

Análisis Espacial de Vecindad Continua y regresión logística en el modelado espacial de probabilidad de ocurrencia de deslizamientos

Application of the Continuum Neighborhood Spatial Analysis and Logistic Regression in the Spatial Modeling of Probability of Occurrence of Landslides

Rutilio Castro Miguel* y Gabriel Legorreta Paulín**

Recibido: 09/08/2018. Aprobado: 05/02/2019. Publicado en línea: 29/03/2019.

Resumen. Este trabajo presenta un análisis comparativo de dos modelos estadísticos de probabilidad de procesos gravitacionales (PG) con la aplicación de regresión logística (RL), utilizando únicamente la variable pendiente del terreno. En un primer modelo se analizó información *in situ* de lugares con deslizamientos y áreas estables; en el segundo se analizó la información de los mismos sitios mediante el Análisis Espacial de Vecindad Continua (AEVC). La precisión que reportaron ambos modelos (*in situ* y AEVC) se evaluó estadísticamente con la medida de ajuste de -2 Logaritmo de la Verosimilitud (-2LL).

Para la calibración de los modelos se utilizó un inventario de deslizamientos y el Continuo de Elevación Mexicano, versión 3.0 (CEM 3.0), del Instituto Nacional de Estadística y Geografía (Inegi).

Los resultados muestran que, al utilizar la información de las áreas de vecindad, se obtiene un mayor nivel de ajuste de la ecuación en comparación con el modelo elaborado utilizando la información *in situ*.

El valor de -2LL para el modelo con datos de vecindad fue 264.312, y para los datos *in situ*, 269.573. Del mismo modo, la tabla de clasificación global del modelo de vecindad reportó 58.5%, mientras que para el análisis *in situ* fue de 51.8%. Lo anterior muestra un aumento de la correcta

clasificación en el modelo estadístico de 6.7% al utilizar el análisis de vecindad.

El área de estudio es la cuenca del río La Ciénega, ubicada en la ladera este del volcán Nevado de Toluca, en el Estado de México.

Palabras clave: Análisis de vecindad, regresión logística, probabilidad, procesos gravitacionales, deslizamientos.

Abstract. Spatial models of probability based on the Logistic Regression (RL) usually collect data for model calibration directly from the location of the sampling site. This data collection method involves the isolation of the site, leading to loss of information, as the neighborhood area is not considered; therefore, the LR model may be less representative of reality.

Aiming to construct spatial models of higher accuracy when using the RL statistical model, this work addresses the analysis and integration of data on independent variables for areas surrounding the sampling sites used for the calibration of the statistical model.

A few works have conducted a statistical evaluation of how the neighborhood areas to calibration sites may yield

* Posgrado en Geografía, Universidad Nacional Autónoma de México, 04510, Ciudad Universitaria, Ciudad de México. Email: rutilio.cm@gmail.com

** Departamento de Geografía Física, Instituto de Geografía, Universidad Nacional Autónoma de México, 04510, Ciudad Universitaria, Ciudad de México. Email: legorretag@igg.unam.mx

a higher relationship with the occurrence of landslides processes, leading to higher precision in the classification of areas based on the probability of occurrence, as compared to *in-situ* data collection at the sampling site. Hence the importance of considering the relationship between the sampling site and its neighborhood area when gathering information for calibrating the probability model.

This paper reports a comparative analysis of two statistical models of probability of occurrence of gravitational processes (PG) involving the application of RL and using terrain slope as the independent variable. A first model analyzed data collected *in situ* on the independent variable from sampling sites with landslides and in stable areas; the second analyzed information for these same sites using Spatial Analysis of Continuum Neighborhood (AEVC) to derive information about the terrain slope variable.

The implementation of AEVC for the elaboration of the statistical model provided information for a detailed assessment of how the area surrounding sampling sites is statistically related to the process studied. The neighborhood area was estimated by using a circular shape centered in the sampling point, the radius of which was increased gradually in 1-pixel increments to 20 pixels.

The data of the terrain slope variable were analyzed separately for the site location (*in situ*) and for each of the neighboring areas, from a distance of 1 to 20 pixels in radio. This approach was used for calibrating the RL statistical models for each distance analyzed, which were then evaluated in statistical terms aiming to identify the model(s) that yield the best classification level.

The precision of *in-situ* and AEVC models was evaluated using -2 Logarithm of Likelihood (-2LL) as a fit measure.

INTRODUCCIÓN

Los PG son resultado del movimiento de materiales no consolidados o bloques de rocas sobre las laderas por efecto de la fuerza de gravedad (Lugo, 2011), principalmente. Debido a la complejidad y los factores involucrados en este tipo de procesos, existe una amplia variedad de clasificaciones basadas en la diversidad de perspectivas con las que se aborda el tema (Alcántara, 2000).

El modelado de los PG se ha realizado desde varios enfoques, lo que ha dado lugar al desarrollo de trabajos con metodologías diversas (determinísticos, heurísticos y estadísticos). Algunos de estos han sido elaborados en la ladera sur de la ciudad de Tuxtla Gutiérrez, Chiapas, y la ladera suroeste del volcán Pico de Orizaba (Legorreta *et al.*, 2014b; Aceves *et al.*, 2016; Legorreta *et al.*, 2017; Paz *et al.*, 2017). Entre los más comunes

This measure facilitates the comparison of two models, where the difference between the values obtained represents the shift in prediction level between models. A lower value of -2LL indicates better goodness of fit of the model; therefore, the size of the neighborhood area analyzed and the value of -2LL were both used for selecting the area for which the terrain slope contributed to better goodness of fit of the probability model.

Models were calibrated using an inventory of landslides, and the terrain slope variable was derived from the *Continuo de Elevación Mexicano* version 3.0 (CMS 3.0). The results show that using data for neighboring areas yields higher goodness of fit of the equation relative to the model developed using *in-situ* data.

The value of -2LL for the model was 264.3 using neighborhood data and 269.5 using *in-situ* data. The table on overall classification reported 58.5 % for the neighborhood model and 51.8 % for the *in-situ* analysis, showing a 6.7 % increase in the classification of the statistical model when the neighborhood analysis is used.

The information used for the selection of the optimal distance for AEVC and the calibration of the statistical model can be depicted spatially; therefore, the results from the LR model can be represented in a map of the distribution of probability of landslides in the study area.

The study area is the La Ciénega river basin located on the eastern slope of the Nevado de Toluca volcano, in the State of Mexico.

Keywords: Neighborhood area, sampling site, probability, gravitational processes, spatial analysis.

se encuentra el análisis multivariante y, en particular, la RL (Ayalew y Yamagishi, 2005; Esper, 2013; Legorreta *et al.*, 2014a; Mirnazari *et al.*, 2014), la cual ha comprobado, en varias ocasiones, ser un método adecuado para la estimación de áreas propensas a estos procesos (Akbari *et al.*, 2014; Legorreta *et al.*, 2014a; Huang *et al.*, 2015).

Existe un gran número de posibles variables que favorecen el desarrollo de los PG en terrenos volcánicos; una de las que más aparecen en los estudios relacionados con el tema es la variable de pendiente del terreno (Nourani y Ghaffari, 2012; Feizizadeh y Blaschke, 2014; Legorreta *et al.*, 2014a; MatebieMeten *et al.*, 2014). En el presente trabajo se utilizó la variable de pendiente del terreno para elaborar modelos parciales de probabilidad a deslizamientos, los cuales permitieron evaluar dos métodos de toma de información.

El estudio de los PG tiene gran importancia, ya que a nivel mundial son uno de los principales peligros (Nourani y Ghaffari, 2012; MatebieMeten *et al.*, 2014), lo que los convierte en una de las causas frecuentes de pérdida de vidas humanas y materiales (Rodríguez *et al.*, 2006; Mirnazari *et al.*, 2014).

El estudio de este tipo de procesos en la cuenca La Ciénega, que se ubica en la ladera este del volcán Nevado de Toluca, toma relevancia debido a que es una zona propensa a PG del tipo de deslizamientos (Aceves *et al.*, 2014b), que en fechas pasadas han causado pérdidas económicas y humanas, como lo sucedido el 28 de junio de 1940, cuando en la zona se presentaron lluvias atípicas en la parte alta de la cuenca, lo que provocó un flujo de material volcánico que arrasó por completo la población de Santa Cruz (Aparicio y Pérez, 2014). El pueblo se edificó nuevamente en el mismo sitio y ahora se llama Santa Cruz Pueblo Nuevo. Esta reubicación no planificada pone en peligro constante el asentamiento humano ante los PG.

El uso de tecnologías informáticas para el manejo y análisis de información geográfica ha tenido una evolución constante desde la década de 1970 hasta la actualidad. Durante estos años se han desarrollado desde métodos para la elaboración de mapas en computadora hasta modelos espaciales de probabilidad mediante la geomática (Berry, 2013). Entre las principales tareas de la geomática se encuentra la obtención de información de campo precisa y de calidad, además de la integración, el análisis, el despliegue y la difusión de la información geográfica (Ghilani y Wolf, 2012). La aplicación de la geomática a estudios geográficos posibilita la implementación de técnicas y métodos de análisis de la información espacial y estadística a los modelos hasta ahora generados.

Generalmente, la toma de información de las variables independientes para la elaboración de modelos de probabilidad a PG, con el uso de la RL, se realiza en el sitio donde se especifica el punto de muestreo. De manera estricta esta forma de recabar la información de las variables aísla los datos coleccionados; al obtener solo lecturas *in situ* significa que el entorno de la toma de información no se está evaluando y, por lo tanto, no se integra

al momento de realizar el modelo de probabilidad. Este aislamiento del entorno provoca una pérdida de información que podría dar, en algunos casos, una mayor precisión al cálculo de modelos probabilísticos de los PG. Sin embargo, poco trabajo se ha hecho para evaluar estadísticamente la manera en que los datos del entorno de los sitios de muestreo pueden tener mayor relación con los PG, en comparación con la toma de información *in situ* del sitio de muestreo.

Entre las técnicas de análisis espacial se encuentra el de vecindad, el cual asocia un sitio con su entorno (Aronoff, 1989); al aplicarlo aumentando el área de vecindad de manera continua, se podrá identificar con precisión cuál es el área que estadísticamente reporta una mayor relación con los deslizamientos (Castro, 2014).

Tobler (1970: 236) menciona que “todo está relacionado con todo, sin embargo, las cosas cercanas están más relacionadas que las cosas lejanas”. Con base en lo anterior se supondría que, tomando la información de alguna variable relacionada con los PG, como el caso de la pendiente del terreno, justo en el lugar donde se ubica el proceso, se adquiriría el dato más relacionado. En caso de que la zona de estudio fuera un área homogénea en la cual la variable analizada presente valores uniformes, la influencia de dicha variable estaría en función únicamente de la distancia, pero ¿cómo afecta una distribución heterogénea de valores de una variable el sitio en que se tome la información? y ¿hasta qué distancia, si la hay, la relación de un sitio con su entorno reporta una mayor relación con el proceso que se analiza?

Cuando los valores de una variable son heterogéneos alrededor de un sitio de muestreo, la influencia que tales datos tienen sobre el sitio en cuestión varía en función de la distancia a la que estos se encuentren en combinación con la diversidad e intensidad de los valores (Figura 1).

La integración de la información de las zonas vecinas a los sitios de muestreo en áreas inestables y estables, y su posterior análisis estadístico con el uso de la RL para evaluar la relación entre las variables dependiente e independiente, proporciona la información necesaria para determinar la distancia óptima a la que las variables

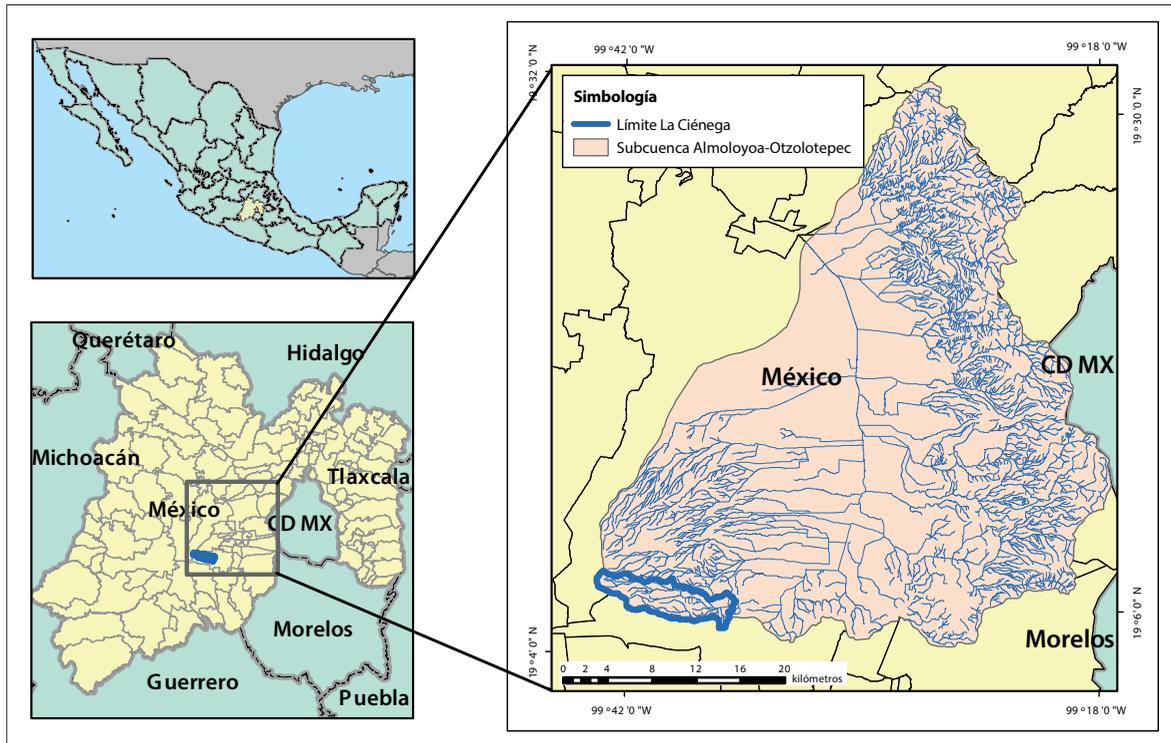


Figura 2. Ubicación de la cuenca del río La Ciénega, en el Estado de México. Fuente: elaboración propia con información del Inegi (2010; 2016).

vecindad (Berry, 1993). En el análisis de vecindad se cuantifican los elementos del entorno de una localidad (sitio de muestreo). El área de vecindad que se analiza puede tener varias formas: circular, cuadrada y rectangular, entre otras (ESRI, 2016). El análisis de los datos dentro del área de vecindad puede ser de varios tipos; entre ellos se cuantifica la variabilidad de valores, la suma de estos y el promedio en toda el área, entre los principales (Aronoff, 1989) (Figura 3).

El análisis espacial de vecindad fue realizado con el programa de Sistema de Información Geográfica (SIG) Arc/info en su versión 10.0.

Para cuantificar la fuerza de la relación que existe entre el sitio de muestreo y su entorno, se realizó una evaluación estadística de los valores de vecindad de la variable pendiente del terreno, la cual se generó a partir del CEM 3.0 (Inegi, 2013b), con la aplicación de la función *slope* del programa ArcGis 10.2.2., la cual utiliza un análisis

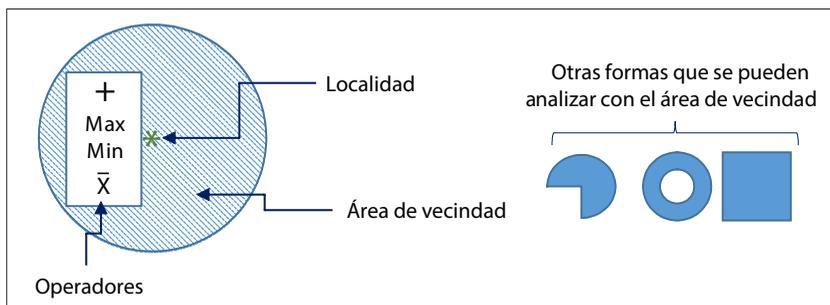


Figura 3. Configuración del análisis de vecindad. Fuente: elaboración propia.

de vecindad de 3 x 3 celdas, y a partir del cálculo de la tasa de cambio en la horizontal y vertical a las ocho celdas vecinas, se determina la pendiente de la celda central (ESRI, 2016). Dado que la forma geométrica más adecuada para el análisis de vecindad es el polígono regular (y, de este, el círculo es el polígono regular ideal) (Fuenzalida *et al.*, 2015), se utilizó una forma circular del área de vecindad para realizar una eficiente selección de las superficies a analizar en función de los sitios de muestreo. Se aplicó un AEVC que se inició con la toma de información del sitio de muestreo, y se aumentó de manera constante hasta una distancia de 20 píxeles, aplicando la función promedio a los valores del área de vecindad. Con ello se obtuvo la información necesaria para evaluar cuál era el tamaño del área de vecindad que reportó un mejor valor para el indicador de ajuste al modelo de RL.

El modelo estadístico de RL es idóneo para trabajar procesos en los que la variable dependiente puede tomar dos posibles valores, que son excluyentes uno del otro (Wayne, 2004). Tal es el caso en los modelados de RL aplicados a PG, en los que la variable dependiente toma los valores de áreas estables e inestables.

A continuación se presentan las ecuaciones de la RL para una sola variable independiente.

$$Y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}} \quad (1)$$

En los casos en que se tienen dos o más variables independientes, la ecuación tiene la siguiente forma:

$$Y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}} \quad (2)$$

Donde:

Y es la probabilidad de que se presente el proceso.

X es la variable independiente que forma parte del modelo.

β_0 es el coeficiente de la constante en el modelo.

β_1 a β_k es el coeficiente de las variables independientes.

La RL utiliza la función de máxima verosimilitud transformándola con el logaritmo neperiano,

esto limita los resultados a valores entre 0 y 1, con el propósito de calcular la pertenencia de la variable dependiente a alguno de los dos grupos posibles (Hair *et al.*, 1999).

Para la evaluación de los modelos generados con el uso de la RL, se utilizan varios indicadores con los cuales se aprecia el nivel de ajuste, entre ellos se encuentra el indicador global de ajuste del modelo -2 veces el logaritmo de la verosimilitud y se representa como -2LL. Este indicador determina el nivel de probabilidad de los resultados en función del modelo obtenido; se calcula contrastando el número de casos observados con los casos esperados al aplicar el modelo (Martín *et al.*, 2008). Cuando un modelo de RL tiene un buen ajuste reporta un valor pequeño para -2LL, donde 0 es el valor mínimo que este puede alcanzar. Una cualidad de tal indicador es que puede ser comparado entre dos ecuaciones, donde la diferencia entre los valores reportados para -2LL representa el cambio en el ajuste predictivo de un modelo a otro (Hair *et al.*, 1999).

Para el presente trabajo se actualizó el inventario de deslizamientos elaborado por Álvarez (2015), que cubre el periodo de 1983 a 2014. Este se basó en la interpretación visual de fotografías aéreas de 1983 en escala 1:37 000 y de imágenes de satélite Landsat 8, de 2014, con una resolución espacial de 15 m para la banda pancromática y de 30 m para las bandas multiespectrales, así como trabajo de campo llevado a cabo de 2012 a 2014 (Álvarez, 2015). En este primer inventario se contabilizaron un total de 154 PG; su actualización se realizó con trabajo de campo durante 2015 y 2016. Se contabilizaron 43 procesos más, lo que en total resulta en 197 PG inventariados en la cuenca La Ciénega en el periodo de 1983 a 2016. Durante los recorridos de campo y la interpretación de fotografías e imágenes se observó que la gran mayoría de deslizamientos identificados se desarrollan a las orillas de los cauces de los ríos (Figura 4); esto se debe a los tipos de materiales que componen las laderas de la cuenca, en su mayoría piroclastos poco consolidados (García *et al.*, 2018) que, aunados a la erosión fluvial, favorecen el desarrollo de los deslizamientos.

Los procesos fueron clasificados en tres tipos de deslizamiento de acuerdo con el Washington State Department of Natural Resources (Departamento

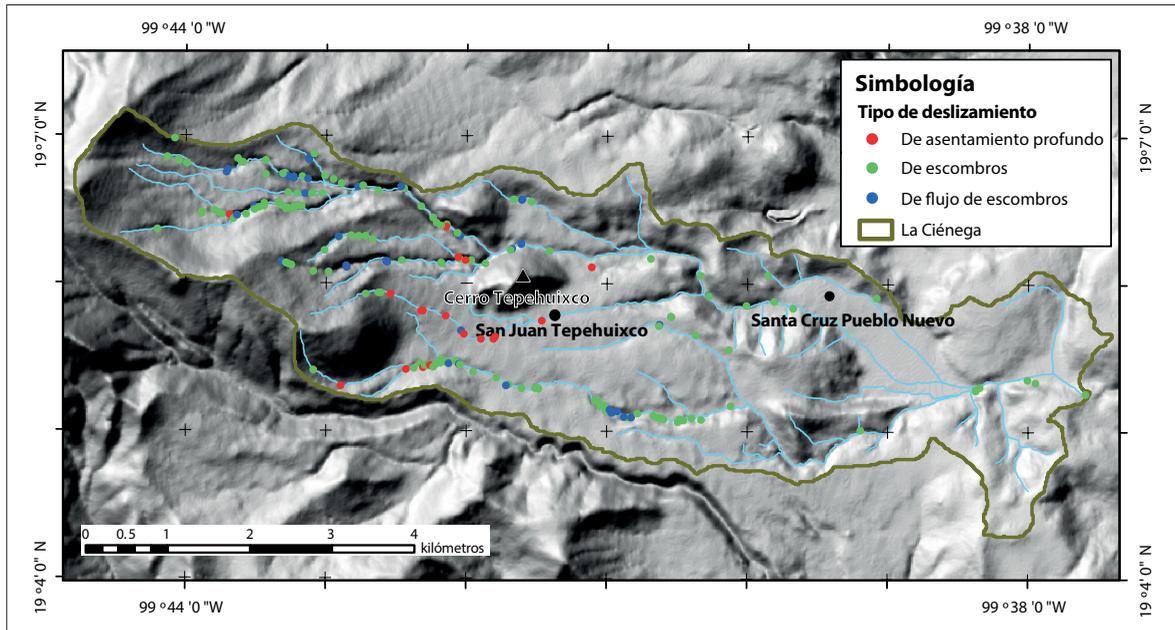


Figura 4. Distribución de los tipos de deslizamientos presentes en la zona de estudio. Fuente: elaboración propia con trabajos de campo e información del Inegi (2013b) y Álvarez (2015).

de Recursos Naturales del Estado de Washington) (WSDNR, 2006).

Dos de ellos se encuentran dentro de la categoría de deslizamientos superficiales, los cuales están definidos por una superficie de falla, generalmente a una profundidad menor a 3 m y sobre el lecho rocoso. Están compuestos por materiales como suelo, regolita, coluvión, aluvión y otros sedimentos, también por depósitos superficiales de baja permeabilidad. El deslizamiento suele ser rápido y de corto tiempo. Los deslizamientos superficiales se subdividen en cuatro tipos; los presentes en la zona de estudio son:

- 1) Deslizamiento de escombros. Se genera de la separación de material en una pendiente pronunciada e involucra un movimiento rápido de suelo y regolita sobre el lecho rocoso.
- 2) Flujo de escombros. Deslizamiento que fluye dentro de un canal formado por las paredes de un río tributario o por un dique producto del mismo deslizamiento, contiene suelo, agua y cantidades variables de residuos leñosos, es caracterizado por un canal de flujo y a menudo

cuenta con un camino largo de salida (WSDNR, 2016).

- 3) Deslizamiento de asentamiento profundo. Se caracteriza porque la mayor parte del plano de deslizamiento o falla está por debajo de la profundidad de las raíces de los árboles, que puede ir desde algunos metros hasta cientos. Pueden estar asociados a depósitos glaciares, regolita profunda, roca erosionada, y/o lecho rocoso, así como a la formación de suelo residual (WSDNR, 2016) (Figura 4).

Para la delimitación de la cuenca La Ciénega y la elaboración del mapa de pendientes del terreno se utilizó el CEM 3.0 (Inegi, 2013b) a una resolución espacial de 15 m por píxel (Figura 5).

Las pendientes del terreno reportadas en la zona de estudio van desde los 0° en la zona centro-este, en donde se encuentra establecido el poblado de Santa Cruz Pueblo Nuevo, hasta los 48° de pendiente, correspondientes al cerro Tepehuixco, ubicado al centro-oeste de la cuenca.

Para el cálculo del tamaño de la muestra se consideraron como individuos de la población

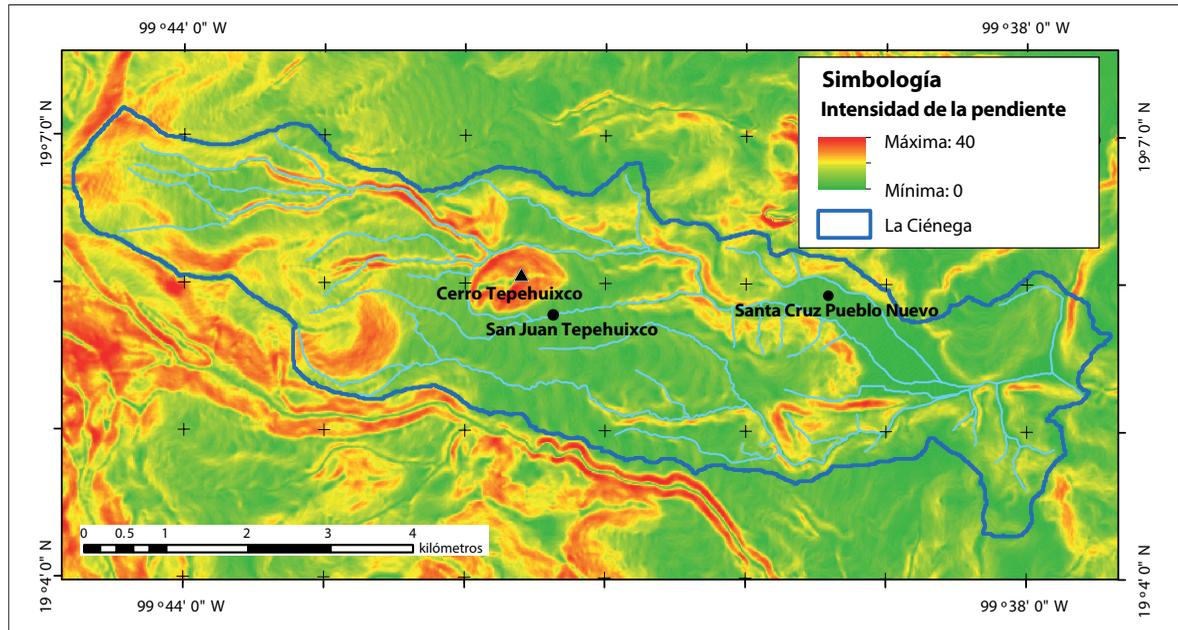


Figura 5. Intensidad de la pendiente en grados. Fuente: elaboración propia con información del Inegi (2013b).

cada uno de los píxeles dentro de la cuenca; se utilizaron los parámetros de nivel de certidumbre de 95% y probabilidad de ocurrencia del proceso de 5%, con lo cual se obtuvo un tamaño de muestra de 196 individuos. Para ello se utilizó la siguiente fórmula (Aguilar-Barojas, 2005):

$$n = \frac{Z^2 pq}{d^2} \quad (3)$$

Donde:

n = Tamaño de la muestra

Z^2 = Coeficiente del nivel de confianza al cuadrado

p = Proporción esperada

$q = 1-p$

d = Margen de error máximo permitido

La selección de la muestra se realizó de manera estratificada y de forma aleatoria (Lincoln, 2002). En total se seleccionó 50% del tamaño de la muestra en zonas estables y el otro 50% en zonas correspondientes a las cabeceras de los deslizamientos.

A partir de la selección de los sitios de muestreo, se realizó el AEVC para toda la muestra. Se inició

con la toma de datos donde se encuentra el punto de muestreo y se continuó con el análisis aumentando de manera continua el tamaño del área de vecindad: se comenzó a un píxel de distancia del sitio de muestreo y se aumentó su tamaño de píxel en píxel hasta llegar a la distancia de 20 píxeles de radio (Figura 6). Esta distancia máxima se determinó con base en estudios previos realizados en la elaboración de modelos de deforestación durante 2011 y 2014 (Castro, 2011; 2014), en los cuales se ha podido determinar una distancia de mayor influencia con el proceso de deforestación dentro de un área de vecindad menor a los 20 píxeles.

La función utilizada dentro del área de vecindad para cuantificar los valores que esta contiene fue el promedio (Figura 7).

El AEVC fue realizado para todos los sitios de la muestra, con ello se obtuvo una matriz de datos en la que, para cada sitio, se asocia el promedio de la intensidad de la pendiente dentro del área de vecindad analizada, es decir, se obtuvieron un total de 21 matrices de datos correspondientes a la ubicación *in situ* hasta el tamaño de 20 píxeles del área de vecindad (Tabla 1).

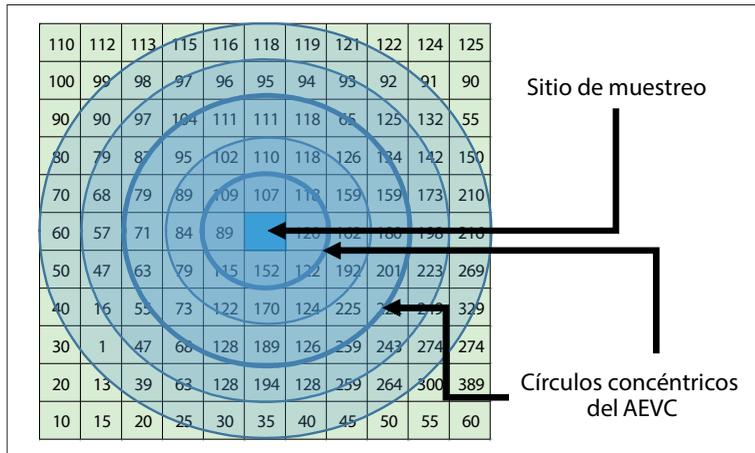


Figura 6. El AEVC se realizó utilizando círculos concéntricos a partir del sitio de muestreo. Fuente: elaboración propia.

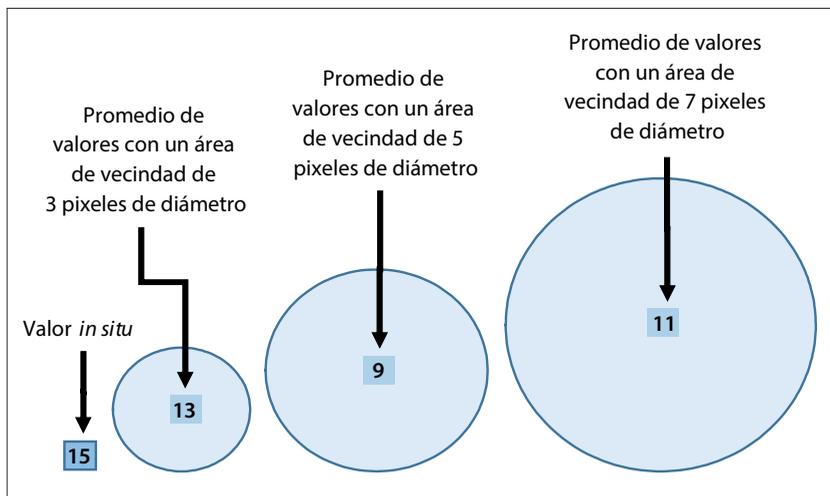


Figura 7. En el lugar de muestreo se realiza la toma del valor de la pendiente del terreno *in situ*, posteriormente, conforme se aplica el AEVC se utiliza la función del promedio de los valores. Fuente: elaboración propia.

Tabla 1. Integración de la matriz con la información de AEVC en función del promedio del valor de la pendiente del terreno (se presenta solo una sección de la matriz).

Registro	Tipo Estable (0) / Deslizamiento (1)	Pendiente <i>in situ</i>	Pendiente promedio AEVC distancia 1	Pendiente promedio AEVC distancia 2	Pendiente promedio AEVC distancia 3	Pendiente promedio AEVC distancia n
1	0	11	12	13	12	11
2	0	16	14	12	10	8
3	1	5	4	4	4	6
4	1	5	10	10	10	9
5	0	4	5	5	5	5
6	1	12	13	12	12	10
7	1	11	12	14	14	14
8	1	18	16	15	15	14

Tabla 1. Continuación.

Registro	Tipo Estable (0) / Deslizamiento (1)	Pendiente <i>in situ</i>	Pendiente promedio AEVC distancia 1	Pendiente promedio AEVC distancia 2	Pendiente promedio AEVC distancia 3	Pendiente promedio AEVC distancia n
9	1	6	8	12	12	15
10	0	7	10	14	14	16
11	0	6	5	6	6	6
12	0	16	11	12	12	12

Fuente: elaboración propia.

La primera columna de la Tabla 1 corresponde a la numeración de los sitios de muestreo (registro); la segunda corresponde a la codificación del sitio de muestreo que se trata (estable = 0, con deslizamiento = 1); la tercera contiene la información referente a la intensidad de la pendiente, inicia con el valor de la pendiente *in situ* (distancia 0); las subsiguientes columnas presentan la aplicación del AEVC a una distancia desde 1 píxel hasta llegar a la distancia máxima de 20 píxeles de radio de área de vecindad.

A cada toma de información, desde el sitio *in situ* hasta el dato correspondiente a los 20 píxeles de áreas de vecindad, junto con la condición del

sitio (columna 2), se les aplicó la RL con una sola variable independiente (ecuación 1). Con ello se obtuvieron una serie de indicadores de ajuste del modelo estadístico, entre los cuales se encuentra el -2LL.

RESULTADOS

En la Tabla 2 se presentan los valores para -2LL al aplicar la RL (ecuación 1) al resultado del AEVC (Tabla 1). El valor de -2LL para cada distancia de análisis de vecindad representa el ajuste del modelo de RL para valores promedio de la pendiente en las

Tabla 2. Valores del indicador -2LL *in situ* y para el AEVC de 1 a 20 píxeles de radio.

Distancia AEVC	-2LL	Distancia AEVC	-2LL
0	269.573	11	265.574
1	268.714	12	266.137
2	267.603	13	266.014
3	266.495	14	266.459
4	265.484	15	266.944
5	264.565	16	267.244
6	264.648	17	267.739
7	264.443	18	268.005
8	264.312	19	267.682
9	264.807	20	267.846
10	265.362		

Fuente: elaboración propia.

21 áreas de vecindad analizadas. Como se observa a la distancia de 8 píxeles de radio, el indicador -2LL reporta el valor más bajo. Dado que el valor más bajo de -2LL representa una mejoría en el ajuste del modelo estadístico, se considera que, con los datos obtenidos a esta distancia, al aplicar el AEVC se obtendrá un mejor modelo probabilístico de ocurrencia de deslizamientos en la zona de estudio.

A partir de la toma de información *in situ*, correspondiente a la distancia 0 hasta llegar a la distancia de 8 píxeles de radio, los valores de -2LL disminuyen de manera casi continua. Después de alcanzar su valor mínimo, y a partir de la distancia de 9 píxeles, los valores de -2LL comienzan a ser mayores hasta llegar a la distancia de 20 píxeles del AEVC.

El dato para -2LL cuando la información es tomada en el lugar en donde se ubica el sitio de muestreo reporta un valor de 269.573, mientras que a una distancia del AEVC de 8 píxeles, el valor de -2LL pasa a 264.312; este último es el más bajo dentro de todo el rango de áreas de vecindad analizadas.

La Figura 8 muestra la gráfica de los valores reportados por el indicador de -2LL, en ella se aprecia cómo, a la distancia de 8 píxeles, se llega al valor mínimo de -2LL. Esto indica que el modelo correspondiente al AEVC de 8 píxeles reporta un mejor ajuste al proceso de deslizamientos, y dado que este valor puede ser comparable entre ecuaciones de RL, señala que, de las 21 ecuaciones, esta es

la que reporta el mejor ajuste global del modelo de probabilidad en la cuenca La Ciénega.

Por último, es importante mencionar que otros indicadores de ajuste del modelo de RL, como el chi-cuadrado y la tabla de contingencia de Hosmer-Lemeshow, mostraron una mejoría en sus valores cuando se analizó el AEVC de 8 píxeles en comparación con los datos *in situ* (Tabla 3).

Como se observa en la Tabla 3, el valor para chi-cuadrado correspondiente al AEVC de 8 píxeles es menor que el reportado cuando se trabaja el dato *in situ*, mientras que la significancia tiene un valor mayor con el AEVC de 8, lo cual refleja un mejor ajuste del modelo a los datos utilizados. La tabla de contingencia a una distancia de 8 píxeles presenta mejores resultados, pues hay una menor discrepancia entre los valores observados y los valores esperados, así como una mayor uniformidad en el número de casos por decil de riesgo (Pardo y Ruíz, 2002).

La tabla de clasificación elaborada con el AEVC de 8 píxeles reporta un nivel de correcta clasificación a nivel global de 58.5%, 6.7%, mayor al valor reportado por la tabla de clasificación de la información tomada en el lugar del sitio de muestreo (Tabla 4).

En la Tabla 5 se muestran los valores de la constante y el coeficiente, resultado de la RL para los análisis *in situ* y del AEVC. Estos valores aplicados en la fórmula 1 completan el modelo estadístico de probabilidad a deslizamientos en la cuenca del río La Ciénega.

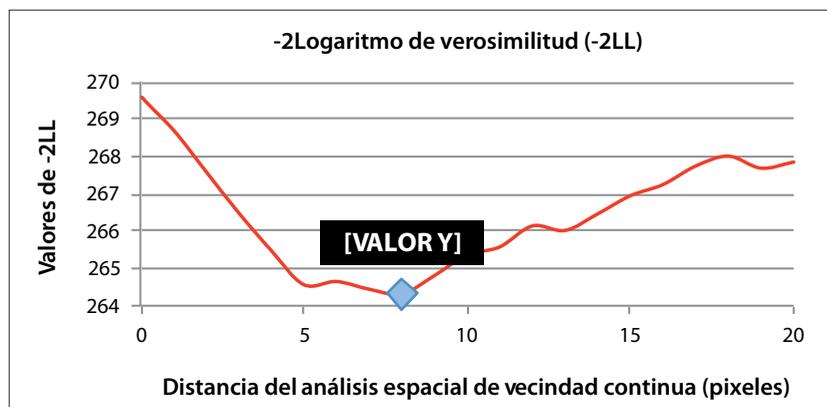


Figura 8. Gráfica del comportamiento de los valores de -2LL. Fuente: elaboración propia.

Tabla 3. Resultados de chi-cuadrado y la tabla de contingencia de Hosmer-Lemeshow para el análisis de datos *in situ* y para el AEVC a 8 píxeles.

<i>In situ</i>					AEVC a 8 píxeles				
Prueba Hosmer-Lemeshow					Prueba Hosmer-Lemeshow				
Paso	chi-cuadrado	gl	Sig.		Paso	chi-cuadrado	gl	Sig.	
1	14.352	8	0.073		1	10.759	8	0.216	

Contingencia Hosmer-Lemeshow						Contingencia Hosmer-Lemeshow							
def = 0			def = 1			Total	def = 0			def = 1			Total
Observado	Esperado		Observado	Esperado			Observado	Esperado		Observado	Esperado		
1	18	12.052	4	9.948	22	1	15	10.321	1	5.679	16		
2	5	10.267	14	8.733	19	2	10	10.476	7	6.524	17		
3	11	13.891	15	12.109	26	3	10	12.409	11	8.591	21		
4	8	7.920	7	7.080	15	4	9	10.245	9	7.755	18		
5	11	11.476	11	10.524	22	5	12	11.264	9	9.736	21		
6	12	10.730	9	10.270	21	6	10	10.140	10	9.860	20		
7	10	8.992	8	9.008	18	7	9	9.674	11	10.326	20		
8	10	9.844	10	10.156	20	8	7	8.697	12	10.303	19		
9	10	9.448	10	10.552	20	9	7	8.794	14	12.206	21		
10	5	5.380	7	6.620	12	10	11	7.980	11	14.020	22		

Fuente: elaboración propia.

Tabla 4. Tablas de clasificación de información *in situ* y del AEVC de 8 píxeles.

Tabla de clasificación <i>in situ</i>					Tabla de clasificación AEVC a 8				
Observado	Predicción				Observado	Predicción			
	def			% correcto		def			% correcto
	0	1				0	1		
0	68	32		68.0	0	66	34		66.0
def 1	62	33		34.7	def 1	47	48		50.5
% Total	El punto de corte es 0.5			51.8	% Total	El punto de corte es 0.5			58.5

Fuente: Elaboración propia.

Tabla 5. Constantes y coeficientes (β) para los modelos *in situ* (a) y a 8 píxeles de AEVC (b).

(a)		(b)	
β	0.012	β	0.049
Constante	-0.218	Constante	-0.739

Fuente: elaboración propia.

CONCLUSIONES

Al integrar la información que rodea los sitios de muestreo, se aporta al análisis de RL información complementaria, lo que permite una mejora en el ajuste del modelo en comparación con un modelo de RL que solo usa información *in situ*. Esto se ve reflejado en la disminución del valor de -2LL. Así, en este estudio, el valor -2LL fue de 269.573 al elaborar la RL con la información de los datos en donde se ubican los sitios muestreados, y decreció a 264.312 al aplicar la RL con los datos del AEVC a una distancia de 8 píxeles (Tabla 2). Esto permitirá crear un modelo de probabilidad de deslizamientos más preciso.

Al realizar el AEVC hasta una distancia de 20 píxeles se aseguró, para este trabajo, incluir las distancias de análisis necesarias para obtener, dentro de este rango, el área de vecindad que tiene una mayor relación, con base en la variable de pendiente del terreno, con el proceso de deslizamientos en la cuenca La Ciénega.

Al realizar el análisis de las áreas de vecindad de una manera continua, con el uso del AEVC se proporciona información constante del promedio de la intensidad de la pendiente, lo que permite establecer de manera precisa cuál es el tamaño del área de vecindad que tiene una mayor relación estadística con el proceso de deslizamientos utilizando la RL.

La disminución del valor obtenido para el indicador de -2LL a una distancia de 8 píxeles de radio, comparado con el valor de -2LL con los datos tomados *in situ* (Tabla 2), muestra estadísticamente cómo, al integrar la información del área de vecindad del sitio de muestreo al modelo de RL, genera una mejora en el ajuste de la ecuación del modelo estadístico. Por consiguiente, el modelo de RL y el uso de los valores del interceptor y los coeficientes β obtenidos para la variable a una distancia de 8 píxeles de vecindad, producirá un mejor modelo de probabilidad a deslizamientos para esta área que el generado con valores *in situ*.

El AEVC puede ser aplicado a prácticamente cualquier variable independiente que se desee incluir en un modelo de regresión logística, con lo que se podría establecer para cada una de ellas

una distancia óptima con relación al proceso que se estudie.

En este trabajo se utilizó la función promedio para evaluar los valores de la variable pendiente del terreno dentro del área de vecindad. Futuros estudios explorarán el análisis aquí expuesto mediante el uso de alguna otra de las funciones que se pueden trabajar con los valores del área de vecindad, lo que podría traer mejoras en el ajuste del modelo de RL.

La distancia a la que el indicador -2LL reporte un valor mínimo dentro del rango del AEVC podrá cambiar de acuerdo a la variable que se analice, a la función de vecindad que se aplique y a la heterogeneidad de los valores que la variable en estudio presente dentro de la zona que se analice; es decir, la distancia óptima del AEVC podría ser distinta de una región a otra, lo que resultaría en la elaboración de modelos a la medida de cada zona de estudio con la aplicación de la misma metodología.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (Conacyt), por la beca económica otorgada para realizar los estudios de doctorado en Geografía, de los cuales forma parte esta investigación; a la Coordinación de Estudios de Posgrado, por el apoyo económico para realizar trabajos de campo durante noviembre de 2017, y al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) #IN10215, por su apoyo para presentar parte de la investigación en la X Reunión Nacional de Geomorfología, en noviembre de 2016.

REFERENCIAS

- Aceves, Q., Legorreta, P. y Álvarez-Ruiz, Y. (2014a). Cartografía geomorfológica para el inventario de procesos gravitacionales en la cuenca endorreica del arroyo La Ciénega, flanco oriental del volcán Nevado de Toluca. *Boletín de la Sociedad Geológica Mexicana*, 66(2), 329-342.
- Aceves, Q., Legorreta, P. y Álvarez-Ruiz, Y. (2014b). Gravitational Processes on the Eastern Flank of the Nevado de Toluca, México. *Zeitschrift für Geomorphologie*, 58(2), 185-200.

- Aceves, Q., Legorreta, P., Lugo, H., Umaña, R. y Legorreta, C. (2016). Sistemas de información geográfica y cartografía geomorfológica aplicados al inventario de deslizamientos y cartografía de susceptibilidad en la cuenca del río El Estado, Pico de Orizaba, México. *Investigaciones Geográficas, Boletín del Instituto de Geografía*, 91, 43-55. DOI: 10.14350/rig.46503
- Aguilar-Barojas, S. (2005). Fórmulas para el cálculo de la muestra en investigaciones de salud. *Salud en Tabasco*, 11(1-2), 333-338.
- Akbari, A., Yahaya, F., Azamirad, M. y Fanodi, M. (2014). Landslide Susceptibility Mapping Using Logistics Regression Analysis and GIS Tools. *Electronic Journal of Geotechnical Engineering*, 19 H, 1687-1696.
- Alcántara, I. (2000). *Landslides: ¿deslizamientos o movimientos del terreno? Definición, clasificaciones y terminología*. *Investigaciones Geográficas, Boletín del Instituto de Geografía*, 41, 7-25. Recuperado de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0188-4611200000100002&lng=es&tlng=es
- Álvarez, Y. R. (2015). *Relación entre los deslizamientos y cambios en la cubierta de uso de suelo y vegetación entre los años 1983 y 2014 en el arroyo la Ciénega, volcán Nevado de Toluca, México*. Tesis de Licenciatura en Biología. Facultad de Ciencias, UNAM, México.
- Aparicio, A. y Pérez, V. (2014). Representaciones sociales del desastre de 1940 en Santa Cruz Pueblo Nuevo, Estado de México. *Investigaciones Geográficas, Boletín del Instituto de Geografía*, 83, 88-101. DOI: 10.14350/rig.36918
- Aronoff, S. (1989). *Geographic Information System: A Management Perspective*. Ottawa: WDL Publications.
- 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*, 65, 15-31. DOI: 10.1016/j.geomorph.2004.06.010
- Berry, J. (1993). *Beyond Mapping: Concepts, Algorithms, and Issues in GIS*. Fort Collins: GIS World.
- Berry, J. (2013). *Beyond Mapping III. Compilation of Beyond Mapping columns appearing in GeoWorld magazine 1996 to 2013*. Recuperado de <http://www.innovativegis.com/basis/MapAnalysis/Default.htm>
- Castro, R. (2011). *Elaboración de un modelo espacial de riesgo de deforestación en bosques*. Tesis de Licenciatura en Geografía. Facultad de Filosofía y Letras, UNAM, México.
- Castro, R. (2014). *Modelo especial de probabilidad a la deforestación en bosques para el estado de Oaxaca*. Tesis de Maestría en Geografía. Posgrado en Geografía, UNAM, México.
- Esper, M. Y. (2013). Debris Flow Susceptibility Mapping in a Portion of the Andes and Preandes of San Juan, Argentina Using Frequency Ratio and Logistic Regression Models. *Earth Sciences Research Journal*, 17(2), 159-167.
- ESRI (Environmental Systems Research Institute) (2016). *Cómo funciona. Estadísticas focalizadas*. Recuperado de <http://desktop.arcgis.com/es/arcmap/10.3/tools/spatial-analyst-toolbox/how-focal-statistics-works.htm>
- Feizizadeh, B. y Blaschke, T. (2014). An Uncertainty and Sensitivity Analysis Approach for GIS-Based Multicriteria Landslide Susceptibility Mapping. *International Journal of Geographical Information Science*, 28(3), 610-638. DOI: 10.1080/13658816.2013.869821
- Fuenzalida, M., Buzai, G. D., Moreno, A. y García, A. (2015). *Geografía, geotecnología y análisis espacial: Tendencias, métodos y aplicaciones*. Santiago de Chile: Triángulo.
- García, R., Legorreta, P. y Aceves, Q. (2018). Susceptibilidad geomorfológica a procesos gravitacionales en arroyos secundarios del sur de la cuenca La Ciénega, volcán Nevado de Toluca. *Investigaciones Geográficas, Boletín del Instituto de Geografía*, 96, 1-14. DOI: 10.14350/rig.59613
- Ghilani, D. y Wolf, P. (2012). *Elementary Surveying. An Introduction to Geomatics*. New Jersey: Prentice Hall-Pearson Education, Inc.
- Hair, J., Anderson, E., Tatham, R. y Black W. (1999). *Análisis multivariante*. Madrid: Pearson.
- Huang, J., Zhou, Q. y Wang, 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. DOI: 10.1080/19475683.2014.992373
- Inegi (Instituto Nacional de Estadística y Geografía) (2010). *Red hidrográfica escala 1:50 000. Documento técnico descriptivo de la red hidrográfica escala 1:50 000. Edición 2.0*. Recuperado de <http://www.inegi.org.mx>
- Inegi (Instituto Nacional de Estadística y Geografía) (2013a). *Censo de Población y Vivienda 2010*. Recuperado de <http://www.inegi.org.mx>
- Inegi (Instituto Nacional de Estadística y Geografía) (2013b). *Continuo de Elevación Mexicano 3.0*. Recuperado de <http://www.inegi.org.mx>
- Inegi (Instituto Nacional de Estadística y Geografía) (2016). *Marco geoestadístico, versión junio de 2016*. Recuperado de <http://www.inegi.org.mx>
- Legorreta, P., Bursik, M., Aceves, F., Paredes, L. M., Hubp, J. L., Ayala, I. y López, J. (2014a). *Temas selectos de geomática: Métodos y aplicaciones. Evaluación práctica y teórica de los modelos de regresión logística múltiple y cartográfica-hidrográfica para la predicción de deslizamientos de tierra*. México: Instituto de Geografía-UNAM.

- Legorreta, P., Bursik, M., Solene, P., Hubp, J. L., Paredes, L. M. y Aceves, F. (2014b). Inventario multitemporal, análisis de susceptibilidad y estimación de volumen de deslizamiento en el flanco SW del volcán Pico de Orizaba, Puebla-Veracruz. *Boletín de la Sociedad Geológica Mexicana*, 66(2), 343-354.
- Legorreta, P., Alanis, A., Arana, S., y Aceves, Q. (2017). Evaluación de la eficiencia de predicción a deslizamientos usando un modelo cartográfico-hidroológico: caso de estudio cuenca La Carbonera, flanco SE del Volcán Pico de Orizaba. *Investigaciones Geográficas, Boletín del Instituto de Geografía*, 95, 1-15. DOI: 10.14350/rig.58387
- Lincoln, C. (2002). *Introducción a la estadística*. México: Editorial Continental.
- Lugo, J. (2011). *Diccionario geomorfológico*. México: Instituto de Geografía-UNAM.
- Macías, J. L. (2005). Geología e historia eruptiva de algunos de los grandes volcanes activos de México. *Boletín de la Sociedad Geológica Mexicana (Volumen conmemorativo del centenario)*, 3, 379-424.
- Martín, Q., Cabero, M. T. y Paz, Y. R. (2008). *Tratamiento estadístico de datos con SPSS*. Madrid: Thomson.
- MatebieMeten, PrakashBhandary, N. y RyuichiYatabe (noviembre de 2014). The Application of Weights of Evidence Modeling for Landslide Susceptibility Mapping of Dejan-GohaTsiyon Transect in the Blue Nile Gorge, Central Ethiopia. Goehazard: Science, Engineering and Management. Simposio internacional llevado a cabo en Kathmandu, Nepal.
- Mirnazari, J., Ahmad, B., Mojaradi, B. y Sattari, F. (2014). Using Frequency Ratio for Spatial Landslide Prediction. *Research Journal of Applied Sciences, Engineering and Technology*, 7(15), 3174 -3180.
- Nourani, V. y Ghaffari, H. (octubre de 2012). Landslide Susceptibility Mapping with Analytical Hierarchical Process and GIS Techniques (Case Study: Zonouz, East Azerbaijan, Iran). 10th International Congress on Advances in Civil Engineering. Congreso llevado a cabo en la Middle East Technical University, Turkey.
- Pardo, A. y Ruíz, M. A. (2002). *SPSS 11*. Madrid: McGraw Hill.
- Paz, T., González, H., Gómez, R. y Velasco, H. (2017). Metodología para la elaboración de mapas de susceptibilidad a procesos de remoción en masa, análisis del caso ladera sur de Tuxtla Gutiérrez, Chiapas. *Investigaciones Geográficas, Boletín del Instituto de Geografía*, 92, 129-143. DOI: 10.14350/rig.52822
- Rodríguez, S., Mora-González, I. y Murrieta, H. (2006). Flujos de baja concentración asociados con lluvias de intensidad extraordinaria en el flanco sur del volcán Pico de Orizaba (Citlaltepétl), México. *Boletín de la Sociedad Geológica Mexicana*, LVIII(2), 223-236.
- Tobler R., W. (1970). A Computer Movie Simulation Urban Growth in the Detroit Region. *Economic Geography*, 46, 234-240. DOI: 10.2307/143141
- Washington State Department of Natural Resources (WSDNR) (2006). *Landslide Hazard Zonation Project Protocol, Version 2.1*. Recuperado de https://www.dnr.wa.gov/publication/fp_lhz_protocol_v2_1_final.pdf?ee5umyf
- Wayne, W. D. (2004): *Bioestadística: Base para el análisis de las ciencias de la salud*. México: Limusa.