

# Uso de modelos multinivel para incorporar la heterogeneidad espacial en la evaluación de la susceptibilidad por movimientos en masa

## Using multilevel models to incorporate spatial heterogeneity in landslide susceptibility assessment

Edier Aristizábal

evaristizabal@unal.edu.co. Departamento de Geociencias y Medio Ambiente, Universidad Nacional de Colombia, Carrera 80 # 65-223, Medellín, Colombia

### Abstract

Landslide susceptibility is defined as the likelihood of landslide occurrence under favorable terrain conditions. Traditional logistic regression models often fall short by assuming spatial homogeneity, ignoring the inherent variability of physical and environmental factors. To overcome this limitation, multilevel models that incorporate both fixed and random effects to capture variability among different spatial units are used. In this study, a multilevel logistic regression model is employed to assess landslide susceptibility in the Colombian Andes, incorporating both natural hydrological divisions and clustering-based spatial techniques. This approach provides a more nuanced representation of susceptibility by addressing geographic heterogeneity. Two regionalization approaches are tested: natural basin divisions (Atrato, Cauca, Magdalena), and spatial clustering derived from morphometric characteristics. The results reveal that basin-based regionalization achieves a better model fit based on Akaike (AIC) and Bayesian (BIC) information criteria metrics compared to the spatio-clustering model. This suggests that hydrological boundaries are more consistent in capturing landslide dynamics in the study area, likely due to their alignment with geomorphological processes. The fixed and random effect coefficients illustrate that variables such as elevation, relief, and area have significant variability depending on the regionalization method used. A major advantage of using multilevel models is their ability to incorporate spatial variability without artificially dividing the study area and training separate models for each subregion. By using shared information between different spatial units, these models can improve the estimation of both global and local landslide effects.

### Resumen

La susceptibilidad por movimientos en masa se define como la probabilidad de que ocurra un deslizamiento bajo condiciones de terreno favorables. Los modelos de regresión logística tradicionales son a menudo insuficientes, dado que asumen homogeneidad espacial, ignorando la variabilidad inherente de los factores físicos y ambientales. Para superar esta limitación, se emplean modelos multinivel que incorporan efectos tanto fijos como aleatorios para capturar así la variabilidad entre diferentes unidades espaciales. En este estudio, se presenta un modelo de regresión logística multinivel para evaluar la susceptibilidad por movimientos en masa en los Andes colombianos, incorporando tanto divisiones hidrológicas naturales como una regionalización espacial basada en agrupamiento (*clustering*). Este enfoque proporciona una representación más matizada de la susceptibilidad al abordar la heterogeneidad geográfica. Se probaron dos enfoques de regionalización: divisiones por cuencas naturales (Atrato, Cauca y Magdalena) y un agrupamiento espacial derivado de características morfométricas. Los resultados revelan que la regionalización por cuencas logra un mejor ajuste del modelo (basado en las métricas de Akaike y Bayesiana) en comparación con el modelo de *clustering*. Esto sugiere que los límites hidrológicos son más consistentes para capturar la dinámica de los movimientos en masa en el área de estudio, probablemente debido a su alineación con procesos geomorfológicos. Los coeficientes de efectos fijos y aleatorios muestran que variables como la elevación, el relieve y el área exhiben una variabilidad significativa dependiendo del método de regionalización utilizado. Una de las principales ventajas de utilizar modelos multinivel es su capacidad para incorporar la variabilidad espacial sin tener que dividir artificialmente el área de estudio ni entrenar modelos separados para cada subregión. Al utilizar información compartida entre las diferentes unidades espaciales, estos modelos pueden mejorar la estimación de los efectos provocados por eventos de movimientos en masa a escala local y global.

**Keywords:** landslide; susceptibility; logistic regression; spatial heterogeneity

**Palabras clave:** deslizamientos, susceptibilidad, regresión logística, heterogeneidad especial.

## 1. Introducción

La susceptibilidad a movimientos en masa se define como la probabilidad que ocurra un movimiento en masa en un área dada (Brabb, 1984; Fell et al., 2008; Corominas et al., 2014). Esta probabilidad está vinculada con la condición natural del terreno, la cual depende de diversos factores geológicos, hidrológicos y geomorfológicos, entre otros (Soeters y Van Westen, 1996; Reichenbach et al., 2018). A diferencia del concepto de amenaza, que incluye la componente temporal, la susceptibilidad se centra exclusivamente en las condiciones del terreno y su entorno que lo hacen favorable a la ocurrencia de movimientos en masa (Fell et al., 2008; Corominas et al., 2014).

Para la evaluación de la susceptibilidad por movimientos en masa se han empleado tradicionalmente diversos enfoques, incluidos métodos basados en el conocimiento (heurísticos) (Barredo et al., 2000), en datos (estadísticos) (Korup y Stolle, 2014) y con base física (Montgomery y Dietrich, 1994). Los modelos heurísticos se basan en observaciones previas y en la experiencia del analista, lo que los hace prácticos pero limitados en cuanto a reproducibilidad y objetividad (Huang et al., 2020). Los modelos con base física intentan representar los procesos que conducen al fallamiento del terreno (Palacio et al., 2020; Sannino et al., 2024); sin embargo, su aplicación suele verse limitada por la disponibilidad de datos detallados y la intensidad del cálculo requerido (Aristizábal et al., 2016; Marín et al., 2021). Los modelos estadísticos, por otro lado, se han popularizado debido a su capacidad para analizar grandes volúmenes de datos e identificar patrones que pueden predecir la ocurrencia de movimientos en masa (Tehrani et al., 2022). Estos modelos utilizan una variable respuesta, también denominada dependiente, que representa la ocurrencia de movimientos en masa, junto con un conjunto de variables predictoras, también denominadas como independientes, condicionantes o covariables, que describen las condiciones que favorecen su ocurrencia, como la pendiente, el tipo de suelo, el uso o cobertura del suelo y la precipitación, entre otros.

En los últimos años, se ha observado un notable aumento en el uso de modelos estadísticos basados en aprendizaje automático (denominados *machine learning*) para la evaluación de la susceptibilidad a movimientos en masa (Tehrani et al., 2022). Estos modelos han mejorado significativamente la capacidad de representar relaciones no lineales complejas entre las variables predictoras y la variable respuesta, logrando un excelente rendimiento en términos de predicción. Sin embargo, este incremento en la capacidad predictiva viene generalmente acompañado de una considerable pérdida en la interpretabilidad (Youssef et al., 2023). Los modelos de aprendizaje automático, especialmente los no-paramétricos, como los basados en redes neuronales, se comportan como ‘cajas negras’, limitándose a proporcionar métricas de predicción sin aportar conocimiento sobre las causas subyacentes de los movimientos en masa (Ermini et al., 2005). Esto representa una limitación para los usuarios, quienes necesitan comprender las interacciones entre los factores que contribuyen a la inestabilidad del terreno para poder tomar decisiones informadas.

Entre los modelos estadísticos tradicionales, la regresión logística se ha consolidado como uno de los enfoques más utilizados para la evaluación de la susceptibilidad a movimientos en masa (Budimir et al., 2015; Lombardo y Mai, 2018; Reichenbach et al., 2018). Este modelo establece una relación entre una variable respuesta binaria, que indica la presencia o ausencia de movimientos en masa, y un conjunto de variables predictoras categóricas o continuas que representan los factores condicionantes (Ayalew y Yamagishi, 2005). La regresión logística estima la probabilidad en cada unidad de mapeo de pertenecer a la clase ‘movimiento en masa’ o ‘no movimiento en masa’ del inventario utilizado, modelándola como un proceso de Bernoulli (Ayalew y Yamagishi, 2005). La relativa simplicidad de su interpretación y la facilidad para identificar el efecto de cada variable predictora han contribuido a su popularidad en el campo de la evaluación de amenazas naturales.

No obstante, una limitación significativa de la regresión logística y de otros modelos estadísticos convencionales, radica en la asignación de coeficientes constantes a las variables predictoras, asumiendo

que la influencia de cada factor es homogénea en toda el área de estudio (Anselin et al., 1996; Cressie, 2015). Este supuesto de homogeneidad es problemático, ya que en la realidad las variables predictoras de los movimientos en masa son inherentemente heterogéneas y su influencia puede variar de manera considerable a lo largo del espacio (Petschko et al., 2012). Estudios previos han demostrado que la influencia de variables como la pendiente, la litología o el uso del suelo no es uniforme, sino que depende de las condiciones específicas de cada subregión del área de estudio (Lombardo et al., 2020).

Para abordar esta heterogeneidad espacial, una de las estrategias más comunes en la evaluación de la susceptibilidad a movimientos en masa consiste en dividir el área de estudio en unidades homogéneas, como cuencas hidrográficas, y aplicar modelos separados en cada una de ellas. Esta estrategia ha sido implementada, por ejemplo, por el Servicio Geológico Colombiano para los mapas de amenaza por movimientos en masa a nivel nacional (SGC, 2013). Aunque esta técnica permite considerar algunas diferencias espaciales, presenta también varias desventajas. En primer lugar, la división del área de estudio conlleva una disminución en la cantidad de datos disponibles para entrenar cada modelo, lo cual puede afectar la robustez de los resultados. Además, se incrementa significativamente la complejidad debido a la necesidad de entrenar y aplicar múltiples modelos independientes. Otro problema importante es la aparición de discontinuidades en los resultados obtenidos, ya que los modelos entrenados por separado tienden a generar inconsistencias en los bordes de las zonas modeladas, lo cual requiere de ajustes y reprocesos.

Ante estas limitaciones, en las últimas décadas se han desarrollado modelos estadísticos espaciales que permiten incorporar la heterogeneidad de una manera más integral (Anselin y Griffith, 1988; LeSage y Pace, 2009; Cressie, 2015; Anselin, 2022). Entre estos modelos se encuentran los modelos jerárquicos o multiniveles (Kumar et al., 2011), así como los modelos de regresión espacial ponderada (Fotheringham et al., 2009). Los modelos multiniveles permiten la incorporación de variaciones en el intercepto o en los coeficientes a lo largo del área de estudio, lo cual es ideal para capturar la heterogeneidad en las variables predictoras. Estos modelos también permiten considerar distintos niveles de agrupación, como cuencas/subcuencas, ciudad/barrio, etc., proporcionando una representación más detallada de la estructura espacial de los datos.

En este trabajo se presenta la aplicación de un modelo multinivel para la evaluación de la susceptibilidad por movimientos en masa a escala de subcuenca en los Andes colombianos, utilizando como grupos homogéneos las cuencas hidrográficas del Atrato, Cauca y Magdalena. El objetivo es presentar cómo los modelos multiniveles permiten una mejor representación de la heterogeneidad espacial, facilitando la incorporación de estas variaciones en la influencia de las variables predictoras a lo largo del área de estudio. De esta manera, se espera contribuir al desarrollo de métodos más precisos y efectivos para la evaluación de la susceptibilidad por movimientos en masa.

## 2. Modelo de regresión logística multinivel

Los modelos multiniveles, también conocidos como modelos jerárquicos, y en algunos casos mixtos o de regímenes espaciales, son un enfoque estadístico que permite la incorporación explícita de la estructura jerárquica o anidada de los datos (Lee y Nelder, 1996). Estos modelos combinan dos tipos de efectos: efectos fijos y efectos aleatorios. Los efectos fijos representan las relaciones generales entre las variables predictoras y la variable respuesta, aplicables a toda el área de estudio (Kumar et al., 2011), y se representan mediante coeficientes constantes que describen la relación promedio entre ambas. Por otro lado, los efectos aleatorios permiten modelar la variabilidad que existe entre diferentes niveles jerárquicos, proporcionando estimaciones específicas para cada unidad espacial o región (Kumar et al., 2011). Estos últimos se modelan como términos adicionales que varían para cada grupo o unidad de nivel superior, permitiendo capturar la variabilidad específica entre los grupos. Esta estructura de efectos fijos y aleatorios es fundamental para capturar adecuadamente la heterogeneidad inherente a los datos espaciales cuando se tiene una organización natural en distintos niveles, donde las unidades de nivel inferior están agrupadas en unidades de nivel superior (Wong y Mason, 1985).

Matemáticamente, un modelo multinivel se puede escribir como una combinación de ecuaciones que representan tanto la variabilidad global como local (Lee y Nelder, 1996). Si se considera, por ejemplo, una situación típica de dos niveles para una unidad de análisis en el nivel 1,  $i$ , que está anidada en la

unidad de nivel 2,  $j$ , se puede definir una variable dependiente binaria  $Y_{ij}$  para el individuo (subcuenca)  $i$  en el grupo (cuenca)  $j$ , donde  $Y_{ij}$  representa un inventario de movimientos en masa. En este caso,  $Y_{ij}$  toma el valor 1 si existe uno o más movimientos en masa y 0 en caso contrario. Se puede utilizar un modelo de regresión logística con la función de enlace *logit*, que representa el logaritmo natural de la relación entre la ocurrencia y no ocurrencia de un evento (*odds*). Esta función transforma la relación lineal de las variables predictoras a una respuesta entre 0 y 1. De esta forma, la probabilidad  $P(Y_{ij}=1)$  se puede expresar como:

$$\text{logit} \left( P(Y_{ij}=1) \right) = \ln \left( \frac{P(Y_{ij}=1)}{1-P(Y_{ij}=1)} \right) = \beta_{0j} + \sum_{j=1}^p \beta_{1j} X_{ij} + \sum_{l=1}^r u_{lj} + \varepsilon_i$$

donde  $\beta_{0j}$  y  $\beta_{1j}$  son los coeficientes fijos que describen la relación entre la variable respuesta  $Y$  y las variables predictoras  $X_{ij}$ . Ambos coeficientes se modelan a nivel de grupo y se puede asumir que siguen una distribución normal, es decir:

$$\beta_{0j} = \gamma_{00} + u_{0j}, u_{0j} \sim N(0, \sigma_{u_0}^2)$$

$$\beta_{1j} = \gamma_{10} + u_{1j}, u_{1j} \sim N(0, \sigma_{u_1}^2)$$

donde  $\gamma_{00}$  es el intercepto promedio entre todos los grupos,  $\gamma_{10}$  es el promedio del coeficiente del predictor  $X_{ij}$  entre todos los grupos, y  $u_{0j}$  y  $u_{1j}$  son los efectos aleatorios a nivel de grupo que capturan la variabilidad entre grupos para el intercepto y el coeficiente de  $X_{ij}$ , respectivamente.  $\sigma^2$  representa la varianza de los efectos aleatorios del intercepto ( $u_0$ ) y del coeficiente ( $u_1$ ).

Combinando los modelos de nivel individual y nivel grupal, el modelo completo se expresa como (Lee y Nelder, 1996):

$$\text{logit}[P_{ij}=1] = (\gamma_{00} + u_{0j}) + (\gamma_{10} + u_{1j})X_{ij}$$

o, de manera equivalente (Wong y Mason, 1985):

$$P(Y_{ij}=1) = \frac{e^{[(\gamma_{00} + u_{0j}) + (\gamma_{10} + u_{1j})X_{ij}]}}{1 + e^{[(\gamma_{00} + u_{0j}) + (\gamma_{10} + u_{1j})X_{ij}]}}$$

18 donde  $P(Y_{ij}=1)$  representa la probabilidad de que el resultado sea 1 para el individuo  $i$  en el grupo  $j$ .  
19

20 En este modelo de regresión logística multinivel, el coeficiente  $\beta_{0j}$  varía entre grupos, lo que permite  
21 capturar diferencias en el nivel base del resultado para cada grupo. El coeficiente  $\beta_{1j}$  del predictor  $X_{ij}$   
22 también varía entre grupos, lo que permite que el efecto o influencia del predictor cambie según el grupo.  
23 Los efectos aleatorios  $u_{0j}$  y  $u_{1j}$  permiten que tanto el intercepto como las pendientes cambien entre grupos,  
24 acomodando la heterogeneidad que no es capturada por los predictores a nivel individual.

25 En el contexto de la evaluación de la susceptibilidad a movimientos en masa, los modelos multiniveles  
26 resultan particularmente beneficiosos debido a la heterogeneidad espacial inherente, es decir, la variación  
27 en las características del terreno y las condiciones ambientales que pueden ocurrir de una ubicación a  
28 otra dentro del área de estudio. Se pueden definir niveles jerárquicos, como unidades de nivel inferior  
29 (por ejemplo, áreas más pequeñas como laderas o barrios) y unidades de nivel superior (como cuencas o  
30 municipios). En general, el nivel inferior se refiere a las unidades más pequeñas y detalladas, mientras  
31 que el nivel superior agrupa a estas unidades en categorías más amplias. Los efectos fijos permiten  
32 modelar cómo los factores (e.g., pendiente, litología, uso del suelo) influyen de manera global en la  
33 susceptibilidad, mientras que los efectos aleatorios reflejan cómo estas relaciones pueden variar entre  
34 diferentes unidades espaciales, capturando así de mejor manera las diferencias locales. Por ejemplo, en  
35 algunas zonas, la pendiente puede tener una influencia predominante en la susceptibilidad a movimientos  
36 en masa debido a la presencia de suelos inestables, mientras que en otras las condiciones climáticas  
37 específicas pueden favorecer el desarrollo de suelos con características distintas, afectando la  
38 susceptibilidad de manera diferenciada.  
39

40 Dependiendo del contexto, se pueden utilizar dos o más niveles, pero se recomienda utilizar un número  
41 limitado (usualmente dos o tres niveles) para evitar complicaciones innecesarias en la modelización y  
42 garantizar resultados robustos (Lee y Nelder, 1996; Kumar et al., 2011). Otra ventaja importante de los  
43 modelos multiniveles es la flexibilidad para incorporar información jerárquica sin necesidad de dividir  
44 el área de estudio de manera arbitraria y entrenar modelos independientes para cada subregión (Lee y  
45 Nelder, 1996). Esto evita los problemas de discontinuidad en los bordes de las zonas y permite utilizar  
46 la información compartida entre las distintas unidades para mejorar la estimación de los efectos globales  
47 y locales.

48 Sin embargo, los modelos multiniveles también presentan ciertas limitaciones. Una de las principales  
49 es su posible mayor complejidad matemática y computacional, lo cual puede dificultar su  
50 implementación inicialmente. Además, la selección de la estructura adecuada del modelo (es decir,  
51 cuántos niveles incluir y cómo definir los efectos aleatorios) puede ser complicada y requiere típicamente  
52 de un conocimiento exhaustivo de la dinámica espacial del fenómeno en estudio (Kumar et al., 2011).  
53 Otro desafío es la interpretación de los efectos aleatorios, ya que estos pueden ser difíciles de comunicar  
54 a audiencias no familiarizadas con el tema.

55 Existen diversos lenguajes de programación y paquetes que permiten trabajar con modelos  
56 multiniveles, cada uno con sus propias ventajas y desventajas.

57 **R (lme4 y nlme):** R es uno de los lenguajes más utilizados para la implementación de modelos  
58 multiniveles, principalmente a través de los paquetes lme4 y nlme. Estos paquetes permiten una gran  
59 flexibilidad para especificar efectos fijos y aleatorios, y ofrecen herramientas avanzadas para la  
60 visualización de resultados. Una ventaja importante de R es su comunidad amplia y activa, lo cual facilita  
61 el acceso a documentación y soporte. Sin embargo, la desventaja principal es que el procesamiento de  
62 grandes volúmenes de datos puede ser lento, lo que limita su uso en estudios de gran escala sin hardware  
63 especializado.

64 **Python (statsmodels y PyMC):** En Python, los modelos multiniveles se pueden implementar  
65 utilizando statsmodels para enfoques más tradicionales o PyMC para un enfoque bayesiano. PyMC es  
66 particularmente útil si se desea explorar la incertidumbre de los modelos mediante técnicas como Monte

67 Carlo. Python es conocido por su versatilidad y su capacidad para integrarse con otros tipos de análisis  
68 y herramientas de aprendizaje automático.

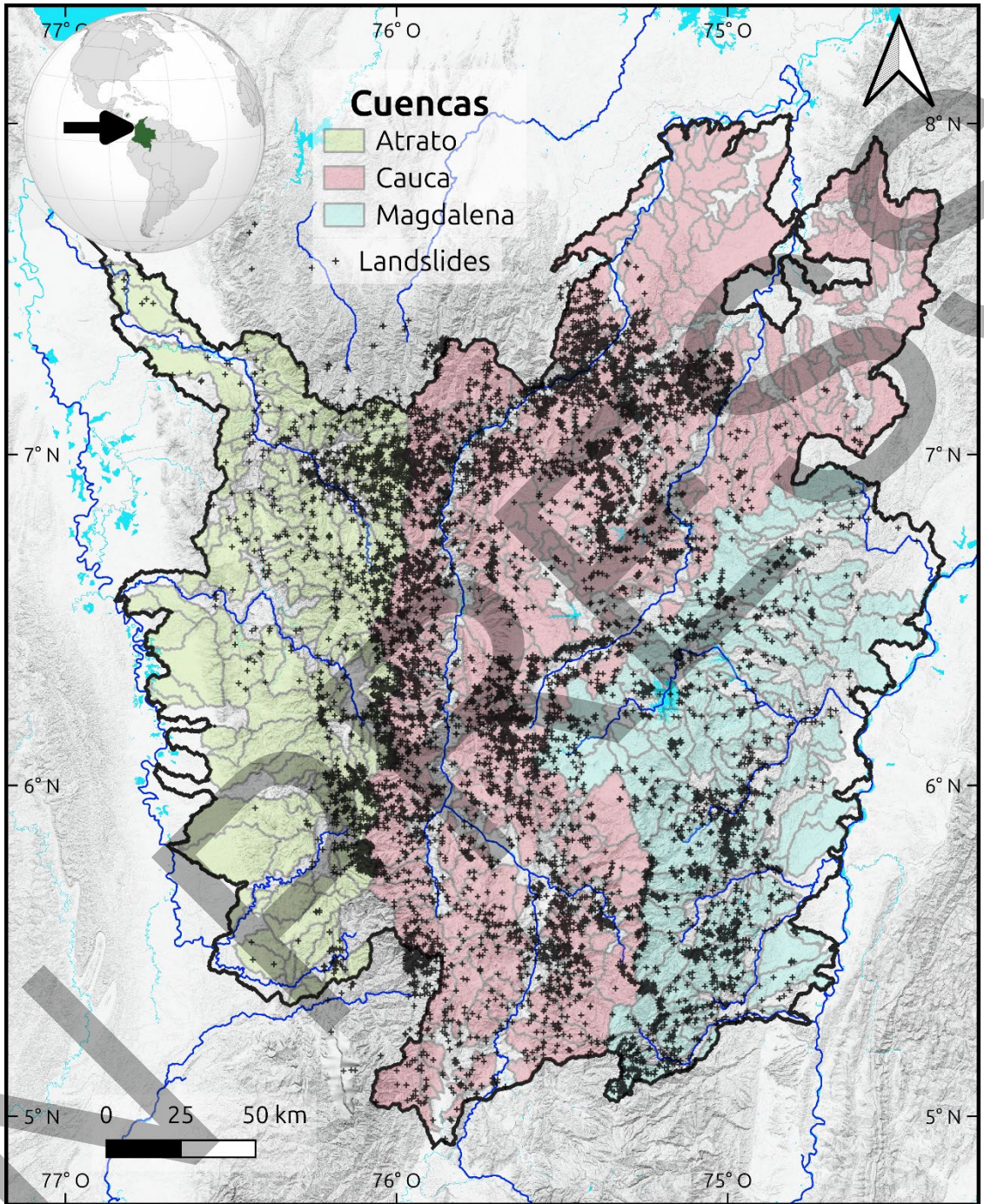
69 **Stan (a través de R o Python):** Stan es un potente lenguaje para realizar inferencia bayesiana y es muy  
70 útil para modelos multiniveles complejos. Se puede acceder a través de interfaces en R (rstanarm) o  
71 Python (pystan). La principal ventaja de Stan es su precisión en la estimación de modelos complejos y  
72 su capacidad para manejar estructuras jerárquicas detalladas. Sin embargo, su principal desventaja radica  
73 en la complejidad computacional y el tiempo necesario para realizar inferencias, especialmente cuando  
74 se aplican modelos con múltiples niveles y muchos parámetros.

75 **SPSS y Stata:** Para aquellos usuarios que prefieren interfaces gráficas, SPSS y Stata también ofrecen  
76 capacidades para trabajar con modelos multiniveles. Estas herramientas son más accesibles para usuarios  
77 con menos experiencia en programación, lo cual facilita la implementación de modelos sin necesidad de  
78 escribir código. La desventaja de estos programas es que son de pago y sus capacidades de  
79 personalización son más limitadas comparadas con las opciones de código abierto como R o Python.

80 La elección del lenguaje o software dependerá del propósito del estudio, las habilidades del  
81 investigador y la disponibilidad de recursos computacionales. Mientras que R y Python ofrecen  
82 opciones poderosas y flexibles para análisis detallados, herramientas como SPSS y Stata pueden ser  
83 más adecuadas para quienes buscan facilidad de uso en detrimento de la capacidad de personalización.

### 84 3. Área de estudio

85  
86 El área de estudio se encuentra al norte de  $5^{\circ}\text{N}$  y abarca las Cordilleras Occidental y Central de los  
87 Andes colombianos, separadas por el cañón del Cauca. La región está delimitada por el río Atrato al  
88 oeste y el río Magdalena al este, cubriendo aproximadamente  $50,000\text{ km}^2$ . El área está dividida en 533  
89 subcuencas dentro de las cuencas hidrográficas del Atrato (25% de las cuencas), Cauca (50%) y  
90 Magdalena (25%). El  $\sim 73\%$  de estas cuencas tienen áreas menores a  $100\text{ km}^2$ , con un área media de  $48$   
91  $\text{km}^2$ .



92  
93  
94  
95  
96  
97

**Fig. 1.** Mapa de localización del área de estudio en las Cordilleras Occidental y Central de los Andes colombianos. Las cruces negras representan los movimientos en masa registrados entre 1970 y 2023. Se muestran los principales ríos y la delimitación de las subcuencas que conforman las cuencas del Atrato (verde), Cauca (rojo) y Magdalena (azul).

98

#### 4. Datos y metodología

99  
100

En este estudio se elaboró un catálogo de 13.777 movimientos en masa superficiales y profundos ocurridos entre 1970 y 2023 (Fig. 1). Todos los eventos corresponden a movimientos en suelo y

101 presentan una alta variabilidad areal, desde unas pocas decenas de metros cuadrados hasta varias  
102 hectáreas de extensión. Cada evento fue mapeado como un punto localizado en el centro de la corona  
103 del deslizamiento, ya que esta zona representa de forma más precisa las condiciones previas a la  
104 ocurrencia del movimiento en masa. La identificación se realizó mediante interpretación visual de  
105 imágenes satelitales, incluyendo imágenes PlanetScope (resolución de 3–4 m) y de Google Earth, desde  
106 1985 hasta 2023, con resolución submétrica desde 2001.

107  
108 Inicialmente, se seleccionaron como variables predictoras para el área de estudio: el área bajo la curva  
109 hipsométrica de la subcuenca, la elevación media, la pendiente media, el relieve local medio, el área de  
110 la subcuenca, la densidad de lineamientos, la precipitación media anual, el número de días con  
111 precipitación diaria acumulada >20 mm, la cobertura del suelo y la geología predominante. La  
112 información geológica fue obtenida a partir de mapas a escala 1:25.000 incluidos en los planes de  
113 ordenamiento territorial de varios municipios y en los planes de ordenación de cuencas hidrográficas.  
114 Complementariamente, se utilizó el mapa geológico nacional a escala 1:100.000 publicado por el  
115 Servicio Geológico Colombiano (SGC, 2013).

116  
117 La elección de estas variables se fundamentó en una revisión crítica de la literatura. Aunque existen  
118 numerosos estudios sobre susceptibilidad a movimientos en masa a escala de ladera (Reichenbach et al.,  
119 2018), a escala de cuenca los trabajos son menos frecuentes y requieren variables que representen  
120 adecuadamente las condiciones generales del paisaje. Por ejemplo, el área bajo la curva hipsométrica es  
121 un indicador del grado de rejuvenecimiento del relieve, siendo más alto en terrenos jóvenes y dinámicos.  
122 La elevación media y el relieve local reflejan la energía potencial del terreno, mientras que la pendiente  
123 media está directamente relacionada con la estabilidad del talud. La precipitación media anual representa  
124 la carga hídrica acumulada sobre las cuencas, y el número de días con lluvias intensas permite  
125 aproximarse a los factores detonantes. La cobertura del suelo aporta información sobre el tipo de  
126 vegetación o uso del suelo que puede modificar la estabilidad de la cuenca, y la geología predominante  
127 y densidad de lineamientos identifican unidades litológicas con diferentes propiedades estructurales y  
128 mecánicas frente a la ocurrencia de movimientos en masa.

129 Los parámetros del terreno se calcularon utilizando el modelo digital de elevación (DEM, por sus siglas  
130 en inglés) del radar de apertura sintética en banda L del satélite de observación terrestre avanzada  
131 (ALOS-PALSAR), con una resolución de 12.5 m (Logan y Karr, 2014).

132 Las métricas de precipitación se derivaron utilizando los datos de CHIRPS (*Climate Hazard Group*  
133 *Infrared Precipitation with Station Data*) (Funk et al., 2015), versión 2.0, para datos con una resolución  
134 de 5 km desde 1981 hasta 2023. Para este cálculo se utilizó la plataforma Google Earth Engine (GEE)  
135 con lenguaje Javascript.

136 Se generó un mapa de cobertura del suelo utilizando datos del programa Copernicus Sentinel-2, con  
137 una resolución de 10 m. Utilizando GEE, se seleccionaron imágenes desde 2018 hasta 2023 con menos  
138 del 10% de cobertura de nubes, eliminando efectos de sombra. El mapa se creó calculando el índice de  
139 vegetación de diferencia normalizada (NDVI) (Kriegler, 1969), el índice de diferencia normalizada de  
140 construcciones (NDBI) (Zha et al., 2003), el índice de diferencia normalizada de agua modificado  
141 (MNDWI) (Xu, 2006), y el índice de suelo desnudo (BSI) (Li et al., 2014). La clasificación de cobertura  
142 del suelo se realizó con un algoritmo de bosque aleatorio (*random forest*) del paquete Classifier de GEE  
143 (Breiman, 2001).

144 Para implementar los modelos multiniveles, se definieron regiones. En este estudio, se utilizaron dos  
145 regiones distintas. El primer régimen corresponde a las divisiones naturales por las principales cuencas:  
146 Atrato, Cauca y Magdalena. El segundo régimen utilizó la clusterización aglomerativa basada en  
147 parámetros morfométricos y una matriz de conectividad espacial, aplicada con diferentes  $K$  vecinos  
148 más cercanos para  $K = 2, 3, 4, 5$  y  $6$ . La agrupación se implementó en Python utilizando la librería  
149 Scikit-learn, guiada por el método del codo (*Elbow method*) y el coeficiente de Silhouette (Pedregosa  
150 et al., 2011). El método del codo ayuda a identificar el número óptimo de grupos trazando la suma de  
151 cuadrados dentro del grupo contra  $K$  para detectar el punto de inflexión (Thorndike, 1953). El  
152 coeficiente de Silhouette evalúa la cohesión dentro de los grupos y la separación entre los grupos, con

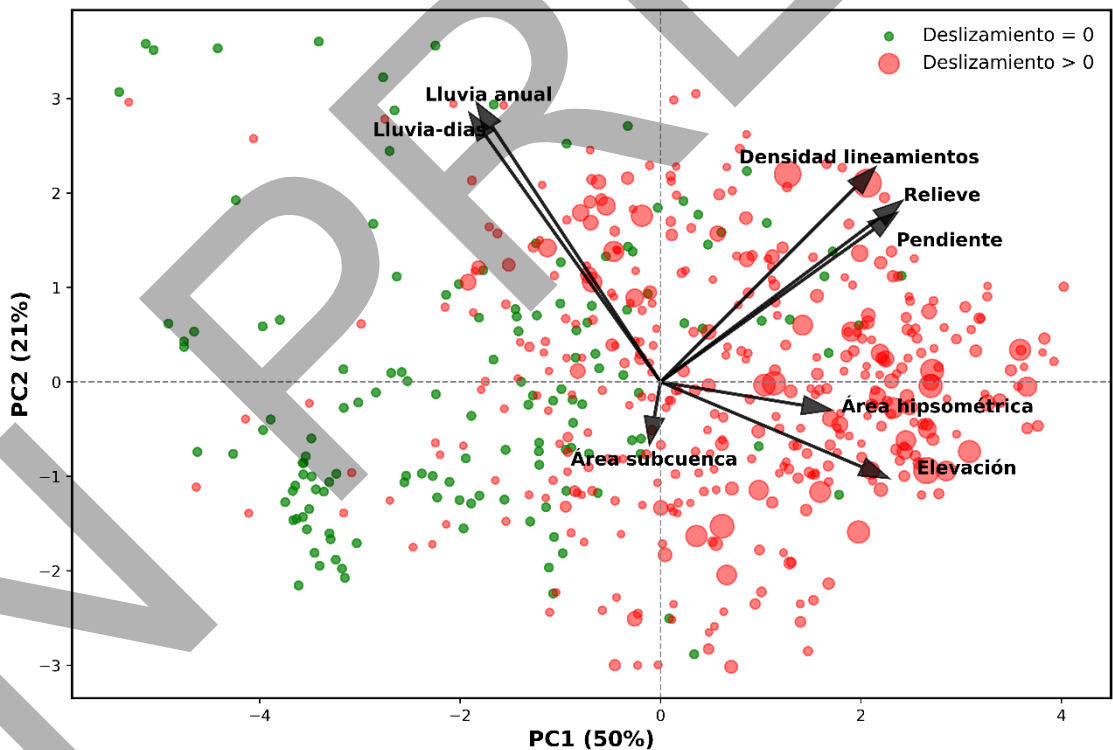
153 un rango de  $-1$  (mala agrupación) a  $+1$  (excelente agrupación) (Rousseeuw, 1987).

154

155 Como paso previo a la modelación, se aplicó un Análisis de Componentes Principales (PCA) para  
156 reducir la dimensionalidad del conjunto de variables y minimizar problemas de colinealidad (Fig. 2). El  
157 análisis se implementó con la librería Scikit-learn en Python. Se obtuvieron ocho componentes  
158 principales (uno por variable). Aunque los tres primeros componentes explican el 50%, 21% y 12% de  
159 la varianza respectivamente, solo se muestran los dos primeros (PC1, PC2) dado que los componentes  
160 restantes explican una proporción muy baja de la varianza y no aportan interpretaciones claras. El análisis  
161 en la figura 2 destaca cuatro grupos de predictores colineales: (1) relieve local medio, densidad de  
162 lineamientos y pendiente promedio, (2) área bajo la curva hipsométrica y elevación media, (3)  
163 precipitación media anual, y (4) área de la subcuenca. A partir de estos resultados, se seleccionó de cada  
164 grupo un predictor tomando en consideración el porcentaje de varianza que explica cada componente y  
165 su relevancia teórica. Así, el relieve local fue escogido como medida de la energía del paisaje; la  
166 elevación media, como indicador general de las condiciones topográficas; la precipitación media anual,  
167 como expresión del régimen climático dominante; y el área de la subcuenca, como una aproximación al  
168 tamaño de la unidad de mapeo y su probabilidad de contener eventos.

169

170 Para modelar la susceptibilidad a movimientos en masa, se emplearon modelos de regresión logística  
171 utilizando el paquete lm4 en R. Se construyeron dos modelos: (1) un modelo de regresión logística simple  
172 sin estructura jerárquica, y (2) un modelo multinivel con intercepto y pendientes aleatorias por región  
173 (cuenca o cluster). Estos modelos permitieron capturar la variabilidad global y local de los factores  
174 condicionantes, evaluando el impacto de la heterogeneidad espacial en los patrones de susceptibilidad.



175

176 **Fig. 2.** Análisis de Componentes Principales (PCA) para evaluar la relación entre las subcuencas con  
177 movimientos en masa y las variables predictoras. Los puntos rojos representan subcuencas con presencia de  
178 movimientos en masa, donde el tamaño de cada punto es proporcional al número de movimientos en masa  
179 registrados en esa subcuenca. Los puntos verdes indican ausencia de movimientos en masa. Las flechas  
180 indican las contribuciones relativas de las variables predictoras en el espacio de los componentes principales.

181

182 En este estudio no se realizó una división del conjunto de datos en subconjuntos de entrenamiento y  
183 validación, ya que el objetivo principal fue evaluar el ajuste y la estructura del modelo de susceptibilidad  
184 a nivel de subcuenca más que desarrollar un modelo predictivo generalizable. Para evaluar el desempeño

185 de los modelos se utilizaron métricas de bondad de ajuste como el área bajo la curva ROC (AUC), así  
186 como los criterios de información de Akaike (AIC) y de Bayes (BIC).  
187

## 188 4. Resultados

189

### 190 4.1 Modelo de regresión logística simple

191 La Tabla 1 presenta un resumen del modelo de regresión logística simple, con las estimaciones de los  
192 coeficientes para las variables predictoras seleccionadas. El intercepto tiene un coeficiente altamente  
193 significativo (valor  $P < 0,001$ ). El valor Z indica cuán lejos está el coeficiente de ser cero en términos de  
194 desviaciones estándar, mientras que el valor de P dice cuán probable es obtener un valor de cero bajo la  
195 hipótesis nula. El coeficiente representa el logaritmo natural de la relación entre la ocurrencia y no  
196 ocurrencia de movimientos en masa cuando todas las variables predictoras son cero.

197

Tabla 1. Modelo de regresión logística simple.

198

Variabes	Estimado	Error Std	Valor Z	Valor P
Intercepto	1,762	0,173	10,164	<0,001
Elevación media	0,794	0,213	3,727	<0,001
Relieve local	1,242	0,214	5,802	<0,001
Área subcuenca	0,680	0,162	4,197	<0,001
Precipitación media anual	-0,508	0,173	-2,935	0,00333
Cox & Snell $R^2$	0,354	NA	NA	NA
AIC	397,828	NA	NA	NA

199

Std: Desviación estándar.

200

AIC: Criterio de Información de Akaike.

201 Las variables predictoras, como la elevación, el relieve, el área y la precipitación anual, presentan  
202 coeficientes estadísticamente significativos (valor  $P < 0,05$ ). La elevación media tiene un coeficiente  
203 positivo, indicando una asociación directa entre mayores altitudes y la ocurrencia de movimientos en  
204 masa. El relieve local también muestra una influencia positiva, sugiriendo que las cuencas de mayor  
205 pendiente son más inestables. Respecto al área de la subcuenca, se observa que aquellas de mayor tamaño  
206 tienden a contener más eventos registrados. Por su parte, la precipitación media anual presenta un  
207 coeficiente negativo. Si bien esto último puede parecer contradictorio, es importante destacar que esta  
208 variable no se interpreta como un desencadenante inmediato, sino como un indicador de las condiciones  
209 climáticas dominantes de largo plazo. En este contexto, una mayor precipitación anual puede estar  
210 asociada a una mayor cobertura vegetal o a suelos más estabilizados por la humedad constante,  
211 reduciendo así la susceptibilidad general a movimientos en masa.

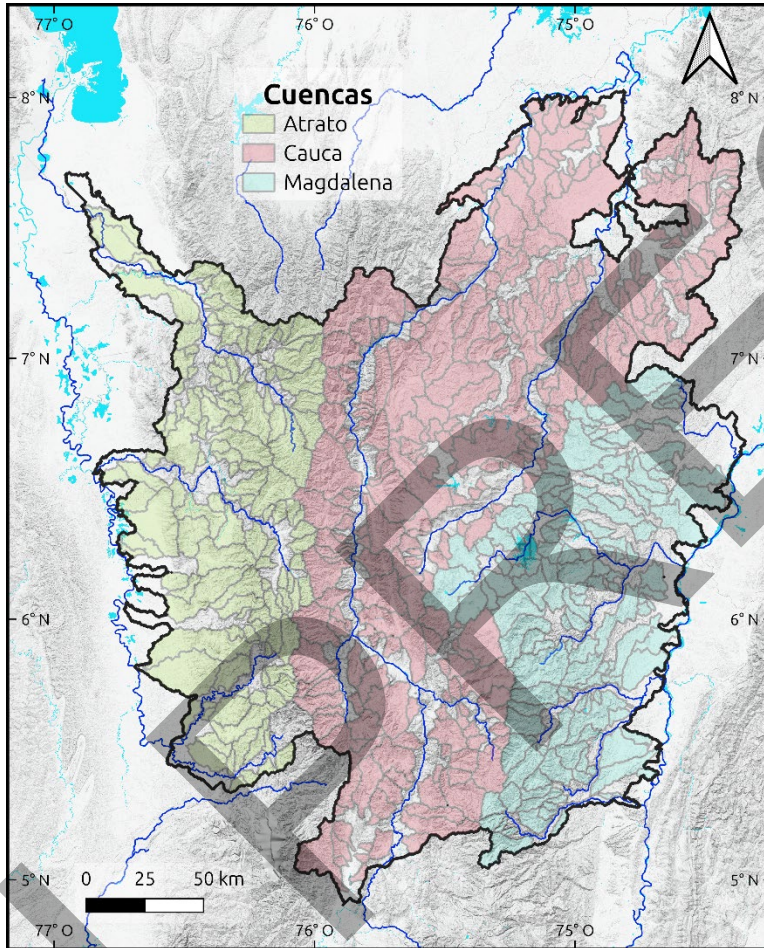
212

### 213 4.2 Modelos de regresión logística multiniveles

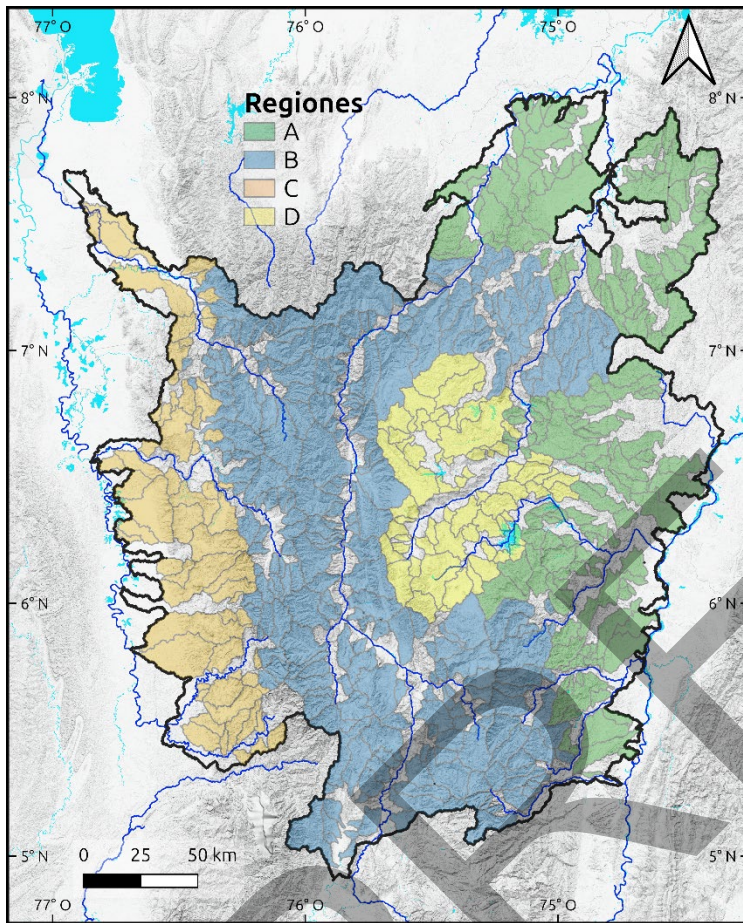
214

215 Para implementar los modelos multinivel es necesario establecer regiones homogéneas. La segmentación  
216 geográfica y natural en cuencas no necesariamente explica la ocurrencia de movimientos en masa. Aunque  
217 el área de estudio se divide naturalmente en las cuencas de los ríos Atrato, Cauca y Magdalena, se  
218 exploraron dos enfoques complementarios para definir regiones espaciales multinivel (Fig. 3). El primero  
219 corresponde a la jerarquía hidrográfica tradicional de cuencas y subcuencas (modelo 'cuencas'), y el  
220 segundo a una agrupación espacial derivada de parámetros morfométricos (modelo 'cluster'). Este último  
221 fue obtenido mediante un análisis de clusterización aglomerativa espacial, utilizando diferentes números

222 de vecinos. El mejor resultado, según el método del codo y el coeficiente de Silhouette, se obtuvo con  
223  $K=5$  vecinos y cuatro grupos. El grupo A corresponde a los terrenos aluviales bajos a lo largo de las  
224 cuencas del Magdalena y Cauca; el grupo B representa los terrenos de alto relieve a lo largo del cañón del  
225 Cauca; el grupo C corresponde a los terrenos aluviales bajos a lo largo de la cuenca del Atrato; y el grupo  
226 D comprende las superficies de alta elevación y bajo relieve en el eje de la Cordillera Central. Así, este  
227 enfoque ofrece una visión alternativa y complementaria para capturar la heterogeneidad espacial del área  
228 de estudio más allá de los límites hidrográficos. La comparación entre ambos modelos se muestra en la  
229 tabla 2.



230



231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241

**Fig. 3.** Comparación de los regímenes espaciales utilizados en el análisis multinivel de susceptibilidad a movimientos en masa. El mapa a la izquierda muestra la división por cuencas hidrográficas (Atrato, Cauca y Magdalena), utilizadas como unidades espaciales de nivel superior. El mapa a la derecha muestra el resultado de la clusterización espacial utilizando cinco vecinos más cercanos, lo cual permitió definir cuatro regiones morfométricamente homogéneas (A, B, C y D).

**Tabla 2.** Comparación de modelos de regresión logística multinivel.

Estadístico	Modelo 'cuencas'	Modelo 'cluster'
AIC	405,5	420,5
BIC	490,8	505,8
Log-verosimilitud	-182,7	-190,2
Desviación	365,5	380,5
<b>Efectos fijos</b>		
Intercepto	2,0011	2,0580
Elevación media	0,8566	0,7834
Relieve local	1,0958	1,2863
Área subcuenca	0,7033	1,2102
Precipitación media anual	-0,3884	-0,4966
<b>Efectos aleatorios (Varianza)</b>		
Intercepto	0,0000	0,0000

Elevación media	0,0033	0,0007
Relieve local	0,8298	0,1016
Área subcuena	0,0075	0,3551
Precipitación media anual	0,2781	0,0007

AIC: Criterio de Información de Akaike.

BIC: Criterio de Información Bayesiano.

242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256

El ajuste de ambos modelos se efectuó mediante el Criterio de Información de Akaike (AIC), el Criterio de Información Bayesiano (BIC), el logaritmo de máxima verosimilitud y la desviación (Tabla 2). El BIC, al igual que el AIC, también mide el ajuste, aunque el primero penaliza más fuertemente la complejidad del modelo. El logaritmo de verosimilitud mide qué tan bien el modelo se ajusta a los datos; valores más altos (i.e., menos negativos) indican un mejor ajuste. La desviación es una medida del ajuste del modelo, similar a la suma de cuadrados residuales en modelos lineales: valores más bajos son mejores. En términos de los estadísticos AIC y BIC, el modelo ‘cuencas’ presenta valores más bajos (AIC: 405,5; BIC: 490,8) en comparación con el modelo ‘cluster’ (AIC: 420,5; BIC: 505,8), lo cual sugiere que el modelo basado en cuencas logra un mejor equilibrio entre ajuste y complejidad del modelo. De manera similar, el logaritmo de verosimilitud y la desviación indican un mejor ajuste para el modelo ‘cuencas’ (-182,7 y 365,5, respectivamente) frente al modelo ‘cluster’ (-190,2 y 380,5).

257  
258  
259  
260  
261  
262  
263  
264  
265

En cuanto a los efectos fijos, los coeficientes representan el efecto promedio de cada variable predictora en todos los grupos (Tabla 2). Los coeficientes de las variables predictoras, por ejemplo, elevación media, indican que un aumento de una unidad en la elevación incrementa el logaritmo de la probabilidad de la ocurrencia de un movimiento en masa en 0,8566 para el modelo ‘cuencas’. Los efectos fijos para ambos modelos muestran coeficientes relativamente similares, pero con algunas diferencias: el relieve tiene un coeficiente más alto en el modelo ‘cluster’ (1,2863) en comparación con el modelo ‘cuencas’ (1,0958), mientras que el área de la subcuena muestra un incremento significativo en su influencia en el modelo ‘cluster’ (1,2102) frente al otro modelo (0,7033); en cambio, la precipitación anual tiene un efecto negativo en ambos modelos, pero es más pronunciado en el modelo ‘cluster’ (-0,4966).

266  
267  
268  
269  
270  
271  
272

En cuanto a los efectos aleatorios, la varianza del intercepto es cero en ambos modelos, indicando que no hay una variabilidad considerable entre las regiones en términos del valor base. Sin embargo, la varianza del relieve es mucho mayor en el modelo ‘cuencas’ (0,8298) comparado con el modelo ‘cluster’ (0,1016), lo cual sugiere una mayor heterogeneidad en el efecto del relieve entre las cuencas. De manera similar, la varianza del área de la subcuena es mayor en el modelo ‘cluster’ (0,3551) en comparación con el otro modelo (0,0075), indicando que la variabilidad espacial de esta variable se captura mejor en el análisis basado en grupos.

273  
274  
275  
276  
277  
278  
279  
280

En la tabla 3 se presentan los coeficientes de los efectos aleatorios para cada cuenca utilizada como variable aleatoria en el modelo multinivel. El intercepto para cada cuenca es 0, lo cual implica que el punto de partida de la predicción es consistente entre todas las cuencas. En las cuencas del Atrato y del Magdalena, el coeficiente de la elevación media es negativo, lo cual indica que la elevación tiene un efecto ligeramente reducido sobre la susceptibilidad a movimientos en masa comparado con el valor general del modelo. En la cuenca del Cauca, en tanto, este coeficiente es positivo, indicando que la elevación media aumenta la probabilidad de movimientos en masa y que su efecto es mayor que el promedio.

281  
282  
283  
284  
285  
286

Respecto al coeficiente de relieve, en el Atrato este es marcadamente negativo (-1,017), lo cual indica que el relieve medio tiene un efecto considerablemente menor en esta cuenca y que otras características del terreno estarían moderando el impacto del relieve en la susceptibilidad. En el Cauca, este coeficiente es marcadamente positivo (1,118), por lo que el relieve medio tendría una influencia fuerte y positiva en la probabilidad de movimientos en masa. Mientras que en el Magdalena este valor es bajo (-0,133), sugiriendo que el efecto del relieve en esta cuenca es limitado.

287  
288

El área de la subcuena muestra valores cercanos a cero para todas las cuencas, por lo que el área no contribuiría de manera importante a aumentar la susceptibilidad. En el Atrato, el coeficiente de

289 precipitación anual es negativo (-0,510), lo que indica que mayores niveles de lluvia anual tienden a  
 290 disminuir la susceptibilidad a movimientos en masa. Para el Cauca el valor de este coeficiente también  
 291 es negativo, aunque más pequeño (-0,079), mientras que en el Magdalena es positivo (0,600), lo que  
 292 indica una relación directa entre la lluvia anual y la susceptibilidad a movimientos en masa.

293  
 294 **Tabla 3.** Efectos aleatorios para cada cuenca. Los coeficientes para cada cuenca (Atrato, Cauca y  
 295 Magdalena) muestran la variabilidad en el efecto de las variables predictoras en el modelo multinivel.

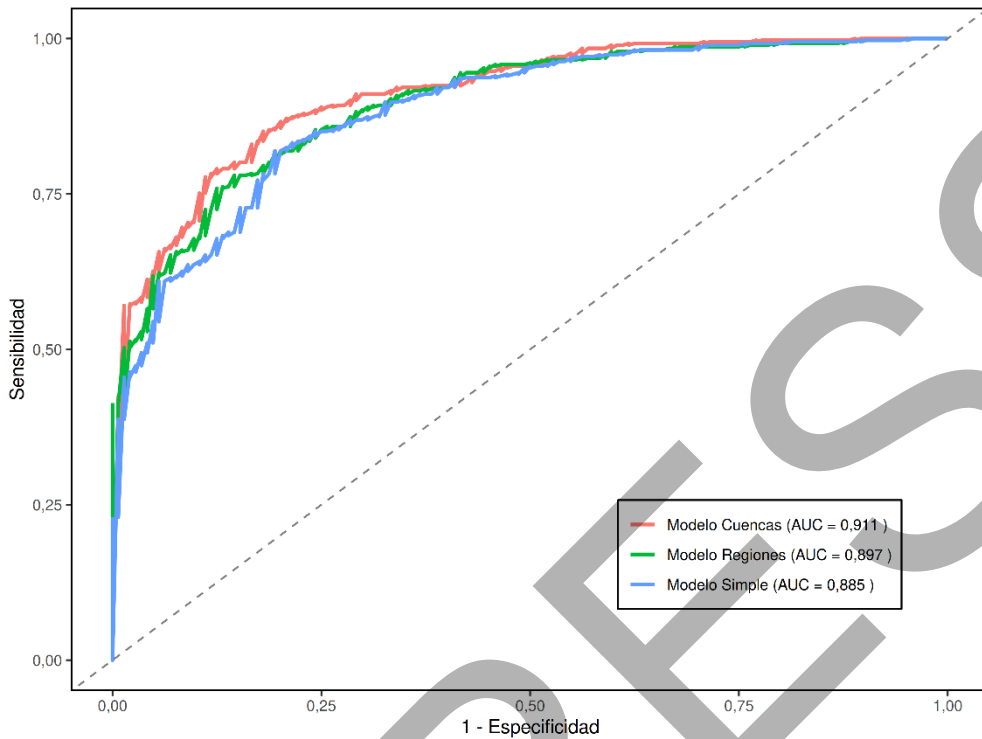
Cuenca	Intercepto	Elevación media	Relieve local	Área subcuenca	Precipitación media anual
Atrato	0	-0,0322642	-1,0171946	0,1127104	-0,5098330
Cauca	0	0,0731148	1,1184392	-0,0475954	-0,0786611
Magdalena	0	-0,0434983	-0,1330732	-0,0649104	0,6003614

297  
 298  
 299 La tabla 4 muestra los coeficientes de los efectos aleatorios para cada una de las regiones definidas  
 300 utilizando el método de agrupamiento. En el grupo A, el relieve medio presenta un coeficiente positivo  
 301 significativo, lo cual sugiere que las zonas con mayor relieve son más susceptibles a movimientos en  
 302 masa; la lluvia anual tiene un efecto mínimo y positivo sobre la susceptibilidad, mientras que la elevación  
 303 media tiene un efecto negativo. Para el grupo B, la elevación y el relieve medios se comportan de manera  
 304 opuesta con respecto al grupo A, y la lluvia anual posee un coeficiente negativo leve, sugiriendo una  
 305 pequeña reducción en la susceptibilidad. En el grupo C, los coeficientes son relativamente pequeños para  
 306 todas las variables, por lo que el efecto de todas ellas en la susceptibilidad es bajo; lo mismo ocurre para  
 307 el grupo D.

308  
 309 **Tabla 4.** Efectos aleatorios para regiones por grupos. Los coeficientes para cada grupo (A, B, C y D)  
 310 muestran la variabilidad en el efecto de las variables predictoras en el modelo multinivel.

Región	Intercepto	Elevación media	Relieve local	Área subcuenca	Precipitación media anual
A	0	-0,0324473	0,3944349	-0,7375476	0,0328496
B	0	0,0316389	-0,3846077	0,7191720	-0,0320312
C	0	0,0051221	-0,0622649	0,1164282	-0,0051856
D	0	-0,0034080	0,0414280	-0,0774655	0,0034502

312  
 313  
 314 La figura 4 presenta las curvas ROC (*receiver operating characteristic curve*) para el modelo de  
 315 regresión logística simple (azul) y los dos modelos multivariados de regresión logística con multiniveles.  
 316 Tanto el modelo de cuencas (representado en color rojo) como el modelo de clusterización (en color  
 317 verde) muestran un rendimiento predictivo superior al modelo simple, con valores de área bajo la curva  
 318 (AUC) de 0,911 y 0,897, respectivamente. La proximidad de ambas curvas refleja una capacidad similar  
 319 para predecir el resultado de la variable dependiente, aunque el modelo de cuencas presenta un leve  
 320 incremento en su AUC, lo que sugiere un mejor desempeño predictivo. Es importante aclarar que estos  
 321 valores elevados de AUC se refieren únicamente a la capacidad del modelo para discriminar entre  
 322 subcuencas con presencia y ausencia de movimientos en masa conforme al planteamiento de un modelo  
 323 logístico binario. Dado que el análisis se realiza a nivel de unidad hidrográfica, no implica  
 324 necesariamente un alto nivel de precisión espacial como sí ocurriría, por ejemplo, con modelos a escala  
 325 de ladera. Este tipo de modelado debe, por lo tanto, entenderse como una aproximación regional inicial,  
 326 de utilidad para identificar patrones generales (regionales) de susceptibilidad sin sustituir análisis  
 327 detallados a escala de ladera.



328

329

330

331

332

333

334

335

336

337

338

339

340

341

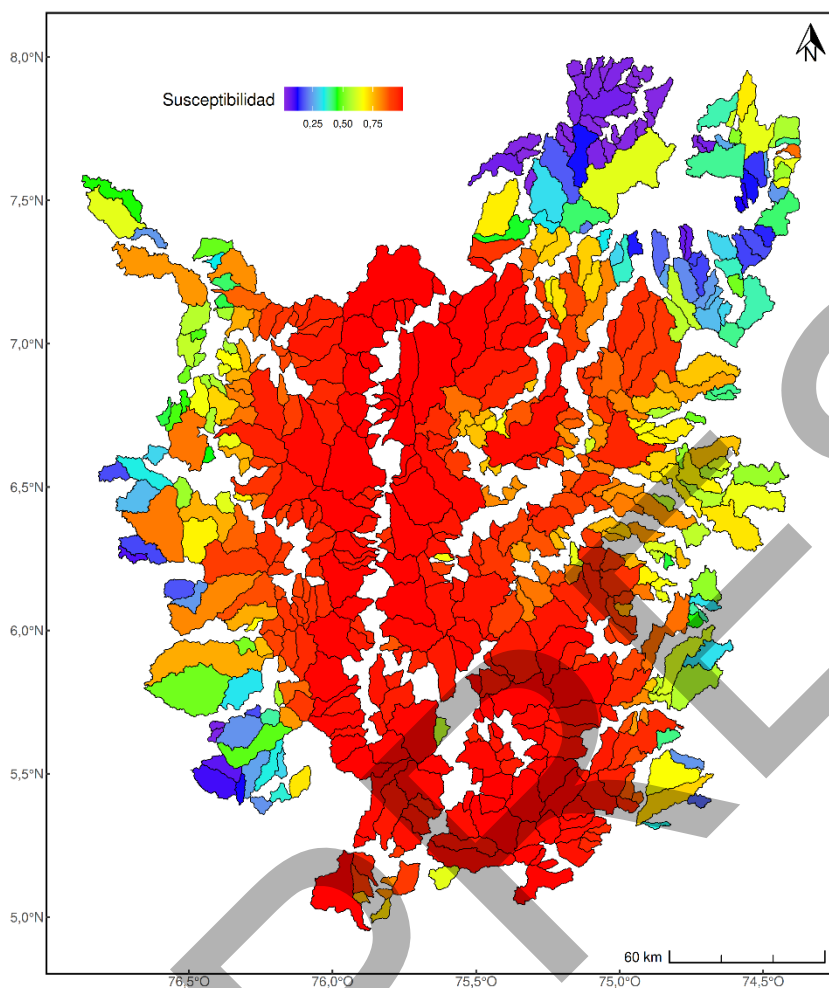
342

343

344

**Fig. 4.** Curvas ROC (*receiver operating characteristic curve*) para los modelos de regresión logística simple (azul) y multinivel (rojo, verde). Cada curva muestra el rendimiento predictivo de cada modelo en términos de sensibilidad (verdadero positivo) frente a 1 - especificidad (falso positivo). Se incluyen los valores del área bajo la curva (AUC) para cada modelo.

Tomando en consideración los resultados de la figura 4 y los valores de AUC, la distribución espacial de la susceptibilidad a movimientos en masa dentro del área de estudio se estimó utilizando el modelo de regresión logística multinivel con cuencas como regiones (Fig. 5). La zona de alta susceptibilidad se presenta para el sector central, en cuencas con laderas de fuertes pendientes, alto relieve y procesos geomorfológicos activos. Los sectores este y noroeste presentan una susceptibilidad comparativamente menor, caracterizada por cuencas con pendientes más suaves y menor elevación, mientras que los menores valores de susceptibilidad se obtuvieron para los sectores noreste y suroeste. Es importante aclarar que valores bajos de susceptibilidad no implica necesariamente la ausencia de movimientos en masa, sino que refleja una menor probabilidad relativa de ocurrencia en comparación con otras zonas dentro del área de estudio.



345  
346  
347  
348

**Fig. 5.** Distribución espacial de la susceptibilidad a movimientos en masa en el área de estudio utilizando el modelo de regresión logística multinivel agrupada en cuencas hidrográficas (modelo 'cuencas'). La susceptibilidad se expresa en una escala continua que varía desde muy baja (morado) hasta muy alta (rojo).

349

## 5. Discusión

350  
351  
352  
353  
354  
355  
356  
357  
358  
359

La evolución de los modelos de regresión para evaluar la susceptibilidad a movimientos en masa ha avanzado considerablemente, evolucionando de modelos lineales tradicionales a la familia de Modelos Lineales Generalizados (GLM) (Fox, 2015; Chen et al., 2019; Patané et al., 2025). Esta última incluye funciones de enlace empleadas en regresión logística, como la función *logit*, facilitando así la representación de relaciones no lineales entre la variable dependiente (presencia de deslizamientos) y variables predictoras (e.g., pendiente, geología, cobertura del suelo). Estas relaciones son críticas en regiones de alta complejidad topográfica y geológica como los Andes colombianos, donde la orogenia activa ha dado lugar a un relieve montañoso con laderas de fuerte pendiente. La combinación de pendientes pronunciadas y lluvias intensas inducidas por la topografía crea un escenario propenso a la ocurrencia recurrente de movimientos en masa (Aristizábal y Korup, 2025).

360  
361  
362  
363  
364

La incorporación de la heterogeneidad geológica y geomorfológica del terreno, así como de otros patrones espaciales, son importantes para evaluar adecuadamente la susceptibilidad a movimientos en masa. Por el contrario, asumir que las variables predictivas tienen una influencia homogénea en toda la región puede conducir a interpretaciones erróneas y a generalizaciones que no son adecuadas (Anselin, 1990; LeSage y Pace, 2009). Por ejemplo, en una región montañosa, la pendiente puede tener una

365 influencia mucho mayor en la ocurrencia de movimientos en masa en comparación con una región de  
366 pendientes más suaves, donde el uso del suelo o sus características físicas serían factores más  
367 determinantes. Los modelos multiniveles proporcionan una solución robusta para capturar y manejar  
368 estas heterogeneidades, permitiendo modelar la variación de los coeficientes de las diferentes unidades  
369 espaciales sin incrementar significativamente la complejidad computacional (Wong y Mason, 1985). De  
370 esta manera, los modelos multiniveles pueden ofrecer una representación más precisa de los patrones de  
371 susceptibilidad a movimientos en masa, lo cual es particularmente relevante en zonas topográfica y  
372 geológicamente complejas como los Andes colombianos (Aristizábal et al., 2015).

373 Los coeficientes estimados del modelo permiten una lectura geológicamente coherente de los procesos  
374 condicionantes a movimientos en masa en la región. En primer lugar, la elevación muestra un efecto  
375 positivo sobre la probabilidad de movimientos en masa, lo cual puede explicarse por la asociación de  
376 mayores altitudes con zonas montañosas jóvenes. Estas zonas presentan una mayor energía del relieve  
377 y, a su vez, pendientes más marcadas, características que incrementan la inestabilidad del terreno. El  
378 relieve local, por otra parte, presenta el coeficiente fijo más alto del modelo (1,0958) y una varianza  
379 aleatoria también elevada (0,8298), lo cual refleja con claridad que la amplitud vertical del terreno es un  
380 determinante clave en la susceptibilidad. Las diferencias locales en esta variable, registradas por su  
381 varianza, revelan que en algunas regiones el relieve actúa como un fuerte disparador de inestabilidad,  
382 mientras que en otras su influencia es más moderada, reflejando la heterogeneidad del paisaje andino.  
383 En tercer lugar, el coeficiente positivo del área como factor predictor señala que subcuencas de mayor  
384 tamaño tienden a registrar más movimientos en masa, lo cual puede estar vinculado a una mayor  
385 diversidad topográfica y litológica, así como a una mayor probabilidad de incluir zonas con condiciones  
386 críticas de pendiente, cobertura vegetal y acumulación hídrica. Respecto a la precipitación media anual,  
387 su coeficiente negativo asociado a una alta variabilidad espacial indica un efecto complejo. Aunque la  
388 lluvia es reconocida como el principal detonante de los movimientos en masa en la región andina  
389 (Aristizábal y Sanchez, 2020), en este caso su inclusión en el modelo responde a su papel como variable  
390 condicionante y no gatillante (van Westen et al., 2008). En este trabajo, la precipitación media anual no  
391 representa la intensidad o recurrencia de eventos extremos, sino condiciones climáticas dominantes de  
392 largo plazo que pueden estar asociadas a laderas más estables mediante una mayor cobertura vegetal.  
393 Así, el resultado negativo no significa una menor influencia de la lluvia como disparador, sino que refleja  
394 cómo las condiciones promedio de humedad pueden incidir en la configuración de laderas menos  
395 propensas a fallar. Finalmente, el modelo muestra que los efectos aleatorios para relieve y precipitación  
396 son sustanciales, lo que confirma que estas variables no ejercen una influencia homogénea en toda la  
397 región, sino que su efecto depende de las características geológicas y geomorfológicas locales, como el  
398 tipo de roca, la cobertura edáfica, la historia tectónica reciente o el grado de consolidación de los  
399 materiales.

400 La mayoría de los estudios de susceptibilidad emplean celdas regulares en mapas ráster (Arnone et al.,  
401 2016), lo cual, si bien facilita el análisis y la aplicación de técnicas computacionales, no necesariamente  
402 refleja las características naturales del paisaje. Esta representación arbitraria puede resultar inadecuada  
403 para capturar la dinámica de los movimientos en masa, dado que no respeta las divisiones naturales del  
404 terreno. En contraposición, estudios recientes han sugerido el uso de unidades de mapeo más  
405 representativas, como subcuencas para estudios regionales o unidades de ladera para estudios locales  
406 (Erener y Düzgün, 2012). Estos enfoques permiten una evaluación más precisa de la susceptibilidad y  
407 mejoran la capacidad de los modelos para identificar áreas de susceptibilidad o amenaza. Es por ello que  
408 la definición adecuada de las regiones de análisis es una componente esencial en el éxito de los modelos  
409 de susceptibilidad a movimientos en masa. Una regionalización incorrecta puede resultar en la pérdida  
410 de información relevante o en la generación de límites que no capturarían de manera adecuada las  
411 características geomorfológicas y dinámicas de la zona, llevando así a un ajuste deficiente del modelo.  
412 En este estudio, se exploraron dos enfoques de regionalización: una natural basada en cuencas y otra  
413 basada en un análisis de agrupamiento espacial mediante técnicas de clusterización. Ambos  
414 agrupamientos presentan ventajas diferentes. La primera utiliza cuencas hidrográficas, las cuales  
415 representan divisiones geográficas naturales alineadas con procesos geomorfológicos, por lo que  
416 proporcionan una estructura lógica para el análisis de la dinámica hidrológica. Por otro lado, los grupos  
417 definidos mediante agrupamiento espacial permiten crear regiones homogéneas desde un punto de vista  
418 morfométrico independientemente de las divisiones hidrológicas del terreno (Erener y Düzgün, 2010).

419 Los resultados obtenidos muestran que la regionalización basada en cuencas presentó un mejor ajuste  
420 general. Esto podría estar vinculado con la coherencia de los límites hidrográficos respecto a la dinámica  
421 de los movimientos en masa. Futuros trabajos podrían explorar otras estrategias de regionalización para  
422 determinar si existe una forma de mejorar la capacidad predictiva del modelo. Finalmente, el uso de  
423 subcuencas hidrográficas como unidades de mapeo favorece la construcción de modelos jerárquicos  
424 hábiles en capturar las variaciones regionales de los componentes considerados como importantes.

425 La regresión logística ha sido ampliamente utilizada para la evaluación de la susceptibilidad a  
426 movimientos en masa (Budimir et al., 2015; Lombardo y Mai, 2018; Reichenbach et al., 2018). Sin  
427 embargo, este enfoque presenta limitaciones significativas, particularmente cuando se intenta modelar la  
428 ocurrencia de múltiples movimientos en masa en áreas discretas de mayor extensión a celdas regulares  
429 (Lombardo et al., 2018). La naturaleza binaria de la variable dependiente no permite distinguir entre la  
430 ocurrencia de un solo evento o múltiples eventos. Es decir, en una unidad de mapeo solo se puede  
431 especificar la ocurrencia (1) o no ocurrencia (0) de movimientos en masa. Esta limitación es  
432 especialmente relevante cuando se trabaja con unidades de análisis grandes, como subcuencas, en las  
433 cuales pueden registrarse múltiples eventos que no pueden ser diferenciados en un modelo logístico  
434 binario. En contraste, cuando se emplean celdas de resolución más fina, el tamaño de la celda puede ser  
435 comparable al de los deslizamientos, reduciendo así esta limitación; por tanto, la escala y resolución del  
436 análisis condicionan la precisión con la que se puede representar la frecuencia de eventos en el modelado  
437 de la susceptibilidad. Pero para unidades de mapeo con áreas mayores a las de los movimientos en masa,  
438 cada unidad de mapeo puede tener más de un evento, lo cual puede resultar en una subestimación de la  
439 susceptibilidad. La regresión de tipo Poisson, también perteneciente a la familia de los GLM, ofrece la  
440 posibilidad de modelar el número de movimientos en masa esperados, mejorando la interpretación del  
441 modelo cuando se trata de áreas donde la frecuencia de movimientos en masa varía (Lombardo et al.,  
442 2020). Este último enfoque proporciona una forma más robusta de capturar la naturaleza del fenómeno  
443 y de reflejar la dinámica de los movimientos en masa en áreas más complejas.

444 Un aspecto que requiere ser considerado con más frecuencia en estudios de susceptibilidad a  
445 movimientos en masa es la evaluación rigurosa del desempeño estadístico de los modelos (Goetz et al.,  
446 2015; Steger et al., 2016). En muchos casos, la evaluación del modelo se limita únicamente a su  
447 capacidad predictiva, utilizando herramientas como la curva ROC y el área bajo la curva (AUC). Aunque  
448 estas métricas son útiles, no aportan información sobre la validez estadística. Para asegurar la robustez y  
449 consistencia de los modelos, es crucial evaluar la significancia estadística de las variables predictoras, la  
450 aleatoriedad de los residuos, la ausencia de colinealidad y la homogeneidad de la varianza. La  
451 significancia estadística asegura que las variables predictoras contribuyan de manera relevante al  
452 modelo; la aleatoriedad de los residuos garantiza que el modelo no esté dejando patrones sin explicar; la  
453 ausencia de colinealidad evita problemas de redundancia entre variables predictoras, que podrían sesgar  
454 los resultados; y la homogeneidad de la varianza indica que el modelo se comporta de manera consistente  
455 a lo largo de toda la muestra. Estas validaciones no solo permiten seleccionar el mejor modelo, sino  
456 también asegurar que las relaciones entre las variables predictoras y la variable respuesta sean coherentes  
457 y representen adecuadamente el comportamiento del fenómeno.

## 458 6. Conclusiones

459 En este estudio se ha implementado un modelo de regresión logística multinivel para evaluar la  
460 susceptibilidad a movimientos en masa en los Andes colombianos, utilizando tanto divisiones naturales  
461 en cuencas hidrográficas como un enfoque basado en clusterización espacial. Los resultados muestran  
462 que la consideración de la heterogeneidad espacial y la incorporación de efectos aleatorios mediante  
463 modelos multiniveles proporcionan una representación más precisa de la susceptibilidad a movimientos  
464 en masa, comparada con los modelos tradicionales de regresión logística, los cuales asumen una  
465 homogeneidad espacial considerada inadecuada para este tipo de fenómenos.

466 En el análisis comparativo de los enfoques de regionalización, se observó que la regionalización basada  
467 en cuencas hidrográficas logró un mejor ajuste global según las métricas de Akaike (AIC) y Bayesiana  
468 (BIC), en comparación con el modelo basado en grupos derivados del agrupamiento espacial. Esto  
469 sugiere que las divisiones geográficas naturales, al menos en la zona de estudio, proporcionan límites

470 más consistentes para la dinámica de los movimientos en masa, alineados con los procesos  
471 geomorfológicos e hidrológicos subyacentes.

472 En cuanto a la aplicación de modelos multiniveles, la utilización de efectos aleatorios para capturar la  
473 heterogeneidad espacial es importante para mejorar la interpretación y la capacidad predictiva de los  
474 modelos de susceptibilidad a movimientos en masa. Esto permite no solo entender mejor cómo varía la  
475 susceptibilidad entre diferentes regiones, sino que también ofrece una representación más robusta de los  
476 patrones que influyen en la inestabilidad del terreno. De esta manera, los modelos multiniveles ofrecen  
477 una metodología flexible y detallada que puede adaptarse a la naturaleza jerárquica y compleja de los  
478 fenómenos geológicos en ambientes montañosos.

## 479 Agradecimientos

480 El presente trabajo se realizó con el apoyo de la Fundación Alexander von Humboldt  
481 (<https://www.humboldt-foundation.de/>) a través del programa *Georg Forster Research Fellowship* para  
482 investigadores. Se agradece a D. Bertin, M. Lara y a un revisor anónimo por sus valiosos comentarios,  
483 los cuales contribuyeron significativamente a la mejora de este artículo. Los datos y códigos de este  
484 trabajo se encuentran disponibles en Github (<https://github.com/edieraristizabal/ModeloMultinivel>).

## 485 Referencias

- 486 Anselin, L. 1990. Spatial dependence and spatial structural instability in applied regression analysis. *Journal of*  
487 *Regional Science* 30 (2): 185-207.
- 488 Anselin, L. 2022. Spatial econometrics. In *Handbook of spatial analysis in the social sciences* (Fisher, M.M.; Nijkamp,  
489 P.; editors). Edward Elgar Publishing: 101-122. Cheltenham.
- 490 Anselin, L.; Bera, A.K. 1996. Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*  
491 26 (1): 77-104.
- 492 Anselin, L.; Griffith, D.A. 1988. Do spatial effects really matter in regression analysis? *Papers in Regional Science* 65  
493 (1): 11-34.
- 494 Aristizábal, E.; Korup, O. 2025. Linking landslide patterns to transient landscapes in the northern Colombian Andes.  
495 *Journal of Geophysical Research: Earth Surface* 130 (3): e2024JF008027.
- 496 Aristizábal, E.; Sánchez, O. 2020. Spatial and temporal patterns and the socioeconomic impacts of landslides in the  
497 tropical and mountainous Colombian Andes. *Disasters* 44 (3): 596-618.
- 498 Aristizábal, E.; Velez, J.; Martinez, H.; Jaboyedoff, M. 2016. SHIA\_Landslide: a distributed conceptual and  
499 physically based model to forecast the temporal and spatial occurrence of shallow landslides triggered by rainfall in  
500 tropical and mountainous basins. *Landslides* 13 (3): 497-517.
- 501 Aristizábal, E.; García, E.; Martínez, C. 2015. Susceptibility assessment of shallow landslides triggered by rainfall in  
502 tropical basins and mountainous terrains. *Natural Hazards* 78 (1): 621-634.
- 503 Arnone, E. 2016. Effect of raster resolution and polygon-conversion algorithm on landslide susceptibility mapping.  
504 *Environmental Modelling & Software* 84: 467-481.
- 505 Ayalew, L.; Yamagishi, H. 2005. The application of GIS-based logistic regression for landslide susceptibility mapping  
506 in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* 65 (1-2): 15-31.
- 507 Barredo, J.I. 2000. Comparing heuristic landslide hazard assessment techniques using GIS in the Tirajana basin, Gran  
508 Canaria Island, Spain. *International Journal of Applied Earth Observation and Geoinformation* 2 (1): 9-23.
- 509 Brabb, E.E. 1984. Innovative approaches to landslide hazard and risk mapping. In *Proceedings of the 4th*  
510 *International Symposium on Landslides*, Vol. 1: 307-324. Toronto.
- 511 Breiman, L. 2001. Random Forests. *Machine Learning* 45 (1): 5-32.

- 512 Budimir, M.E.A.; Atkinson, P.M.; Lewis, H.G. 2015. A systematic review of landslide probability mapping using  
513 logistic regression. *Landslides* 12 (3): 419-436.
- 514 Chen, W.; Pourghasemi, H.R.; Zhang, S.; Wang, J. 2019. A comparative study of functional data analysis and  
515 generalized linear model data-mining techniques for landslide spatial modeling. In *Spatial Modeling in GIS and R for*  
516 *Earth and Environmental Sciences* (Pourghasemi, H.R.; Rossi, M.; editors). Elsevier: 467-484.
- 517 Corominas, J. 2014. Recommendations for the quantitative analysis of landslide risk. *Bulletin of Engineering Geology*  
518 *and the Environment* 73 (2): 209-263.
- 519 Cressie, N. 2015. *Statistics for spatial data*. John Wiley & Sons: 928 p. New York.
- 520 Erener, A.; Düzgün, H.S.B. 2010. Improvement of statistical landslide susceptibility mapping by using spatial and  
521 global regression methods in the case of More and Romsdal (Norway). *Landslides* 7 (1): 55-68.
- 522 Erener, A.; Düzgün, H.S.B. 2012. Landslide susceptibility assessment: what are the effects of mapping unit and  
523 mapping method? *Environmental Earth Sciences* 66 (3): 859-877.
- 524 Ermini, L.; Catani, F.; Casagli, N. 2005. Artificial neural networks applied to landslide susceptibility assessment.  
525 *Geomorphology* 66 (1-4): 327-343.
- 526 Fell, R. 2008. Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Engineering*  
527 *Geology* 102 (3-4): 85-98.
- 528 Fotheringham, A.S.; Brunson, C.; Charlton, M.E. 2009. Geographically weighted regression. In *The Sage Handbook*  
529 *of Spatial Analysis* (Fotheringham, A.S.; Rogerson, P.A.; editors). Sage Publications: 243-254. London.
- 530 Fox, J. 2015. *Applied regression analysis and generalized linear models*. Sage Publications: 816 p. Thousand Oaks.
- 531 Funk, C.; Peterson, P.; Landsfeld, M. 2015. The Climate Hazard Group InfraRed Precipitation with Station Data  
532 (CHIRPS). *Bulletin of the American Meteorological Society* 96 (9): 1401-1419.
- 533 Goetz, J.N.; Brenning, A.; Petschko, H.; Leopold, P. 2015. Evaluating machine learning and statistical prediction  
534 techniques for landslide susceptibility modeling. *Computers & Geosciences* 81: 1-11.
- 535 Huang, F. 2020. Comparisons of heuristic, general statistical and machine learning models for landslide  
536 susceptibility prediction and mapping. *Catena* 191: 104580.
- 537 Korup, O.; Stolle, A. 2014. Landslide prediction from machine learning. *Geology Today* 30 (1): 26-33.
- 538 Krieglner, F.J. 1969. Normalized Difference Vegetation Index (NDVI). In *Proceedings of the 6th International*  
539 *Symposium on Remote Sensing of Environment*: 112-116. Ann Arbor.
- 540 Kumar, N.; Mastrangelo, C.; Montgomery, D. 2011. Hierarchical modeling using generalized linear models. *Quality*  
541 *and Reliability Engineering International* 27 (6): 835-842.
- 542 Lee, Y.; Nelder, J.A. 1996. Hierarchical generalized linear models. *Journal of the Royal Statistical Society Series B:*  
543 *Statistical Methodology* 58 (4): 619-656.
- 544 LeSage, J.; Pace, R.K. 2009. *Introduction to spatial econometrics*. Chapman and Hall/CRC: 374 p. Boca Raton.
- 545 Li, P.; Shen, Q.; Guo, Y. 2014. Bare Soil Index for Land Cover Analysis. *Journal of Applied Remote Sensing* 8 (1):  
546 083605.
- 547 Logan, T.L.; Karr, C.L. 2014. Advanced Land Observing Satellite (ALOS) - Phased Array-Type L-Band Synthetic  
548 Aperture Radar (PALSAR). *Journal of Geospatial Research* 7: 102-119.
- 549 Lombardo, L.; Mai, P.M. 2018. Presenting logistic regression-based landslide susceptibility results. *Engineering*

550 Geology 244: 14-24.

551 Lombardo, L.; Opitz, T.; Ardizzone, F. 2020. Space-time landslide predictive modelling. *Earth-Science Reviews* 209:  
552 103318.

553 Lombardo, L.; Opitz, T.; Huser, R. 2018. Point process-based modeling of multiple debris flow landslides using INLA:  
554 an application to the 2009 Messina disaster. *Stochastic Environmental Research and Risk Assessment* 32 (7): 2179-  
555 2198.

556 Marin, R.; Garcia, E.; Aristizabal, E. 2021. Assessing the effectiveness of TRIGRS for predicting unstable areas in a  
557 tropical mountain basin (Colombian Andes). *Geotechnical and Geological Engineering* 39 (3): 2329-2346.

558 Montgomery, D.R.; Dietrich, W.E. 1994. A physically based model for the topographic control on shallow landsliding.  
559 *Water Resources Research* 30 (4): 1153-1171.

560 Palacio, J.; Mergili, M.; Aristizabal, E. 2020. Probabilistic landslide susceptibility analysis in tropical mountainous  
561 terrain using the physically based r.slope.stability model. *Natural Hazards and Earth System Sciences* 20 (3): 815-  
562 829.

563 Patanè, G.; Bortolotti, T.; Yordanov, V.; Biagi, L.G.A.; Brovelli, M.A.; Truong, X.Q.; Vantini, S. 2025. An  
564 interpretable and transferable model for shallow landslides detachment combining spatial Poisson point processes  
565 and generalized additive models. *Stochastic Environmental Research and Risk Assessment* 39 (4): 1723-1740.

566 Pedregosa, F. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12: 2825-2830.

567 Petschko, H. 2012. Landslide susceptibility modeling with generalized additive models-facing the heterogeneity of  
568 large regions. In *Landslides and Engineered Slopes, Protecting Society through Improved Understanding* (Eberhardt,  
569 E.; editor). CRC Press: 769-777. Boca Raton.

570 Reichenbach, P. 2018. A review of statistically-based landslide susceptibility models. *Earth-Science Reviews* 180:  
571 60-91.

572 Rousseeuw, P.J. 1987. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal*  
573 *of Computational and Applied Mathematics* 20: 53-65.

574 Sannino, G. et al. 2024. Deterministic Physically Based Distributed Models for Rainfall-Induced Shallow Landslides.  
575 *Geosciences* 14 (10): 255.

576 SGC Servicio Geológico Colombiano. 2013. Documento metodológico de la zonificación de susceptibilidad y amenaza  
577 por movimientos en masa escala 1: 100.000. Servicio Geológico Colombiano: 120 p. Bogotá DC.

578 Soeters, R.; Van Westen, C.J. 1996. Slope instability recognition, analysis and zonation. In *Landslides: Investigation*  
579 *and Mitigation* (Turner, A.K.; Schuster, R.L.; editors). National Academy Press: 129-177. Washington DC.

580 Steger, S.; Brenning, A.; Bell, R.; Petschko, H.; Glade, T. 2016. Exploring discrepancies between quantitative  
581 validation results and the geomorphic plausibility of statistical landslide susceptibility maps. *Geomorphology* 262: 8-  
582 23.

583 Tehrani, F.S.; Calvello, M.; Liu, Z.; Zhang, L.; Lacasse, S. 2022. Machine learning and landslide studies: recent  
584 advances and applications. *Natural Hazards* 114 (2): 1197-1245.

585 Thorndike, R.L. 1953. Who Belongs in the Family? *Psychometrika* 18 (4): 267-276.

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

588 Wong, G.Y.; Mason, W.M. 1985. The hierarchical logistic regression model for multilevel analysis. *Journal of the*  
589 *American Statistical Association* 80 (391): 513-524.

590 Xu, H. 2006. Modification of Normalized Difference Water Index (MNDWI) to Enhance Open Water Features in  
591 Remotely Sensed Imagery. *International Journal of Remote Sensing* 27 (14): 3025-3033.

592 Youssef, K. et al. 2023. Landslide susceptibility modeling by interpretable neural network. *Communications Earth &*  
593 *Environment* 4 (1): 162.

594 Zha, Y.; Gao, J.; Ni, S. 2003. Use of Normalized Difference Built-up Index in Automated Urban Area Mapping.  
595 *Remote Sensing of Environment* 87 (4): 467-480.

596

IN PRESS