Environmental critical thresholds based on statistical analysis for modelling landslide susceptibility in Continental Basaltic Provinces

Jennifer Fortes Cavalcante Renk[1], Tatiana Sussel Gonçalves Mendes[2], Silvio Jorge Coelho Simões[1], Marcio Roberto Magalhães de Andrade[3], Luana Albertani Pampuch Bortolozo[2], Adriano Martins Junqueira[4], Melina Almeida Silva[2]

[1] Graduate Program in Natural Disasters (UNESP/CEMADEN), São José dos Campos - SP, Brazil - jennifer.renk@unesp.br, silvio.simoes@unesp.br
[2] Institute of Science and Technology, São Paulo State University (UNESP), São José dos Campos - SP, Brazil – tatiana.mendes@unesp.br, luana.pampuch@unesp.br, melina.almeida@unesp.br
[3] National Center for Monitoring and Early Warning of Natural Disasters (CEMADEN), São José dos Campos - SP, Brazil – marcio.andrade@cemaden.gov.br
[4] School of Engineering, São Paulo State University (UNESP), Guaratinguetá – SP, Brazil - adriano.junqueira@unesp.br

Keywords: Natural Disasters, Landslides Susceptibility, Frequency Ratio, Pearson's Linear Correlation Coefficient, Continental Basaltic Provinces

Abstract

The study aims to estimate the environmental critical thresholds using statistical approaches to understand the landslide conditioning factors that can trigger landslides in the Continental Basaltic Provinces a landslide-prone area, using as reference the landslides that occurred in an extreme rainfall event. The study area is a region that was the scene of an extreme hydrological event in January 2017, with an accumulated volume of rain of 163.9 mm in 8 hours, causing a widespread event of shallow planar landslides with more than 400 scars detected. Hydrological, anthropic, geological, geomorphological, and topographical features of this region were analyzed considering landslides and non-landslides samples set, and their influence in the event was carried out using the Frequency Ratio method, followed by Pearson's Linear Correlation Coefficient and Linear Regression. The results showed that this process helped us to understand environmental critical thresholds based on classes of conditioning factors that have a greater influence on rainfall-triggered landslide occurrences and, consequently, higher predictive capacity in the landslide susceptibility models with the same geoenvironmental parameters which is a valuable insight for risk management.

1. Introduction

Shallow planar landslides are often triggered by extreme rainfall and are related in particular to saprolitic or lateritic residual soils that form a weaker layer than the underlying material on steep slopes, forming a layer that can persist deposited under the cementation formed by negative pore pressure due to incomplete unsaturation or vegetation root reinforcement, but can be destroyed by the downward advance of a wetting front that is conducive to triggering avalanches or debris flows and Landslides (Hungr et al., 2014).

The Serra Gaúcha, in southern Brazil and southeastern South America, was hit by a rainfall event on January 5, 2017, that lasted approximately 4 hours and showed rainfall accumulation similar to that accumulated throughout the month (SEMA, 2017), which caused widespread shallow landslides on steep slopes in the Mascarada River basin, resulting in widespread mass flow that mobilized a large volume of sediment and showed high destructive potential.

The region affected by the extreme weather event corresponds to the Mascarada River basin, which is arranged in residual plateau edge levels and is marked by a structural step with three cascades in the middle-upper part where is the epicenter of the shallow landslides occurred. After the 2017 event, in the middle-lower third of the basin, traces of flows were found that led to the formation of a debris fan and a mud delta.

This region is based on geological units of ancient basaltic flows in thick Paleo-Mesozoic sedimentary sequence and the geomorphological characteristics with mountain ranges and plateau escarpments which have a saprophytic or lateritic residual soil that are substantially weaker than the underlying original material are very prone to landslides.

Landslide Susceptibility Mapping (LSM) is an important component of geological disaster prevention and control (Yong et al., 2022). It is a procedure used to predict the probability of landslide occurrence based on the spatial distribution of the sample set and the Landslide Conditioning Factors (LCF) (Poddar and Roy, 2024) related to the characteristics of the physical or geo-environmental conditions (Moghimi et al., 2024). Susceptibility models can be classified into deterministic or data-driven models (Huang et al., 2020; Poddar and Roy, 2024). Based on data, statistical approaches aim to analyze the spatial distribution of past landslides to indicate the probability of likely areas of new landslides under similar conditions (Reichenbach et al., 2018; Segue et al., 2024; Moghimi et al., 2024).

These LCF can be classified as dynamic or static data depending on the duration, frequency of acquisition, and processing of the data to be integrated into models. LCF are database features calculated from hydrological, anthropic, geological, geomorphological, and topographical data considered in this research. Furthermore, other researchers consider other features such as human activities (Chen et al., 2018), anthropogenic factors, climate, radar features, soil permeability conditions, earthquakes (Moghimi et al., 2024), or Normalized Difference Vegetation Index (NDVI) to assess vegetation coverage (Poddar and Roy, 2024).

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-2024-463-2024 | © Author(s) 2024. CC BY 4.0 License.

A landslide scar inventory is necessary for statistical analysis, which can be obtained manually by vectoring using satellite imagery as a reference or using more sophisticated approaches methods such as object-based image analysis (OBIA) and Deep Learning (DL) that can update the inventory map faster regarding the new events landslide (Ghorbanzadeh et al., 2022; Moghimi et al., 2024).

Using a river basin as a unit of analysis, the study aims to estimate the Environmental Critical Thresholds (ECT) through statistical analysis of Frequency Ratio (FR), Pearson's Linear Correlation Coefficient, and Linear Regression to understand the LCF that can trigger landslides in Continental Basaltic Provinces, using as a reference an event that occurred, in 2017.

The ECT or weight approach create model flexibility to integrate Machine Learning models (ML) with Geographic Information System (GIS), Remote Sensing (SR) and Artificial Intelligence (AI) to develop innovative LSM approaches (Moghimi et al., 2024) and it is important to understand the LCF that can trigger landslides (Poddar and Roy, 2024).

2. Study Area and Data Set

The Mascarada River Basin (Figure 1) covers around 295.89 km² in the northern region of the Porto Alegre Metropolitan, in the state of Rio Grande do Sul. The area is situated in the Morphostructural Unit of the Campos Gerais Plateau around 38,863.15 km² on the edge of the Paraná Sedimentary Basin.

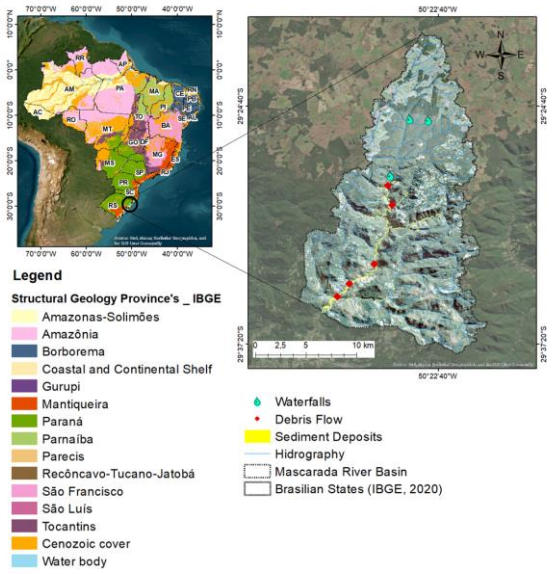

Figure 1. The Mascarada River Basin in the geological structural provinces.

The area of interest is based on geological units of ancient Basalt Paleo-Mesozoic thickness sedimentary sequences on mountain ranges and plateau escarpments and are transition areas between plateaus and depressions that have saprophytic or lateritic residual soils that are substantially weaker than the original underlying material, which geomorphological characteristics are very conducive to high-intensity landslides.

On January 5, 2017, an extreme rainfall event (Figure 2) characterized as a convective system occurred in the region of the municipality of Rolante, causing hundreds of landslides in the Mascarada River basin. The event was analyzed using rainfall data from the National Center for Monitoring and Natural Disasters (Cemaden) and radar data that made it possible to obtain the dynamics of the rain event in relation to the duration and distribution of rainfall in the region where the landslides began.

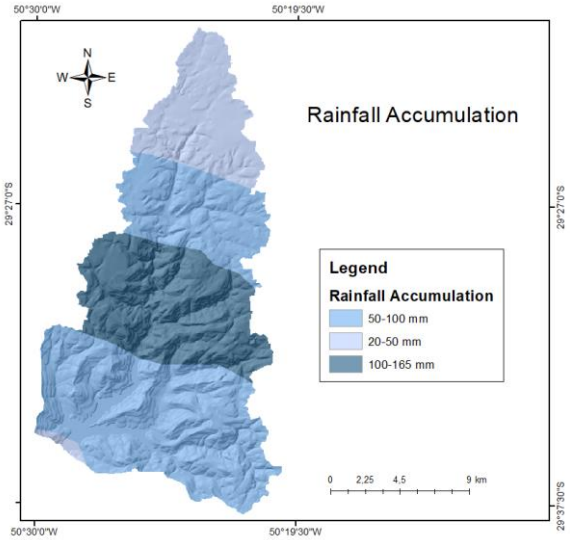

Figure 2. The extreme rainfall event in Mascarada River Basin.

This convective system reached a radius of up to 46 km away, with a more intense nucleus coinciding with the center of the landslides and with a diameter varying around 16 to 28 km. The radar data showed that the center of the convective phenomenon detected in the area with the highest rainfall coincides perfectly with the center of landslide concentration, where it reached a volume of 163.9 mm in 8 hours. The phenomenon clearly demonstrates a strong relationship with the Serra Gaúcha escarpment on the border with the Santa Catarina plateau.

To study this phenomenon together with other LCF related to landslides, the following data were used:

- Meteorological radar data from Morro da Igreja (DTCEA) on January 5, 2017, with the accumulated volume of rain in the 8 hours of the event from 3 pm by value classes and data from Cemaden rainfall stations.
- Planet images of a spatial resolution of 3.5 m from before the event (December 06, 2016). The image mosaic was used to generate a Land Use and Land Cover (LULC) map on a 1:10,000 scale.
- RapidEye images of a spatial resolution of 5 m after the event (March 13, 2017). The image was used to extract a landslide scar inventory and sedimentary deposits on a 1:10,000 scale.
- Digital Elevation Model (DEM), generated with 10 m spatial resolution, derived from the continuous vector cartographic base of Rio Grande do Sul – scale 1:50,000 (Hasenack and Weber, 2010).

For this research, the software available at the Geoprocessing and computer laboratory at ICT/UNESP and the Scikit-Learning library (available at https://scikit-learn.org/stable/; Pedregosa et al., 2011), used in Anaconda Navigator development environment with the Jupyter Notebook application. Among the software used, there are:

- ArcMAP and ArcGIS Pro (ESRI): for pre-processing data and generating information plans related to LCF.

- MATLAB® version 9.13.0 (R2023a): for statistical correlation analyses.

3. Methodology

The methodology consists of manually mapping landslide scars to create training and test sample sets representing landslides and non-landslides, generating LCF, determining the ECT based on FR, and performing correlation analysis of these sample sets with the LCF. This best represents the weight approach to develop the LSM models.

3.1 Inventory Map and Samples Set

The landslide scars were manually obtained as interpreted polygons based on a scene from the RapidEye satellite images after the event (Figure 3) which resulted on the total amount of 306 landslide scars extracted in the study area.

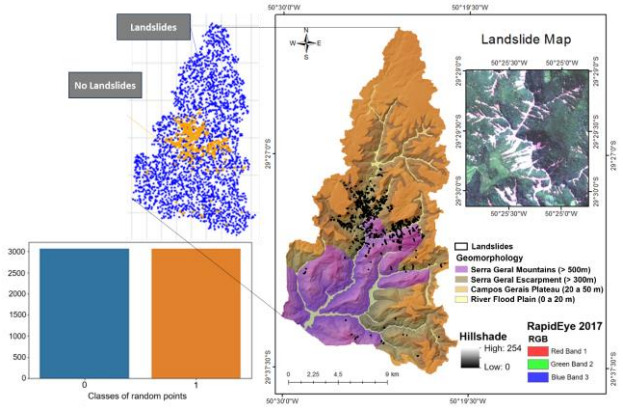

Figure 3. Landslide scars from the January 5, 2017, event in the Mascarada River basin and points of landslide and non-landslide randomly sampled.

Landslides points were randomly obtained from the limit of the landslide’s scars coverage, totaling an amount of 3,060 points. The same number of points was obtained randomly for the remaining Mascarada River basin, where no landslide events occurred. In total, 6,120 points of landslides (value 1) and non-landslides (value 0) were produced to use as a sample (Figure 3).

3.2 Landslide conditioning factors

Choosing the LCF is an important step for the rainfall-triggered landslide susceptibility model, as the selection of these factors influences the predictive capacity of the models. In this work, 15 were selected based on the landslide modeling literature (Catani et al., 2013; Tien Bui et al., 2016; Park and Kim, 2019; Dang et al., 2020; Zhou et al., 2021).

The LCF can be subdivided into meteorological, anthropic, geological, geomorphological, hydrological, and topographic. Figure 4 presents the LCF divided according to these typologies and a brief description is presented as follows.

The rainfall accumulation map was generated from 8-hour accumulated data from the Morro da Igreja meteorological radar DTCEA).

The LULC map was produced from a mosaic of Planet images from 2016 (before the event) through supervised classification using the Support Vector Machine algorithm followed by manual editing.

The structural lineaments density map was obtained based on lineaments extracted using shaded relief derived from the DEM with azimuths equal to 315º and 45º and a solar angle of 45º. Lineaments as aligned ridges and floodplain downgraded valleys were mapped to analyze the role of the geological substrate in controlling the development of concave relief forms (Avelar and Neto, 1992), which explain the hydrological and erosive responses of the landscape and define potential areas of slope instability.

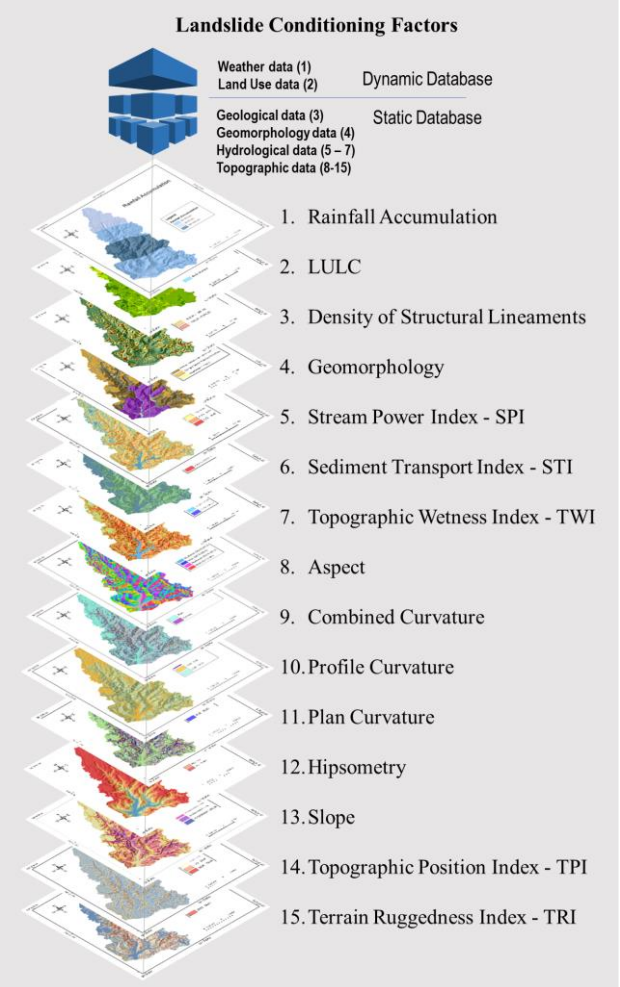

Figure 4. Characterization of the conditioning factors used according to their typology.

The structural lineaments density map was obtained based on lineaments extracted using shaded relief derived from the DEM with azimuths equal to 315º and 45º and a solar angle of 45º. Lineaments as aligned ridges and floodplain downgraded valleys were mapped to analyze the role of the geological substrate in controlling the development of concave relief forms (Avelar and Neto, 1992), which explain the hydrological and erosive responses of the landscape and define potential areas of slope instability.

The geomorphological map was developed based on the elevation range, slope, and shaded relief derived from the DEM to represent the morphology, structure, genesis, and chronology

of the landscape. The geomorphological classes mapped were based on the Geodiversity map of Rio Grande do Sul (Vieiro, 2010) and the Relief Pattern Library (Dantas, 2016), both produced by the Mineral Resources Research Company (CPRM, in Portuguese).

The following LCF were generated from the DEM. The Stream Power Index (SPI) characterizes the process of erosion and sediment deposition, which reflects the topographic effects of soil loss according to Burrough and MCDonnell (1998), which is similar to the slope factor included in the Universal Soil Loss Equation (USLE) (Wischmeier and Smith, 1978) and was applied to three-dimensional surfaces by Moore et al. (1991).

The Sediment Transport Index (STI) represents the runoff model (Beven and Kirkby, 1979) that describes the sediment transfer capacity in channel streams and has been used as a measure of the erosive power of flowing water (Moore et al., 1991) based on the assumption that discharge is proportional to the specific catchment area about the net erosion in areas of profile with convexity and tangential concavity (flow acceleration and convergence zones) and net deposition in areas of profile concavity (zones of decreasing flow velocity) (Gallant and Wilson, 2000).

The Topographic Wetness Index (TWI) represents the spatial distribution of humidity and extension of saturation zones for runoff generation as a function of the slope contribution area, soil transmissivity, and slope gradient, therefore assuming, steady-state conditions, and uniform soil properties (Gallant and Wilson, 2000). This variable has been useful for quantifying the effects of topography on hydrological processes and indicating the balance between water accumulation and drainage conditions according to Bannari et al. (2017).

The slope orientation map (aspect) aims to evaluate the angle of solar incidence on the topographic surface, evapotranspiration tendency, and the distribution and abundance of flora and fauna (Gallant and Wilson, 2000).

The combined curvature map has a three-dimensional approach to slope systems as a basis for the pedogenetic evolution of the relief. Thus, the geometric relationships (plan versus profile) coupled with the subsurface structure aim to individualize the Soil-Landscape system (Hugget, 1975).

The profile curvature map was generated from DEM and characterizes the maximum slope along a flow line and the degree of downward acceleration or deceleration in the landscape (Gallant and Wilson, 2000) as negative for the convex flow profile, typical of higher slopes, and positive for the concave profile, typical of lower slopes. The objective is to indicate the relief forms of the “ramp complexes” that represent the reworking of subsurface materials associated with episodes of formation of colluvial deposits (Meis and Monteiro, 1979).

The plan curvature map represents the areas perpendicular to the direction of the slope and the curvature of the plane that characterizes the degree of convergence or divergence of the flow in the landscape to represent the amount of water in the soil and the characteristics of the soil, being negative for divergent flow along ridges and positive for convergent areas such as along valley bottoms (Gallant and Wilson, 2000).

The hypsometric map represents the cumulative distribution of elevations in a geographic area to demonstrate climatic significance for vegetation and measure the energy potential of the environment (Gallant and Wilson, 2000).

The slope map indicates the slope gradient in degrees (ranging from 0° to 90°) for each grid cell to express the relationship between the difference in height between two points and the horizontal distance between these points (Valeriano, 2002) and how it influence the terrestrial and underground flow, the speed, the rate of runoff, the distribution of vegetation, landform geomorphology and the amount of water in the soil (Gallant and Wilson, 2000).

The Topographic Position Index (TPI) compares the average elevation of a specific neighborhood around a cell to isolate specific landscape features such as canyon bottoms, steep slopes, or other topographic positions (Guisan et al., 1999).

The Terrain Ruggedness Index (TRI) characterizes a measure of the local topographic relief regarding its height (Riley et al., 1999) according to the calculation of the mean squared deviation for each grid cell and its eight neighbors.

3.3 Frequency Ratio to Estimate the Environmental Critical Thresholds

The FR is a bivariate statistical probabilistic method based on the relationship between the distribution of landslides and the LCF in the study area that reveal the landslide density within a specific class (Pradhan, 2010; Park and Kim 2019; Sahin et al., 2020, Yang et al., 2020, Poddar and Roy, 2024).

The FR can be described by Equation 1, which will be used for all classes of landslide LCF based on their relationships with the scar inventory (Oh et al., 2011). This method was applied based on landslide scars as a raster image.

$$FR = \frac{(A/B)}{(C/D)} \quad (1)$$

where A is the number of landslide pixels for each class LCF; B is the total number of landslide pixels in the study area; C is the number of pixels in the class area of each LCF, and D is the total number of pixels in the study area.

Then, the RF was normalized to FR on a scale of probability values from 0 to 1 using Equation 2:

$$LRF = \frac{Factor\ class\ FR}{\sum Factor\ class\ FR} \quad (2)$$

The normalization aims for a bivariate probability analysis, which if equal to 1, has a strong correlation between the incidence of landslides and each class of LCF, and if the probability is less than 1, a weaker correlation (Tien Bui et al., 2016; Park and Kim, 2019, Poddar and Roy, 2024).

This method explains the landslide event and can be used as an attribute selection factor according to the calculated weights and to estimate the ECTs that determine the predisposition of each class of the landslide LCF. Also, it can be useful in calibrating landslide susceptibility models.

3.4 Pearson's Linear Correlation Coefficient

Pearson's Linear Correlation Coefficient is an indicator that reflects the degree of correlation between two variables (Gu et

al., 2023), describing how strong the linear relationship is between the two variables and their direction (positive or negative). The Pearson's Linear Correlation Coefficient is used to find the correlation between the landslides and non-landslides samples set with each LCF.

That way, the *X* columns correspond to the observations and the *Y* columns correspond to the variables or LCF. For column *X* and column *Y*, Pearson's linear correlation coefficient *r(X,Y)* is defined by Equation 3.

$$r(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \quad (3)$$

Where *n* is the number of landslides and non-landslides samples set, X_i and Y_i are the values of the *i-th* sample data in *X* and *Y*, respectively, and $\bar{X}$ and $\bar{Y}$ are the averages of *X* and *Y*, respectively.

Pearson Correlation Coefficient values range from –1 to +1, where a value of –1 indicates perfect negative correlation, while a value of +1 indicates perfect positive correlation, and a value of 0 indicates that there is no correlation between the columns (MATLAB® version 9.13.0).

Pearson's Linear Correlation Coefficient returns the p-value (*pval*) to test the null hypothesis, which assumes that the statistical result was obtained by chance. P-values range from 0 to 1, where values close to 0 correspond to a significant correlation and a low probability of observing the null hypothesis. A low p-value (below a significance level, in this case 0.05) suggests that observed correlation is statistically significant, meaning that it is unlikely to have occurred by chance.

Afterward, the linear adjustment was determined through the Coefficient of determination (R^2) which is the square of the correlation coefficient. R^2 is a measure of how much of the variance in the dependent variable is explained by the independent variables X in the linear regression model. It ranges from 0 to 1. The larger the R-squared is, the more variability is explained by the linear regression model.

4. Results

The ECT admitted relating to the landslide event analyzed in the study area are obtained by the highest weight according to the FR are available in Figures 5 and 6 and detailed below:

- In the rainfall accumulation data, the class that triggered the landslide event has about 97% of the scars in the 100-165 mm range and 3% in the 50-100 mm range.
- In LULC, scars are predominantly located in the classes of Native Tree Vegetation (66%), and Agricultural Culture (10%).
- The scars are mainly found in moderate to high classes of the structural lineament’s density, mainly in areas of channelized and non-channelized embedded valley bottoms formed from the development of concave relief forms, which explain hydrological responses and erosion of the landscape and define potential zones of slope instability.
- By geomorphological data, the scars at the epicenter of the event are located mainly in altitude ranges between > 300 m (51%) and > 500 m (25%) referring to the morphology of the Serra Geral escarpment, which has shallow soils of the Neossols type and slope > 45º in relation to the river plain areas.
- SPI indicated that the scars have an epicenter around moderate to very high indices, with around 96%, which represent areas with convexity and tangential concavity where flow acceleration and convergence zones predominate.
- STI indicated that the scars have an epicenter around moderate to high rates (around 84%) that indicate higher soil losses.
- TWI indicated that the scars have an epicenter around the very low indices below (around 60%) and very high (around 21%, indicating the effects of topography on hydrological processes and accumulation of water in saturation zones to generate flow.

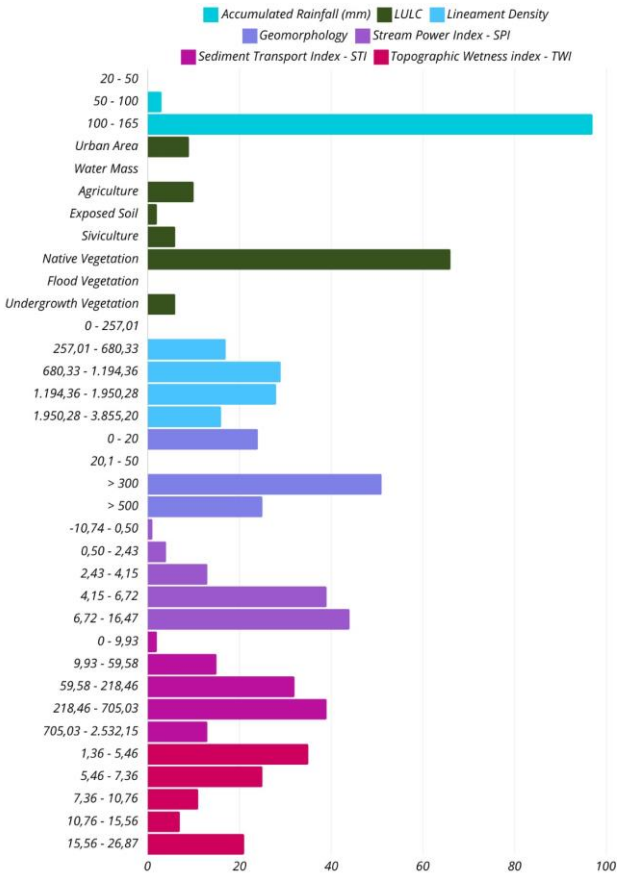

Figure 5. Graphic showing the FR results for conditioning factors (Accumulated rainfall, LULC, Lineament Density, Geomorphology, SPI, STI and TWI) and their corresponding ECT.

- In Aspect, the scars are located in structures facing mainly N-L orientation (more than 50% of them), which seems to correlate with the front of the meteorological phenomenon.
- In combined curvature, scars occurred around 63% in concave areas of the relief.
- In profile curvature, the scars occurred mainly, around 64%, in negative areas with a convex flow profile, typical of higher slopes, and around 22%, in positive areas for the concave profile, typical of lower slopes that have the objective indicate the relief forms of the

"ramp complexes" that represent the reworking of subsurface materials associated with episodes of formation of colluvial deposits (Meis and Monteiro, 1979).

- In the plan curvature, the scars occurred mainly in the convergent areas, along the valley bottoms, and in negative areas with about 41% for the divergent flow along the crests.
- The hypsometry demonstrated that the scars developed at an amplitude of 300 to 750 m.
- The slope demonstrated that the epicenter of the scars is located on slopes > 45º (52%), and 35-45º (27%).
- TPI indicated that the scars have an epicenter around very low to moderate indices (around 93%) which characterize areas with medium elevation as canyon bottoms and steep slopes.
- TRI indicated that scars have an epicenter around moderate to very high indices, with around 43% of high roughness of relief and elevation.

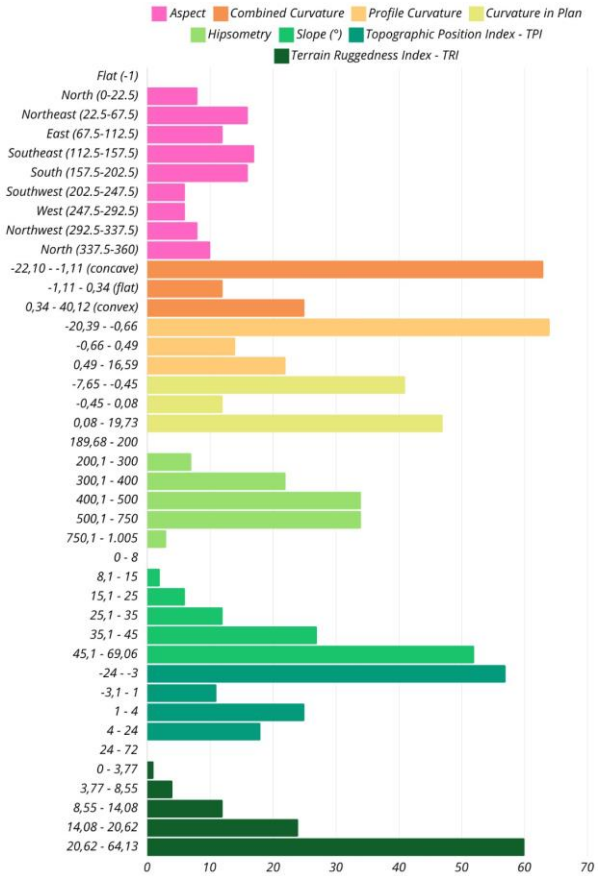

Figure 6. Graphic showing the FR results for conditioning factors (Aspect, Combined Curvature, Profile Curbature, Curvature in Plan, Hipsometry, Slope, TPI and TRI) and their corresponding ECT.

Pearson's Linear Correlation Coefficient (rho) between the LCF (matrix with stacked rasters) and the landslides and non-landslides samples set indicated a positive correlation between the two columns since no value was lower than 0. Thus, all the LCF presented a positive correlation as shown in Figure 7. The rainfall accumulation was the one that have the greatest correlation (0.68), indicating a moderate correlation. TRI, Slope, Geomorphology, hypsometry, STI, and SPI also present moderate correlation, with pho ranging 0.40 to 0.47.

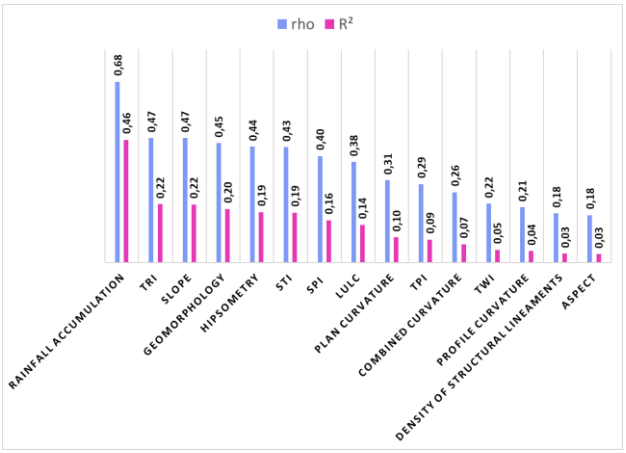

Figure 7. Graphics represent the Pearson's Linear Correlation Coefficient (rho) between the conditioning factors and coefficient of determination R^2.

Figure 7 also demonstrated a very weak correlation (rho lower 0.20) for aspect and density of structures lineaments LCF based on the distribution of random samples selected.

The p-value analysis indicated that all LCF present a significant correlation (with a significance level of 0.05) since the p-value close to 0 corresponds to a correlation in R^2 and a low probability of observing the null hypothesis.

The coefficient of determination R^2 (Figure 7) indicates that 46% of the set of landslide and non-landslide samples were triggered by the extreme rainfall events of 2017, with rainfall accumulation being the LCF. While TWI, profile curvature, density of structure lineaments, and aspect explain values smaller than 5% of the sample set.

5. Discussion

Of the total number of landslide scars interpreted and extracted by the image, around 74.26% are in the Plateau Edge Escarpment geomorphological domain, with a relief range greater than 300 m and a slope greater than 45°. Of the remaining scars, around 25.74% are in the Serra Geral geomorphological domain, which has a relief amplitude of over 500 m and a slope above 45°. It is worth noting that both reliefs have shallow Neossolo Litólico soils.

High slopes are not always indicative of high susceptibility to landslides (Gameiro et al., 2022). Landslides on slopes above 30° are not as common, however, extreme rainfall events such as occurred in this region can induce landslides on greater slopes (Riffel et al., 2021). Gameiro et al. (2019) also report landslides with slopes above the average in the study for the region.

It is worth considering that the LCF anthropic represented by LULC has so far been little explored in studies related to these landslide events. In this work, it was found that 66% of landslides are related to the native vegetation class, showing that anthropized areas did not have a greater influence on the triggering of the event.

The rainfall accumulation was the one that have the greatest influence on landslide occurrences, with a correlation value (rho) of 0.68 and a coefficient of determination R^2 equal to 0.46

(46%). This shows that most landslides coincide with the epicentre of the hydrological event.

Others LCF such as slope, TWI, geomorphology, hypsometry, STI, and SPI also showed a moderate correlation with the sample set, indicating their importance in the susceptibility model.

LCF least correlated with the landslide and non-landslide samples set were TWI, profile curvature, density of structure lineaments, and aspect. Gameiro et al. (2019) also verified that aspect or orientation and TWI presented the lowest correlation with the event in the study area.

6. Conclusion

ECT for landslide events that occurred in 2017 at the Mascarada River basin were investigated based on 15 LCF (meteorological, anthropic, geological, geomorphological, hydrological, and topographic) using the FR method and correlation analysis to identify the integration of geoenvironmental parameters and landslide prone area for Continental Basaltic Provinces.

ECT derived by the FR method can be tested in other Basaltic Provinces outcrops found in various places around the world or can be used in mass movement prediction models in areas of interest with geological and similar geomorphological characteristics also, which is a valuable insight for risk management.

The Pearson Correlation analysis between the LCF and the samples set is an important phase antecedent to the landslide susceptibility model allows finding LCF more correlated and best explain landslide events, as well as those that do not contribute to their understanding, and may even degrade the results in the susceptibility model.

This sample set of landslides and non-landslides will be tested in LSM using ML algorithms, such as Random Forest. Moreover, it could be used in hybrid models for innovative LSM

References

Avelar, A.S., Netto, A.L.C., 1992. Fraturas e Desenvolvimento de Unidades Geomorfológicas Côncavas no Médio Vale do Rio Paraíba do Sul. *Revista Brasileira de Geociências*, 22(2), 222-227.

Bannari, A., Ghadeer, A., El-Battay, A., Hameed, N., Rouai, M., 2017. Detection of Areas Associated with Flash Floods and Erosion Caused by Rainfall Storm Using Topographic Attributes, Hydrologic Indices, and GIS. *Global Changes and Natural Disaster Management: Geo-Information Technologies*, p. 155–174. doi:10.1007/978-3-319-51844-2_13.

Beven, K.J., Kirkby, M.J.A., 1979. Physically Based, Variable Contributing Area Model of Basin Hydrology. Un modèle à base physique de zone d’appel variable de l’hydrologie du bassin versant. *Hydrological Sciences Bulletin*, 24, 43-69.

Burrough, P.A., McDonnell, R.A., 1998. Principles of Geographical Information Systems. *Oxford*, Oxford University.

Chen, W., Peng, J., Hong, H., Shahabi, H., Pradhan, B., Liu, J., Zhu, A.-X., Pei, X., Duan, Z., 2018. Landslide susceptibility modelling using GISbased machine learning techniques for Chongren county, Jiangxi province, China. *Science of the Total Environment*, 626, 1121–1135.

Catani, F., Lagormarsino, D., Segoni, S., Tofani, V., 2013. Landslide susceptibility estimation by random forest technique: sensitivity and scaling issues. *Natural Hazards and Earth* System Sciences, 13(11), 2815–2831. doi:10.5194/nhess-13-2815.

Dantas, M.E., 2016. Biblioteca de padrões de relevo: carta de suscetibilidade a movimentos gravitacionais de massa e inundação. Software. Open Source Geospatial Foundation. http://rigeo.cprm.gov.br/jspui/handle/doc/16589 (21 Jun. 2023).

Dang, V.H., Hoang, N.D., Nguyen, L.M.D., Bui, D.T., Samui, P., 2020. A Novel GIS-Based Random Forest Machine Algorithm for the Spatial Prediction of Shallow Landslide Susceptibility. *Forests*, 11(1), 118. doi:10.3390/f11010118

Gallant, J.C., Wilson, J.P., 2000. *Primary topographic attributes.* In. Terrain Analysis: Principles and Applications. 51-86.

Gameiro, S., de Oliveira, G.G., Guasselli, L.A., 2022. The influence of sampling on landslide susceptibility mapping using artificial neural networks. *Geocarto International*, 1-23. https://doi.org/10.1080/10106049.2022.2144475

Gameiro, S., Quevedo, R.P., Oliveira, G.G., Ruiz, L.F.C., Guasselli, L.A. 2019. Análise e correlação de atributos morfométricos e sua influência nos movimentos de massa ocorridos na Bacia do Rio Rolante, RS. *In:* SBSR Ann. of XIX Simpósio Brasileiro de Sensoriamento Remoto.

Gu, T., Li, J., Wang, M., Duan, P., Zhang, Y., Cheng, L., 2023. Study on landslide susceptibility mapping with different factor screening methods and random forest models. *PLoS one*, 18 (10).

Guisan, A., Weiss, S.B., Weiss, A.D., 1990. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology*, 143(1), 107–122. doi.org/10.1023/A:1009841519580.

Ghorbanzadeh, O., Shahabi, H., Crivellari, A., Homayouni, S., Blaschke, T., Ghamisi, P., 2022. Landslide detection using deep learning and object-based image analysis. *Landslides*, 19, 929–939. doi.org/10.1007/s10346-021-01843-x.

Hasenack, H., Weber, E. (org.), 2010. Base cartográfica vetorial contínua do Rio Grande do Sul - escala 1:50.000. Porto Alegre: UFRGS Centro de Ecologia, Open Source Geospatial Foundation.
https://www.ufrgs.br/labgeo/index.php/publicacoes/serie-geoprocessamento/ (01 Jan. 2022).

Huang. F., Cao, Z., Guo, J., Jiang, S-H., Li, S., Guo, Z., 2020. Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *Catena*, 191, 104580.

Hugget, R.J., 1975. Soil Landscape System: a model of soil genesis. *Geoderma*, 75, 1-22.

Hungr, O., Lerqueil, S., Picarelli, L., 2014. The Varnes classification of landslide types, an update. *Landslides*, 11, 167–194. doi.org/10.1007/s10346-013-0436-y.

MATLAB®. Linear or rank correlation. Available at: <https://www.mathworks.com/help/stats/corr.html> Accessed on 26 nov. 2023.

Marzoli, A., 1999. Extensive 200-Million-Year-Old Continental Flood Basalts of the Central Atlantic Magmatic Province. *Science*, 284(5414), 616–618. doi:10.1126/science.284.5414.616

Meis, M.R.M., Monteiro, A.M.F., 1979. Upper Quaternary "Rampas": Doce River Valley, SE Brazilian Plateau. Z. *Geomorphology*, 23(2), 132-151.

Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrological Processes*, 5(1), 3-30.

Moghimi, A., Singha, C., Fathi, M., Pirasteh, S., Mohammadzadeh, A., Varshosaz, M., Huang, J., Li, H., 2024. Hybridizing genetic random forest and self-attention based CNN-LSTM algorithms for landslide susceptibility mapping in Darjiling and Kurseong, India. *Quaternary Science Advances*, 14. 100187. ISSN 2666-0334.doi.org/10.1016/j.qsa.2024.100187.

Oh H.-J., Kim Y.-S., Choi J.-K., Park E., Lee S., 2011 GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea. Journal of Hydrology, 399(3), 158–172.

Park, S., Kim, J., 2019. Landslide Susceptibility Mapping Based on Random Forest and Boosted Regression Tree Models, and a Comparison of Their Performance. *Applied Sciences*, 9(5), 942. doi:10.3390/app9050942

Pedregosa, F., Varoquaux, G., Gramfort, A. et al., 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(2011), 2825-2830.

Pradhan, B., 2010. Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. *Journal of the Indian Society of Remote Sensing*, 38(2), 301–320. doi:10.1007/s12524-010-0020-z.

Poddar, I., Roy, R., 2024. Application of GIS-based data-driven bivariate statistical models for landslide prediction: a case study of highly affected landslide prone areas of Teesta River basin, *Quaternary Science Advances*, 13, 100150. https://doi.org/10.1016/j.qsa.2023.100150

Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M., Guzzetti, F.A., 2018. Review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, 180, 60–91. https://doi.org/10.1016/j.earscirev.2018.03.001

Riffel, E.S., Guasselli, L.A., Ruiz, L.F.C., Gameiro, S., 2021. Relação entre ponto de ruptura e padrão morfométrico em deslizamentos, bacia hidrográfica do Rio Rolante-RS. *Revista do Departamento de Geografia*, 41, 181554-181554.

Riley, S.J., Degloria, A.S.D., Elliot, R.A., 1999. Terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences*, 5, 23–27.

SEMA (Secretaria do Ambiente e Desenvolvimento Sustentável), 2017. *Diagnóstico Preliminar: Descritivo dos eventos ocorridos no dia 5 de janeiro de 2017 entre as regiões dos municípios de São Francisco de Paula e Rolante/RS*. Secretaria do Ambiente e Desenvolvimento Sustentável, Porto Alegre.

Segue, W.S., Njilah, I.K., Fossi, D.H., Nsangou, D., 2024. Advancements in mapping landslide susceptibility in Bafoussam and its surroundings area using multi-criteria decision analysis, statistical methods, and machine learning models. *Journal of African Earth Sciences*, 213, 105237.

Tien Bui, D., Tuan, T.A., Klempe, H. et al., 2016. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides*, 13, 361–378. https://doi.org/10.1007/s10346-015-0557-6

Valeriano, M.M., 2002. Programação do cálculo da declividade em SIG pelo método de vetores ortogonais. *Espaço e Geografia* (UnB), 5(1), 69-85.

Vieira, S.M., Sousa, J.M.C., Kaymak, U., 2012. Fuzzy criteria for feature selection. *Fuzzy Sets and Systems.* 189(1), 1-18. https://doi.org/10.1016/j.fss.2011.09.009.

Acknowledgements

All authors thank FINEP (Financiadora de Estudos e Projetos) for financing the REDEGEO project (Carta Convite MCTI/FINEP/FNDCT 01/2016) and all the support of those involved (UNESP and CEMADEN), in conducting and developing this research.