

EVALUACIÓN DE LA SUSCEPTIBILIDAD AL DESLIZAMIENTO EN MASA, USANDO REGRESIÓN LOGÍSTICA Y REDES BAYESIANAS EN PAISAJES MONTAÑOSOS DE LA REGIÓN CENTRO NORTE VENEZUELA

María Corina Pineda^A, Alvaro Viloría^B, Jesús Viloría^A

^A Facultad de Agronomía, Instituto de Edafología, Universidad Central de Venezuela, Maracay, Estado Aragua, Venezuela, Email pinedac@agr.ucv.ve, viloriaj@agr.ucv.ve

^B Universidad Central de Venezuela, Facultad de Ciencias, Escuela de Computación, Caracas, Venezuela, Email alvarovb@gmail.com

RESUMEN

Los movimientos en masa son el resultado de la interacción entre variables intrínsecas y activadoras. La variación espacial de las variables intrínsecas determina la distribución geográfica de la susceptibilidad al deslizamiento en masa. En este estudio se relacionó un mapa de cicatrices de deslizamiento con mapas de variables intrínsecas que incluyen variables categóricas, variables morfométricas, distancias a la red de drenaje y la diferencia normalizada del índice de vegetación (NDVI). Tanto la regresión logística (RL) como la red bayesiana (RB) mostraron que los atributos más asociados con la susceptibilidad a los deslizamientos en la cuenca del río Caramacate son: la forma de la pendiente, la distancia a la red de drenaje, el índice topográfico de humedad, el NDVI, el tipo de relieve y la litología. El modelo de RB mostró claramente la interacción entre las variables intrínsecas, mientras que los resultados de la RL permiten representar la distribución espacial de la susceptibilidad a deslizamientos.

Palabra clave: Sistema de Información Geográfica, factores de inestabilidad, análisis de máxima verosimilitud, WEKA, variables morfométricas

INTRODUCCIÓN

Los movimientos en masa son el resultado de la acción conjunta de diferentes factores que pueden agruparse en dos categorías: a) variables intrínsecas, tales como las condiciones geológicas, las propiedades del suelo y la pendiente, y b) variables externas o de activación como las precipitaciones, los sismos y las actividades humanas. La variación espacial de las variables intrínsecas determina la distribución geográfica de la susceptibilidad relativa a los movimientos de masas en una región determinada (Huaibin *et al.* 2005). Por lo tanto, este tipo de fenómeno no ocurre al azar en el paisaje, por el contrario, tiende a seguir una distribución geográfica que responde a una combinación particular de factores de control. En consecuencia, la probabilidad de ocurrencia de deslizamientos puede ser estimada por las relaciones estadísticas entre las cicatrices producidas por los deslizamientos de tierra antiguos y un conjunto de datos espaciales de variables de inestabilidad (Can *et al.* 2005). Diferentes métodos que combinan el análisis estadístico con las herramientas de sistema de información geográfica (SIG) se han aplicado para asociar los factores de inestabilidad o variables intrínsecas a los movimientos en masa. Estos métodos incluyen el uso de regresión logística, por ejemplo: Dai y Lee (2002); Ohlmacher y Davis (2003); Martínez-Casasnovas *et al.* (2004); Can *et al.* (2005) y Greco *et al.* (2007) y los modelos bayesianos

de probabilidad, por ejemplo: Pistocchi *et al.* (2002); Lee *et al.* (2002);.. Ermini *et al.* (2005); Demoulin y Chung (2007); Van Den Eeckhaut *et al.* (2009), entre otras técnicas. El objetivo de este estudio ha sido comparar los resultados de la aplicación de regresión logística y redes bayesianas como medio para identificar las variables intrínsecas que controlan los deslizamientos e inferir las zonas sensibles, en la cuenca del río Caramacate, en Venezuela

METODOLOGÍA

El área de estudio (6760 ha) se encuentra en una zona de rocas metasedimentarias metavolcánicas en la región fisiográfica Cordillera de la Costa Central de Venezuela. Este sector se ha dividido en cuatro unidades litogeomorfológicas: metatobas de "El Chino-El Caño" (OCSCN), metalavas de "El Carmen" (OCSCA), sedimentos aluviales (OCSCQ) y sedimentos coluvio-aluviales (OCSCC) (Pineda, 2008). El relieve es montañoso, con una altitud de 334 a 1405 m sobre el nivel del mar y una pendiente media de 40%. La precipitación media anual es de 1100 mm y la temperatura media anual es de 22 ° C aproximadamente. En la región se presenta un pastoreo extensivo, razón por la cual la cobertura predominante es herbácea, sólo interrumpida por corredores de bosques de galería y parches de bosque perenne en las tierras más altas. El suelo tiende a ser poco profundo, excepto en las zonas que no han sido afectadas por la erosión o la deposición de los sedimentos (Pineda, 2008).

Para evaluar la susceptibilidad a los deslizamientos en el área de se usaron fotografías aéreas a escala 1:25 000, un modelo de elevación digital (DEM) en formato raster con celdas de 20 m de lado, construida a partir de curvas de nivel cada 20 m obtenidas a partir de un mapa topográfico a escala 1: 25 000, una imagen de satélite SPOT-4 y un mapa de unidades litogeomorfológicas y unidades de tipo de relieve (coropléticos). El procedimiento aplicado comprendió las siguientes etapas:

1. Creación de un mapa de deslizamientos actuales, el cual incluyó: a) 214 cicatrices de deslizamiento observados en fotografías aéreas y posteriormente validadas en campo, b) 233 puntos de no deslizamiento seleccionados al azar en lugares situados fuera de una zona de amortiguación de 50 metros alrededor de cada una de las cicatriz de deslizamiento (Dai y Lee 2002). Los puntos correspondientes a cicatrices de deslizamiento se identificaron con uno (1) y los puntos de no-deslizamiento se identificaron como cero (0), para crear una variable binaria llamada erosión en masas (EM). Del total de puntos se tomaron 412 (197 cicatrices de deslizamiento y 215 sitios de no-deslizamiento) para generar el modelo de inferencia y 40 puntos para validarlo. .

2. Producción de mapas temáticos relacionados con las características del terreno y potencialmente relacionados con los deslizamientos (variables intrínsecas) incluyó: a) variables derivadas del mapa geomorfológico tales como los mapas de unidad litogeomorfológica y tipo de relieve (cresta y viga o laderas), b) variables morfométricas calculadas a partir del MDE, que incluyeron: altitud (m), gradiente de pendiente (%), orientación geográfica (radianes), perfil de curvatura (m/m²), plano de curvatura (m/m²), curvagrid o forma de la pendiente (relación entre el plano y perfil de curvatura), área de captación (As- área de drenaje que contribuye a cada celda específica del MDE), y el índice topográfico de humedad calculado como $\ln(As/\tan \beta)$ donde $\tan \beta$ es la pendiente local en grados, c) la distancia a la red de drenaje obtenida por la generación de curvas de contorno

cada 50 metros a partir de las líneas de drenaje (Dai y Lee 2002), d) cobertura vegetal y la diferencia normalizada del índice de vegetación (NDVI) determinada a partir de la imagen de satélite.

3. Por medio de un análisis de componentes principales se identificaron y eliminaron variables redundantes. En consecuencia, sólo las siguientes variables fueron consideradas en el análisis: curvagrid (CD), gradiente de pendiente (GP), la distancia a la red de drenaje (DR), el índice topográfico de humedad (IH), NDVI, unidad litogeomorfológica (N5) y el tipo de relieve (FT).

4. El análisis de regresión logística se realizó entre la erosión en masa (EM), como variable dependiente, y las otras variables temáticas como variables independientes, utilizando el programa SPSS versión 12 (SPSS Inc., Chicago, IL, EE.UU.). Este programa creo variables ficticias para las variables categóricas unidad litogeomorfológica (N5) y tipo de relieve (FT), las cuales fueron codificadas como se muestra en los cuadros 1 y 2, respectivamente. El modelo de regresión fue generado por el análisis de máxima verosimilitud (estadístico de Wald). La probabilidad de deslizamientos fue determinada por medio de la ecuación ($P = 1 / 1 + e^{-\hat{g}}$), donde P es la probabilidad y \hat{g} es la ecuación de regresión logística generada por el modelo. Debido a que el factor tiempo no se toma en cuenta, esta probabilidad debe ser interpretada como una susceptibilidad a los deslizamientos y no como una probabilidad de ocurrencia de deslizamientos (Can *et al.* 2005).

5. El modelo de red bayesiana se generó para representar las relaciones probabilísticas entre los deslizamientos y las variables intrínsecas. Para este fin se utilizaron el algoritmo de búsqueda K2 y el estimador de parámetros simples implementados en el software “Waikato Environment for Knowledge Analysis (WEKA) (Hall *et al.* 2009). Dado que estos algoritmos utilizan variables discretas, las variables seleccionadas fueron clasificadas, como se muestra en el Cuadro 3. Para cada variable fija como una clase se estimó una red bayesiana diferente y la red con mayor poder de predicción fue seleccionada.

6. El poder predictivo de ambos modelos (regresión logística y la red bayesiana) se evaluó comparando los valores observados y predichos usando los datos de validación.

RESULTADOS Y DISCUSIÓN

La regresión logística produjo la siguiente ecuación con un poder de predicción 80,8%:

$$\hat{g} (EM) = - 17,788 - 0,238 (CD) - 0,015 (RD) + 0,15 (IH) - 13,353 (NDVI) - 2,367 (FT)_{(1)} - 21,857 (FT)_{(2)} + 21,39 (N5)_{(1)} + 19,958 (N5)_{(2)} + 21,768 (N5)_{(3)} \quad (1)$$

donde CD = curvagrid o forma de la pendiente, RD = distancia a la red de drenaje, IH= índice topográfico de humedad, NDVI = índice de vegetación de diferencia normalizada. FT₍₁₎, FT₍₂₎ y N5₍₁₎, N5₍₂₎ y N5₍₃₎ corresponden a los códigos de las variables categóricas N5 (unidad litogeomorfológica) y FT (tipo de relieve) como se muestra en los cuadros 2 y 3.

De acuerdo a esa ecuación la ocurrencia de deslizamientos disminuye a medida que los valores de CD, RD, y NDVI aumentan y el valor de IH disminuye. El efecto de las variables de CD, RD y NDVI coincide con los resultados reportados por D'Amato *et al.* (2004) y Federici *et al.* (2006). Sin embargo, según Ohlmacher (2007), el efecto de la variable CD puede variar en función del tipo de erosión en masa considerado. Para las

variables discretas el modelo de regresión indica que la incidencia de los deslizamientos es más alta en la unidad litogeomorfológica "Metatobas de El Chino-El Caño" (N54) y en las laderas (FT3).

Cuadro 1. Codificación de la variable categórica unidad litogeomorfológica (N5).

			Código			Frecuencia
	Descripción		N5 ₍₁₎	N5 ₍₂₎	N5 ₍₃₎	
N51	OCSCA	Metalavas de "El Carmen"	1	0	0	107
N52	OCSCC	Sedimentos coluvio-aluviales	0	1	0	12
N53	OCSCN	Metatobas de "El Chino-El Caño"	0	0	1	292
N54	OCSCQ	Sedimentos aluviales	0	0	0	1

Cuadro 2. Codificación de la variable categórica tipo de relieve (FT)

		Código		Frecuencia
FT	Descripción	FT ₍₁₎	FT ₍₂₎	
FT1	Cresta y viga	1	0	57
FT2	Cresta de piedemonte	0	1	7
FT3	Ladera	0	0	348

Cuadro 3. Categorización de las variables incluidas en la red Bayesiana

CD	GP (%)	IH	N5	RD (m)	FT	NDVI
Cóncava (-4,64 a -0,49)	29,33	Bajo 2,9-4,3	OCS CA	Bajo 50-99	Cresta y viga	Bajo 0,01-0,19
Plano (-0,50 a 0,49)	43,63	Moderadamente bajo 4,4-4,8	OCS CC	Moderadamente bajo 100-149	Cresta de piedemonte	Moderado 0,20-0,39
Convexa (0,50 a 7,58)	56,28	Moderadamente alto 4,9-5,5	OCS CN	Moderadamente alto 150 -199	Ladera	Alto 0,40-0,63
-	2,01	Alto 5,6-13,4	OCS CQ	2 Alto 00 -250	-	-

CD= Forma de la pendiente; GP=gradiente; IH= índice topográfico de humedad N5= unidad litogeomorfológica RD= distancia a la red de drenaje ; FT= tipo de relieve

La Figura 1 muestra el modelo de red bayesiana con el más alto poder de predicción la cual utiliza la variable de FT fija en la raíz de la red. El Cuadro 4 muestra la interacción entre las variables intrínsecas (nodos de la red bayesiana). La probabilidad de ocurrencia de deslizamientos es mayor cerca de las líneas de drenaje (<50 m) en las laderas de la unidad litogeomorfológica "Metatobas de El Chino-El Caño".

La regresión logística clasificó correctamente 80% de los valores observados en los datos de validación, mientras que la red bayesiana clasificó correctamente el 87,5% de estos valores.

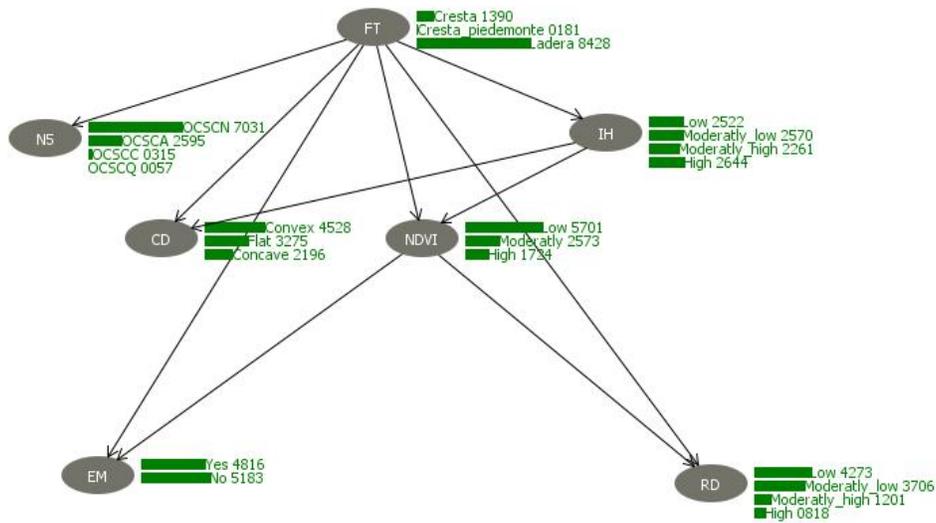


Figura 1: modelo de red bayesiana con la variable de FT fija en la raíz de la red.

Cuadro 4. Resumen de la distribución de probabilidades de la red Bayesiana.

FT	Probabilidad individual para FT	Probabilidad Combinada							
		Clases de N5				RD			
		N51	N52	N53	N54	1	2	3	4
Cresta y viga	0,13	0,36	0,01	0,62	0,01	0,35	0,46	0,09	0,10
Cresta de piedemonte	0,02	0,17	0,28	0,50	0,06	0,28	0,61	0,06	0,06
Laderas	0,85	0,24	0,03	0,73	0,00	0,45	0,34	0,13	0,08

CONCLUSIONES

Tanto el modelo de regresión logística como la red bayesiana mostraron que en el área de estudio, los atributos que más se asocian con la ocurrencia de deslizamientos son: la forma del terreno o curvagrid, la distancia a la red de drenaje, el índice topográfico de humedad, el NDVI (como indicador de la cobertura vegetal), el tipo de relieve y la unidad litogeomorfológica. Los resultados obtenidos de la aplicación de ambas técnicas son complementarios. Por un lado, el modelo de red bayesiana mostró más claramente el efecto de la interacción entre las variables relacionadas con los deslizamientos y por otro el modelo de regresión logística muestra la distribución espacial de la susceptibilidad a los deslizamientos. En general, el área de estudio muestra una sensibilidad alta o muy alta a los deslizamientos. Ambos métodos, la regresión logística y la red bayesiana, no consideraron al gradiente de pendiente como una variable predictora de la susceptibilidad a deslizamientos. Esto obedece a que la casi totalidad de los puntos considerados en el estudio (90%) se encuentran en el mismo rango de pendiente (20%).

BIBLIOGRAFÍA

- CAN T., NEFESLIOGLU H., GOKCEOGLU C., SONMEZ H., DUMAN T. Y. (2005). Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses. *Geomorphology* 72, 250–271.
- DAI FC., LEE CF. (2002). Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology* 42, 213–228.
- D'AMATO AVANZI G., GIANNECCHINI R., PUCCINELLI A. (2004). The influence of the geological and geomorphological settings on shallow landslides. An example in a temperate climate environment: the June 19, 1996 event in northwestern Tuscany (Italia). *Engineering Geology* 73, 215–228.
- DEMOULIN A., CHUNG C. (2007). Mapping landslide susceptibility from small datasets: A case study in the Pays de Herve (E Belgium). *Geomorphology* 89, 391–404.
- ERMINI L., CATANI F., CASAGLI N. (2005). Artificial Neural Networks applied to landslide susceptibility assessment. *Geomorphology* 66, 327–343.
- FEDERICI P.R., PUCCINELLI R.A., CANTARELLI E., CASAROSA N., D'AMATO AVANZI G., FALASCHI F., GIANNECCHINI R., POCHINI A., RIBOLINI A., BOTTAI M., SALVATI N., TESTI C. (2006). Multidisciplinary investigations in evaluating landslide susceptibility An example in the Serchio River valley (Italia). *Quaternary Internacional*. 171-172, 52-63.
- GRECO R., SORRISO-VALVO M., CATALANO E. (2007). Logistic Regression analysis in the evaluation of mass movements susceptibility: The Aspromonte case study, Calabria, Italy. *Engineering Geology* 89, 47–66.
- HALL M., FRANK E., HOLMES G., PFAHRINGER B., REUTEMANN P., WITTEN IH. (2009). The WEKA Data Mining Software: An Update; SIGKDD Explorations, 11 (1).
- HUABIN W., GANGJUN L., WEIYA X., GONGHUI W. (2005). GIS-based landslide hazard assessment: an overview. *Progress in Physical Geography* 29(4), 548–567.
- LEE S., CHOI J., MIN K. (2002) Landslide susceptibility analysis and verification using the Bayesian probability model. *Environmental Geology* 43, 120–131.
- MARTÍNEZ-CASSASNOVAS J.A., RAMOS MC., POESEN J. (2004). Assessment of sidewall erosion in large gullies using multi-temporal dems and logistic regression analysis. *Geomorphology*. 58, 305-321.
- OHLMACHER G.C., DAVIS J. C. (2003). Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Engineering Geology* 69,331–343.
- OHLMACHER G.C. (2007). Plan curvature and landslide probability in regions dominated by earth flows and earth slides. *Engineering Geology* 91 (2-4),117-134
- PINEDA MC. (2008). Determinación de áreas susceptibles a movimientos en masa y erosión laminar en unidades de paisaje del río Caramacate, localizada en la cuenca alta del río Guárico. Trabajo de Ascenso a Asistente. Universidad Central de Venezuela. Facultad de Agronomía. Departamento de Edafología. Maracay, Venezuela. 332p.

- PISTOCCHI A., LUZI L., NAPOLITANO P. (2002). The use of predictive modeling techniques for optimal exploitation of spatial databases: a case study in landslide hazard mapping with expert system-like methods *Environmental Geology* 41, 765–775.
- VAN DEN EECKHAUT M., MOEYERSONS J., NYSSSEN J., ABRAHA A., POESEN J., HAILE M., DECKERS J. (2009). Spatial patterns of old, deep-seated landslides: A case-study in the northern Ethiopian highlands. *Geomorphology* 105, 239–252