltration, greater convergence of water flow, and the presence of more fragile soils to erosive processes (SCHIETTECATTE et al., 2005; BARROS E RIBEIRO, 2009; ARAGÃO et al., 2019; HIPÓLITO et al., 2019). In addition, in rainy periods, the increase in surface water flow can compromise the stability of the micro-dam and cause ruptures, with severe environmental damage (HIPÓLITO et al., 2019).

Figure 3. Potential areas for building micro-dam by methods: (a) multi-criteria analysis (AHP); (b) Random Forest (RF); (c) combination AHP + RF.



Comparing the AHP and RF maps, with some limiting factors for micro-dam construction, it was evident that in the multicriteria analysis, there were many areas with high potential (levels 4 and 5) in steep slope areas (Figure 5a). Overestimating the high potential levels for micro-dam in hillslope can be a critical problem, since these zones coincide with the dominance of the Inceptisols, soils with higher silt content and more susceptible to erosion (FU et al., 2011; FONSECA et al., 2017). On the other hand, the RF did not overestimate the high and very high potential levels in hillslopes. However, RF underestimated very low potential level, and there was a predominance of the low and medium potential (Figure 5b). In general, the values underwent normalization in the prediction by the RF method, which is a characteristic of the algorithm (BREIMAN, 2001). Sometimes, this normalization plays an essential role because eliminating values of low representativeness makes it possible to raise R² and reduce RMSE (CHI et al., 2008). However, in the prediction of micro-dams, it generates maps that are not consistent for underestimating the restrictive classes and suitable for micro-dam allocation.

Another aspect is the distribution of potential levels in PPA; in the AHP method, the very high potential area occupies 24% of the PPA, while in the RF model, this same level occupies 0.12%. Given these scenarios, the alternative is to combine the two maps; combining the two maps (AHP + RF), it allows a better definition of potential areas for micro-dam construction (Figure 3c). The isolation of the low potential level and associating it with the RF reduced the area of the most restrictive class for micro-dam (i.e., very high) allocation over PPA zones (Figure 5d). Besides, generated a negative linear correlation with the slope (Pearson: -0,56), and low potential levels predominate in areas of low slope (Figure 5c).

A configuration maintained in the combination of maps that may be considered inappropriate is the increment of medium potential areas in PPA (Figure 5d). The prevalence of this class over PPA was due to the effect of data normalization practiced by the RF. Also, some river PPA have characteristics of landforms that are in principle favorable to the allocation of micro-dam, for example, low slope. However, river PPA has other environmental functions for preserving watercourses and maintaining fauna and flora, including preservation guaranteed by law (BARROS E RIBEIRO, 2009).

Figure 4. Area of potential levels for the construction of micro-dam by methods: AHP: multicriteria analysis; RF: Random Forest; AHP + RF: Multicriteria analysis and Random Forest. Potential levels: 1 Very low, 2 Low, 3 medium, 4 High, 5 Very high.

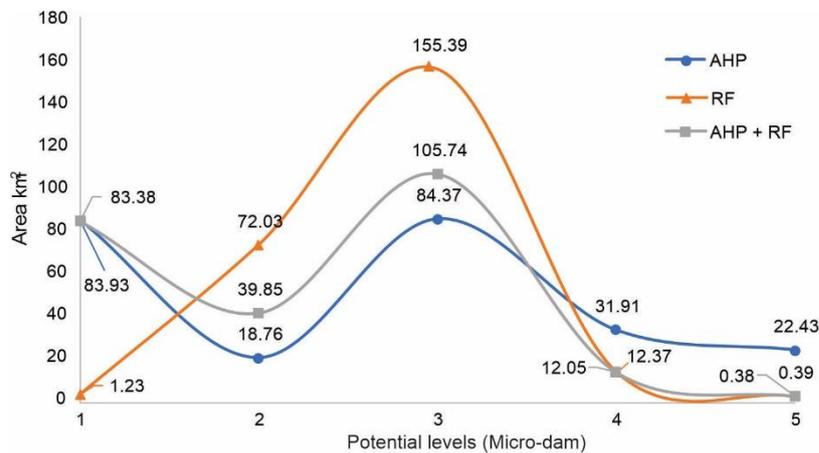
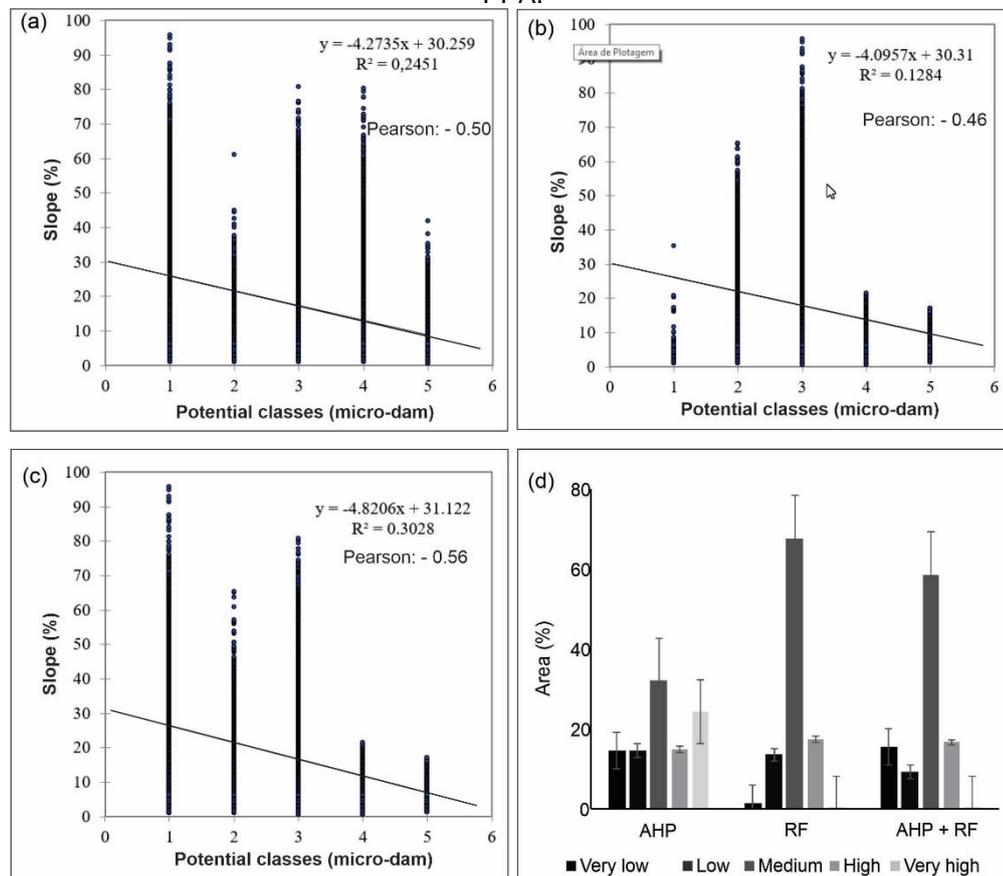


Figure 5. Scatter plots between slope values and potentiality levels for the construction of micro-dam by methods: (a) AHP, (b) RF, and (c) AHP + RF. (d) potential levels for micro-dam construction in PPA.



In general, combining the micro-dam potential map by the AHP+RF method generates better spatial results, reducing the effect of underestimating areas for micro-dam allocation in restricted areas when applying the methods in isolation. Notably, the advantage of ML is the inclusion of covariates in prediction studies (KUHN E JOHNSON, 2013; CHAGAS et al., 2018; SOUZA et al., 2018; GOMES et al., 2019). The insertion of covariates opens up vast possibilities for future studies on areas for micro-dam allocation, especially by inserting covariates of hydrological and climatic factors, which are important factors in micro-dam allocation (SCHIETTECATTE et al., 2005; HIPÓLITO et al., 2019; MESHARAM et al., 2020). However, it was not evaluated in this study due to data limitation in an adequate scale.

CONCLUSIONS

The micro dam potential map by multicriteria analysis overestimates the high potential class in environmentally restrictive areas for micro-dam allocation.

The RF method generates a map and error and precision measurements and selects important covariates. The R^2 was 0.43 and with low R^2 variation in training and validation, indicating good performance in the process by the low overfitting effect. Besides, seven covariates were important in the prediction, and are data that are related to hydrological flows. However, the RF-generated map underestimates the extreme level classes for micro-dam allocation (i.e., very low and very high potentiality classes).

The overlay of the low potential level of the AHP map over the RF model map generates a more consistent result in the definition of potential areas for micro-dam, with a reduction of high levels of potential in sloping areas and Permanent Preservation Area (PPA).

ACKNOWLEDGMENT

The research was supported by the Municipality of Visconde de Rio Branco. For the project: Instrumentation, monitoring and recovery in the Serra da Piedade in Visconde do Rio Branco – MG. The authors thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for the support. Process number: 88882.315083/2019-01.

REFERENCES

- AB'SÁBER, A.N., **Províncias geológicas e domínios morfoclimáticos no Brasil**. São Paulo: Universidade de São Paulo, Instituto de Geografia, 1970. 26 p.
- AMIRI, M., POURGHASEMI, H.R., GHANBARIAN, G.A., AFZALI, S.F. Assessment of the importance of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping using three machine learning algorithms. **Geoderma**, v.340. p.55-69. 2019.
- ANACHE, J.A.A., WENDLAND, E.C., OLIVEIRA, P.T.S., FLANAGAN, D.C., NEARING, M.A. Runoff and soil erosion plot-scale studies under natural rainfall: A meta-analysis of the Brazilian experience. **Catena**, v.152. p.29-39. 2017.
- ARAGÃO, V.R., DE SOUZA, A.C.S., PAOLIELLO, T., SILVA, T.G., LIMA, F.A.J.H.E. Identificação de áreas aptas a barraginhas na Bacia do Rio São Lamberto. **Holos environment**, v.19. n. 2, p.304-319. 2019.
- ASSESSMENT, M.E., **Ecosystems and Human Well-being: Opportunities and Challenges for Business and Industry**. Washington: World Resources Institute, 2005. 31.
- ASSIS, L.M., FRANCELINO, M.R., DAHER, M., FERNANDES-FILHO, E.I., VELOSO, G.V., GOMES, L.C., SCHAEFER, C.E.G.R. Modeling regolith thickness in iron formations using machine learning techniques. **Catena**, v.207. p.105629. 2021.
- AYER, J.E.B., OLIVETTI, D., MINCATO, R.L., SILVA, M.L.N. Erosão hídrica em Latossolos Vermelhos distróficos **Pesqui Agropecu Trop**, v.45. n. 2, p.180-191. 2015.
- BARROS, L., RIBEIRO, P.D.A., **Barraginhas: água de chuva para todos**. Brasília: Embrapa - Empresa Brasileira de Pesquisa Agropecuária, 2009. 49 p.
- BREIMAN, L. Random Forests. **Machine Learning**, v.45. n. 1, p.5-32. 2001.
- CHAGAS, C.D.S., JÚNIOR, W.D.C., PINHEIRO, H.S.K., XAVIER, P.A.M., BHERING, S.B., PEREIRA, N.R., FILHO, B.C. Mapping Soil Cation Exchange Capacity in a Semiarid Region through Predictive Models and Covariates from Remote Sensing Data. **Rev. Bras. Ciên. Solo** v.42. 2018.
- CHEN, Y., YU, J., KHAN, S. Spatial sensitivity analysis of multi-criteria weights in GIS-based land suitability evaluation. **Environ. Model. Software**, v.25. n. 12, p.1582-1591. 2010.
- CHI, M., FENG, R., BRUZZONE, L. Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem. **Adv. Space Res.**, v.41. n. 11, p.1793-1799. 2008.
- CHUNG, J.Y., LEE, S. Dropout early warning systems for high school students using machine learning. **Children and Youth Services Review**, v.96. p.346-353. 2019.
- DEUSCHLE, D., MINELLA, J.P.G., HÖRBE, T.D.A.N., LONDERO, A.L., SCHNEIDER, F.J.A. Erosion and hydrological response in no-tillage subjected to crop rotation intensification in southern Brazil. **Geoderma**, v.340. p.157-163. 2019.
- DIDONÉ, E.J., MINELLA, J.P.G., MERTEN, G.H. Quantifying soil erosion and sediment yield in a catchment in southern Brazil and implications for land conservation. **J. Soils Sediments**, v.15. n. 11, p.2334-2346. 2015.
- FONSECA, L.D.M., LANI, J.L., FILHO, E.I.F., SANTOS, G.R., FERREIRA, W.P.M., SANTOS, A.M.R.T. Variability in soil physical properties in landslide-prone areas **Acta Sci-Agron**, v.39. n. 1, p.109-118. 2017.
- FU, Z., LI, Z., CAI, C., SHI, Z., XU, Q., WANG, X. Soil thickness effect on hydrological and erosion characteristics under sloping lands: A hydrogeological perspective. **Geoderma**, v.167-168. p.41-53. 2011.

- GALLANT, J.C., AUSTIN, J.M. Derivation of terrain covariates for digital soil mapping in Australia %J Soil Research. v.53. n. 8, p.895-906. 2015.
- GALLANT, J.C., DOWLING, T.I. A multiresolution index of valley bottom flatness for mapping depositional areas. **Water Resources Research**, v.39. n. 12, 2003.
- GARCÍA-RUIZ, J.M., BEGUERÍA, S., NADAL-ROMERO, E., GONZÁLEZ-HIDALGO, J.C., LANA-RENAULT, N., SANJUÁN, Y. A meta-analysis of soil erosion rates across the world. **Geomorphology**, v.239. p.160-173. 2015.
- GEOMORFOLOGIA-UFV. Mapa de uso da terra e Áreas de Preservação Permanente - Município Visconde de Rio Branco. **Projeto: Instrumentação, monitoramento e recuperação na Serra da Piedade em Visconde do Rio Branco - MG**, 2019.
- GOMES, L.C., FARIA, R.M., SOUZA, E., VELOSO, G.V., SCHAEFER, C.E.G.R., FERNANDES FILHO, E.I. Modelling and mapping soil organic carbon stocks in Brazil. **Geoderma**, v.340. p.337-350. 2019.
- GRANITTO, P.M., FURLANELLO, C., BIASIOLI, F., GASPERI, F. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. **Chemometrics and Intelligent Laboratory Systems**, v.83. n. 2, p.83-90. 2006.
- GRUBER, S., PECKHAM, S., Chapter 7 Land-Surface Parameters and Objects in Hydrology. In: T. Hengl, H.I. Reuter (Eds.) **Developments in Soil Science**. Elsevier, 2009. 171-194.
- HIPÓLITO, M.T., COSTA, T., BARROS, L., VIANA, A.J.E.M.E.S.S.L., BRAZIL. Alocação de Barraginhas com uso de Modelagem Hidrológica e Geoprocessamento. **Boletim de Pesquisa e Desenvolvimento 190 - Empresa Brasileira de Pesquisa Agropecuária**, p.1679-0154. 2019.
- HOLANDA, F.S.R., ROCHA, I.P.D., OLIVEIRA, V.S. Estabilização de taludes marginais com técnicas de bioengenharia de solos no Baixo São Francisco **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.12. n. 6, p.570-575. 2008.
- JACKSON JR., L., KOSTASCHUK, R., MACDONALD, G., Identification of debris flow hazard on alluvial fans in the Canadian Rocky Mountains. In: J. Costa, G. Wiczeoreck (Eds.) **Debris Flows/Avalanches Recognition and Mitigation**. Boulder, Colorado: The Geological Society of America, 1987. 115.
- JUNIOR, W.C., CALDERANO FILHO, B., DA SILVA CHAGAS, C., BHERING, S.B., PEREIRA, N.R., PINHEIRO, H.S.K. Regressão linear múltipla e modelo Random Forest para estimar a densidade do solo em áreas montanhosas. **Pesquisa Agropecuária Brasileira**, v.51. n. 9, p.1428-1437. 2016.
- KOPECKÝ, M., ČÍŽKOVÁ, Š. Using topographic wetness index in vegetation ecology: does the algorithm matter? v.13. n. 4, p.450-459. 2010.
- KUHN, M., JOHNSON, K., **Applied predictive modeling**. New York: Springer, 2013. 600.
- LABRIÈRE, N., LOCATELLI, B., LAUMONIER, Y., FREYCON, V., BERNOUX, M. Soil erosion in the humid tropics: A systematic quantitative review. **Agriculture, Ecosystems & Environment**, v.203. p.127-139. 2015.
- MALCZEWSKI, J. GIS-based multicriteria decision analysis: a survey of the literature. **International Journal of Geographical Information Science**, v.20. n. 7, p.703-726. 2006.
- MESHAM, D., GORANTIWAR, S.D., WADNE, S.S., ARUN KUMAR, K.C., Planning, Designing and Construction of Series of Check Dams for Soil and Water Conservation in a Micro-watershed of Gujarat, India. In: P.K. Shit, H.R. Pourghasemi, G.S. Bhunia (Eds.) **Gully Erosion Studies from India and Surrounding Regions**. Cham: Springer International Publishing, 2020. 337-343.
- MOHAMMADY, M., POURGHASEMI, H.R., AMIRI, M. Land subsidence susceptibility assessment using random forest machine learning algorithm. **Environmental Earth Sciences**, v.78. n. 16, p.503. 2019.
- NOBRE, C.A., MARENGO, J.A., SELUCHI, M.E., CUARTAS, L.A., ALVES, L.M. Some characteristics and impacts of the drought and water crisis in Southeastern Brazil during 2014 and 2015. **Journal of Water Resource Protection**, v.8. n. 2, p.252-262. 2016.
- OLIVEIRA, P.T.S., WENDLAND, E., NEARING, M.A. Rainfall erosivity in Brazil: A review. **Catena**, v.100. p.139-147. 2013.
- PAL, M. Random forest classifier for remote sensing classification. **International Journal of Remote Sensing**, v.26. n. 1, p.217-222. 2005.

- PIMENTEL, D. Soil Erosion: A Food and Environmental Threat. **Environment, Development and Sustainability**, v.8. n. 1, p.119-137. 2006.
- PROSSER, I.P., ABERNETHY, B. Predicting the Topographic Limits to a Gully Network Using a Digital Terrain Model and Process Thresholds. **Water Resources Research**, v.32. n. 7, p.2289-2298. 1996.
- SAATY, T.L. Exploring optimization through hierarchies and ratio scales. **Socio-Economic Planning Sciences**, v.20. n. 6, p.355-360. 1986.
- SAATY, T.L. The Modern Science of Multicriteria Decision Making and Its Practical Applications: The AHP/ANP Approach. **Operational Research**, v.61. n. 5, p.1101-1118. 2013.
- SAATY, T.L., VARGAS, L.G., The Analytic Network Process. In: T.L. Saaty, L.G. Vargas (Eds.) **Decision Making with the Analytic Network Process: Economic, Political, Social and Technological Applications with Benefits, Opportunities, Costs and Risks**. Boston, MA: Springer US, 2013. 1-40.
- SAKUNO, N.R.R., GUIÇARDI, A.C.F., SPALEVIC, V., AVANZI, J.C., SILVA, M.L.N., MINCATO, R.L. Adaptation and application of the erosion potential method for tropical soils. **Revista Ciência Agronômica**, v.51. n. 1, p.e20186545. 2020.
- SCHIETTECATTE, W., OUESSAR, M., GABRIELS, D., TANGHE, S., HEIRMAN, S., ABDELLI, F. Impact of water harvesting techniques on soil and water conservation: a case study on a micro catchment in southeastern Tunisia. **Journal of Arid Environments**, v.61. n. 2, p.297-313. 2005.
- SENA, N.C., VELOSO, G.V., FILHO, E.I.F., FRANCELINO, M.R., SCHAEFER, C.E.G.R. Analysis of terrain attributes in different spatial resolutions for digital soil mapping application in Southeastern Brazil. **Geoderma Regional**, p.e00268. 2020.
- SICAT, R.S., CARRANZA, E.J.M., NIDUMOLU, U.B. Fuzzy modeling of farmers' knowledge for land suitability classification. **Agricultural Systems**, v.83. n. 1, p.49-75. 2005.
- SOUZA, C.M.P., FIGUEIREDO, N.A., COSTA, L.M., VELOSO, G.V., ALMEIDA, M.I.S., FERREIRA, E.J. Machine learning algorithm in the prediction of geomorphic indices for appraisal the influence of landscape structure on fluvial systems, southeastern - Brazil. **Revista Brasileira de Geomorfologia**, v.21. n. 2, p.363-378. 2020.
- SOUZA, C.M.P., THOMAZINI, A., SCHAEFER, C.E.G.R., VELOSO, G.V., MOREIRA, G.M., FERNANDES FILHO, E.I. Multivariate analysis and machine learning in properties of ultisols (Argissolos) of brazilian Amazon. **Revista Brasileira de Ciência do Solo** v.42. p.1 - 20. 2018.
- VALENTIN, C., POESEN, J., LI, Y. Gully erosion: Impacts, factors and control. **Catena**, v.63. n. 2, p.132-153. 2005.
- VAN REMORTEL, R.D., HAMILTON, M.E., HICKEY, R.J. Estimating the LS Factor for RUSLE through Iterative Slope Length Processing of Digital Elevation Data within ArcInfo Grid. **Cartography**, v.30. n. 1, p.27-35. 2001.
- WITTEN, I.H., FRANK, E., HALL, M.A., PAL, C.J., **Data Mining: Practical machine learning tools and techniques**. Cambridge: Morgan Kaufmann, 2016. 525.
- WUEPPER, D., BORRELLI, P., FINGER, R. Countries and the global rate of soil erosion. **Nature Sustainability**, v.3. n. 1, p.51-55. 2020.
- XU, J., WEI, W., BAO, H., ZHANG, K., LAN, H., YAN, C., SUN, W.J.B.O.E.G., ENVIRONMENT, T. Failure models of a loess stacked dam: a case study in the Ansai Area (China). **Bulletin of Engineering Geology and the Environment**, 2019.
- YESILNACAR, E., TOPAL, T. Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). **Engineering Geology**, v.79. n. 3, p.251-266. 2005.

Recebido em: 08/06/2020

Aceito para publicação em: 24/09/2021