

Análisis de modelos de regresión logística para la identificación de susceptibilidad a deslizamientos en la cuenca hidrográfica del río Guayuriba, Colombia

Analysis of Logistic Regression Models for the Identification of Landslide Susceptibility. Guayuriba River Basin, Colombia

Edgar Libardo Barahona Cantor¹ 

Resumen

Los deslizamientos son los fenómenos naturales que más pérdidas económicas y de vidas generan en Colombia, afectan la conectividad y economía entre las grandes urbes generando miles de millones en pérdidas, como es el caso del territorio que forma la cuenca hidrográfica del río Guayuriba, afectado constantemente por deslizamientos. Se planteó mapear la susceptibilidad a deslizamientos en esta zona mediante el uso de dos modelos de regresión logística (ModeloDummy y ModeloCon), para probar la efectividad del método bajo diferentes formas de entrenamiento. El mapa de deslizamientos, el preprocesamiento de variables y el modelado de susceptibilidad se realizó usando las plataformas GIS, mientras que el análisis estadístico, entrenamiento y validación de los modelos se realizó en Python. La validación usando curvas ROC presentó una precisión para las predicciones del ModeloCon del 93,2% y 91,7% para el ModeloDummy, mientras que la precisión general (*accuracy*) por medio de matrices de confusión mostro valores de 85,77% para el ModeloCon y 85,03% para el ModeloDummy respectivamente. El estudio demostró la efectividad de la regresión logística para el mapeo de la susceptibilidad a deslizamientos y permitió validar la superioridad de los modelos elaborados a partir de variables continuas frente a modelos elaborados usando variables dummy.

Palabras clave:

amenaza natural, análisis de riesgo, deslizamiento de tierra, regresión logística, sistemas de información geográfica.

Cómo citar

Barahona Cantor, E. L. (2024). Análisis de modelos de regresión logística para la identificación de susceptibilidad a deslizamientos en la cuenca hidrográfica del río Guayuriba, Colombia. *Análisis Geográficos*, 57(1), 8 - 27.

Abstract

Landslides are the natural phenomena that generate more economic losses and loss of life in Colombia, affecting the connectivity and economy between large cities, generating billions in losses, as is the case of the territory that forms the watershed of the Guayuriba river, constantly affected by landslides. In this work, it was proposed to map the landslide susceptibility in the area by using two logistic regression models (Dummy Model and ModelCon), to test the effectiveness of the method under different training methodologies. Landslide mapping, variable preprocessing and susceptibility modeling were performed using GIS platforms, while statistical analysis, training and validation of the models were performed in Python. The validation using ROC curves presented an accuracy for the ModelCon predictions of 93.2% and 91.7% for the DummyModel, while the overall accuracy by means of confusion matrices showed values of 85.77% for the ModelCon and 85.03% for the DummyModel respectively. The study demonstrated the effectiveness of logistic regression for mapping landslide susceptibility and validated the superiority of the models elaborated from continuous variables versus models elaborated using dummy variables.

Keyword:

natural hazard, risk analysis, landslides, logistic regression, geographic information systems.

Introducción

Los deslizamientos son uno de los fenómenos naturales más destructivos, pues causan impactos ambientales, daños a la infraestructura y cuestan vidas humanas. Para el territorio colombiano este fenómeno es uno de los que más afectaciones generan, después de las inundaciones según la Comisión Económica para América Latina y el Caribe (Cepal, 2012).

Un movimiento en masa o proceso de talud es el movimiento hacia fuera y hacia abajo de un talud bajo la influencia de la gravedad, que resulta del fallo de una sección del talud a lo largo de un plano de fractura. Estos pueden ser afectados por diversos estímulos naturales externos como lluvias intensas y/o prolongadas, eventos sísmicos, erupciones volcánicas y deshielo. Estos procesos naturales tienen diferentes nombres como “procesos de ladera”, “procesos de movimiento de masas”, o también como “deslizamientos” (Alcántara, 2000).

La susceptibilidad a los deslizamientos se refiere la propensión de ocurrencia en una zona determinada, en función de las características del terreno, sin tener en cuenta el periodo de retorno, ni la probabilidad de que se produzcan fenómenos de inestabilidad (Soeters y Van Westen, 1996). En general, se trata de una propiedad del terreno que indica lo favorable o desfavorable que son sus condiciones para que se produzcan desprendimientos (Suárez, 2009).

Los mapas de susceptibilidad a los deslizamientos describen las zonas potencialmente propensas a estos mediante el análisis de algunos de los principales factores que contribuyen a su ocurrencia (Basu y Pal, 2017). Estos mapas intentan identificar dónde pueden producirse deslizamientos de tierra en una región determinada basándose en un conjunto de características medioambientales relevantes (Guzzetti et al., 1999).

En el estudio de la susceptibilidad a los deslizamientos, los diferentes enfoques se caracterizan dentro de dos directrices metodológicas principales, denominadas cualitativa y cuantitativa (Aleotti y Chowdhury, 1999; Guzzetti et al., 1999). Los enfoques cualitativos se basan en el juicio del conocimiento experto (Yalcin, 2008; Huang et al., 2015) y el enfoque de la persona o personas que

realizan la evaluación de la susceptibilidad (Aleotti y Chowdhury, 1999), tales métodos son subjetivos y determinan la susceptibilidad de forma heurística, con la definición del nivel de susceptibilidad mediante el uso de términos descriptivos (Reichenbach et al., 2018).

Por otro lado, los enfoques cuantitativos se caracterizan en dos tipos diferentes de métodos: deterministas y probabilísticos (Yalcin, 2008), los cuales se diferencian en la forma en que manejan la incertidumbre, pues los métodos deterministas, a pesar de ser predictivos y precisos, no tienen en cuenta la incertidumbre, mientras que los métodos probabilísticos la incluyen mediante el uso de modelos estadísticos. Estos enfoques se basan en expresiones matemáticas que describen la relación entre los factores condicionantes y los deslizamientos (Yalcin, 2008; Mandal y Mandal, 2018). Generalmente, producen estimaciones numéricas de las probabilidades de ocurrencia de fenómenos de deslizamiento en cualquier zona de riesgo (Guzzetti et al., 1999).

Usualmente, el enfoque y las variables seleccionadas para la identificación de la susceptibilidad del territorio están directamente relacionados con la escala de trabajo (Soeters y Van Westen, 1996), y la fiabilidad de los mapas de susceptibilidad a los deslizamientos depende de la cantidad y calidad de los datos disponibles, la escala del trabajo, la selección de la metodología de análisis y modelización adecuada (Yalcin, 2008).

La regresión logística es una herramienta de amplia utilidad para la evaluación y el mapeo de la susceptibilidad del territorio ante la ocurrencia de deslizamientos, y es uno de los métodos cuantitativos más populares y efectivos (Yilmaz, 2009; Huang et al., 2015).

En este estudio se implementó el método de regresión logística (LR) para mapear la susceptibilidad a deslizamientos dentro de la cuenca hidrográfica del río Guayuriba, territorio que conecta las ciudades de Bogotá y Villavicencio mediante la vía nacional 40 y se ha visto históricamente afectado por la ocurrencia de deslizamientos que restringen la conectividad entre ambas urbes y generan pérdidas económicas millonarias.

Con la implementación del método se buscó generar modelos de susceptibilidad a escala regional, que puedan ser implementados como insumo base en planes de ordenamiento territorial

y gestión del riesgo, dada su capacidad de proporcionar información crucial sobre las zonas propensas a la ocurrencia de deslizamientos y que pueden generar afectaciones económicas y humanas. La aplicación del método se apoyó en el uso de sistemas de información geográfica (SIG), que permiten la extracción de parámetros morfológicos e hidrológicos que son difíciles de obtener (Yalcin, 2008) y que son de vital importancia para el modelado de susceptibilidad a deslizamientos.

Área de estudio

El área de estudio, la cuenca hidrográfica del río Guayuriba, que fue objeto de análisis se muestra en la Figura 1. Se ubica sobre la cordillera oriental; región andina de Colombia, entre los 342 y 4.415 m.s.n.m., tiene un área territorial de 3.198,97 km² de los cuales 2.483,04 km² correspondientes al 77,6% del área total se encuentran en jurisdicción del departamento de Cundinamarca y los restantes 715,93 km² (22,4%) se encuentran en jurisdicción del departamento del Meta. La precipitación media anual en la cuenca del río Guayuriba presenta valores que varían entre 720 a 5.200 mm/año, presentando las menores acumulaciones en su parte alta, hacia el páramo de Sumapaz y zona nororiental de la cuenca y mayores acumulaciones hacia la zona del piedemonte llanero (Fondo de Adaptación, 2018).

La cuenca del río Guayuriba está conformada por rocas cuyo origen geológico pertenecen a dos estadios de tiempo. El primero se da desde el periodo Ordovícico (485,4 Ma atrás) hasta el periodo Carbonífero (298,9 Ma atrás), durante el Paleozoico. Se caracteriza por la conformación predominante de rocas metamórficas relacionadas con una cuenca de *rift* asociada a la separación entre Norteamérica y Sudamérica (Cooper et al., 1995). En la cuenca estas formaciones se ubican en una franja que recorre de sur a oriente, abarcando parte de los municipios de Guayabetal, Gutiérrez, Quetame y Acacias.

El segundo estadio, data del Cretácico inferior (~145 Ma) hasta el periodo Cuaternario (0,0042 Ma); entre el Mesozoico y Cenozoico, se caracteriza por la formación predominante de rocas sedimentarias producto de procesos regresión y transgresión marina. Estas formaciones predominan en la cuenca y se ubican en dos franjas sobre esta, la primera franja cubre el occidente y norte de la cuenca en los municipios de Uña, Ubaque, Chiriquá, Choachí, Fomeque, Cáqueza, Fosca, Fomeque, La Calera y Guasca. La segunda franja se ubica al suroccidente de la cuenca en los municipios de Acacias y Villavicencio.

Materiales y métodos

Plan de trabajo

El plan de trabajo consta de cuatro etapas principales: la primera etapa fue la adquisición de datos, después se llevó a cabo una etapa de procesamiento y producción de factores (variables), las cuales se generaron en formato ráster con una resolución espacial de tres metros. Posteriormente, usando codificación de Python 3.10, se procesaron las variables de insumo para el entrenamiento y validación de los modelos de regresión logística, así como los análisis de multicolinealidad y la validación del modelo mediante curvas de receptor operativo (ROC), por último, se generó el mapa final de susceptibilidad nuevamente en la plataforma ArcGISPro-3.1 (Figura 2).

Adquisición y preprocesamiento de datos

Se adquirieron imágenes de satélite de alta resolución PlanetScope, así como también información de fuentes oficiales del Instituto geográfico Agustín Codazzi (IGAC), el Servicio Geológico Colombiano (SGC), el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) y el Alaska Satellite Facility (ASF). Los datos adquiridos, sus características y uso en la investigación se muestran en la Tabla 1.

Figura 2 - Plan de trabajo

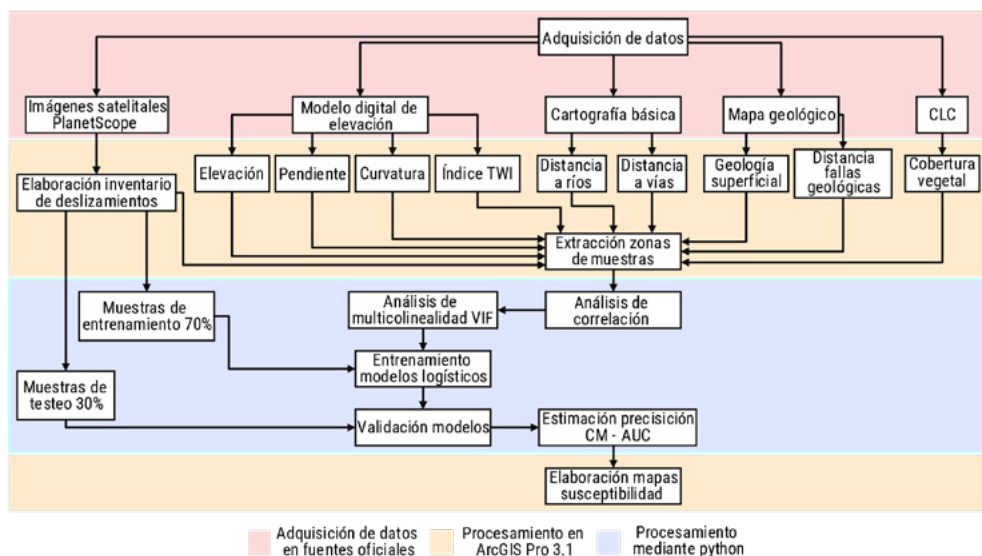
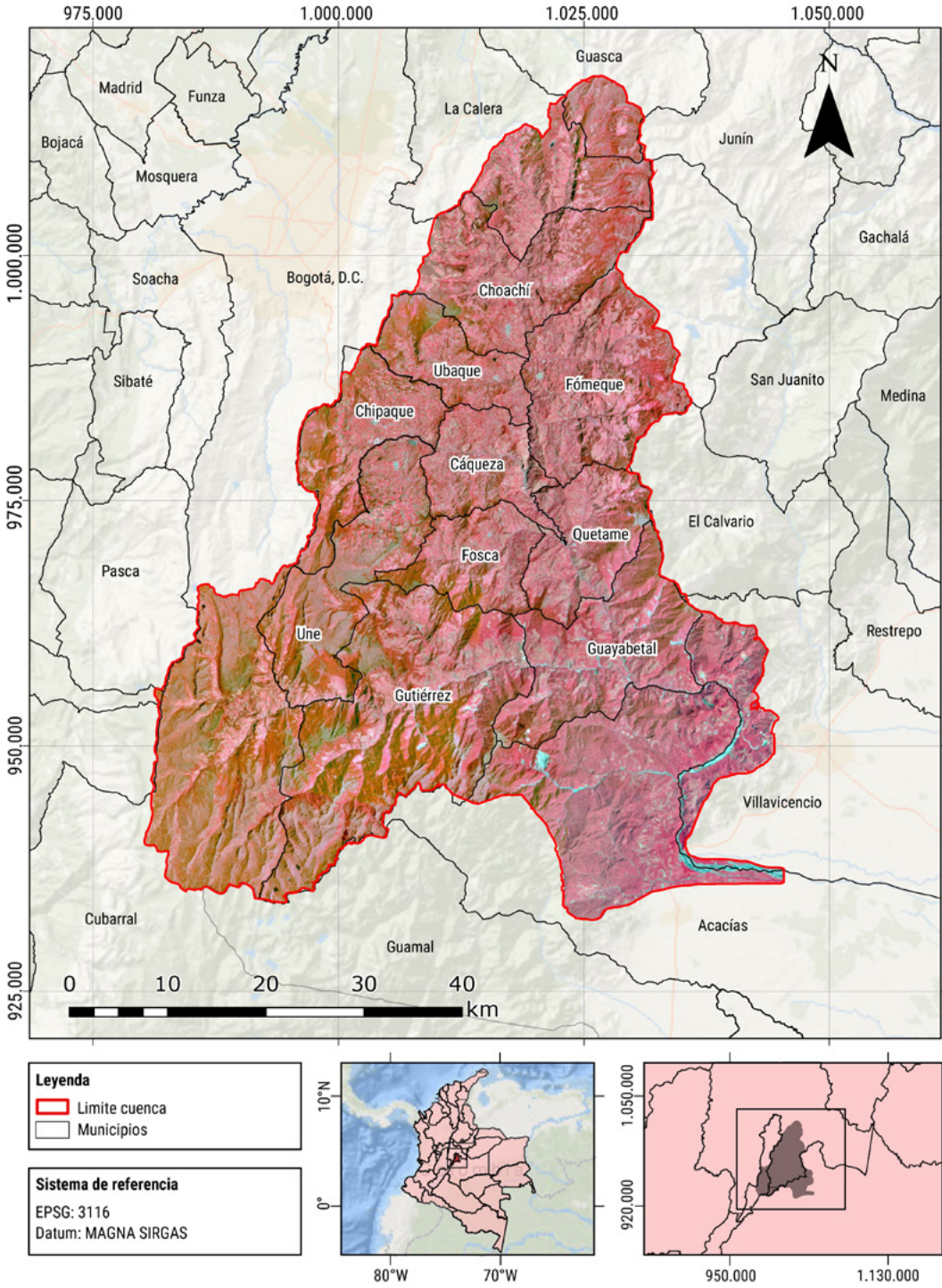


Figura 1 - Ubicación de la cuenca hidrográfica del río Guayuriba



Fuente: elaboración propia a partir de imágenes PlanetScope, 2023.

Tabla 1 - Características de los insumos usados para el modelado de susceptibilidad

Tipo de datos	Plataforma	Resolución espacial o escala	Uso	Fuente
Imágenes satelitales	PlanetScope	3 metros	Inventario deslizamiento	PlanetScope
DEM	STRM resampleado	125 metros	Pendiente/ TWI/curvatura/ elevación	Alaska Satellite Facility
Cartografía básica	-	1:25.000	Distancia a ríos/distancia a vías	IGAC
Mapa geológico 2020	-	1:100.000	Distancia a fallas geológicas/ geología superficial.	SGC
CLC 2018 Colombia	-	1:100.000	Cobertura vegetal	IDEAM

Preparación del inventario de deslizamientos

El mapa de inventario de deslizamientos es esencial y es un punto de partida clave para estudiar la relación entre los deslizamientos y los factores condicionantes (Huang et al., 2015). En este estudio, el inventario de deslizamientos fue elaborado a partir de la identificación visual de zonas deslizadas usando como apoyo un mosaico de imágenes PlanetScope, que se elaboró usando 16 imágenes capturadas entre el 26 de enero y el 3 de febrero de 2023, el cual cuenta con 0% de nubosidad y una resolución espacial de tres metros.

Se identificaron en total 1964 zonas deslizadas dentro de la cuenca hidrográfica del río Guayuriba, la mayoría ubicadas en la parte sur oriental de la cuenca, especialmente en los municipios de Gutiérrez, Guayabetal y Acacias. Para complementar el inventario, se capturan 2346 zonas de áreas no deslizadas; estas fueron capturadas usando un rectángulo con un área igual a la media de las zonas de deslizamiento y a una distancia no menor de 200 metros de cualquier zona deslizada con la finalidad de no capturar zonas de no deslizamiento en áreas afectadas por deslizamientos cercanos. La Tabla 2 muestra las estadísticas del inventario de deslizamientos, mientras que la Figura 3 muestra en ejemplo del proceso de captura de deslizamientos.

Tabla 2 - Estadísticas de las zonas capturadas de deslizamiento y no deslizamiento identificadas

	Zonas deslizadas	Zonas no deslizadas
Numero de muestras	1964	2346
Suma	968.487 ha	1.148.430 ha
Media	0,490 ha	0,489 ha
Máximo	187.226 ha	0,489 ha
Mínimo	0,0085 ha	0,489 ha

Determinación de factores condicionantes

Factores geomorfológicos

Estos generalmente se derivan de modelos de elevación digital (MDE) y describen las características geomorfológicas del terreno y su influencia en la ocurrencia de deslizamientos. Un factor comúnmente usado para el mapeo de susceptibilidad a deslizamientos es la elevación (Conforti et al., 2014; Huang et al., 2015; Zhao et al., 2019), pues esta controla la variación espacial de las condiciones hidrológicas y la estabilidad de las laderas (Yilmaz, 2009). También se implementó la pendiente del terreno (Figura 4), que es uno de los condicionantes más importantes en la ocurrencia de deslizamientos (Chauhan et al., 2010; Reichenbach et al., 2018), esta presenta una correlación positiva con la probabilidad de ocurrencia de estos (Chen y Li, 2020), dado a que a medida que aumenta la pendiente también aumenta el nivel de esfuerzo cortante inducido por la gravedad sobre los suelos (Dai et al., 2001); es decir, la inestabilidad del suelo y las rocas tiende a aumentar con el incremento de la pendiente (Guillard y Zezere, 2012).

La curvatura del terreno se tuvo en cuenta (Figura 5), dado que realiza una contribución importante en la inestabilidad de la pendiente del terreno (Huang et al., 2015). Este parámetro representa la variación espacial del gradiente de la pendiente del terreno (Trigila et al., 2015), y está relacionado con el flujo de escurrimiento superficial y subsuperficial que converge cuando la pendiente es cóncava y diverge cuando es convexa (Guillard y Zezere, 2012); es decir, que controla la convergencia o divergencia de los escombros y el agua del deslizamiento en la dirección de su movimiento (Huang et al., 2015). Por último, se tuvo en cuenta el índice de humedad topográfica (TWI), este indica el efecto de la topografía sobre la ubicación y el tamaño de las zonas saturadas por efecto de la escorrentía (Conforti et al., 2014; Trigila et al., 2015); vale decir, mide el control topográfico sobre el flujo y la acumulación de aguas subterráneas (Zhou et al., 2020). En general, los

Figura 3 - Imágenes municipio Guayabetal, a) imagen zonas de deslizamientos, b) polígonos de deslizamientos capturados, c) vista 3D de las zonas deslizadas en ArcGIS Pro



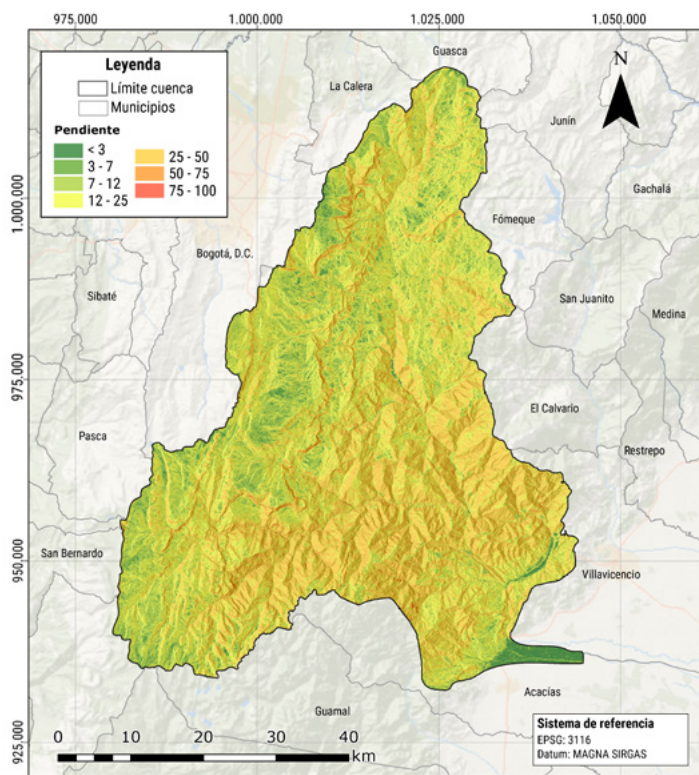


Figura 4 - Factor pendiente del terreno

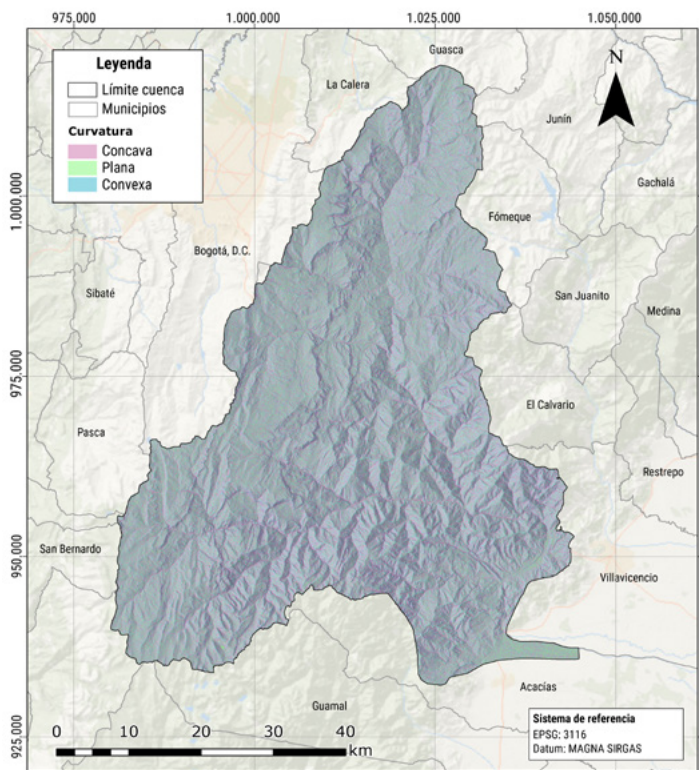


Figura 5 - Factor curvatura

terrenos con valores de TWI más altos tienen más probabilidades de saturarse durante eventos de precipitación (Zhou et al., 2020; Goyez-Peñañiel y Hernández-Rojas, 2021), lo que indica que son más susceptibles a la ocurrencia de deslizamientos.

La elevación (Figura 6), la pendiente, la curvatura y el índice TWI (Figura 7) fueron calculados partir de un modelo de elevación STRM resampleado a una resolución espacial de 12,5 metros. La elevación se clasificó en cinco rangos usando el método de cortes naturales, el ángulo de la pendiente, se caracterizó según la clasificación de pendientes del IGAC en grupos como sigue <

3°, 3-7°, 7-12°, 12-25°, 25-50°, 50-75°, 75-100°, la curvatura se definió en tres clases distintas; cóncava (< 0), plana (= 0) y convexa (> 0). Por otra parte, el índice de humedad topográfica se clasificó en cinco rangos como sigue 0,18-4,33; 4,33-5,80; 5,80-7,55; 7,55-10,04; 10,04-23,66. Los cálculos anteriores se realizaron usando los geoprocenos disponibles en la plataforma ARCGIS Pro-3.1.

Factores obtenidos a partir del análisis de distancias

Este tipo de factores buscan medir la influencia de un elemento natural o antrópico en la ocurrencia de un deslizamiento en función de su distancia. En este estudio se tuvieron en cuenta tres factores:

Figura 6 - Factor elevación

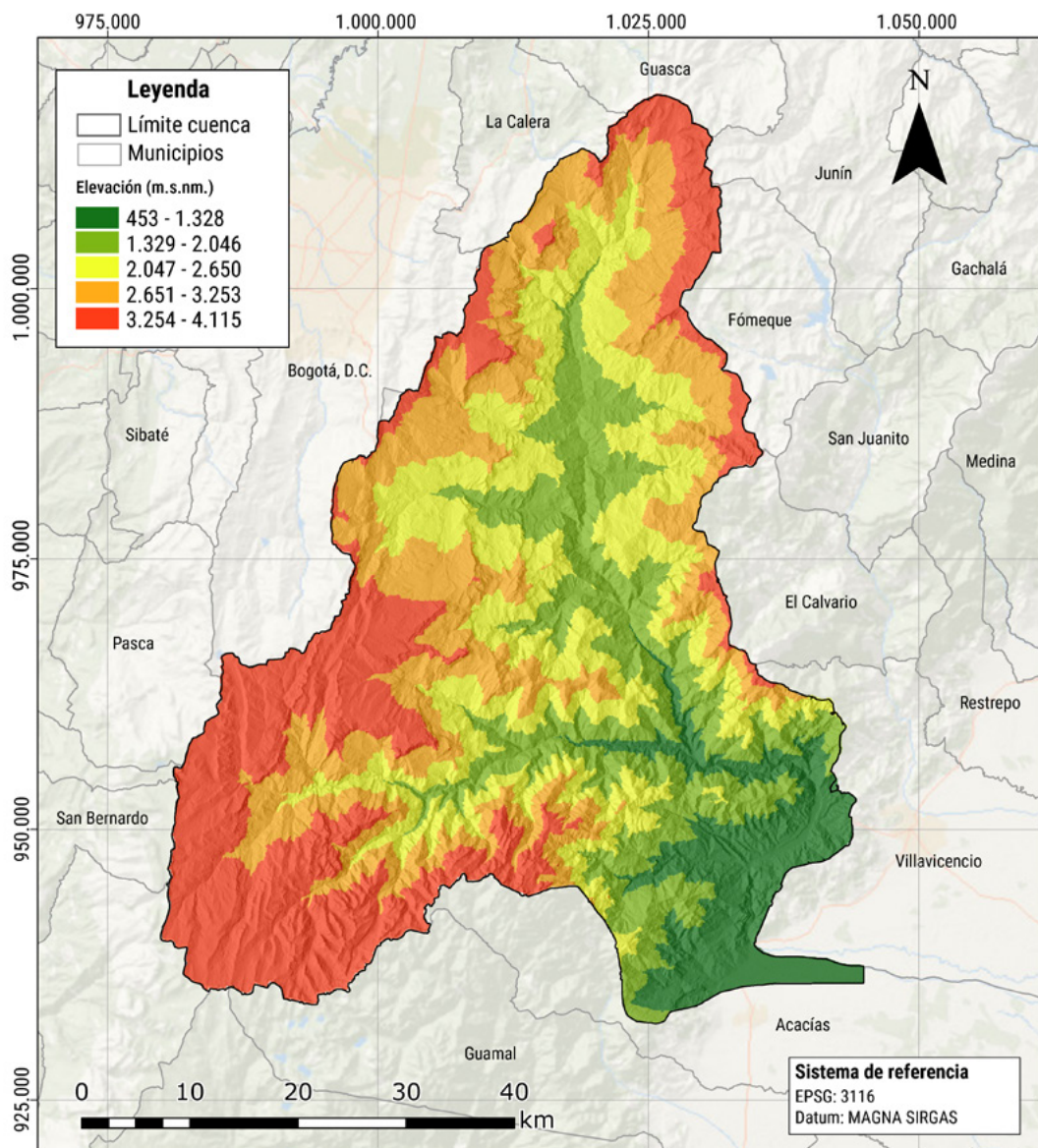
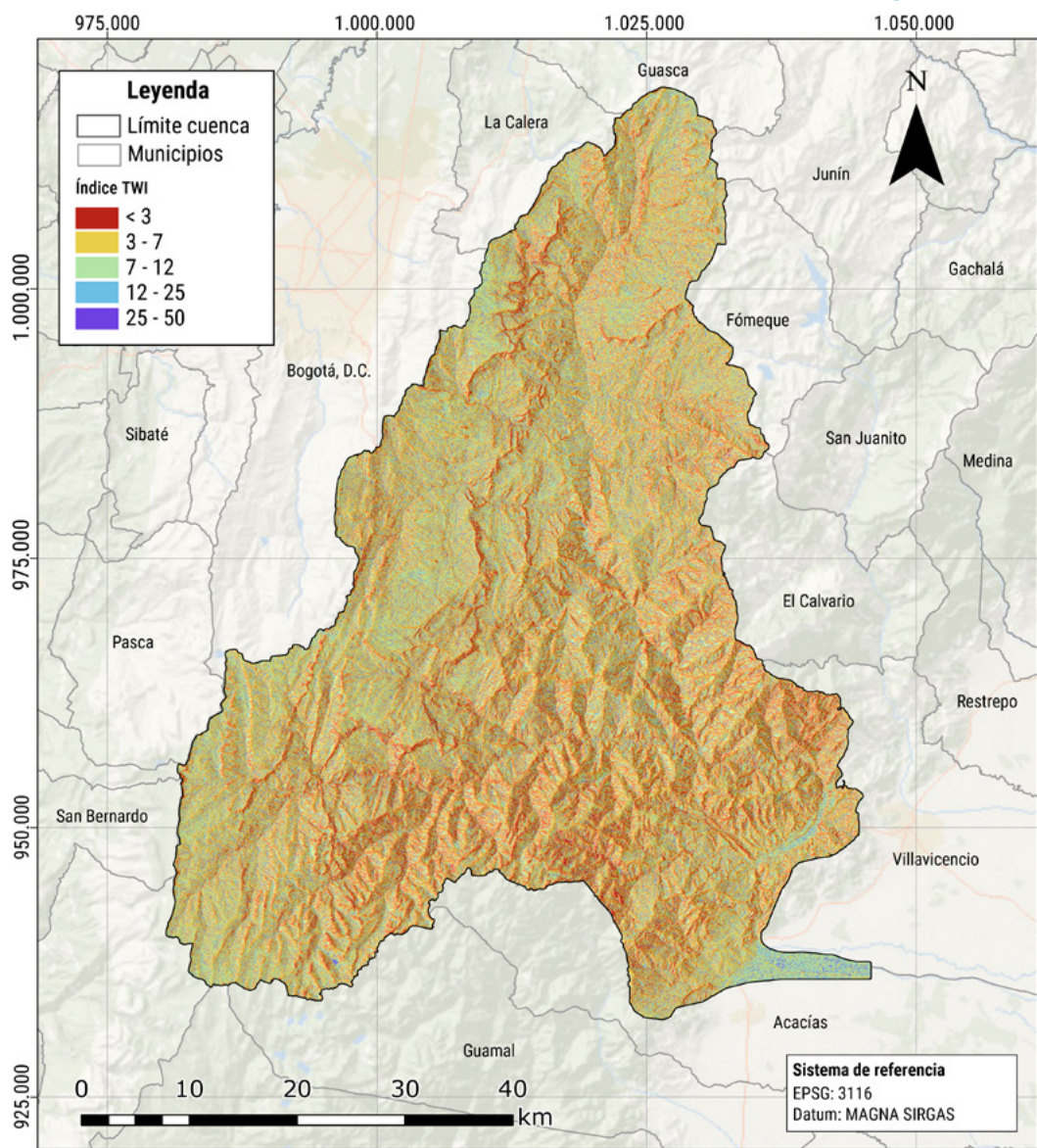


Figura 7 - Factor índice TWI

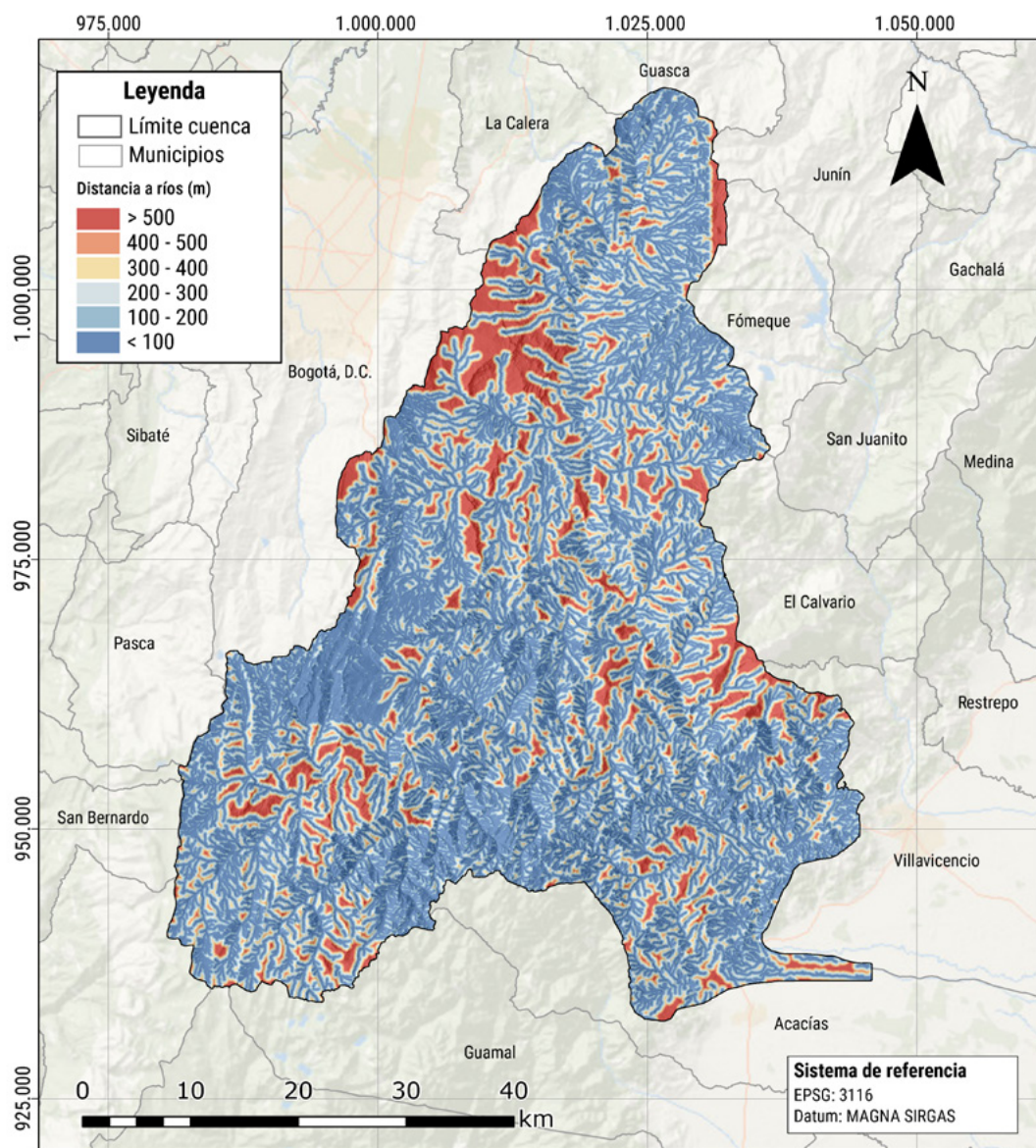


El primero es la distancia a los ríos, se sabe que deslizamientos en zonas montañosas también pueden producirse debido a la actividad erosiva asociada a los drenajes (Chauhan et al., 2010), ello debido a que la inestabilidad de las laderas se ve afectada por la erosión de los ríos que satura y socaba la base de estas (Dai et al., 2001). La distancia a los drenajes capta las condiciones hidrogeológicas relevantes menos favorables para la estabilidad de las laderas debido a la concentración del flujo de agua subterránea, y al efecto desestabilizador de la incisión del río (Reichenbach et al., 2018), por lo que se considera un factor de control en los

deslizamientos causados por la lluvia (Huang et al., 2015) (Figura 8).

El segundo es la distancia a fallas geológicas, estas producen una zona de debilidad de varios metros a cada lado y en el caso de las fallas de gran magnitud, de varios cientos de metros (Suárez, 2009), estas discontinuidades estructurales forman líneas de debilidad en las rocas que tienen probabilidad de ser áreas con alta permeabilidad, erosión profunda y una incidencia particularmente alta de pendientes inestables (Conforti et al., 2014) (Figura 9).

El tercero es la distancia a las vías, estas suelen ser sitios de inestabilidad inducida antropológicamente que, dependiendo

Figura 8 - Factor de distancia a los ríos/drenajes

de su ubicación en las montañas, generalmente sirve como fuente de deslizamientos (Ayalew y Yamagishi, 2005), estas cambian la topografía provocando un aumento de tensión en la base de la pendiente y la generación de grietas sobre esta (Zhou et al., 2020). A menudo, las vías en zonas de pendiente pronunciada como los sistemas montañosos se relacionan con la ocurrencia de deslizamientos y erosión (Mauri et al., 2022), y en general la distancia a estas suele tener gran influencia en la ocurrencia de deslizamientos (Chen y Li, 2020) (Figura 10). La distancia a los ríos, fallas geológicas y vías fue calculada por medio de factores de amortiguamiento o zonas de influencia y

organizada en cinco clases. Para la distancia a los ríos, zonas que varían en intervalos de cien metros entre 0-500 metros, la distancia a las fallas geológicas se definió en intervalos de 200 metros entre cero a mil metros, por último, la distancia a las fallas geológicas se catalogó en intervalos de 300 metros entre 0-1500 metros.

Factor de cobertura de la tierra

Este factor representa un papel importante en la ocurrencia de deslizamientos ya que la vegetación condiciona la resistencia natural que presentan los suelos ante los deslizamientos, esta aporta efectos mecánicos e hidrológicos que proporcionan esta

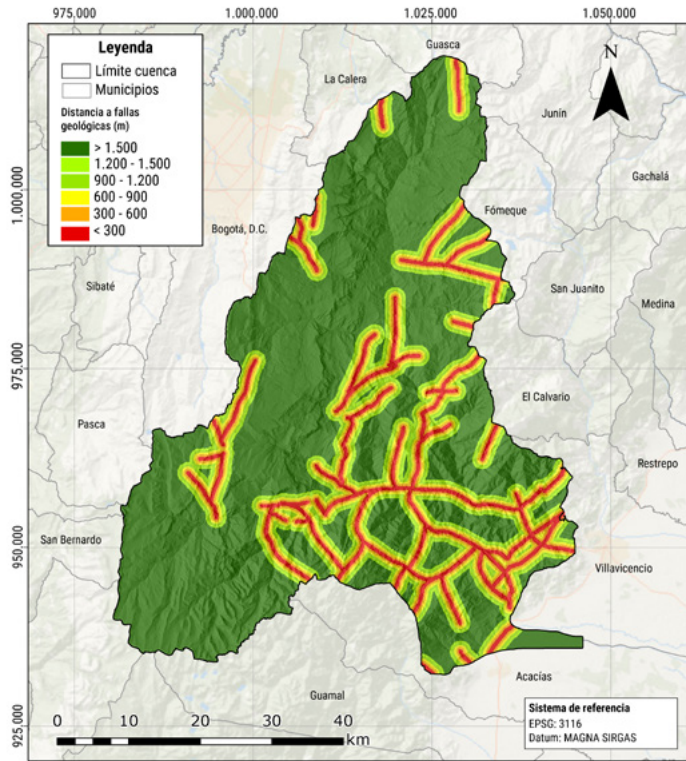


Figura 9 - Factor distancia a fallas geológicas

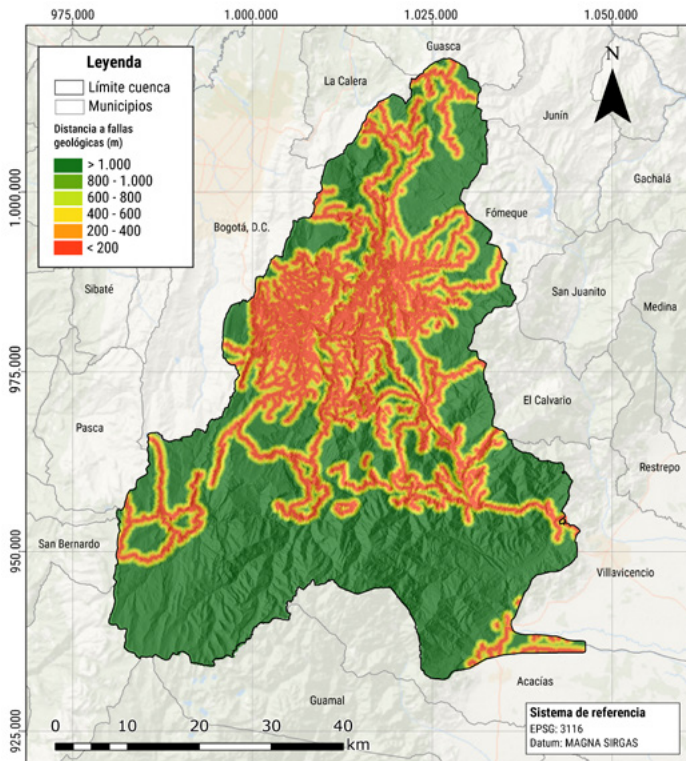


Figura 10 - Factor distancia a las vías

bilidad estructural en suelos con alto grado de pendiente (Huang et al., 2015), cuando es de raíces profundas proporciona resistencia cohesiva a las capas del suelo, facilitando el drenaje subterráneo y reduciendo la posibilidad de deslizamientos superficiales (Suárez, 2001) y aumentando la resistencia del suelo al cizallamiento (Trigila et al., 2015). Por el contrario, las zonas desérticas o en barbecho tienden a desestabilizar el terreno (Chauhan et al., 2010). En este caso, la cobertura del vegetal se representó por medio del mapa de coberturas Corine Land Cover Nivel 2 para Colombia elaborado en el año 2018 por del IDEAM. (Figura 11).

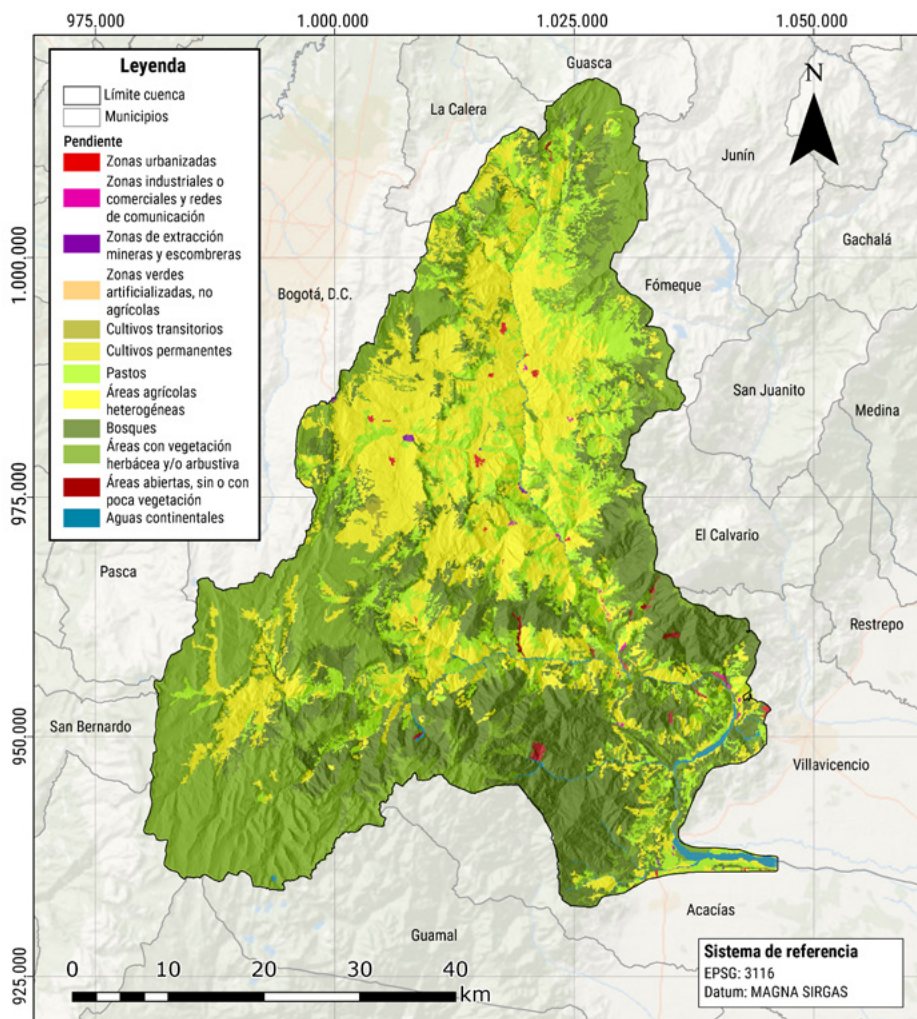
Factor de litología

Este factor mide el grado de influencia de los materiales, sedimentos y rocas en la generación de deslizamientos (Quesada y Feoli-Boraschi, 2018). La litología juega un papel fundamental en la distribución de los deslizamientos y se correlaciona con

las propiedades del suelo como la resistencia y la estructura de las rocas (Bai et al., 2014), esta influye en gran medida en la ocurrencia de deslizamientos de tierra, porque las variaciones litológicas y estructurales a menudo conducen a una diferencia en la resistencia y la permeabilidad de las rocas y los suelos (Ayalew y Yamagishi, 2005; Chauhan et al., 2010), por lo tanto, es factible hacer inferencias sobre la ocurrencia de deslizamientos mediante el análisis de las características litológica, dado que la naturaleza y la velocidad de los procesos de deslizamiento dependen tanto de la litología como de las características de los materiales subyacentes (Dai et al., 2001). El factor de litología fue elaborado usando las unidades cronoestratigráficas presentes del mapa geológico de Colombia 2018, que describe la edad y el componente litológico de las rocas en la zona de estudios (Figura 12).

Los diferentes factores elaborados fueron convertidos a formato ráster con una resolución de 3×3 metros, y sobre

Figura 11 - Factor de cobertura de la tierra



Fuente: elaboración propia a partir de datos del IDEAM, 2018.

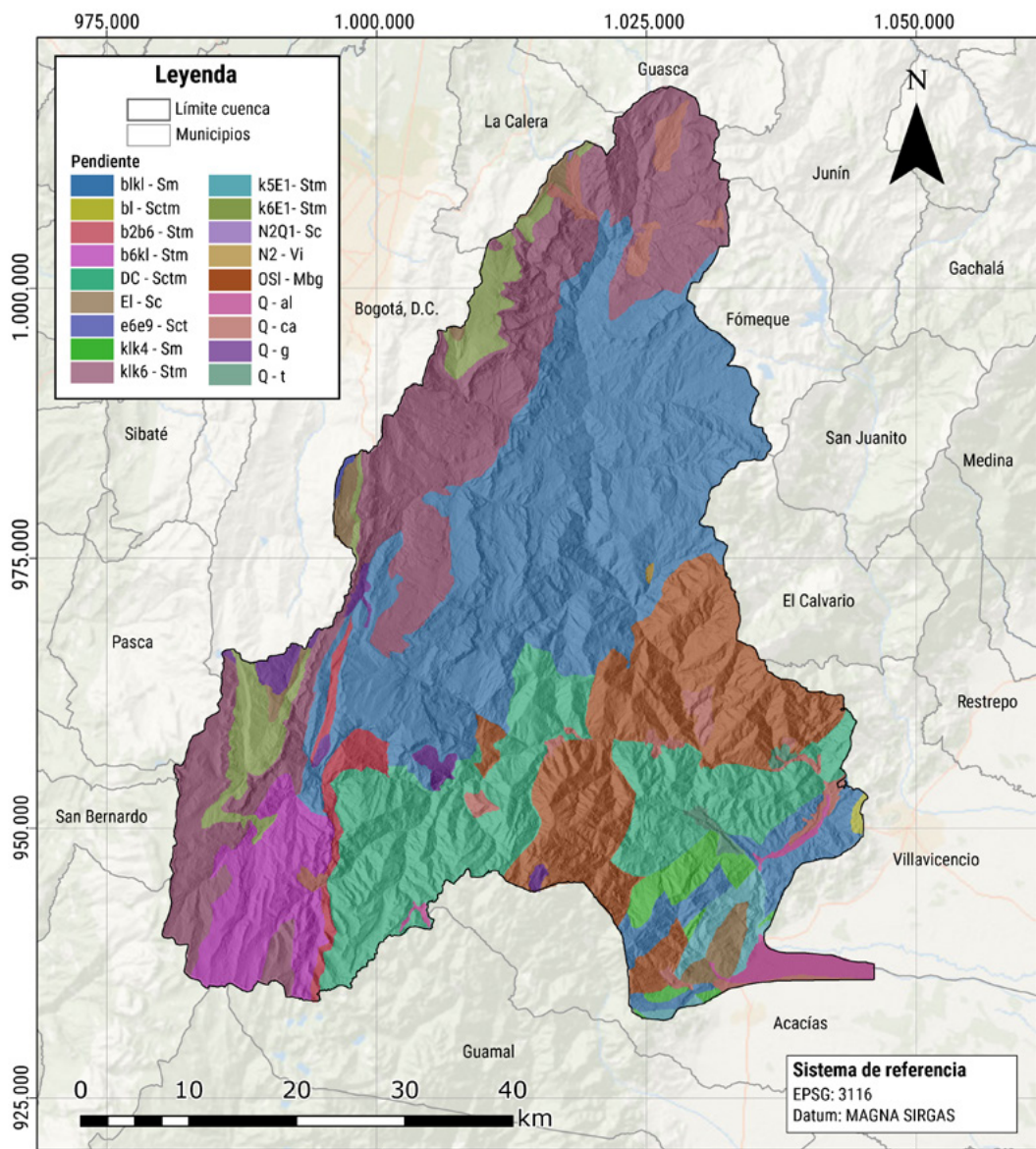
estos se extrajeron los píxeles al interior de las zonas del inventario de deslizamientos, generando una nueva capa ráster que únicamente contiene información en las zonas extraídas y que será implementada como variables de entrenamiento.

Modelado de susceptibilidad mediante regresión logística

La regresión logística es el método estadístico cuantitativo más

ampliamente utilizado para el análisis de susceptibilidad a deslizamientos (Huang et al., 2015; Goyez-Peñafiel y Hernández-Rojas, 2021), su objetivo es encontrar el modelo adecuado para describir la asociación entre la presencia de deslizamientos (variable dependiente) y un conjunto de parámetros independientes (Ayalew y Yamagishi, 2005), mediante la creación de una relación de regresión multivariante entre estas (Zhao et al., 2019). La regresión logística

Figura 12 - Factor de litología



modela la ocurrencia de deslizamientos mediante una variable dicotómica en la que los valores 0 y 1 representan ocurrencia y no ocurrencia de deslizamientos respectivamente. La probabilidad de ocurrencia de un deslizamiento puede definirse por la ecuación 1:

$$P(y=1) = \frac{1}{1+e^{-Z}} \quad 1$$

Donde P es la probabilidad de ocurrencia de un deslizamiento, valor que varía entre 0 y 1. Z por su parte, es una combinación lineal de las variables independientes x_i , los coeficientes de regresión β_i y el intercepto o también llamado sesgo β_0 (Goyez-Peñafiel y Hernández-Rojas, 2021). Esta representa la presencia o ausencia de deslizamientos (condicionales actuales) en las variables independientes (condiciones previas al deslizamiento) (Bai S. et al., 2014), y se define por la siguiente ecuación (2):

$$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n \quad 2$$

La relación entre P y Z puede ser expresada usando la ecuación 3 (Huang et al., 2015):

$$\logit(P) = \ln\left(\frac{P}{1-P}\right) = Z \quad 3$$

Donde $P/(1-P)$ se conoce como odds *ratio*. En esta investigación para la identificación de la susceptibilidad a los deslizamientos se definieron dos modelos de regresión logística. El primero (ModeloDummy) enfocado a medir el rendimiento del método usando variables continuas y variables dummy² el segundo (ModeloCon) está orientado a medir el rendimiento del método únicamente usando variables continuas. Para ello se elaboró una *script* en un entorno de Python 3.10, donde los archivos ráster que representan las zonas de muestras de entrenamiento fueron convertidos a vectores planos usando la librería GDAL, para después ser alojados en *dataset* como vectores columnas, en el que se codificaron sus clases y eliminaron datos nulos en los casos. El dataset final cuenta con un total de 2.191.775 registros y 35 columnas.

Análisis de multicolinealidad

Un problema en el ajuste de modelos de regresión logística es la sensibilidad del método a la presencia de colinealidad entre variables independientes (Hosmer y Lameshow, 2000), lo que hace necesario usar métricas que permitan su identificación. Las medidas más comunes para evaluar la multicolinealidad son el valor de tolerancia (TOL) y el factor de inflación de la varianza (VIF) las cuales miden el grado en el que cada variable independiente se explica por otras variables independientes (Hair et al., 1999). En este caso, a pesar de que no existe una regla que defina como deben ser interpretados los VIF, en este estudio se definió el umbral de corte en un valor de $VIF > 2$ y $Tol < 0.4$, el cual ha sido empleado en múltiples ocasiones la definición de multicolinealidad (Bai S. et al., 2010; Bai S. et al., 2014).

Usando el dataset elaborado con anterioridad, se calcularon los valores de los factores de inflación de varianza (VIF) y tolerancia (TOL) para las de variables de cada modelo, con lo que se identifica que no es necesario remover variable por problemas

relacionados con multicolinealidad. El cálculo de los valores VIF, TOL y el tipo de variable usada para el entrenamiento de los modelos se muestra en la Tabla 3.

Tabla 3 - Resultado del análisis de multicolinealidad para las variables de los modelos en entrenamiento

Variable	ModeloDummy		ModeloCon		Tipo
	VIF	TOL	VIF	TOL	
Distancia a ríos	1,0968	0,9118	1,0485	0,9537	Continua
Distancia a fallas geológicas	2,0768	0,4815	1,5215	0,6572	Continua
Pendiente del terreno	1,7658	0,5663	1,6488	0,6064	Continua
Curvatura	1,0689	0,9356	1,0676	0,9366	Continua
TW1	1,2294	0,8134	1,2211	0,8188	Continua
Elevación	2,2366	0,4471	1,6422	0,6089	Continua
Distancia a las vías	1,8704	0,5346	1,5188	0,6584	Continua
Densidad deslizamiento litología	-	-	2,2890	0,4368	Continua
Densidad deslizamiento CLC	-	-	1,4798	0,6757	Continua
Lito_DC-Sctm	1,5057	0,6641	-	-	Dummy
Lito_Et-Sc	1,1326	0,8829	-	-	Dummy
Lito_N2-Vi	1,0035	0,9965	-	-	Dummy
Lito_Q-al	1,2146	0,8233	-	-	Dummy
Lito_Q-ca	1,1808	0,8469	-	-	Dummy
Lito_Q-g	1,0651	0,9389	-	-	Dummy
Lito_Q-t	1,0691	0,9354	-	-	Dummy
Lito_bt-Sctm	1,4379	0,6955	-	-	Dummy
Lito_btk-Sm	2,1871	0,4572	-	-	Dummy
Lito_bzb6-Stm	1,0542	0,9486	-	-	Dummy
Lito_b6k1-Stm	1,5990	0,6254	-	-	Dummy
Lito_e6e9-Sct	1,0113	0,9888	-	-	Dummy
Lito_k1k4-Sm	1,1449	0,8734	-	-	Dummy
Lito_k1k6-Stm	2,4854	0,4023	-	-	Dummy
Lito_k5Et-Stm	1,1072	0,9032	-	-	Dummy
Lito_k6Et-Stm	1,3051	0,7662	-	-	Dummy
CLC_11 Zonas urbanizadas	1,0196	0,9807	-	-	Dummy
CLC_12 Zonas industriales o comerciales	1,0221	0,9783	-	-	Dummy
CLC_13 Zonas de extracción mineras y...	1,0113	0,9888	-	-	Dummy
CLC_21 Cultivos transitorios	1,0084	0,9917	-	-	Dummy
CLC_22 Cultivos permanentes	1,0301	0,9708	-	-	Dummy
CLC_23 Pastos	1,6222	0,6165	-	-	Dummy
CLC_24 Áreas agrícolas heterogéneas	1,8452	0,5420	-	-	Dummy
CLC_32 Áreas con vegetación herbácea...	1,7777	0,5625	-	-	Dummy
CLC_33 Áreas abiertas, sin o con poca...	1,5135	0,6607	-	-	Dummy
CLC_51 Aguas continentales	1,4919	0,6703	-	-	Dummy

Entrenamiento de los modelos de regresión logística

El dataset de variables se dividió en dos grupos: un dataset de entrenamiento y un dataset testeo, los cuales contienen el 70% y 30% del total de los datos respectivamente. El primer modelo (ModeloDummy) se entrenó usando como independientes variables categorizadas y variables dummy elaboradas a partir de los factores de cobertura de la tierra y litología, mediante un proceso de codificación *one-hot*, partir del cálculo donde cada categoría de la variable categórica se representa por un vector de *bits* (uno y cero). Por otra parte, el segundo modelo (ModeloCon) se entrenó usando como insumos únicamente variables continuas y categorizadas, las últimas elaboradas a partir del cálculo de densidad de área deslizada por categoría para las variables de

2. Las cuales son variables binarias usadas para representar una categoría determinada.

cobertura de la tierra y litología.

Los modelos se entrenaron usando la librería Statsmodels-v1.24.0 que ofrece herramienta orientadas al entrenamiento de modelos de regresión lineal múltiple. El proceso de entrenamiento de cada modelo consistió en una primera etapa de ajuste usando todas las variables, sobre este se verificó la significancia estadística de cada variable, para finalmente ajustar el modelo con las variables significativas. Con los modelos entrenados, se probó cada uno usando los datos disponibles en el dataset de testeo, ello para validar el rendimiento de las predicciones realizadas con los modelos mencionados.

La Tabla 4 muestra el resultado del entrenamiento de los parámetros asociados a cada variable usada en los modelos planteados, de las variables implementadas originalmente en ModeloDummy siete fueron eliminadas por no ser estadísticamente representativas. De forma somera el valor del Pseudo R2 indica que el ModeloCon ajusto los datos de entrenamiento de mejor forma que y con un numero menos de iteraciones que el ModeloDummy.

Tabla 4 - Resultados de la estimación de pesos para el modelado de susceptibilidad por medio de los modelos planteado

Modelo Entrenado	ModeloDummy	ModeloCon
Parámetros	Estimado	Estimado
Distancia a ríos	0,4286	0,4811
Distancia a fallas geológicas	0,2138	0,2020
Pendiente del terreno	0,8185	0,7505
Curvatura	-0,0932	-0,1055
TWI	0,0603	0,0582
Elevación	-0,4569	-0,2257
Distancia a las vías	-0,2342	-0,0548
Densidad deslizamiento litología		0,9398
Densidad deslizamiento CLC		0,8515
Lito_DC-Sctm	-0,8189	
Lito_EI-Sc	-1,0770	
Lito_N2-Vi	-2,2013	
Lito_Q-al	-5,2254	
Lito_Q-ca	-0,5486	
Lito_Q-g	-3,8445	
Lito_Q-t	-0,2177	
Lito_b1k1-sm	-2,6918	
Lito_b2b6-Stm	-3,1485	
Lito_b6k1-stm	-1,5880	
Lito_k1k4-sm	-3,1132	
Lito_k1k6-Stm	-3,6790	
Lito_k5E1-Stm	-1,9172	
Lito_k6E1-Stm	-2,4610	
CLC.1.2. Zonas industriales o comerciales y redes de ...	-0,7764	
CLC.1.3. Zonas de extracción mineras y escombreras	3,7575	
CLC.2.3. Pastos	0,0666	
CLC.3.2. Áreas con vegetación herbácea y/o arbustiva	0,5107	
Intercepto	-2,9179	-10,1004
Numero de iteraciones	213	7
Pseudo R2	0,4849	0,5261

Validación de los modelos de regresión logística

Para validar la precisión de los modelos entrenados, se utilizó la curva de características operativas del receptor ROC por sus siglas en inglés, esta es una técnica para visualizar, organizar y seleccionar

clasificadores en función de su rendimiento (Fawcett, 2006). Generalmente se utiliza para cuantificar la precisión de los modelos de susceptibilidad a deslizamientos (Bai et al., 2014; Chauhan et al., 2010; Guillard y Zezere, 2012; Basu y Pal, 2017; Vakhshoori et al., 2019).

El área bajo la curva ROC o AUC por sus siglas en inglés, se usa como una métrica escalar que representa el rendimiento de los modelos (Fawcett, 2006), el cálculo de la curva ROC se basa en la construcción de un número definido de umbrales donde se identifican los valores asociados a la sensibilidad y especificidad, que en este caso vienen definidos por la ecuación 4 y ecuación 5 (Vakhshoori et al., 2019):

$$\text{Sensibilidad} = \frac{L_{>i}}{T_L} \quad 4$$

$$\text{Especificidad} = \frac{S_{>i}}{T_S} \quad 5$$

Donde L_i es el número de píxeles de deslizamiento con un valor superior al umbral, y S_i es el número de píxeles estables con un valor inferior al umbral. T_L y T_S son el número total de píxeles de deslizamiento y píxeles estables respectivamente. La curva ROC se crea al graficar los valores de *sensibilidad* y de *especificidad* para cada umbral definido en el eje x y en el eje y respectivamente, el valor del AUC para la curva se calcula mediante la ecuación 5 (Guillard y Zezere, 2012; Vakhshoori et al., 2019):

$$AUC = \sum_{i=1}^{n+1} \frac{1}{2} \sqrt{(x_i - x_{i+1})^2 \cdot (y_i + y_{i+1})} \quad 6$$

Donde x_i y y_i son los umbrales definidos. Adicional para cada modelo se calculó la matriz de confusión, la cual permite comparar las observaciones realizadas en el inventario de deslizamiento con las predicciones realizadas por los modelos (Aristizabal et al., 2019), con lo que es posible hacer inferencias sobre la exactitud de los modelos planteados.

Resultados

Resultados modelos de regresión logística

La Figura 13 muestra las curvas ROC calculadas para estimar la precisión de las predicciones para el ModeloDummy y el ModeloCon. El valor de AUC para el primero es igual a 0,91730 que corresponde a una precisión del 91,73% en las predicciones realizadas con este, por otra parte, el segundo modelo tiene un AUC de 0,93280, mostrando una precisión en las predicciones realizadas del 93,28%.

Por otra parte, los valores de verdaderos positivos (TP) y de verdaderos negativos (TN) que se muestran en la matriz de confusión de la Figura 14 indican que el ModeloDummy clasificó de forma adecuada el 43,5% de los píxeles como zonas de deslizamiento y el 41,5% como zonas de no deslizamiento, indicando con ello una exactitud del 85,03% para las predicciones del modelo, los valores de los falsos positivos (FP) y falsos negativos (FN) aunque bajos, indican que el modelo tiende sobre estimar las zonas de deslizamiento en el 9,4% de los casos mientras que sobre estima las zonas de no deslizamiento el 5,6% de los casos.

En cuanto al ModeloCon, los valores de falsos positivos (FP) y falsos negativos (FN) en la Figura 13c muestran la tendencia de

Figura 13 - Curvas ROC calculadas para estimar la precisión de las mediciones

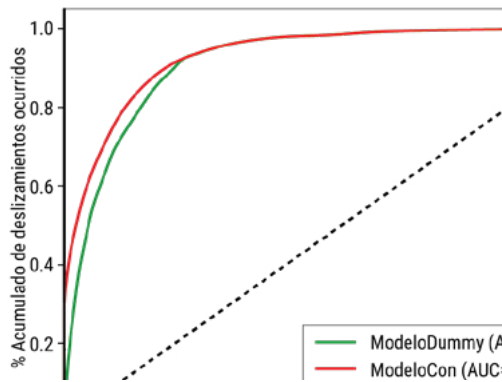
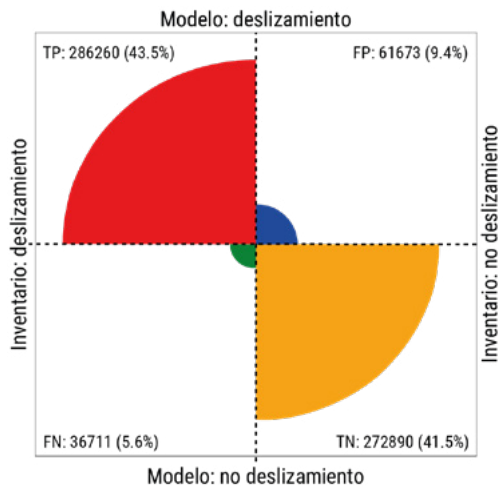


Figura 14 - Matriz de confusión ModeloDummy

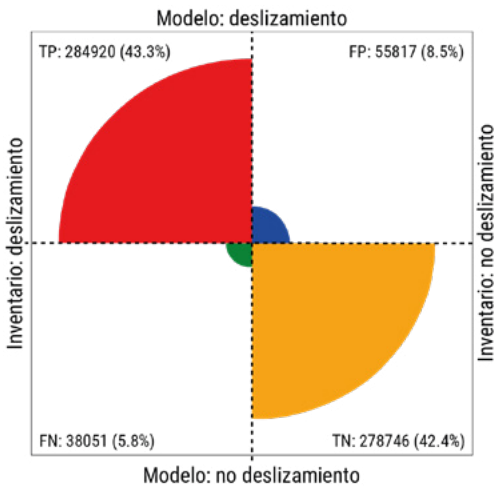


este a sobreestimar las zonas no deslizadas y deslizadas en el 5,8% y 8,5% de los casos respectivamente, el valor de verdaderos positivos (TP) y verdaderos negativos (TN) indican que el modelo clasificó de forma adecuada el 43,4% de los píxeles como zonas de deslizamiento y el 42,4% como zonas de no deslizamiento, mostrando una precisión en las predicciones del 85,77% (Figura 15).

Resultado de los mapas de susceptibilidad

Una vez entrenados los modelos y evaluada su capacidad de predicción, el siguiente paso consistió en la elaboración de los mapas de susceptibilidad a deslizamientos. Un aspecto a tener en cuenta es la necesidad de clasificación del nivel susceptibilidad definida para cada uno de los mapas, la cual varía entre 0 a 1. En este caso se caracterizó la susceptibilidad en cinco rangos iguales, según se muestra en la Tabla 5, para el proceso de creación de los mapas de susceptibilidad se usaron geoprocetos de las herramientas de la plataforma ArcGIS Pro-3.1.

Figura 15 - Matriz de confusión ModeloCon



Probabilidad	Grado de susceptibilidad
0,0-0,2	Muy bajo
0,2-0,4	Bajo

Tabla 5 - Clasificación de susceptibilidad de los valores de probabilidad estimados por modelos de regresión logística

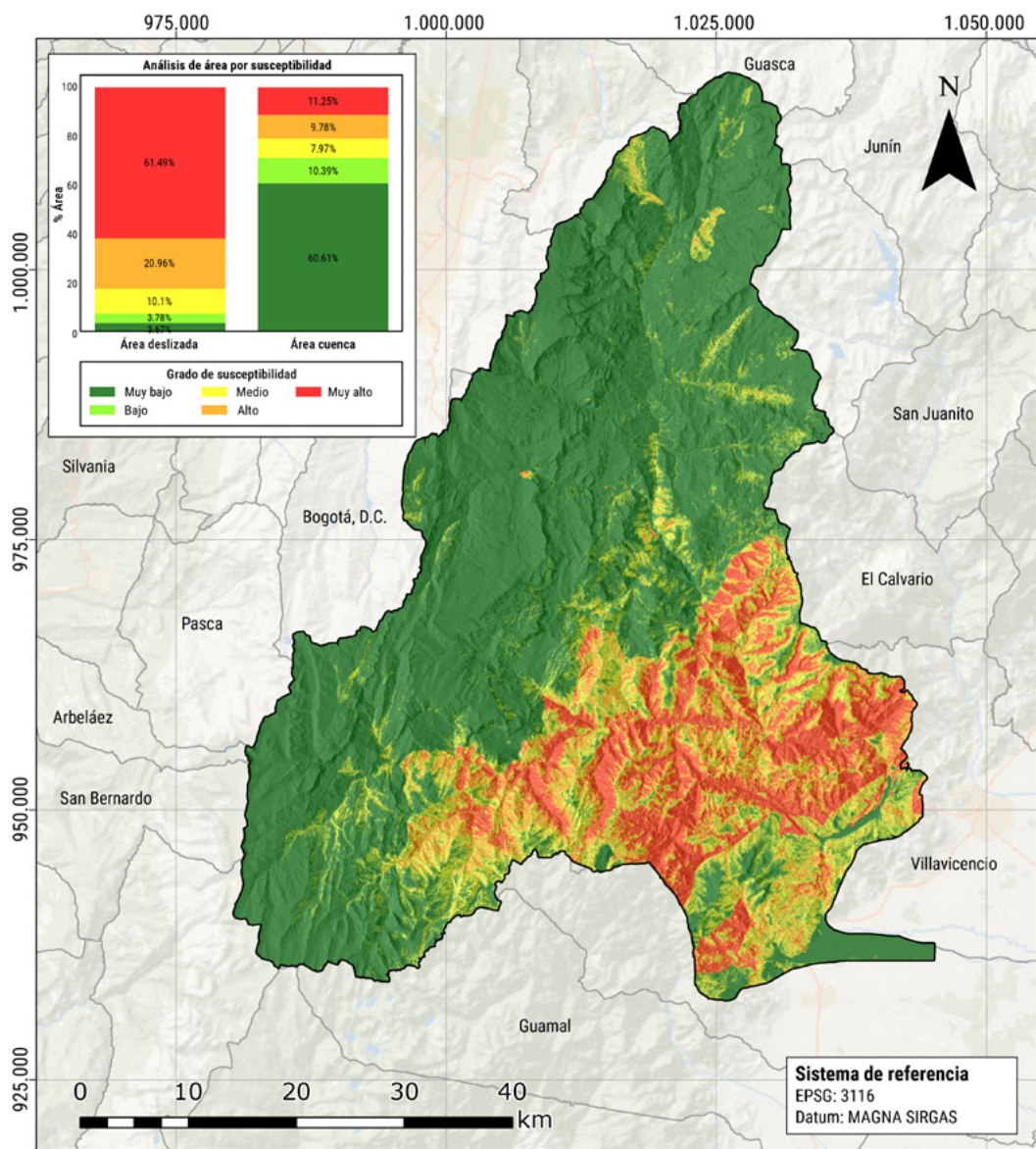
Probabilidad	Grado de susceptibilidad
0,4-0,6	Medio
0,6-0,8	Alto
0,8-1,0	Muy alto

El resultado del mapa de susceptibilidad a deslizamientos elaborado a partir del ModeloDummy que se muestra en la Figura 16, permitió identificar que el 3,67% de los deslizamientos se presentó en zonas con grado de susceptibilidad muy bajo las cuales representan el 60,61% del área de la cuenca, tales zonas presentan un porcentaje de deslizamiento por hectárea de 0,019%. Por otra parte, en las zonas con grado de susceptibilidad bajo que representan el 10,39% del área total de la cuenca, se presentaron el 3,78% de los deslizamientos, mostrando un porcentaje de deslizamiento por hectárea de 0,12%.

En las zonas definidas como grado de susceptibilidad medio se presenta el 10,1% de los deslizamientos, estas cubren el 79,7% de área de la cuenca y presentan un porcentaje de deslizamientos por hectárea de 0,41%. Las zonas con grado de susceptibilidad alto ocupan el 20,98% de la cuenca, en estas se presentan el 20,96% de los deslizamientos mapeados, indicando un porcentaje de deslizamientos por hectárea de 0,41%. Por ultimo las zonas cuyo grado de susceptibilidad es muy alto se presentan en el 11,25% del área de la cuenca y dentro de tales se concentra el 61,49% de los deslizamientos ocurridos, presentando un porcentaje de deslizamiento por hectárea de 1,8%.

En cuanto al mapa de susceptibilidad a deslizamientos elaborado a partir del ModeloCon que se muestra en la Figura 17, se identificó que zonas con grado de susceptibilidad muy

Figura 16 - Resultado del mapa de susceptibilidad usando el ModeloDummy



bajo representan el 64,48% del área de la cuenca y contienen el 4,47% de los deslizamientos, mostrando un porcentaje de deslizamiento por hectárea de 0,022%. Las zonas con grado de susceptibilidad bajo contienen el 3,78% de los deslizamientos y ocupan el 8,84% del área de la cuenca con un porcentaje de deslizamiento por hectárea del 0,14%.

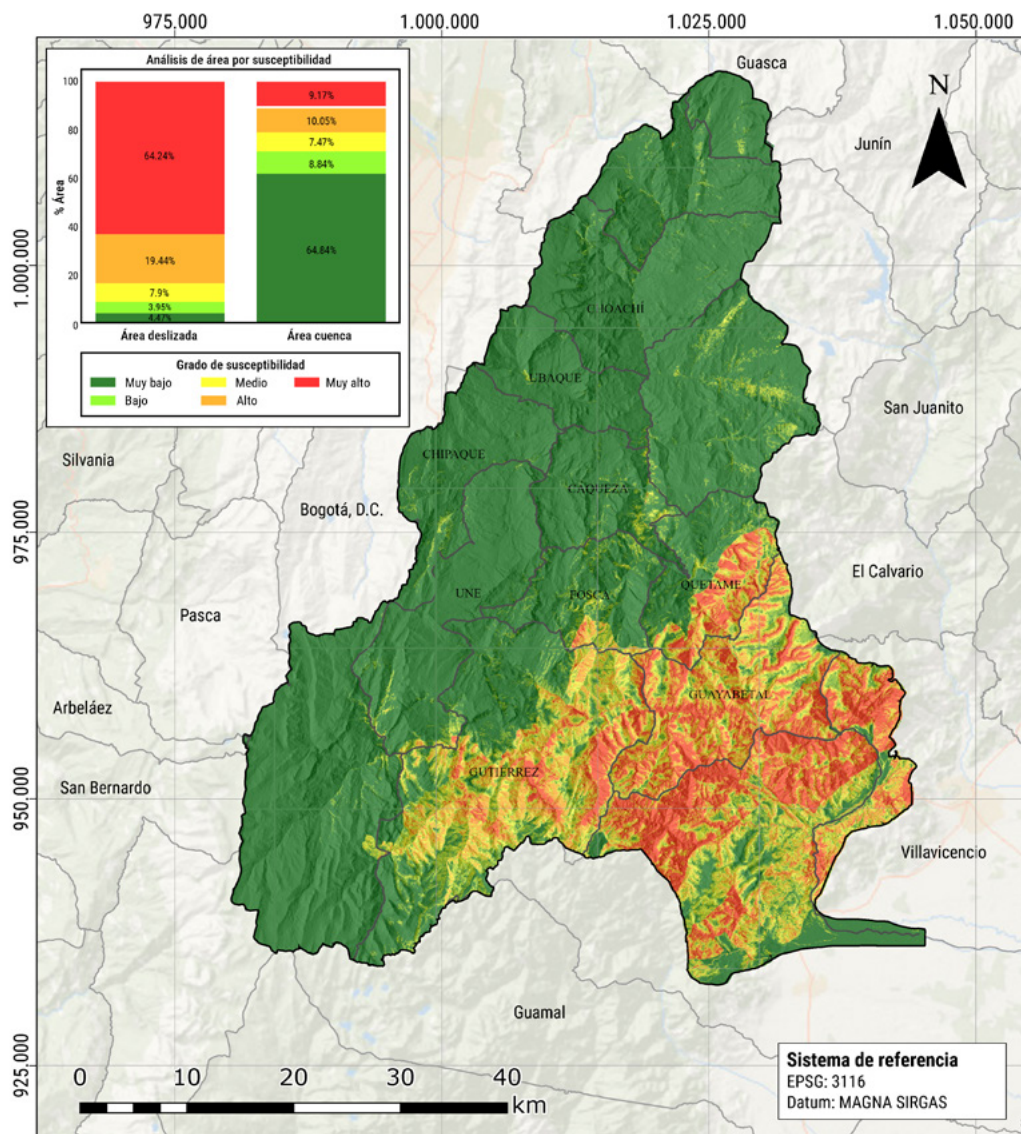
Dentro de las zonas con grado de susceptibilidad medio que cubren el 79,7% de la cuenca se presenta el 79,0% de los deslizamientos presentando un porcentaje de deslizamientos por hectárea de 0,34%. En las zonas definidas como grado de susceptibilidad alto se presenta el 19,44% de los deslizamientos, estas

cubren el 10,05% de área de la cuenca y presentan un porcentaje de deslizamientos por hectárea de 0,63%. Por último, las zonas con grado de susceptibilidad muy alto ocupan el 9,17% de la cuenca, en estas se ocurre el 64,24% de los deslizamientos mapeados, indicando un porcentaje de deslizamientos por hectárea de 2,3%.

Discusión

Las métricas de evaluación y entrenamiento descritas mostraron que el ModeloCon presentó un mejor ajuste al fenómeno mapeado, requiriendo tiempos de preparación de factores y entrenamiento de modelos menores, así como menor coste computacional. Según

Figura 17 - Resultado del mapa de susceptibilidad usando el ModeloCon



los valores de AUC que se muestran en la Figura 13, el ModeloCon presenta un desempeño en su capacidad de predicción que supera aproximadamente en 15% al mostrado por el ModeloDummy, esto concuerda con los valores de pseudo R^2 obtenidos para el entrenamiento de los modelos (véase Tabla 4). Igualmente, el resultado de los mapas de susceptibilidad (Figura 14 y Figura 15) muestra que el mapa elaborado a partir de los coeficientes entrenados en el ModeloCon presentó un mejor ajuste de los deslizamientos ocurridos en relación de la zona de susceptibilidad muy alta, dado que en el 9,17% del total del área de la cuenca se enmarcan el 64,24% de los deslizamientos ocurridos, mientras que en el ModeloDummy el 61,49% de los deslizamientos se presenta en el 11,25% del área de la cuenca.

El resultado de los modelos de susceptibilidad elaborados

mostró una tendencia a la ocurrencia de deslizamientos en la cuenca a lo largo de un cinturón al sur oriente, sobre la cordillera oriental y el piedemonte llanero, especialmente en las zonas de los municipios de Guayabetal, Gutiérrez, Quetame, Acacias y Villavicencio que se encuentran dentro de la cuenca. Estas zonas presentan relación con rocas sedimentarias de ambiente transicional marino y rocas metamórficas de bajo grado las cuales datan de los periodos Devónico y Carbonífero (419 a 298 Ma) y periodo Ordovícico a Silúrico (485,4 a 419,2 Ma) respectivamente, pertenecientes al grupo Quetame y Farallones.

En cuanto a los factores implementados, los valores de los coeficientes obtenidos durante el entrenamiento (Tabla 4) mostraron que, para los dos modelos, factores como la distancia a

las vías y ríos, la pendiente e índice TWI tienen una correlación positiva con la ocurrencia de deslizamientos, mientras que la curvatura, la elevación y la distancia a las vías presentan una correlación negativa. El valor del coeficiente entrenado para la distancia a las vías permite inferir que a nivel regional estas no juegan un papel predisponente en la ocurrencia a deslizamientos, dado que a medida que aumenta la cercanía a las vías se reduce la probabilidad de ocurrencia de deslizamientos, esta afirmación claramente contrastante con la realidad de los eventos presentados sobre la vía nacional 40, se puede justificar debido a que en la zona de estudio (Figura 12) las vías se encuentran concentradas en zonas donde la ocurrencia de deslizamientos es mínima, lo que infiere que esta variable podría ser omitida en los modelos de regresión logística.

Como muestran Reichenbach et al. (2018), los factores que se tuvieron en cuenta, se encuentran dentro los criterios que con mayor frecuencia se implementan en el mapeo de la susceptibilidad a los deslizamientos. Cabe señalar que este hecho no indica que los mismos sean los más apropiados para caracterizar la existencia de deslizamientos en la cuenca, aunque, por otra parte, el resultado validado mediante el AUC y matrices de confusión, muestran la capacidad de los factores usados en los modelos, para describir el fenómeno en estudio.

En comparación con estudios como los llevados a cabo por autores como Huang et al. (2015), Zhao et al. (2019) y Aristizabal et al. (2019), los cuales se tiene en cuenta varios factores condicionantes en común con el presente estudio (elevación, pendiente, curvatura, litología, distancia a las fallas geológicas y ríos) y en los que los deslizamientos son representados mediante puntos, los modelos aquí presentados lograron un mejor rendimiento, mostrando un valor de AUC superior hasta en 13%. Por otra parte, al comparar con estudios como los realizados por Yilmaz (2009) y Ayalew y Yamagishi (2005), en los cuales las zonas deslizadas se describen por medio de polígonos capturados sobre imágenes con un tamaño de pixel de 10x10 metros. Los modelos aquí entrenados siguen presentando un mejor ajuste, de hasta un 8% en relación con los estudios mencionados.

La diferencia en el rendimiento de los modelos aquí planteados con relación a los estudios mencionados, puede verse explicada entre otros factores, por un mayor número de elementos (registros) para el entrenamiento de estos, ello dado a que en el presente estudio las zonas deslizadas fueron caracterizadas como polígonos usando un mosaico de imágenes satelitales con un tamaño de pixel de 3x3 metros, lo que implica una mejor descripción de tales zonas y un mayor número de píxeles por cada zona; el uso de un amplio dataset de entrenamiento lleva a que la convergencia de los modelos sea más acelerada, logrando niveles de precisión altos, lo que se reflejan en modelos robustos y confiables.

Conclusiones

Los modelos de susceptibilidad generados demostraron ser de gran utilidad para la identificación y mapeo de las zonas con potencial a presentar deslizamientos a nivel regional, la información obtenida a partir de estos puede ser vital para la seguridad de la vida humana y servir como insumo en planes de ordenamiento y gestión del riesgo. En general, las variables y la metodología utilizadas

presentan un ajuste adecuado a la realidad del fenómeno dentro de la cuenca del río Guayuriba, lo que se ve reflejado el valor del área bajo la curva (AUC) y la matriz de confusión que muestran un ajuste excelente para los modelos elaborados.

La resolución espacial de las variables implementadas y la caracterización de las zonas de deslizamiento como polígonos, permitió obtener un amplio número de píxeles para el entrenamiento, garantizando una buena representación de la variabilidad espacial y la relación entre los factores predisponentes y las zonas deslizadas, lo que se ve evidenciado en el rendimiento y la capacidad predictiva de los modelos. Las métricas de evaluación de estos confirman el alto rendimiento de la técnica en términos de predicción de posibles zonas de deslizamiento, la comparación entre modelos entrenados usando diferente metodología permitió validar que los modelos entrenados con variables continuas presentaron un mejor desempeño requiriendo menor coste computacional.

Agradecimientos

Se agradece a Procalculo Prosis S.A. por suministrar acceso a las imágenes satelitales de alta resolución PlanetScope, el acceso al sistema de información geográfica ArcGIS Pro y por apoyar las iniciativas de investigación de sus colaboradores promoviendo el desarrollo académico.

Referencias

- Alcántara, Irasema. (2000). Landslides: ¿deslizamientos o movimientos del terreno? Definición, clasificaciones y terminología. *Investigaciones Geográficas*, 1(41). <https://doi.org/10.14350/riig59101>
- Aleotti, P. y Chowdhury, R. (1999). Landslide Hazard Assessment: Summary Review and New Perspectives. *Bulletin of Engineering Geology and the Environment*, (58), 21-44. <https://doi.org/https://doi.org/10.1007/s100640050066>
- Aristizabal, E., Vásquez, M. y Ruíz, D. (2019). Métodos estadísticos para la evaluación de la susceptibilidad por movimientos en masa. *Tecnológicas*, 22(46), 39-60. <https://doi.org/https://doi.org/10.22430/22565337.1247>
- Ayalew, L. y Yamagishi, H. (2005). The Application of GIS-based Logistic Regression for Landslide Susceptibility Mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology*, (1-2), 15-31. <https://doi.org/10.1016/j.geomorph.2004.06.010>
- Bai, S., Wang, J., Lu, G. N., Zhou, P. g., Hou, S. S. y Xu, S. N. (2010). GIS-based Logistic Regression for Landslide Susceptibility Mapping of the Zhongxian Segment in the Three Gorges Area, China. *Geomorphology*, 115(1-2), 23-31. <https://doi.org/https://doi.org/10.1016/j.geomorph.2009.09.025>
- Bai, S., Wang, J., Thiebes, B., Cheng, C. y Chang, Z. (2014). Susceptibility Assessments of the Wenchuan Earthquake-triggered Landslides in Longnan Using Logistic Regression. *Environmental Earth Sciences*, (71), 731-743. <https://doi.org/https://doi.org/10.1007/s12665-013-2475-z>

- Basu, T. y Pal, S. (2017). Exploring Landslide Susceptible Zones by Analytic Hierarchy Process (AHP) for the Gish River Basin, West Bengal, India. *Spatial Information Research*, 25, 665-675. <https://doi.org/https://doi.org/10.1007/s41324-017-0134-2>
- Chauhan, S., Sharma, M., Arora, M. y Gupta, N. (2010). Landslide Susceptibility Zonation Through Ratings Derived from Artificial Neural Network. *International Journal of Applied Earth Observation and Geoinformation*, 12(5), 340-350. <https://doi.org/10.1016/j.jag.2010.04.006>
- Chen, W. y Li, Y. (2020). GIS-based Evaluation of Landslide Susceptibility Using Hybrid Computational Intelligence Models. *CATENA*, (195), 104777. <https://doi.org/10.1016/j.catena.2020.104777>
- Comisión Económica para América Latina y el Caribe. (2012). *Valoración de daños y pérdidas: ola invernal en Colombia, 2010-2011*. Comisión Económica para América Latina y el Caribe-Cepal y el Banco Interamericano de Desarrollo-BID. <https://repositorio.cepal.org/server/api/core/bits-treams/8cf425f2-2b9b-4362-b5b5-d33e21321125/content>
- Conforti, M., Pascale, S., Robustelli, G. y Sdao, F. (2014). Evaluation of Prediction Capability of the Artificial Neural Networks for Mapping Landslide Susceptibility in the Turbolo River Catchment (Northern Calabria, Italy). *CATENA*, (113), 236-250. <https://doi.org/10.1016/j.catena.2013.08.006>
- Cooper, M. A., Addison, F. T., Alvarez, R., Coral, M., Graham, R. H., Hayward, A. B. y Howe, S. (1995). Basin Development and Tectonic History of the Llanos Basin, Eastern Cordillera, and Middle Magdalena Valley, Colombia. *AAPG Bulletin*, 79(10), 1421-1442. <https://doi.org/10.1306/7834D9F4721-11D7-8645000102C1865D>
- Dai, F., Lee, C. y Xu, Z. (2001). Assessment of Landslide Susceptibility on the Natural Terrain of Lantau Island, Hong Kong. *Environmental Geology*, (40), 381-391. <https://doi.org/10.1007/s002540000163>
- Fawcett, T. (2006). An Introduction to ROC Analysis. *Pattern Recognition Letters*, 27(8), 861-874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Fondo de Adaptación. (2018). Capítulo 4. Climatología. En F. d. adaptación, *POMCA Río Guayuríba Villavicencio* : CORMACARENA.
- Goyez-Peñafiel, P. y Hernández-Rojas, A. (2021). Landslide Susceptibility Index Based on the Integration of Logistic Regression and Weights of Evidence: A Case Study in Popayan, Colombia. *Engineering Geology*, (280), 105958. <https://doi.org/10.1016/j.enggeo.2020.105958>
- Guillard, C. y Zezere, J. (2012). Landslide Susceptibility Assessment and Validation in the Framework of Municipal Planning in Portugal: The Case of Loures Municipality. *Environmental Management*, 50(4), 721-735. <http://doi.org/10.1007/s00267-012-9921-7>
- Guzzetti, F., Carrara, A., Cardinali, M. y Reichenbach, P. (1999). Landslide Hazard Evaluation: A Review of Current Techniques and Their Application in a Multi-scale Study, Central Italy. *Geomorphology*, 31(1-4), 181-216. [https://doi.org/10.1016/S0169-555X\(99\)00078-1](https://doi.org/10.1016/S0169-555X(99)00078-1)
- Hair, J. F., Anderson, R. E., Tatham, R. L. y Black, W. (1999). *Análisis multivariante*. (5ª edición). Prentice Hall.
- Hosmer, D. W. y Lemeshow, S. (2000). *Applied Logistic Regression*. (3ª edición). Wiley.
- Huang, J., Zhou, Q. y Wnag, F. (2015). Mapping the Landslide Susceptibility in Lantau Island, Hong Kong, by Frequency Ratio and Logistic Regression Model. *Annals of GIS*, 21(3), 191-208. <http://doi.org/10.1080/19475683.2014.992373>
- Mandal, B. y Mandal, S. (2018). Analytical Hierarchy Process (AHP) Based Landslide Susceptibility Mapping of Lish River Basin of Eastern Darjeeling Himalaya, India. *Advances in Space Research*, 62(11), 3114-3132. <https://doi.org/10.1016/j.asr.2018.08.008>
- Mauri, L., Cucchiaro, S., Grigolato, S., Dalla Fontana, G. y Tarolli, P. (2022). Evaluating the Interaction Between Snowmelt Runoff and Road in the Occurrence of Hillslope Instabilities Affecting a Landslide-prone Mountain Basin: A Multi-modeling Approach. *Journal of Hydrology*, 612, 128200. <https://doi.org/10.1016/j.jhydrol.2022.128200>
- Quesada, A. y Feoli-Borasci, S. (2018). Comparación de la metodología Mora-Vahrson y el método morfométrico para determinar áreas susceptibles a deslizamientos en la microcuenca del río Macho, Costa Rica. *Revista Geográfica de América Central*, 2(61), 17-45. <https://doi.org/10.15359/rgac.61-2.1>
- Reichenbach, P., Rossi, M., Malamud, B., Mihir, M. y Guzzetti, F. (2018). A Review of Statistically-based Landslide Susceptibility Models. *Earth-Science Reviews*, 180, 60-91. <https://doi.org/10.1016/j.earscirev.2018.03.001>
- Soeters, R. y Van Westen, C. J. (1996). Slope Instability Recognition, Analysis and Zonation. En A. Keith Turner y R. L. Schuster (Ed.), *Landslides, Investigation and Mitigation. Transportation Research Board, National Research Council, Special Report 247* (pp. 129-177). National Academy Press. <https://onlinepubs.trb.org/Onlinepubs/sr/sr247/sr247-008.pdf>
- Suárez, J. (2001). *Control de erosión en zonas tropicales*. (Vol. 1). Bucaramanga: Universidad Industrial de Santander. <https://edicionesuis.edu.co/index.php/publicacionesuis/catalog/book/124>
- Suárez, J. (2009). *Deslizamientos: análisis geotécnico*. (Vol. 1). Bucaramanga: Universidad Industrial de Sanatander. <https://edicionesuis.edu.co/index.php/publicacionesuis/catalog/book/107>
- Trigila, A., Iadanza, C., Esposito, C. y Scarascia-Mugnozza, G. (2015). Comparison of Logistic Regression and Random Forests Techniques for Shallow Landslide Susceptibility Assessment in Giampilieri (NE Sicily, Italy). *Geomorphology*, (249), 119-136. <https://doi.org/10.1016/j.geomorph.2015.06.001>
- Vakhshoori, V., Reza Pourghasemi, H., Zare, M. y Blaschke, T. (2019). Landslide Susceptibility Mapping Using GIS-Based Data Mining Algorithms. *Water*, 11(11), 2292. <https://doi.org/10.3390/w11112292>
- Yalcin, A. (2008). GIS-based Landslide Susceptibility Mapping Using Analytical Hierarchy Process and Bivariate Statistics in Ardesen (Turkey): Comparisons of Results and Confirmations. *CATENA*, 72(1), 1-12. <https://doi.org/10.1016/j.catena.2007.01.003>
- Yilmaz, I. (2009). Landslide Susceptibility Mapping Using Frequency Ratio, Logistic Regression, Artificial Neural Networks and Their Comparison:

A Case Study From Kat Landslides (Tokat-Turkey). *Computer & Geosciences*, 35(6), 1125-1138. <https://doi.org/10.1016/j.cageo.2008.08.007>

Zhao, Y., Wang, R., Jiang, Y., Liu, H. y Wei, Z. (2019). GIS-based Logistic Regression for Rainfall-induced Landslide Susceptibility Mapping Under Different Grid Sizes in Yueqing, Southeastern China. *Engineering Geology*, 259(4), 105147. <https://doi.org/10.1016/j.enggeo.2019.105147>

Zhou, S., Zhou, S. y Tan, X. (2020). Nationwide Susceptibility Mapping of Landslides in Kenya Using the Fuzzy Analytic Hierarchy Process Model. *Land*, 9(12), 535. <https://doi.org/10.3390/land9120535>

