DYNA
http://dyna.medellin.unal.edu.co/

UNIVERSIDAD NACIONAL DE COLOMBIA

Integrated Landslide Susceptibility and Hazard Assessment Using Multi-Method Modeling in the Tributarios de Río Grande Basin, Colombia

Dubán Uribe Góez

[a] Universidad Nacional de Colombia, sede Medellín, Facultad de Minas, Medellín, Colombia duuribeg@unal.edu.co

Abstract

This study presents a multi-method landslide susceptibility and hazard assessment in the Tributarios de Río Grande basin, located in Santa Rosa de Osos, Antioquia, Colombia. The basin covers an area of 38.83 km² and is characterized by steep topography and intense land use pressures. Using high-resolution input data (1 m spatial resolution), we developed and compared heuristic, parametric, semi-parametric, and non-parametric susceptibility models to identify zones prone to landslides. The models incorporated key conditioning factors such as slope, lithology, hydrological accumulation, and morphometric variables. Additionally, a probabilistic hazard component was introduced by applying a Poisson-based temporal model, considering the recurrence of historical landslide events. This integrative approach not only enhances the spatial accuracy of susceptibility maps but also quantifies temporal probability, contributing to a more comprehensive risk management strategy for the region. The methodology and findings offer valuable insights for decision-makers, territorial planning, and the implementation of early warning systems in mountainous tropical environments.

1. Introduction

Landslide hazard assessment in tropical mountainous regions demands high-resolution analysis and flexible modeling approaches to account for complex terrain and variable conditioning factors. In Colombia, the combination of steep slopes, intense rainfall, and land use pressures makes localized studies essential for effective risk management. This work applies and compares multiple modeling techniques—heuristic, parametric, semi-parametric, and non-parametric—alongside a temporal hazard analysis, to evaluate landslide occurrence in a small Andean basin. To assess and compare the predictive performance of the models, Receiver Operating Characteristic (ROC) curves were also generated, providing a quantitative measure of how well each approach distinguishes between landslide and non-landslide areas. The integration of spatial and temporal components seeks to improve the precision and applicability of susceptibility and hazard maps for planning and early warning in data-rich environments.

2. Input Data and Secondary Sources

When developing any type of model—whether for delineating landslide susceptibility or hazard—it is essential to gather a comprehensive set of data at both local and regional scales. The following section outlines the most relevant information compiled for this study.

A morphodynamic process inventory was used as a foundational input. Its construction required a high-resolution digital elevation model (DEM) and a hillshade, both obtained from Cartoantioquia, with a spatial resolution of 1 m/pixel and dated from 2009. In addition, multiple optical sources were consulted to enhance the accuracy and temporal depth of the inventory, including orthophotos from Cartoantioquia (2009), Google Earth imagery (04/06/2010 and 08/12/2023), PlanetScope images (16/01/2016, 23/05/2022, and 10/03/2024), Sentinel-2 imagery (04/01/2019), and Bing Maps (20/12/2016).

For model construction, custom geomorphological mapping units (UGI, *Unidades Geomorfológicas Integradas*) were developed. Additionally, a series of derived topographic maps—such as flow accumulation, slope, and aspect—were generated using the Cartoantioquia DEM to serve as key explanatory variables within the modeling framework.

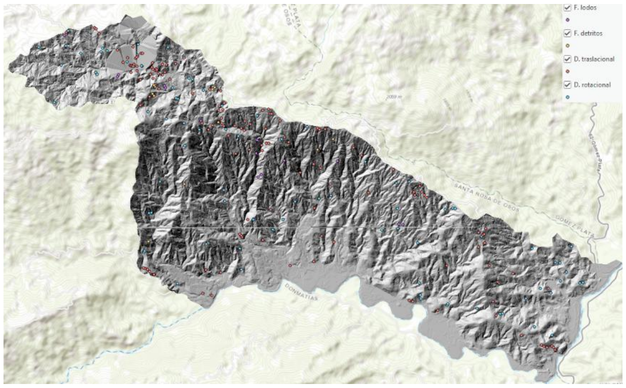

Fig. 1. Spatial Distribution of Mapped Mass Movement Events.

Fig. 2. Examples of Mass Movement Types: (A) Translational Landslide, (B) Debris Flow.

3. Heuristic Model

Heuristic models are expert-based approaches used to assess landslide susceptibility by qualitatively evaluating the influence of conditioning factors. Unlike statistical or machine learning models, heuristic methods do not require a historical frequency of events or model calibration; instead, they rely on empirical knowledge and visual data analysis to guide interpretation and variable selection. These models are particularly valuable in preliminary assessments or when working with limited datasets.

In this study, the heuristic model was developed using seven conditioning variables: slope, aspect, vertical and horizontal curvature, flow accumulation, elevation, geology, and land cover. To evaluate their potential contribution to landslide susceptibility, histograms were generated to analyze the distribution of each variable across the study area. Additionally, comparative diagrams were created to observe how these variables behave in areas with and without mapped mass movements (based on our landslide inventory). This visual and comparative analysis allowed us to identify which variables showed meaningful contrasts between stable and unstable zones, helping to determine their statistical relevance for inclusion in the susceptibility model.

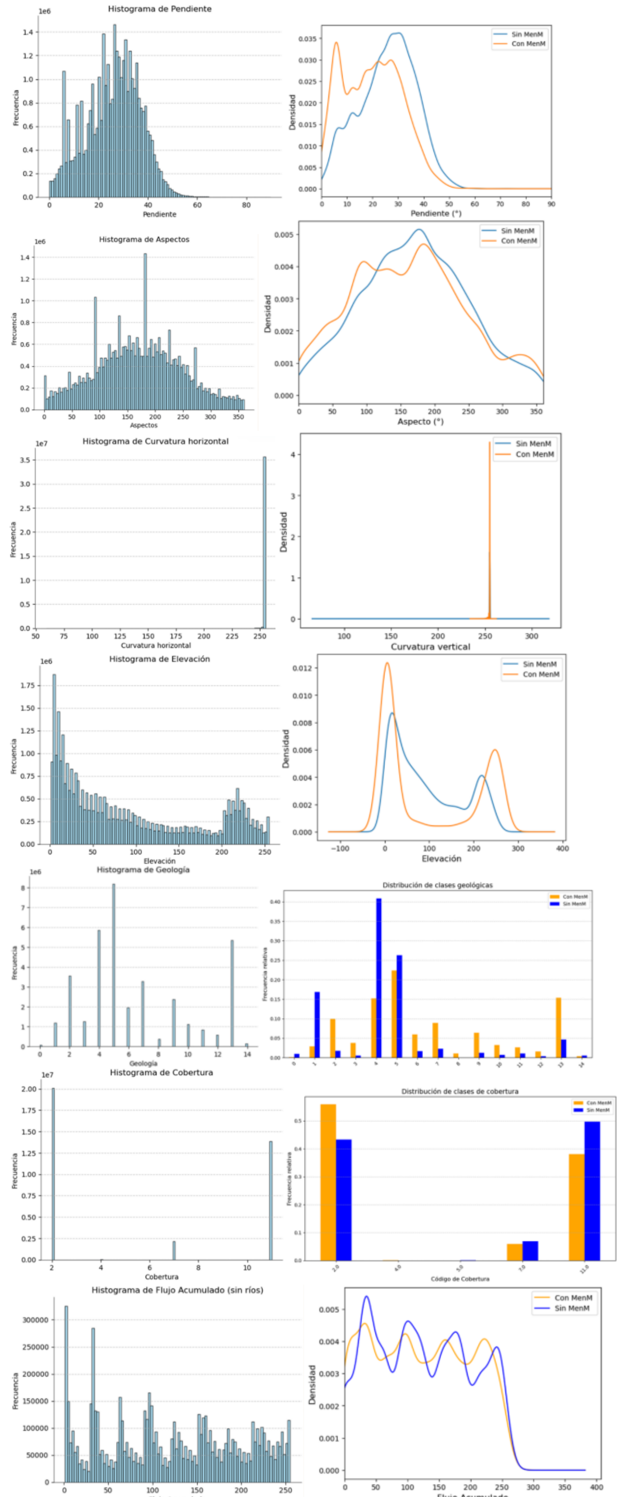

Fig. 3. Left: Histogram showing the overall distribution of each variable across the basin. Right: Comparative plots showing the behavior of each variable in areas with and without mapped mass movements.

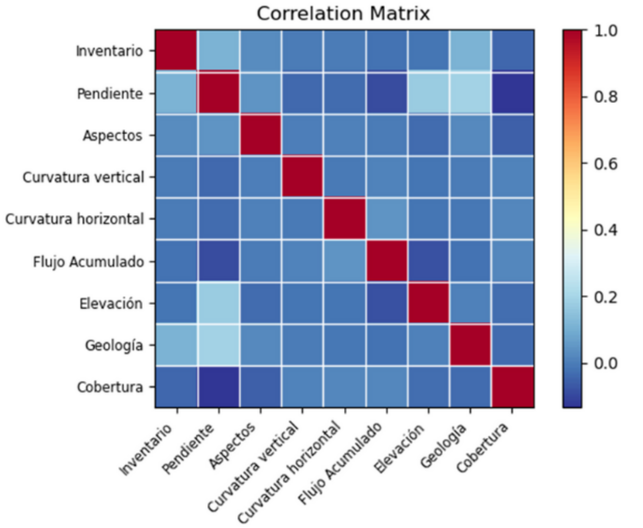

Fig. 4. Correlation Matrix of Conditioning Variables Used in the Susceptibility Model.

The matrix shows the pairwise correlation coefficients between the variables considered for landslide susceptibility analysis. High correlations may indicate redundancy and are considered when selecting input variables.

An examination of the comparative diagrams and the correlation matrix revealed that slope, elevation, aspect, and flow accumulation were the most influential variables for characterizing landslide susceptibility in the study area. These variables showed clear behavioral differences between areas with and without mapped mass movements, making them relevant for the heuristic model.

In contrast, curvature (both vertical and horizontal) displayed similar patterns in both landslide and non-landslide areas, indicating no meaningful influence. Geology showed only slight differentiation, and, in some cases, landslides appeared in areas classified as bedrock, reducing its reliability. Land cover exhibited limited variability across the basin, offering little discriminatory power. For these reasons, curvature, geology, and land cover were excluded from the final set of variables used in the model.

	Pendiente	Altura	Aspecto	Flujo	Valor propio
Pendiente	1,00	1,17	1,75	2,33	0,35
Altura	0,86	1,00	1,50	2,00	0,29
Aspecto	0,67	0,67	1,00	1,33	0,21
Flujo	0,43	0,50	0,75	1,00	0,15

Table. 1. Saaty Pairwise Matrix and Derived Weights for Heuristic Modeling

The matrix shows the relative importance between variables (slope, elevation, aspect, and flow accumulation) based on expert judgment. The last column presents the normalized eigenvalues, which reflect the weight of each variable in the final susceptibility model.

After selecting the most relevant variables, their relative importance in the heuristic model was determined using the eigenvector values derived from the Saaty Pairwise Matrix and Derived Weights for Heuristic Modeling (Table 1). Once the variables were weighted, each one was classified into discrete classes based on the behavior observed in Figure 3, which compares variable responses in areas with and without mapped mass movements. Subsequently, the Weight of Evidence (WoE) method was applied to assign a weight to each class within the variables. This two-step approach—assigning weights to variables using eigenvectors from the Saaty matrix, and to their classes using WoE—resulted in the following heuristic susceptibility model:

$$f(x) = 0.35x_1 + 0.29x_2 + 0.21x_3 + 0.15x_4$$

where:

- x_1 is the WoE-derived weight of the slope class,
- x_2 is the WoE-derived weight of the elevation class,
- x_3 is the WoE-derived weight of the aspect class, and
- x_4 is the WoE-derived weight of the flow accumulation class.

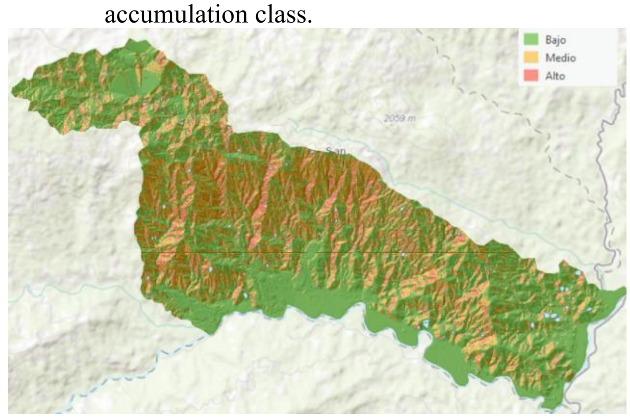

Fig. 5. Landslide Susceptibility Map Based on the Heuristic Model.

The map (Fig. 4) shows the spatial distribution of landslide susceptibility in the study area, obtained using the heuristic model. The susceptibility index was calculated through a weighted linear combination of four conditioning variables: slope, elevation, aspect, and flow accumulation. Variable weights were derived from the Saaty Pairwise Matrix (Table 1), and class weights were assigned using the Weight of Evidence (WoE) method. The final values were categorized into three susceptibility levels: low, medium, and high, providing a simplified and interpretable tool for identifying areas potentially exposed to mass movement processes.

4. Logistic regression (LR)

It is one of several multivariate statistical methods used to analyze the relationship between a categorical dependent variable (in this case, landslide occurrence, coded as 0 for no occurrence and 1 for occurrence) and a set of independent variables, such as terrain-conditioning factors.

In this study, various combinations of variables were tested—slope, aspect, elevation, flow accumulation, and geology. The model with the highest pseudo R-squared value was selected, which included the variables slope, aspect, elevation, and flow accumulation.

```
Optimization terminated successfully.
         Current function value: 0.167928
         Iterations 8
                           Logit Regression Results
==============================================================================
Dep. Variable:             Inventario   No. Observations:             36486974
Model:                          Logit   Df Residuals:                 36486969
Method:                           MLE   Df Model:                            4
Date:                Wed, 09 Jul 2025   Pseudo R-squ.:                 0.04391
Time:                        08:15:47   Log-Likelihood:            -6.1272e+06
converged:                       True   LL-Null:                   -6.4086e+06
Covariance Type:            nonrobust   LLR p-value:                     0.000
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      3.2905      0.001   3458.906      0.000       3.289       3.292
Pendiente      0.5973      0.001    690.298      0.000       0.596       0.599
Aspectos       0.0853      0.001    105.287      0.000       0.084       0.087
Elevación     -0.2114      0.001   -279.253      0.000      -0.213      -0.210
Flujo         -0.0697      0.001    -92.004      0.000      -0.071      -0.068
==============================================================================
```

Fig. 5. Logistic Regression Results.

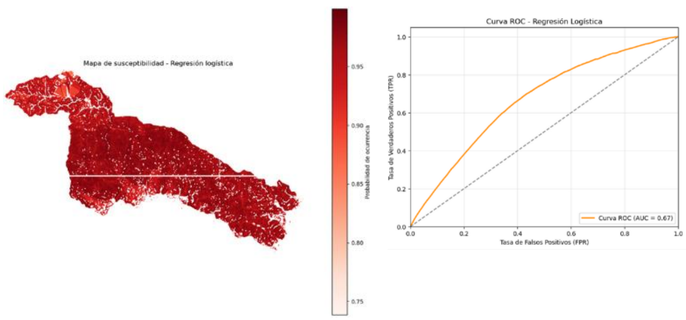

Fig. 6. Landslide susceptibility index map produced using Logistic Regression (left) and its ROC curve for model validation (right).

5. Linear Discriminant Analysis (LDA)

Using the same conditioning variables applied in the Logistic Regression model—slope, aspect, elevation, and flow accumulation—a susceptibility map was also generated using Linear Discriminant Analysis (LDA). LDA is a parametric multivariate statistical method that seeks to find a linear combination of features that best separates two or more classes—in this case, the presence or absence of landslides. The method assumes normal distribution and equal covariance among classes and is commonly used for classification problems due to its interpretability and computational efficiency.

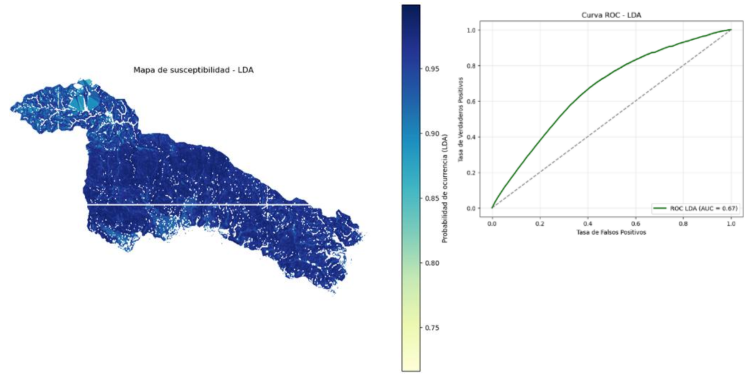

Fig.7. Landslide susceptibility index map produced using LDA (left) and its ROC curve for model validation (right).

6. Neural Network Model

We also implemented a neural network model using the same set of conditioning variables. Neural networks are non-parametric machine learning models capable of capturing complex, non-linear relationships between inputs and outputs without assuming a predefined functional form. They consist of layers of interconnected nodes (neurons) that process information through weighted connections. In this study, we used a simple feedforward neural network architecture with two hidden layers—the first containing two neurons and the second five neurons—to model the probability of landslide occurrence.

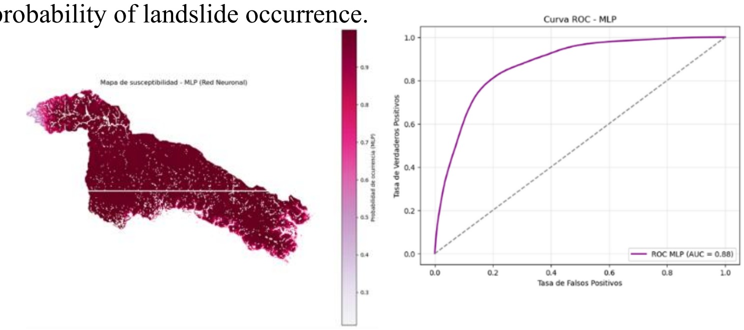

Fig.8. Landslide susceptibility index map produced using Neural Network Model (left) and its ROC curve for model validation (right).

7. Random Forest Model

We also applied a Random Forest model using the same conditioning variables. Random Forest is a non-parametric ensemble learning method based on decision trees. It builds multiple trees during training and aggregates their predictions to improve accuracy and reduce overfitting. This model is particularly effective in handling complex interactions and non-linear relationships between variables, making it suitable for landslide susceptibility analysis in heterogeneous terrains.

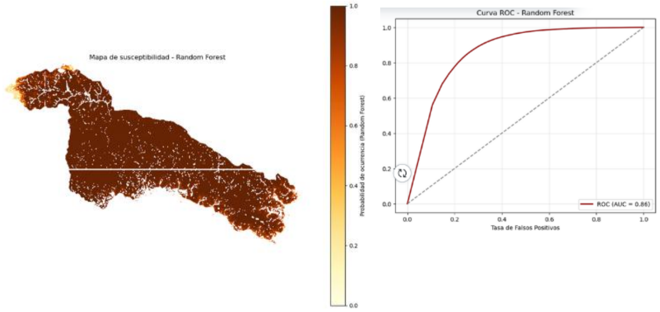

Fig.9. Landslide susceptibility index map produced using Random Forest (left) and its ROC curve for model validation (right).

After analyzing and comparing the ROC curves of all the models, we found that the neural network model performed the best, as it showed the highest predictive capability and the most favorable ROC curve. It was followed by the Random Forest model, which also demonstrated strong performance. The LDA and logistic regression models showed similar predictive power, with overlapping ROC curves, indicating comparable performance in classifying landslide-prone areas.

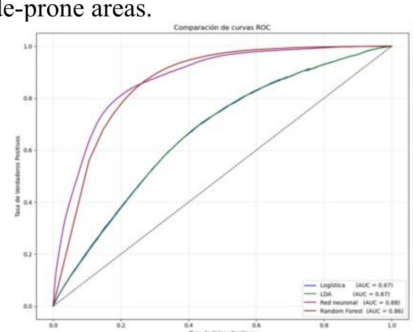

Fig.10. ROC Curve Comparison Among Susceptibility Models: Logistic Regression, LDA, Neural Network, and Random Forest.

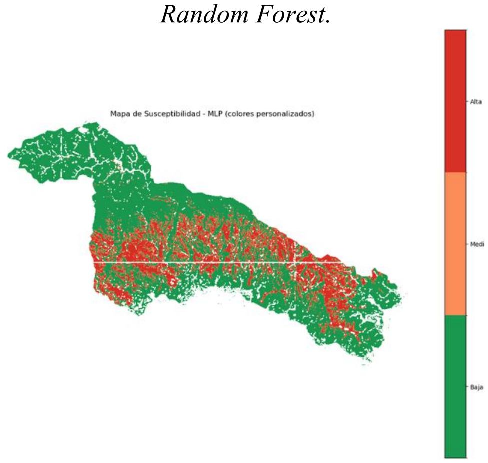

Fig.11. Landslide Susceptibility Classes Derived from the Neural Network Model.

8. Cumulative Landslide Hazard Assessment Using the Poisson Model

To integrate landslide hazard over time, a Poisson model was applied to estimate the cumulative probability of landslide occurrence within a given period. The graph (fig. 12) illustrates how this probability varies depending on different frequencies of occurrence (from at least 1 to 10 landslide years) as a function of the number of years considered. This approach allows for long-term risk assessment and supports preventive planning. For example, over a 6-year period, there is an 80% chance of experiencing at least two years with landslide activity.

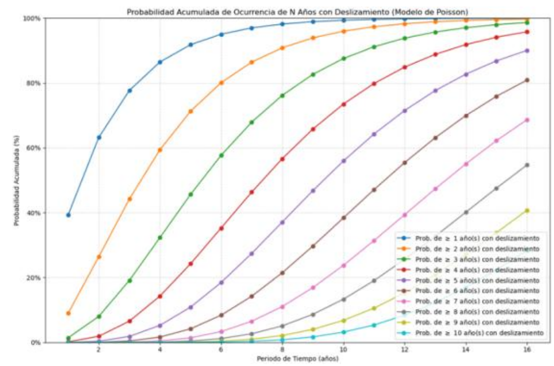

Fig.12. Cumulative Probability of Landslide Occurrence (Poisson Model).

9. Infinite Slope Model

To evaluate slope stability across the study area, we applied the infinite slope model, a widely used approach for analyzing shallow landslides under saturated conditions. The model assumes a planar failure surface parallel to the slope and uniform soil properties along the depth. The factor of safety (FS) is calculated using the following equation:

$$FS = \frac{C + (\gamma_s * Z * \cos(\theta)^2 - \gamma_w * q * \cos(\theta)^2) * \tan \emptyset}{\gamma_s * Z * \sin\theta * \cos\theta}$$

where:

- C is the soil cohesion (kPa),
- γs is the unit weight of the soil (kN/m³),
- Z is the soil thickness (m),
- θ is the slope angle (rad),
- γw is the unit weight of water (9.81 kN/m³),
- qis the steady-state infiltration rate (100 m/h),
- ϕ is the internal friction angle (degrees).

The spatial input parameters required by the model—slope, soil cohesion, internal friction angle, specific weight, and contributing flow—were derived from geomorphological units (UGI) previously constructed. Each parameter was mapped based on characteristic values for each unit. For example, values of cohesion and friction angle were assigned based on regional geotechnical literature (e.g., Mora-Castro et al., 2020; Eraso & Rodríguez, 2020), while

the infiltration rate was estimated from the local hydrological context and typical values reported in similar studies.
To estimate soil thickness (ZZZ), we used an empirical relationship based on slope angle:

$$Z = 10 * e^{-0.06*\theta}$$

where θ is the slope in degrees, and Z is returned in meters (e.g., Montgomery and Dietrich, 1994).

Based on the calculated FS values, the terrain was classified into five stability categories:

- $FS \leq 1.0$: Unstable
- $1.0 < FS \leq 1.3$: Potentially unstable
- $1.3 < FS \leq 1.5$: Conditionally stable
- $1.5 < FS \leq 2.0$: Stable
- $FS > 2.0$: Very stable

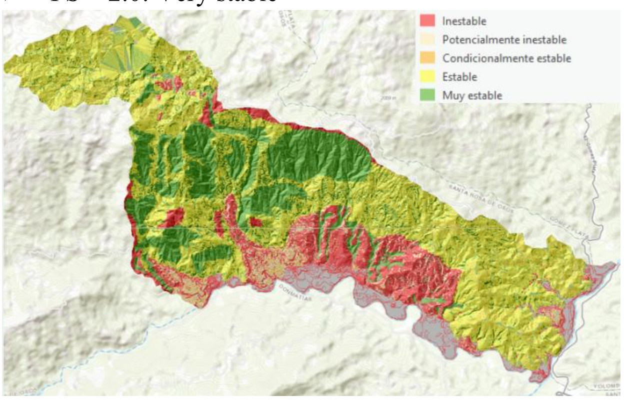

Fig.13. Terrain Stability Classification Based on the Infinite Slope Model.

10. Landslide Propagation Modeling

The landslide propagation was simulated using the Flow-R 2.1 software, which implements empirical spreading models based on digital elevation data. For lateral spreading, the Holmgren modified algorithm was used, with parameters dh = 0.5 and x = 2, allowing for a realistic simulation of material flow paths. The persistence of flow was governed by the Weights algorithm with default parameters.

The friction model was defined using the *Travel angle* approach with an angle of 3°, reflecting low frictional resistance typical of rapidly moving landslide materials. A maximum velocity constraint of 40 m/s was applied to avoid unrealistic speeds in steep terrain. The only output generated was the maximum susceptibility map, which indicates the maximum flow intensity reached in each cell after multiple simulations. This simulation and output were applied exclusively to flow-type landslides.

After obtaining the maximum susceptibility map from the propagation model, a combined susceptibility index was calculated to integrate additional geomorphological factors. The index was computed using the following formula:

$$Index = 0.5 * propagation + 0.3 * \frac{Slope}{90} + 0.2 * \frac{log_{10}(Flow + 1)}{log_{10}(10000 + 1)}$$

Where:

- Propagation is the output susceptibility value from the infinite slope model, normalized between 0 and 1.
- Slope is the terrain slope in degrees.
- Flow represents the accumulated flow, which serves as a proxy for water concentration or drainage convergence.
-

Fig.14. Flow-Type Landslide Susceptibility and Propagation Index Map.

Figure 14 illustrates the potential extent of flow-type landslides by integrating a heuristic susceptibility model with slope angle normalization and a logarithmic transformation of flow accumulation. The index highlights areas where both initiation and propagation of flows are more likely to occur, providing a refined spatial interpretation of landslide hazard.

11. Results

The application of multiple modeling techniques resulted in various susceptibility and hazard maps that reflect spatial patterns of landslide-prone areas in the Río Grande tributaries basin.

Among the susceptibility models, the neural network outperformed all others in terms of predictive accuracy, as demonstrated by its ROC curve, showing the highest area under the curve (AUC). The Random Forest model followed closely, while logistic regression and LDA exhibited similar and lower performance levels. These findings are consistent across the susceptibility maps generated, where high-risk zones were consistently located in steep, concave, and highly drained areas.

The heuristic model, although simpler, provided a useful preliminary classification and showed good agreement with areas of known landslide occurrence.

The Poisson-based hazard analysis revealed that there is a high cumulative probability (>80%) of experiencing at least two years with landslide activity over a 6-year period, emphasizing the relevance of temporal hazard integration.

The infinite slope model allowed for the identification of areas classified from “very stable” to “unstable” under saturated conditions. Flow-type landslides were further analyzed using Flow-R, and a combined susceptibility-propagation index was generated, refining the understanding of areas where both initiation and movement of debris flows are more likely to occur.

Together, these results provide a robust multi-method characterization of landslide susceptibility and hazard, useful for risk-informed planning and early warning strategies in the region.

12. References

Soil cohesion (c) and angle of internal friction (φ) values were assigned based on the Unified Geotechnical Units (UGI) previously developed. These values were adapted from typical ranges found in:

Van Westen, C. J., Castellanos, E., & Kuriakose, S. L. (2008). Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. Engineering Geology, 102(3-4), 112–131.

USDA. (2007). Soil Mechanics Level I, Module 3: USDA Textural Classification. Natural Resources Conservation Service.

Unit weight of soil (γ) values were obtained following:

Das, B. M. (2010). Principles of Geotechnical Engineering (7th ed.). Cengage Learning.

Soil depth (Z) was calculated using an empirical formula proposed by:

Gómez, H. & Kavzoglu, T. (2005). Assessment of shallow landslide susceptibility using an empirical model based on slope and contributing area