

Minería de datos aplicada a la estimación de vulnerabilidad al deslizamiento de taludes a partir de parámetros hidrogeomorfológicos

Data mining applied to slope landslide vulnerability assessment from hydrogeomorphological parameters

Uzcategui-Salazar, Marisela^{1,2*}; Toro-Mora, Rosibeth^{1,3}; Guerrero, Omar Antonio^{1,4}; Jégat, Hervé⁵

¹ Grupo de Investigación TERRA. Escuela de Ingeniería Geológica - Universidad de los Andes- Venezuela.

² Escuela Internacional de Doctorado. Universidad Rey Juan Carlos- España.

³ Facultad de Ciencias-Universidad de Zaragoza- España.

⁴ Escuela de Geociencias- Universidad Técnica Particular de Loja- Ecuador.

⁵ Centro Interamericano de Desarrollo e Investigación Ambiental y Territorial. Universidad de los Andes. Mérida - Venezuela.

*mariselaucateguis@gmail.com

Resumen

La minería de datos es un conjunto de técnicas que permiten explorar grandes bases de datos para descubrir patrones, similitudes y/o conexiones entre los datos (a través de algoritmos) y comprender el comportamiento de los mismos para predecir tendencias. El agrupamiento de datos (clustering) a través de la técnica de k-medias, es una técnica de minería de datos no supervisada en la que el algoritmo ayuda a identificar todo tipo de patrones desconocidos en los datos sin etiquetas predeterminadas, con una supervisión humana mínima o nula. El presente trabajo tiene como finalidad determinar a través del análisis de agrupamiento de datos (clustering) k-medias, la vulnerabilidad al deslizamiento que pueden tener los taludes, analizándolos a partir de datos hidrogeomorfológicos, lo que permite hacer una caracterización preliminar de las zonas potenciales de deslizamiento a fin de generar planes de ordenación territorial y prevenir futuros eventos de deslizamiento. Los parámetros involucrados incluyen la cobertura vegetal, coeficiente de humedad del suelo, precipitación, pendientes del terreno y acumulación de flujo. Todos estos parámetros fueron obtenidos espacialmente para toda el área de estudio y almacenados en una base de datos a través de sistemas de información geográfica (SIG). El área fue discretizada en celdas con tamaño de píxeles de 150x150m en los que fue determinado el valor de cada uno de los parámetros, generando así una gran base de datos. Los resultados mostraron que los parámetros: Precipitación, humedad del suelo y acumulación de flujo contribuyen de manera importante en aumentar la vulnerabilidad de los taludes, mientras que los parámetros cobertura vegetal y pendiente poseen menos influencia. Estos resultados demuestran que las técnicas de minería de datos son una opción confiable, eficiente y económica para la evaluación de la vulnerabilidad de taludes.

Palabras claves: Vulnerabilidad de taludes, minería de datos, k-medias, hidrogeomorfología mezclas asfálticas.

Abstract

Data mining is a set of techniques to explore large databases to discover patterns, similarities and/or connections between data (through algorithms) to understand the behaviour of the data to predict trends. Clustering, through k-means technique, is an unsupervised data mining technique in which the algorithm helps to identify all kinds of unknown patterns in the data without predetermined labels, with a little or no human supervision. The present work aims to determine, through k-means clustering analysis, the landslide vulnerability of slopes by analysing them from hydrogeomorphological data. It allows a preliminary characterisation of potential landslide zones in order to generate land management plans and prevent future landslide events. The parameters involved include vegetation cover, soil moisture coefficient, rainfall, terrain slopes and flow accumulation. All these parameters were obtained for the whole study area and they were stored in a database through geographic information systems (GIS). The area was divided into 150x150m cells size pixels to obtain a large dataset to

evaluate the different parameters in each point of cell. The results showed that the parameters: rainfall, soil moisture and flow accumulation contribute significantly to increase the vulnerability of slopes, while the parameters vegetation cover and slope slope penetration have less influence. These results demonstrate that data mining techniques are a reliable, efficient and economical option for slope vulnerability assessment.

Keywords: Landslide vulnerability, data mining, k-means, hydrogeomorphology.

1 Introducción

El análisis de la estabilidad de los taludes es una herramienta de gran importancia práctica en el diseño de obras civiles y la prevención de deslizamientos naturales que puedan afectar o poner en riesgo las actividades humanas (Gonzalez de Vallejo et al., 2002; Li and Liu, 2005). Para su evaluación es necesario conocer el marco geológico y geomecánico de los materiales que forman el talud y los factores que condicionan o desencadenan la inestabilidad (Gonzalez de Vallejo et al., 2002). Además del conocimiento geológico (geotécnico), es importante conocer la hidrogeología y los aspectos geomorfológicos que condicionan naturalmente el deslizamiento de las laderas. Sin embargo, muchas veces es complicado lograr un análisis de la estabilidad de un talud de manera veraz ya que la dificultad en el conocimiento de parámetros geomecánicos e hidrogeológicos que en muchos casos son inexistentes. Es por esta razón, que un estudio preliminar a gran escala puede permitir enfocar el estudio en las zonas con mayores probabilidades de deslizamiento y concentrar esfuerzos e inversión en lograr los parámetros necesarios para la correcta aplicación de los métodos de equilibrio en el análisis de estabilidad. Esta manera de encontrar preliminarmente los sitios más vulnerables para que se produzca un evento de afectación ambiental ha atraído a muchos investigadores en el área de la inteligencia artificial (Javadi et al., 2017; Li and Liu, 2005; Yoo et al., 2016). La minería de datos comprende tres grandes ramas de la ciencia: la estadística, la inteligencia artificial y el aprendizaje automático (*machine learning*) (Pérez Márques, 2014). La minería de datos, también conocida como descubrimiento de conocimientos en base de datos, ha ganado terreno en casi todas las áreas de la ciencia (Riquelme et al., 2006), siendo una técnica de gran importancia en las aplicaciones de campos donde pueden aportar soluciones rápidas, eficientes y objetivas a fin de tomar decisiones apropiadas de manera oportuna. Dentro de la minería de datos existen varias técnicas o algoritmos desarrollados para identificar patrones en las bases de datos para agruparlos de manera apropiada de acuerdo a sus similitudes (Khillar, 2020). El algoritmo de K-medias, es una técnica de agrupamiento de datos no supervisada que permite clasificar información en diferentes grupos o clusters de acuerdo a sus similitudes, basadas en minimizar la suma de la distancia euclidiana entre los puntos de datos y el centroide del clúster (Dabbura, 2020). En los lugares en los que se cuenta con poca información detallada y solo se tienen datos muy gene-

rales de las condiciones hidrogeomorfológicas de un área, la aplicación de esta técnica puede ser de gran ayuda para redefinir el área de estudio y detallar la información de campo necesaria para futuras aplicaciones en el análisis de estabilidad de taludes.

El objetivo principal de esta investigación es estimar la vulnerabilidad de la estabilidad natural de los taludes, por medio de la aplicación de la técnica de minería de datos (K-medias), a partir de parámetros hidrogeomorfológicos del área, con la finalidad de establecer de manera preliminar las zonas a ser estudiadas a detalle para la posterior aplicación de métodos de equilibrio límite. Esta metodología será aplicada a la microcuenca de Ña Leona, estado Mérida-Venezuela.

2 Área de estudio

2.1 Ubicación

La microcuenca de la quebrada Ña Leona, está situada en la región occidental del estado Mérida, específicamente al este del Municipio Santos Marquina (Fig. 1), tiene un área de 18.6 km² y una altitud que comprende desde su nacimiento 4.200 m.s.n.m. hasta su desembocadura en el río Chama de cota 1.686 m.s.n.m.



Fig. 1. Ubicación del área de estudio. Coordenadas, Datum WGS 84, Zona N19

2.2 Aspectos hidrográficos y fisiográficos

El río Chama, el cual constituye el principal eje fluvial de la región Andina, tiene su nacimiento en el Páramo del Águila y desemboca en el Lago de Maracaibo; en el sector de La Mucuy (Tabay, municipio Santo Marquina), se encuentra interceptado por tres tributarios importantes denominados: Ña leona, Desbarrancadero y Mucututuy; que conforman la microcuenca Ña Leona por ser esta la que tiene mayor longitud 9.5 Km y de orden de jerarquía 3; a su vez, confluye con la quebrada La Mucuy (Camacho and Paredes, 2016).

El relieve de la zona está constituido por áreas montañosas con altitudes desde los ~4.300 m.s.n.m. (Laguna Verde - Sierra Nevada) hasta los ~1.500 m.s.n.m. en el Valle del Chama, con pendientes elevadas (> 20 %). Sólo una pequeña porción comprende áreas planas (acumulaciones aluviales y coluvio-aluviales de los fondos de los valles) con pendientes que oscilan entre 0 y 20 %. Ambos drenes son responsables del origen de un amplio abanico aluvial dispuesto perpendicularmente al curso del río Chama (Ferrer, 1988).

Los procesos recientes del modelado del terreno se encuentran bien atenuados; las evidencias de algunos deslizamientos se presentan cubiertos de vegetación y muchos de los escurrimientos, originan lavado y erosión de los suelos. En aquellas zonas donde la cobertura vegetal protege al suelo, los torrentes son pocos activos, pero en aquellas áreas donde hay altas pendientes con un buen poder de arrastre, se forman conos de deyección tanto individuales como coalescentes (CORPOANDES, 2009).

2.3 Aspectos geológicos

El sistema andino montañoso está constituido estratigráficamente por rocas cuyas edades van desde el Precámbrico hasta el Cuaternario; explicando así que la litología presente en el área de estudio está conformada cronológicamente de manera ascendente por la Asociación Sierra Nevada, que forma el basamento de la Sierra Nevada de Mérida perteneciente a edades Precámbricas, formadas por granitos, esquistos estaurolíticos, gneises cuarzos feldespáticos, con predominio de otras rocas silíceas hacia la parte superior; además, localmente se encuentran cuerpos lenticulares de anfibolitas y esquistos pelíticos y depósitos de material Cuaternario que conforman el relleno de los fondos de los valles, modelados en forma de terrazas, tratándose de sedimentos de origen fluvio-glacial y coluvial. También se conforman los abanicos aluviales, que pueden ser típicos conos de deyección o abanicos de lavas torrenciales. Estas terrazas, tiene un origen climático y están controladas fundamentalmente por la tectónica; las cuales se distinguen por dos secuencias separadas: una Pliocena- Pleistocena de arenas y gravas fluviales pobremente escogidas, en capas bien estratificadas y en forma de grandes terrazas y abanicos, y la otra, Pleistocena de arenas y gravas mal escogidas y mal consolidadas, que forman relleno de valles y abanicos sobre sedimentos de la primera; obteniendo así una discordancia

estratigráfica (Gonzalez de Juana et al., 1980).

La tectónica de manera regional está controlada por el sistema de fallas de Boconó la cual es uno de los rasgos neotectónicos más importantes, caracterizada por ser un sistema activo con movimiento transcurrente dextral (Audemard et al., 2009). En esta zona, la Falla de Boconó se presenta en forma casi rectilínea dentro del valle del río Chama, donde se pueden medir escarpes con saltos verticales que oscilan entre los 3 a 15 m. Así mismo, se pudieron observar marcadores geomorfológicos que refleja un movimiento rumbo deslizante dextral: depósitos cuaternarios deformados, drenajes controlados, drenajes desviados, trincheras de fallas, entre otros (Alvarado et al., 2015). Otra particularidad del área en estudio es que forma parte del segmento denominado Tabay-Parque Los Aleros, dicha traza culmina al sur de la población de Tabay donde es aprovechada por la quebrada La Mucuy para entallar el valle rectilíneo paralelo al valle del río Chama (Alvarado et al., 2015). Además, origina un abanico aluvial que evidencia un ligero quiebre de pendiente en su parte distal y un ligero desplazamiento dextral que se observa justo en el cauce de la quebrada, posiblemente la culminación de este segmento origina esfuerzos transpresivos que dan origen al relieve en forma cómica de rocas metamórficas de la Asociación Sierra Nevada (Alvarado et al., 2015). La Fig. 2, muestra el comportamiento estructural del área de estudio, siendo "F3" la traza principal denominada Falla de Boconó, "F2" la Falla de La Mucuy y "F1" una falla asumida que controla la corona de deslizamiento de la microcuenca Ña Leona, y sólo tiene evidencia en imagen satelital (Alvarado et al., 2015; Camacho and Paredes, 2016).

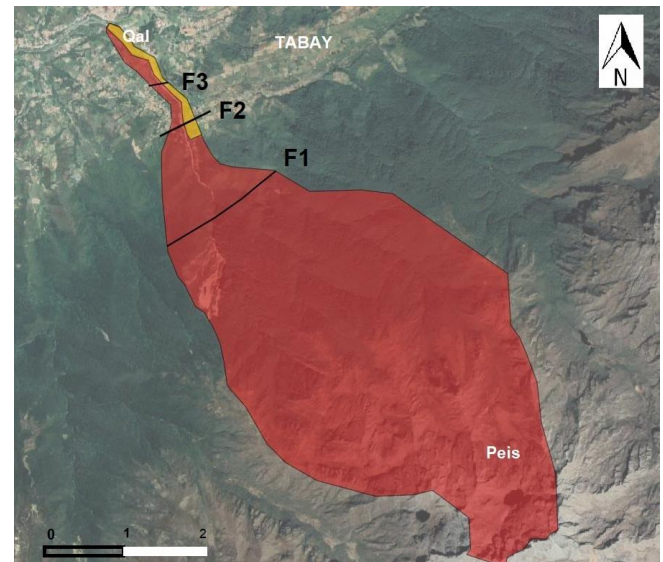


Fig. 2. Aspectos Geológicos Estructurales del Area de Estudio. Peis: Precámbrico, Asociación Sierra Nevada. Qal: Material Cuaternario Aluvial

2.4 Aspectos climáticos

Predomina el clima templado con una temperatura promedio anual de 17.5 °C y precipitaciones de 2.000 mm y balances hídricos de las estaciones la Mucuy I y II positivo con 9 meses húmedos y 3 meses secos y Mucuruba negativo con 6 meses húmedos y 6 meses secos (Guerrero et al., 2013). El régimen pluviométrico es bimodal, con el máximo principal en octubre y el máximo secundario en abril, con una estación bien seca marcada entre diciembre y marzo, con mínimos pluviométricos en el mes de diciembre. Durante los meses de mayo, junio, julio, agosto y septiembre hay una disminución importante de lluvias. Los totales anuales se encuentran entre los 2700mm y 3100mm (Camacho and Paredes, 2016).

3 Materiales y Métodos

3.1 Colección de Datos

Los datos se almacenaron como base de datos geográfica en ArcGIS v10.4.1 y QGIS 3.22. Toda el área de estudio (18.6 Km²) se dividió en 829 píxeles con un tamaño de celda de 150mx150m, con el fin de obtener un gran conjunto de datos para evaluar las diferentes variables en cada punto mediante el análisis de cluster de K-means. Los datos considerados en esta investigación se muestran en la Tabla 1.

Tabla 1. Datos colectados en la investigación

Tipo	Descripción	Fuente	Uso
Cartografía	Mapa topográfico	(MOP, 1975)	Mapa de pendientes Mapa de acumulación de flujo
	Mapa geológico	(MEM, 1980)	Litología, tipo de suelo. Coeficiente de humedad
Datos meteorológicos	Precipitación	(INAMEH, 2000)	Mapa de isoyetas (precipitación), Coeficiente de humedad
Suelo y vegetación	Fotografías aéreas	(BNV, 1980)	Mapa de humedad del suelo
	Imágenes satelitales	(ASTER, 2004)	Mapa de cobertura vegetal

3.2 Metodología

La evaluación se realizó en las siguientes tres etapas:

- Selección de características hidrogeomorfológicas para la estimación de la vulnerabilidad al deslizamiento de taludes.
- Cartografía de los mapas temáticos de las variables seleccionadas
- Estimación de la vulnerabilidad al deslizamiento de taludes mediante análisis de agrupamiento de datos (algoritmo K-medias).

Los parámetros hidrogeomorfológicos considerados en el análisis de la estabilidad de taludes fueron: Pendiente del terreno (*S*), coeficiente de humedad del suelo (*H*), precipitación (*P*), acumulación de flujo (*F*) y cobertura vegetal (*V*).

La pendiente del terreno fue calculada a partir del modelo digital del terreno, utilizando la herramienta de estimación de pendiente de la superficie de ArcGIS.

El coeficiente de humedad del suelo fue obtenido a partir de la información del suelo y vegetación en el área de estudio, estrechamente relacionado con la capacidad de campo y el punto de marchitez permanente a partir de la siguiente ecuación (FCIHS, 2003):

$$HS = (Cc - Pmp) * \gamma_{\text{suelo}} * \text{Profundidad radicular} \quad \text{Ecuación 1}$$

Los valores de precipitación fueron obtenidos a partir de los datos meteorológicos de las estaciones La Mucuy I, La montaña, La Mucuy II, La Aguada y Loma Redonda.

La acumulación de flujo, se determinó a partir del mapa topográfico, en el que se pudo determinar las direcciones de flujo y la posterior acumulación de flujo, calculado con la herramienta en ArcGIS, determinado como el peso acumulado de todas las celdas que fluyen hacia cada celda con pendiente descendente en el ráster de salida (ERSI, 2019).

La cobertura vegetal fue obtenida a partir de las fotografías aéreas e imágenes satelitales del área de estudio. Al tratarse de una variable cuantitativa, fue necesario asignar un valor numérico de acuerdo con la facilidad de deslizamiento de acuerdo al tipo de vegetación (Tabla 2)

Tabla 2. Facilidad de deslizamiento dependiendo de la vegetación

Tipo de vegetación	Facilidad de deslizamiento
Vegetación densa/Bosque	1
Pastizal	2
Arbustos	3
Vegetación escasa	4

El procesamiento de los datos se realizó con RStudio v.4.0.5. Cada parámetro fue normalizado, con el fin de reducir el sesgo (las variables con rangos muy altos predominan sobre las variables con rangos más bajos) (0-1). Se utilizó el método de normalización max-min (Salazar and Del Castillo, 2018) (Ecuación 2).

$$X_n = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad \text{Ecuación 2}$$

Donde *X_i*, indica el valor de los datos en la ubicación *i* del conjunto de datos y *X_{max}* y *X_{min}* son los valores máximo y mínimo del conjunto de datos.

Se utilizó la herramienta de extracción de datos en ArcGIS v.10.4.1 para obtener los datos de cada variable en 829 puntos.

El objetivo de K-medias, es agrupar los puntos de datos con similitudes intrínsecas en un conjunto de datos reducido. Este proceso iterativo se logra mediante el siguiente procedimiento (Dabbura, 2020):

- Creación de un conjunto de datos de matriz $n \times d$, donde n es el número de puntos de datos (829) en un espacio de características d -dimensional (todos los parámetros considerados).
- Selección del número de clústeres: K . El número de clusters se define mediante un índice. El paquete NbCluster en RStudio V.4.0.5 calcula 26 índices (Charrad et al., 2014) para obtener el número óptimo de clústeres y la recomendación es elegir el número de clústeres dado por la regla de la mayoría.
- Cada punto se coloca aleatoriamente en los clústeres iniciales a los que está más cerca. La distancia euclidiana se utiliza para averiguar la distancia de cada punto a un clúster temporal.
- Los clústeres temporales se recalculan con los nuevos centroides en función de los puntos más cercanos situados en ellos. Esto se consigue minimizando la suma de errores al cuadrado de la distancia A , entre cada punto y el centroide de cada cluster mediante la siguiente ecuación (Dabbura, 2020)

$$A = \min \sum_{i=1}^k \sum_{x \in k_i} \|x_k - m_i\|^2 \quad \text{Ecuación 3}$$

Donde $x_k = (x_1, x_2, x_3, \dots, x_n)$ son los datos pertenecientes al cluster k_i ; y m_i es el centroide del cluster k_i (Dabbura, 2020)

$$m_i = \frac{\sum_{k=1}^{N_i} x_k}{N_i} * x_k \in k_i \quad \text{Ecuación 4}$$

Donde N_i es el número de datos en el cluster i .

El procedimiento termina cuando no se reasignan puntos de un cluster a otro o cuando se alcanza el número predefinido de iteraciones.

4 Discusión y Resultados

Los mapas temáticos de los parámetros analizados para la estimación de la vulnerabilidad al deslizamiento de los taludes se pueden observar en la Figura 2.

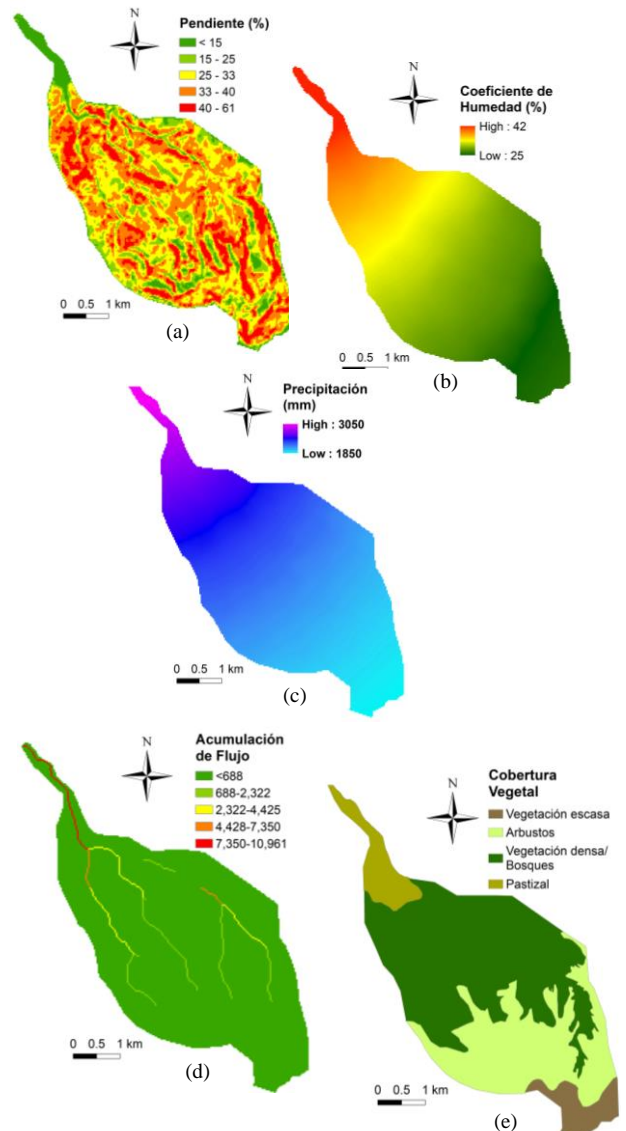


Fig.2 Parámetros utilizados en la determinación de la vulnerabilidad al deslizamiento de taludes por agrupamiento de datos (K-medias). (a) Pendientes, (b) Coeficiente de humedad, (c) Precipitación, (d) Acumulación de flujo, (e) Cobertura vegetal

El algoritmo de K-medias aplicado a los cinco parámetros analizados, arrojó un número de clúster óptimo de tres clústeres (Fig. 3)

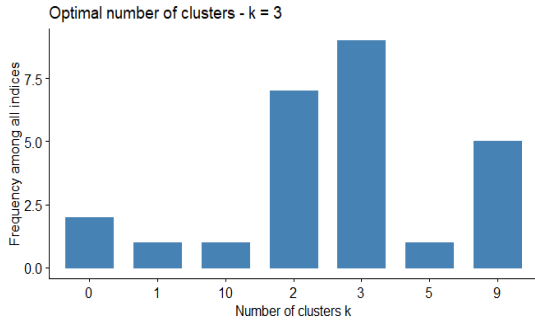


Fig.3 Número de clusters optimo

Los resultados del análisis de agrupamiento de datos con tres clusters, se puede observar en la Tabla 2.

Tabla 2. Rango de variación de los parámetros analizados en cada clúster

Cluster	Puntos	Punto (%)	Area (km2)	Area (%)	V	P (mm)	H (%)	S (%)	F
	829	100	18.6	100			Media		
1	304	37	7.9	42	3	1996	2.7	31	17
2	454	55	8.8	47	1	2250	3	33	105
3	71	9	1.9	10	2	2726	3.7	22	483

La distribución espacial de los clusters y su grado de vulnerabilidad se puede observar en la figura 4.

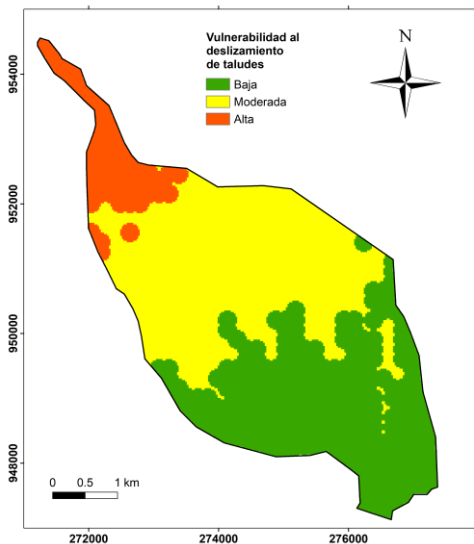


Fig. 4 Vulnerabilidad al deslizamiento por análisis de agrupamiento de datos K-medias

El clúster 1 posee los valores más bajos precipitación (P), acumulación de flujo (F) y humedad de suelo (H); éstas características le confieren al clúster una vulnerabilidad baja al deslizamiento de taludes, a pesar de tener una pendiente del terreno importante, la diferencia de pendientes (S) con el resto de los clústeres es poco significativa. Esta zona se encuentra al sureste del área de estudio y abarca un 42% del

área de estudio. Estos resultados se validan con las características del área, ya que esta zona corresponde con el área de mayor cota en la cuenca Ña Leona, compuesta por rocas de gran calidad (gneises, granitos, esquistos estaurolíticos, cuarzos feldespáticos poco fracturados) y vegetación de páramo alto andino (escaso), lo que facilita el escurrimiento superficial, restringiendo la acumulación de agua en el suelo, haciéndola poca propensa al deslizamiento.

El cluster 2 posee valores elevados de precipitación (P), humedad de suelo (H), pendiente (S) y acumulación de flujo (F), a pesar de no ser los valores más altos de todos los clústers, representan características importantes que facilitan el deslizamiento de los taludes. Por ésta razón, el cluster 2 se corresponde con una vulnerabilidad moderada. Esta zona abarca la mayor parte del área de estudio (47% del área) y está localizada en la zona central, aunque se pueden observar pequeñas franjas hacia el suroeste, que se correlacionan con algunas coronas de deslizamiento observadas en las imágenes satelitales.

El Cluster 3 posee los valores más altos de precipitación (P), humedad de suelo (H), y acumulación de flujo (F), así mismo, la vegetación facilita un poco el deslizamiento y a pesar de no tener valores especialmente elevados de la pendiente (S), esta zona se establece de vulnerabilidad elevada. Esta zona se encuentra hacia la parte baja de la cuenca, ubicada al noroeste del área de estudio abarcando una extensión pequeña del área (10%). Al ser esta zona la parte baja de la cuenca, el material está conformado por algunos sedimentos depositados que forman conos de deyección y abanicos aluviales ayudados por el flujo descendente del agua. La acumulación de flujo en esta zona, verifica el carácter altamente vulnerable al deslizamiento de taludes y en reconocimiento en campo se puede verificar la existencia de cuantiosos deslizamientos en el área.

5 Conclusiones

El análisis de vulnerabilidad asociada a la estabilidad natural de taludes es una herramienta interesante en el diseño preliminar de obras civiles y de prevención de daños ambientales, económicos y sociales. En la presente investigación se determinó a partir de datos de fácil adquisición que contemplan información hidrogeomorfológica los sitios más vulnerables al deslizamiento de taludes, a fin de seleccionar las zonas adecuadas para el emplazamiento de obras civiles, o realizar propuestas de mitigación de en zonas altamente vulnerables. Así mismo, estos resultados, permiten definir las zonas en las que, por su alta vulnerabilidad, requieren un trabajo más exhaustivo de toma de datos y adquisición de parámetros para determinar la estabilidad de los taludes seleccionados con los parámetros y factores de seguridad apropiados. La metodología para la determinación de la vulnerabilidad al deslizamiento de taludes se aplicó a la microcuenca Ña Leona, en el estado Mérida. El procesamiento de los datos se realizó con un novedoso análisis a través de agrupamiento de datos con el algoritmo de

K-medias. La base de datos se obtuvo a partir de cinco parámetros hidrogeomorfológicos seleccionados: Pendiente del terreno (*S*), coeficiente de humedad del suelo (*H*), precipitación (*P*), acumulación de flujo (*F*) y cobertura vegetal (*V*). El análisis de clusters generó tres zonas de vulnerabilidad: baja, moderada y alta. Los resultados mostraron que una pequeña área, localizada en la parte baja de la cuenca (zona norte), posee una alta vulnerabilidad al deslizamiento de taludes, principalmente influenciada por la acumulación de flujos, el coeficiente de almacenamiento del suelo y las altas precipitaciones. En esta zona se han evidenciado numerosos deslizamientos, asociados a conos de deyección y abanicos aluviales, incluida la corona de deslizamiento de la microcuenca Ña Leona, lo que verifica los resultados obtenidos en esta investigación. Así mismo, un área importante localizada en la zona central (parte media de la cuenca), posee una vulnerabilidad moderada. Finalmente, en la zona sur (Parte alta de la cuenca) la vulnerabilidad al deslizamiento resultó baja, concordando con las características geológicas de la zona, compuestas por la Asociación Sierra Nevada, cuya litología está compuesta de rocas sanas de buena calidad (en su mayoría). Los resultados obtenidos, muestran una excelente aplicación de las técnicas de minería de datos en los análisis geotécnicos preliminares, que son de gran ayuda en la obtención de resultados objetivos, confiables y económicos.


Referencias


- Alvarado, M., Cantos, G., Pérez, E., Audemard, F., 2015. Cartografía Neotectónica de la Falla de Boconó entre Tabay y la Toma, Merida, Venezuela. *Boletín de Geología* 37, 47–55.
- ASTER, 2004. Imagen satelital ASTER. Sistema de proyección WGS 84 Banda 1.
- Audemard, Frank, Valladares, R., Aldana, H., Gutiérrez, J., Medina, D., Ferrer, C., 2009. Falla de Boconó. *Atlas de Deformaciones Cuaternarias de Los Andes*. Publicación Geológica Multinacional 7, 259–271.
- BNV, 1980. Fotografías aéreas. Tripletas Misión 010493-171, 172, 173. Escala 1:65.000. Tripletas misión 010473-068, 069, 070 y tripleta de la misión 010486-381, 382 y 383 escala 1:12.500 y 1:5.000 respectivamente.
- Camacho, R., Paredes, A., 2016. Análisis de los Controles Geomorfológicos de la Vertiente Sur del Río Chama: Microcuenca de la Quebrada Ña Leona. Municipio Santos Marquina. Estado Mérida. Universidad de los Andes, Merida, Venezuela.
- CORPOANDES, 2009. Dossier Municipal 2009. Municipio Santos Marquina. Mérida, Venezuela.
- Dabbura, I., 2020. K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks [WWW Document]. Imad Dabbura. Published in Towards Data Science Web Page. URL <https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a> (accessed 5.26.21).
- ERSI, 2019. Cómo funciona Acumulación de flujo—Ayuda | ArcGIS Desktop [WWW Document]. ArcGIS Desktop. URL <https://desktop.arcgis.com/es/arcmap/10.6/tools/spatial-analyst-toolbox/how-flow-accumulation-works.htm> (accessed 9.15.22).
- FCIHS, 2003. Balance de Agua en el Suelo. Cálculo de la Recarga.
- Ferrer, C., 1988. Revisión sobre las características geológicas de un segmento de la zona de la falla de Boconó: Sector Tabay – Estanques Estado Mérida. Venezuela.
- Gonzalez de Juana, C., Iturralde, J., Picard, X., 1980. Geología de Venezuela y sus cuencas petrolíferas, Ediciones Foinves., ed. Caracas, Venezuela.
- Gonzalez de Vallejo, L., Ferrer, M., Ortuno, L., Oteo, C., 2002. Ingeniería Geológica. Pearson. Prentice-Hall, Madrid, España.
- Guerrero, O., Mattie, E., Sanchez, E., Cuevas, R., 2013. Caracterización Geomorfológica y tasa de erosión de los humedales alto-andinos de la cuenca La Mucuy- la Coromoto. *Geominas* 41.
- INAMEH, 2000. Datos de precipitación, Temperatura. Estaciones meteorológicas Mucuy I, II, La Montaña, La Aguada y Loma Redonda. Período 1980-2000.
- Javadi, S., Hashemy, S.M., Mohammadi, K., Howard, K.W.F., Neshat, A., 2017. Classification of aquifer vulnerability using K-means cluster analysis. *Journal of Hydrology* 549, 27–37. <https://doi.org/10.1016/j.jhydrol.2017.03.060>
- Khillar, S., 2020. Difference Between Data Mining Supervised and Unsupervised. Difference Between Similar Terms and Objects. URL <http://www.differencebetween.net/technology/difference-between-data-mining-supervised-and-unsupervised/> (accessed 4.7.22).
- Li, S., Liu, Y., 2005. Data Mining Techniques for Slope Stability Estimation with Probabilistic Neural Networks, in: Li, D., Wang, B. (Eds.), *Artificial Intelligence Applications and Innovations, IFIP — The International Federation for Information Processing*. Springer US, Boston, MA, pp. 491–498. https://doi.org/10.1007/0-387-29295-0_53
- MEM, 1980. Mapa Geológico. Región Mérida. Escala 1:50.000.
- MOP, 1975. Mapa topográfico. Escala 1:25.000. Hoja 5941-I-NE.
- Pérez Márques, M., 2014. Minería de datos a través de ejemplos. RC Libros, Madrid, España.
- Riquelme, J., Roberto, R., Gibert, K., 2006. Minería de Datos: Conceptos y Tendencias. *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial* 10, 11–18.


- Salazar, C., Del Castillo, S., 2018. Fundamentos Básicos de Estadística, 1st ed. Quito, Ecuador.
- Yoo, K., Shukla, S.K., Ahn, J.J., Oh, K., Park, J., 2016. Decision tree-based data mining and rule induction for identifying hydrogeological parameters that influence groundwater pollution sensitivity. *Journal of Cleaner Production* 122, 277–286.
<https://doi.org/10.1016/j.jclepro.2016.01.075>

Recibido: 04 de junio de 2022

Aceptado: 10 de octubre de 2022

Uzcátegui-Salazar, Marisela: Prof. Titular Escuela de Ingeniería Geológica. ULA. PhD. Candidata en Hidrología y Gestión de los Recursos Hídricos. MSc en Desarrollo de los Recursos de Aguas y Tierras. Ingeniero Civil. Hidrogeólogo
 <https://orcid.org/0000-0002-2894-5925>

Toro-Mora, Rosibeth: Especialista en Gestión de Abastecimiento, recolección y tratamiento de agua (CIAT-ULA) y doctorado en Ciencias Geológicas (Zaragoza-España). Profesora en el área de geología de campo y colaborador en Hidrogeología y Geofísica en la Escuela de Ingeniería geológica de la ULA-Mérida.
Correo electrónico: ingrosibethoro@gmail.com
 <https://orcid.org/0000-0002-1252-3584>

Guerrero, Omar: Profesor Titular Escuela de Ingeniería Geológica. ULA. Postgrados y estudios doctorales en ciencias geológicas.
Correo electrónico: omarguerrero1231@gmail.com
 <https://orcid.org/0000-0002-8859-2906>

Hervé, Jégat: Profesor Titular CIDIAT- ULA. PhD. Mecánica de Fluidos. Ingeniero Hidráulico. Profesor Titular Centro Interamericano de Desarrollo e Investigación Ambiental y Territorial. ULA.
Correo electrónico: hjégat@gmail.com
 <https://orcid.org/0000-0003-0185-2873>