



Mapas de susceptibilidad de deslizamientos a partir del modelo de regresión logística en la cuenca del río Oria (Gipuzkoa). Estrategias de tratamiento de variables

Landslide susceptibility maps using logistic regression model for the Oria river catchment (Gipuzkoa Province). Strategies for variables processing

Bornaetxea, T. ⁽¹⁾; Antigüedad, I. ⁽²⁾; Ormaetxea, O. ⁽¹⁾

(1) Departamento de Geografía, Prehistoria y Arqueología, Facultad de Letras, Universidad del País Vasco. c/ Tomás y Valiente, s/n, 01006, Vitoria-Gasteiz. txomin.bornaetxea@ehu.eus

(2) Departamento de Geodinámica de la Universidad del País Vasco UPV/EHU. Barrio Sarriena, s/n 48940, Leioa

Resumen

Este trabajo presenta una investigación en la cuenca del río Oria (882 km²) (Gipuzkoa), en la que se ha realizado un análisis comparativo entre diferentes estrategias de tratamiento de datos de entrada para la confección de mapas de susceptibilidad de deslizamientos utilizando un modelo de regresión logística. Las tres estrategias experimentadas han sido: (i) no se transforma ninguna variable (continua ni categórica); (ii) se transforman las variables continuas en categóricas dividiendo la distribución de los valores en intervalos de igual rango; y (iii) las variables categóricas son convertidas a continuas mediante el cálculo de la densidad de deslizamientos para cada clase. Para los tres modelos así desarrollados, se ha aplicado el mismo inventario de deslizamientos, obtenido a partir de trabajo de campo; y el mismo conjunto de variables seleccionadas a partir de un procedimiento estadístico objetivo. La comparación de los resultados logrados ha servido para determinar la conveniencia de transformar las variables categóricas en continuas para su introducción en el modelo de regresión logística. Sin embargo, el procedimiento estadístico para la selección de variables no ha sido capaz de detectar el sesgo producido por una de las variables explicativas, que ha quedado patente, únicamente, después de generar los mapas de susceptibilidad. Por eso, adicionalmente, se ha producido un cuarto modelo, y su correspondiente mapa, teniendo en cuenta los resultados de los tres primeros modelos.

Palabras clave: mapa de susceptibilidad de deslizamientos; regresión logística; tratamiento de variables explicativas; selección de variables; provincia de Gipuzkoa.



Abstract

We present a study in the Oria River basin (882 km²) (Gipuzkoa), where we carried out a comparative analysis among different input data processing strategies for the study of the susceptibility of landslides using a logistic regression model. The three adopted strategies were: (i) no variable (continuous or categorical) was transformed; (ii) the continuous variables were transformed into categorical by dividing the distribution of the values in equal-range classes; and (iii) the categorical variables were transformed into continuous by calculating the landslide density for each class. Within these three strategies, we considered a landslide inventory obtained from fieldwork and used a common set of variables, selected from an objective statistical procedure. We determined the advantage of transforming categorical variables into continuous by comparing the results of a logistic regression model applied to the different strategies. However, the statistical procedure for the selection of variables was unable to detect the bias produced by one of the explanatory variables, which was highlighted only after generating the susceptibility maps. Therefore, we performed a fourth Logistic Regression model run, and its map, considering the results of the first three.

Key words: landslide susceptibility map; logistic regression; independent variables treatment; variables selection; Gipuzkoa Province.

1. Introducción

De acuerdo con la base de datos internacional “The International Disaster Database” (EM-DAT), en las últimas décadas el número de desastres naturales ha aumentado considerablemente. Tras las inundaciones (44%) y tormentas (35%), los deslizamientos de ladera supusieron el 6% del total de desastres registrados a nivel mundial entre 1970 y 2012 (CRED, 2014) y según este mismo estudio, las pérdidas económicas globales causadas por deslizamientos de ladera para ese periodo ascendieron a 23.900 millones de dólares. Las pérdidas humanas por efecto de estos procesos también son destacables. Como ejemplo, para el periodo 2004-12 y como resultado de deslizamientos no sísmicos (2.620), fueron contabilizados 32.322 fallecidos en todo el mundo (Petley, 2012).

En la provincia de Gipuzkoa los deslizamientos de ladera también son muy frecuentes y, aunque la pérdida de vidas humanas no es significativa (Ayala-Carcedo y Olcina, 2002) sí producen daños en construcciones e infraestructuras (Diario Vasco, 2013) causando importantes pérdidas económicas (Bonachea, 2006). Así, en el mapa de susceptibilidad ELSUS1000 versión I desarrollado por la Comisión Europea para todo el continente

(Günther *et al.*, 2014), este territorio está clasificado como de alta a muy alta susceptibilidad a deslizamientos. Por su parte, a nivel provincial sólo se han realizado, hasta la fecha, dos estudios de susceptibilidad que abarcan el área de estudio (Diputación Foral de Gipuzkoa, 1987; 2007), siendo aplicaciones a menor escala y muy espaciadas en el tiempo. Sin embargo, sí han sido publicados trabajos en áreas específicas de la provincia y cercanas a la cuenca del Oria (Remondo *et al.*, 2008; Bonachea *et al.*, 2009; Felicísimo *et al.*, 2013; Ramírez *et al.*, 2005), aunque con diversas metodologías.

La delimitación de zonas de susceptibilidad de deslizamientos mediante mapas se considera la base para cualquier plan de mitigación y suministra a los planificadores y órganos de decisión la información necesaria (Van Westen, 1993; Van Den Eeckhaut *et al.*, 2012; Nefeslioglu *et al.*, 2008). Tal y como se define en Guzzetti *et al.* (2005), un mapa de susceptibilidad de deslizamiento es “una estimación de dónde se espera que ocurran los deslizamientos y en términos matemáticos se puede definir como la probabilidad espacial de ocurrencia de deslizamientos, considerando un conjunto de condiciones geo-ambientales”.

Existen varios enfoques metodológicos para elaborar estos mapas (Soeters y Van Westen, 1996; Guzzetti *et al.*, 1999; Dai y Lee, 2002; Chung y Fabbri, 2005; Corominas y Mavrouli, 2011). En este estudio, se ha aplicado la regresión logística (en adelante RL), que establece la relación matemática entre un conjunto de variables explicativas y una variable de respuesta dicotómica, como por ejemplo la presencia o ausencia de deslizamientos. De acuerdo con los buenos resultados que diversos autores han obtenido tras utilizar diferentes métodos aplicados al mismo inventario y con las mismas variables (Schicker, 2010; Das *et al.*, 2010; Cascini, 2008; Blais Stevens *et al.*, 2012; Amorim, 2012), se ha considerado este método como óptimo, sobre todo cuando se dispone de un conjunto de datos limitado. Además, entre otras ventajas, la RL ofrece la posibilidad de utilizar variables tanto de carácter categórico como continuo, y a su vez no es estrictamente necesario que las variables presenten una distribución normal (Atkinson y Massari, 1998).

Sin embargo, tal y como se ha constatado en los referentes consultados (Dai y Lee, 2002; Grozavu *et al.*, 2013; Lee, 2005; Duman *et al.*, 2006; Van Den Eeckhaut *et al.*, 2012; Amorim, 2012), esta versatilidad en la utilización de variables permite diferentes opciones a la hora de introducir una misma variable explicativa en el modelo (ver epígrafe 3.4). La misma variable siendo en origen continua, como por ejemplo la pendiente, siempre puede ser transformada a categórica agrupando sus valores en clases. O, por el contrario, una variable, en principio, categórica (p.ej. la litología), podría ser tratada como continua si se asigna a cada clase un valor numérico de acuerdo a algún criterio de ponderación. De modo que, el tipo de tratamiento aplicado a cada variable explicativa ejercerá, lógicamente, un efecto en los resultados del modelo de RL y, en consecuencia, en el mapa de susceptibilidad de deslizamientos. Por ello, a la hora de realizar un mapa de estas características siguiendo el modelo de RL se plantea una primera incertidumbre ante la determinación de cada variable explicativa.

Por otro lado, la elección del conjunto de variables que definirá el modelo de susceptibilidad supone también una segunda incertidumbre. Hoy en día se dispone de una gran cantidad de datos espaciales en formato digital que permiten la confección de numerosas variables que pueden considerarse como factores condicionantes en la implementación del modelo. A pesar de ello, la escala de trabajo, así como el tipo de inventario de deslizamientos utilizado puede condicionar, en este caso, la idoneidad de ciertas variables (Corominas y Mavrouli, 2011), y esto supone la necesidad de asumir una decisión a la hora de decidir qué variables explicativas introducir en la RL.

En consecuencia, este trabajo que tiene por fin último desarrollar un mapa de susceptibilidad de deslizamientos a partir de un modelo de RL para una cuenca hidrográfica de Gipuzkoa (Oria), aborda también esas preguntas relacionadas con las decisiones que se han de tomar inicialmente: *¿cómo se deben tratar las variables explicativas?* y *¿qué variables explicativas se deben introducir en el análisis?* En este marco, se plantean como objetivos para este caso de estudio, la realización de un análisis comparativo entre las diferentes estrategias para el tratamiento de variables explicativas dentro de la RL, y la búsqueda de las variables más significativas relacionadas con la aparición de deslizamientos.

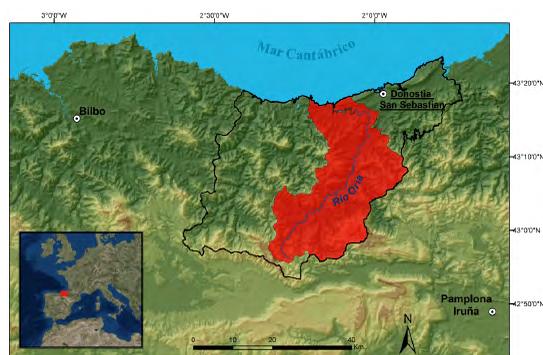


Figura 1: Localización de la cuenca del río Oria (en rojo) mayoritariamente en la provincia de Gipuzkoa (líneas negras).

Figure 1: Location of the Oria river basin (in red) which belongs to the Gipuzkoa Province (black lines).

2. Área de estudio

El 90% de la cuenca de río Oria se ubica dentro de la provincia de Gipuzkoa (Figura 1). Con 65,5 km de longitud, el Oria fluye desde las montañas más altas (Aizkorri 1.528 m y Aralar 1.341 m) hasta la desembocadura con una cuenca vertiente de 882 km² y una densidad de drenaje de 1,6 km·km⁻². Se trata de la mayor cuenca hidrográfica de la provincia, ocupando el 40% de su superficie y con una pendiente superior a 20° en el 60% de la misma (Figura 2 a).

La densidad de población es de 240 hab·km⁻² con una red de infraestructura de transporte muy densa (1,94 km·km⁻²) dadas las características de poblamiento y fisiografía del territorio (Figuras 2 b). Además, en esta cuenca se encuentra uno de los principales ejes industriales de la provincia, el corredor Ordizia-Beasain. Respecto a la ocupación del suelo, las principales coberturas son las plantaciones forestales (36%), los bosques naturales (26%) y los prados (20%) (Figura 2 c).

La cuenca del Oria corresponde a dos unidades geológicas, al sur y al oeste el *Arco Vasco* con materiales sedimentarios marinos mesozoico y cenozoicos, y al nordeste el *Macizo Paleozoico de Cinco Villas* (EVE, 2010) (Figura 2 d). En la cabecera de la cuenca dominan las margas, calizas y calcarenitas con algunos afloramientos de rocas detríticas alternantes. En la parte media aparecen estratos de margas descarbonatadas, calizas impuras, ofitas, arcillas, yesos y materiales detríticos de grano grueso. Al este afloran las pizarras del afloramiento Paleozoico y al norte y tramo final se alternan las margas, calizas impuras, calcarenitas, margocalizas y rocas detríticas de grano grueso (Campos y García-Dueñas, 1972; Campos *et al.*, 1983).

En este ámbito atlántico se registra una precipitación media anual de 1.597 mm (González-Hidalgo *et al.*, 2011) con un mayor cómputo en invierno (30,4%) y con cuatro meses en los que se superan de promedio 150 mm (noviembre, diciembre, enero y abril). Así mismo,

al alto valor de precipitación hay que añadir la persistencia de la lluvia diaria como otra de sus características (Fernández-Arroyabe y Martín-Vide, 2012). Esa cuantiosa y persistente precipitación es el principal mecanismo desencadenante de procesos de ladera en este territorio, favorecido en buena parte de ellos por las modificaciones que desmontes y actividades extractivas forestales realizan (Bonachea, 2006).

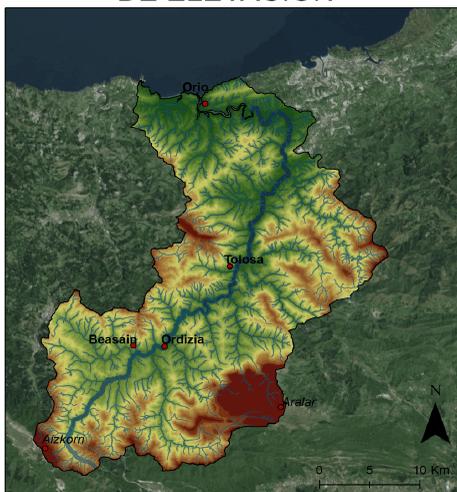
3. Material y métodos

3.1. Inventario de deslizamientos

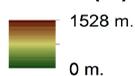
El mapa de inventario de deslizamientos se considera un paso fundamental en cualquier modelo de susceptibilidad (Soeters y Van Westen, 1996) y los métodos estadísticos dependen de su precisión, ya que los valores de las variables explicativas son extraídos atendiendo a su localización. Además, el número de observaciones debe ser lo suficientemente amplio como para cubrir la variabilidad de los parámetros categóricos y continuos en el área de estudio; pero no demasiado extenso para evitar la auto-correlación espacial de los parámetros de entrada, lo que violaría la suposición de independencia de las observaciones como indican diversos autores (Van Den Eeckhaut *et al.*, 2006; Trigila *et al.*, 2015).

En este trabajo, se ha recurrido a la documentación cartográfica y bibliográfica con el fin de obtener un inventario de deslizamientos adecuado a los objetivos de la investigación. Así, se han recopilado bases de datos procedentes del (i) Mapa Geomorfológico de Euskadi (IDE de Euskadi), (ii) Inventario y Análisis de las Áreas sometidas a Riesgo de Inestabilidades del Terreno de la Comunidad Autónoma del País Vasco (CAPV) (INGEMISA, 1996) y (iii) Evaluación y gestión integrada de riesgos geotécnicos en la red de carreteras de Gipuzkoa (Diputación Foral de Gipuzkoa, 2013). La primera fuente ofrece tanto la localización como las dimensiones (polígonos) de numerosas inestabilidades de ladera, pero el mapa presenta grandes extensiones en las cuales

a) MODELO DIGITAL DE ELEVACIÓN

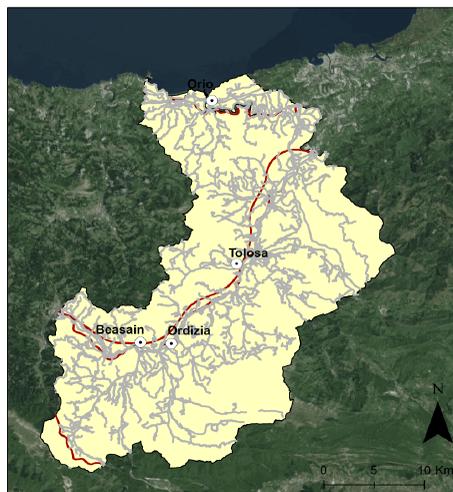


Altitud (m)



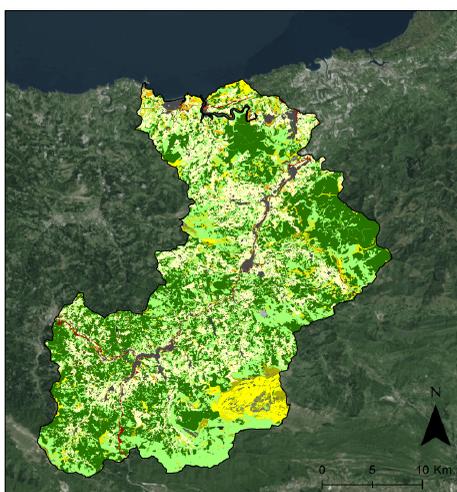
— Red de drenaje
— Río Oría

b) RED DE TRANSPORTES



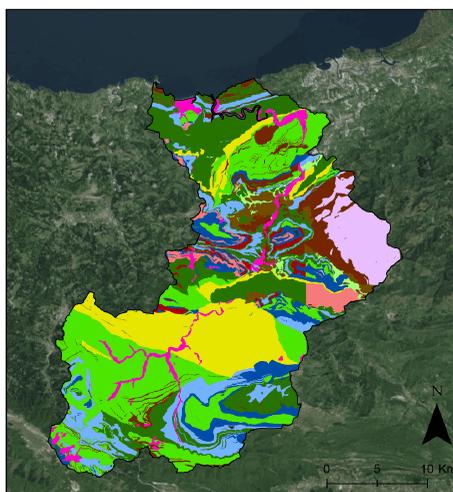
— Carreteras principales
— Vías ferreas

c) USO DEL SUELO



■ Acantilados marinos	■ Urbano
■ Arbustados	■ Mosaico agrícola con artificial
■ Bosque	■ Pastizal-Matorral
■ Bosque de Galería	■ Prados
■ Bosque de Plantación	■ Prados con setos
■ Cultivos	■ Transportes
■ Herbazal-Pastizal	

d) LITOLOGÍA



■ Depósitos superficiales	■ Calizas impuras y calcarenitas
■ Calizas	■ Rocas detriticas de grano grueso (Areniscas). Dominante
■ Ofitas	■ Rocas detriticas de grano fino (Lutitas). Dominante
■ Detriticos alternantes	■ Arcillas con yesos y otras sales
■ Pizarras	■ Alternancia de margocalizas, margas calizas y calcarenitas
■ Margas	■ Margas descarboxatadas

Figura 2: a) Modelo digital de elevaciones, b) Mapa de la red de transportes, c) Mapa de usos del suelo y d) Mapa litológico de la cuenca del río Oría. Fuente: IDE de Euskadi.

Figure 2: a) Digital elevation model, b) Map of the transport network, c) Land use map and d) Lithological map of the Oría river basin. Source: IDE de Euskadi.

hay ausencia de datos. La segunda fuente presenta las coordenadas de las inestabilidades del terreno (puntos) obtenidas por revisión bibliográfica, sobre todo, de mapas geomorfológicos de los municipios, a escala 1:5000 y efectuados en los años 80 (solamente realizados para ciertos municipios). La tercera fuente sí que abarca todo el conjunto del área de estudio, aunque se limita únicamente a las áreas próximas a las carreteras. Además, comprobaciones aleatorias en campo de los puntos de inestabilidad procedentes de las dos primeras fuentes citadas, han mostrado considerables incongruencias entre la geolocalización en el inventario y lo observado sobre el terreno, y por esa razón, se ha optado por la realización de un inventario propio a partir de trabajo de campo.

De esta manera se han inventariado 325 puntos inestables correspondientes a deslizamientos superficiales de tipo rotacional y traslacional (Varnes, 1958). Para cada uno se ha completado una hoja de registro de campo especificando el tipo de movimiento, el estado de actividad y se han tomado fotografías del mismo. También se ha geolocalizado la parte superior y no movilizada del escarpe principal, verificando y corrigiendo las coordenadas de cada punto con la aplicación de Google Earth. La localización de estos puntos en las capas temáticas espaciales ha permitido conocer las características del ámbito anteriores al deslizamiento (Wang *et al.*, 2015). Posteriormente, se ha creado una capa de puntos de distribución espacial con la ayuda del software ArcGIS 10.0. Y finalmente, han sido seleccionados aleatoriamente el 20% de los puntos inventariados en el campo (65 puntos) y han sido separados para el proceso de validación. El resto del inventario (260 puntos) se ha utilizado para la calibración y posterior elaboración de los modelos (Figura 3). Dada la limitada extensión del inventario de deslizamientos disponible, se ha optado por esta división (80% y 20%) para poder asegurar la máxima cantidad de datos en la calibración del modelo estadístico sin comprometer la disponibilidad de números suficiente para su validación (Nefeslioglu *et al.*, 2011).

Para la consecución del trabajo, debido a que la RL requiere un conjunto de datos compuesto por un número casi equilibrado de casos positivos (inestables) y negativos (estable) (Felicísimo *et al.*, 2013; Costanzo *et al.*, 2014), se han seleccionado de forma aleatoria 260 puntos estables utilizando un esquema de muestreo espacialmente uniforme. Además, con el fin de minimizar el impacto del tamaño de los deslizamientos (Dai y Lee, 2002), se ha excluido un buffer de 30 m alrededor de cada punto de inestabilidad para evitar que los puntos estables caigan dentro de las áreas deslizadas. Así, se han considerado aquellos lugares sin ningún deslizamiento de tierra inventariado como espacios libres de inestabilidades (Nefeslioglu *et al.*, 2008; Pourghasemi *et al.*, 2013; Wang *et al.*, 2015).

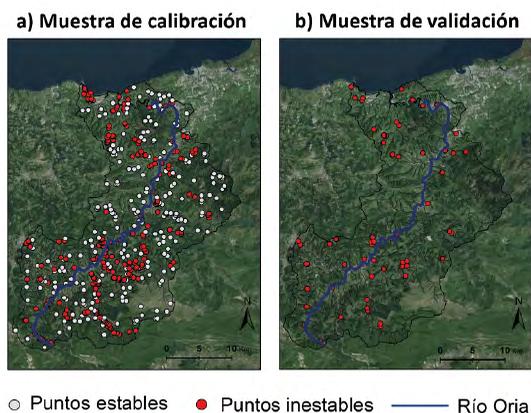


Figura 3: Distribución espacial de las muestras de a) calibración y b) validación.

Figure 3: Spatial distribution of a) calibration and b) validation samples.

3.2. Variables independientes

Siguiendo las propuestas de otros autores (Budimir *et al.*, 2015) se han recopilado datos espaciales temáticos para seleccionar la información ambiental relacionada con el desencadenamiento de deslizamientos (Tabla 1). Estas variables pueden ser continuas, lo que significa que cada píxel de la capa tiene un valor numérico entre el mínimo y máximo del rango de la variable, o categóricas, lo que significa que cada píxel pertenece a una categoría entre las diferentes clases de la variable.

Tabla 1: Descripción de variables consideradas como posibles variables explicativas y resultados de las pruebas estadísticas de Kolmogorov Smirnov (K-S), Mann Whitney y Chi² (nivel de significancia 0.05). En negrita las variables que presentan alta correlación con alguna otra variable y señalados con * las variables que no presentan significancia estadística con la variable dependiente.

*Table 1: Description of variables considered as possible explanatory variables and results of the statistical tests of Kolmogorov Smirnov (K-S), Mann Whitney and Chi² (significance level 0.05). In bold the variables that have a high correlation with some other variable and highlighted with *, variables that do not have any statistical significance with the dependent variable.*

Nombre	Descripción	K-S	Mann-Whitney	Chi2
Continua				
<i>Altitud</i>	La altitud en metros sobre el nivel del mar de cada píxel de 5x5 m. Los datos han sido recogidos por la tecnología LIDAR durante 2013.	0	0	
Pendiente	El gradiente de pendiente en grados derivado del modelo digital de elevación utilizando la función Slope de la herramienta ArcGIS Spatial Analyst.	0	0	
Pendiente senoidal	La función matemática senoidal aplicada a la variable de pendiente (Amorim, 2012).	0	0	
Rugosidad	La relación entre el volumen teórico y la superficie de cada píxel (fuente ArcGIS).	0	0	
<i>Índice de humedad del terreno (TWI)</i>	La distribución espacial de la humedad/saturación del suelo (Yilmaz, 2009).	0	0,01	
* <i>Curvatura</i>	La variación espacial del gradiente de pendiente (fuente ArcGIS).	0	0,41	
* <i>Curvatura planar</i>	Curvatura de la superficie perpendicular a la dirección de la pendiente máxima (fuente ArcGIS).	0	0,26	
* <i>Curvatura de perfil</i>	Curvatura de la superficie en la dirección de la pendiente máxima (fuente ArcGIS).	0	0,34	
Distancia al cauce	La distancia euclídea en metros de cada píxel a la corriente de agua más cercana (Devkota <i>et al.</i> , 2013).	0	0,01	
Distancia a la red de transportes	La distancia euclidiana en metros de cada píxel a la carretera o vía férrea más cercana.	0	0	
<i>Índice diferencial de vegetación normalizada (NDVI)</i>	Se ha calculado mediante las imágenes fusionadas del satélite SPOT 5 (Rouse Jr <i>et al.</i> , 1974).	0	0	
Categoría				
<i>Litología</i>	Las categorías originales han sido reclasificadas por criterios expertos (fuente IDE de Euskadi).			0
* <i>Pemeabilidad</i>	Las categorías originales han sido reclasificadas por criterios expertos (fuente IDE de Euskadi).			0,09
<i>Espesor del regolito</i>	La capa correspondiente al área de estudio se ha obtenido del mapa litológico (fuente IDE de Euskadi).			0,01
Uso del suelo 1	Desarrollado a partir del inventario forestal. Las categorías originales han sido reclasificadas por criterios expertos (fuente IDE de Euskadi).			0
Uso del suelo 2	Se trata de la capa vectorial de la Base Topográfica Armonizada (BTA) (fuente IDE de Euskadi).			0
Distancia al cauce CAT	Es la reclasificación de la variable continua homónima.			0
Distancia a la red de transportes CAT	Es la reclasificación de la variable continua homónima.			0
<i>Orientación</i>	Representa la dirección de la pendiente, medida en grados desde 0 grados (hacia el norte) hasta 360 grados y clasificada en 9 clases (fuente ArcGIS).			0
<i>Orientación N-S</i>	Es la reclasificación de la variable orientación en norte, sur o plano.			0
* <i>Orientación E-O</i>	Es la reclasificación de la variable orientación en este, oeste o plano.			0,52
* <i>Concavidad convexidad</i>	Es la reclasificación de la variable curvatura en cóncava, convexa o plana.			0,93

Desde la Infraestructura de Datos Espaciales de Euskadi (IDE de Euskadi) que recopila la cartografía digital básica, fotográfica y geocientífica de la Comunidad Autónoma del País Vasco, se han descargado las capas vectoriales de *Litología*, *Permeabilidad*, *Espeores del regolito*, *Usos del suelo* (dos capas provenientes de dos fuentes diferentes), *Red de drenaje* y *Red de transportes por carretera y tren*. Así mismo, se ha utilizado el Modelo Digital de Elevaciones (*DEM*) con resolución espacial de 5x5 m para la producción de variables espaciales tales como *Pendiente*, *Pendiente senoidal*, *Orientación*, *Rugosidad*, *Índice topográfico de humedad (TWI)*, *Curvatura*, *Curvatura planar*, *Curvatura de perfil*, *Distancia al cauce* y *Distancia a la red de transportes*. Y para la producción del *Índice diferencial de vegetación normalizado (NDVI)* se han utilizado imágenes multi-espectrales del satélite SPOT 5 correspondientes al verano de 2013.

Teniendo en cuenta la información espacial disponible y las diversas formas de presentación de variables encontradas en la bibliografía, para este trabajo se ha considerado un conjunto de 22 variables explicativas que se describen en la tabla 1.

Las variables a considerar en la RL han sido tratadas con técnicas de estadística descriptiva. Para las variables categóricas se ha aplicado la prueba de χ^2 con un nivel de significancia de 0,05. En este caso, aquellas clases en las que se hayan computado menos de 5 observaciones han sido eliminadas para evitar posibles sobreestimaciones (Rana y Singhal, 2015). En el caso de las variables continuas, primero se ha verificado su distribución mediante la prueba de Kolmogorov-Smirnov. Si la variable presenta una distribución normal, se ha realizado la prueba de T Student. Y si la variable no está distribuida de manera normal, se ha realizado la prueba de Mann-Whitney (De Winter y Dodou, 2010). En consecuencia, un valor de significancia superior al 0,05 significa que la variable independiente no presenta una diferencia estadística significativa entre los puntos estables e inestables, por lo que se debería descartar su utilización en la RL.

Una vez comprobada la diferencia estadística entre los dos grupos estables e inestables, el siguiente paso ha consistido en verificar la independencia de las variables. Una de las condiciones básicas de la RL es la independencia de las variables, porque si no, estaríamos introduciendo información redundante en la regresión; por consiguiente, los resultados estarían sesgados. Con este objetivo, se ha calculado una tabla de correlación en la que las relaciones se han verificado con el estadístico de Spearman. En este caso, la correlación entre dos variables se considera demasiado alta si el valor de significancia es inferior a 0,01 y el coeficiente de correlación es superior a 0,5 (Pardo y Ruiz, 2002). Si se cumplen estas dos condiciones, las dos variables están demasiado relacionadas y, en términos estadísticos, proporcionan la misma información.

Todo este procedimiento estadístico ofrecerá diferentes combinaciones de variables explicativas, de las cuales, se considerará como óptima, aquella que alcance el mayor índice de acierto general en la RL.

3.3. Regresión logística

La RL, al tratarse de un método cuantitativo y multivariante, resulta especialmente apropiado para los casos en los que se desea predecir la presencia o ausencia de una característica o resultado según los valores de un conjunto de indicadores. La siguiente ecuación (1) formula la relación matemática subyacente entre la variable dicotómica dependiente y las variables explicativas.

$$Z = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n \quad (1)$$

En este estudio del Oria, Z representaría la variable dicotómica dependiente que expresa la presencia (representado con el valor 1) o ausencia (representado con el valor 0) de deslizamientos, β_0 sería el intercepto y $\beta_1 \dots \beta_n$ los coeficientes estimados que miden la contribución de las variables independientes $X_1 \dots X_n$. Además, mediante una función de enlace (ecuación 2) se calcula la P , probabilidad de

$Z = 1$, que en este caso representaría la probabilidad estimada de ocurrencia de un deslizamiento superficial (Lee, 2005; Yilmaz, 2009; Kumar y Anbalagan, 2015).

$$P = \frac{1}{1 + \exp^{-Z}} \quad (2)$$

Por medio del paquete estadístico SPSS XXII se han calculado los valores que representan la influencia relativa de cada variable independiente con respecto a la ocurrencia de deslizamientos (Tabla 2). A continuación, estos valores han sido asignados a las capas digitales en formato ráster de cada variable independiente utilizando para ello el software ArcGIS 10.0. Finalmente, los mapas de susceptibilidad de deslizamientos han sido calculados siguiendo la expresión matemática de la ecuación (2) (Trigila *et al.*, 2015).

3.4. Descripción de los modelos

En la literatura se pueden encontrar diferentes formas de tratamiento de las variables explicativas dentro de la RL. Dai y Lee (2002) reclasifican todas las variables en diferentes clases y las tratan como variables categóricas utilizando la denominada codificación de variables de confusión (dummy variables) (Pardo y Ruiz, 2002). Otras propuestas, como la de Grozavu *et al.* (2013), dan un valor numérico a cada clase de variables categóricas en base a su densidad de deslizamientos y después todas las variables son tratadas como variables continuas. Por otra parte, autores como Lee (2005), Duman *et al.* (2006), Van Den Eeckhaut *et al.* (2012) o Amorim (2012) mantienen cada variable sin transformación alguna, tratando las variables continuas con todos sus valores numéricos, y aplicando la estrategia de codificación de variables de confusión a las categóricas.

Sobre estas premisas metodológicas en este trabajo se han preparado cuatro modelos de susceptibilidad de deslizamientos. En tres de ellos únicamente varía la estrategia de tratamiento de las variables explicativas, utilizando el mismo conjunto de puntos estables e

inestables y las mismas variables explicativas (elegidas en la fase de selección de variables). Para el cuarto modelo, las variables explicativas han sido seleccionadas teniendo en cuenta los resultados de los tres primeros modelos y siguiendo el criterio experto. A continuación, se detallan las estrategias aplicadas para cada modelo.

3.4.1. Modelo I

El modelo I ha sido calculado transformando las variables categóricas en variables de confusión de código binario tomando como referencia siempre la primera clase. Más específicamente, para cada una de estas variables, las k categorías han sido reemplazadas por $k-1$ variables dummy, cada una con valores de 1 y 0, lo que indica la presencia o ausencia de una de las categorías $k-1$ (Van Den Eeckhaut *et al.*, 2012). Se trata de un procedimiento automático que proporciona el software SPSS XXII como alternativa para posibilitar la introducción de variables de tipo categórico en la RL (Pardo y Ruiz, 2002). Esto quiere decir que, en este caso, cada clase de las variables categóricas obtendrá su propio coeficiente β . Por otro lado, las variables continuas no han sido transformadas.

3.4.2. Modelo II

En el caso del modelo II, las variables continuas han sido reclasificadas en el SIG dividiendo cada una en 5 categorías de igual rango de valor (Dai y Lee, 2002). Luego se ha realizado la RL aplicando la transformación a variables dummy para todo el conjunto de variables.

3.4.3. Modelo III

En el modelo III, se han calculado las densidades de deslizamientos para las clases de cada variable categórica y estos valores han sido utilizados como valores numéricos relativos para su transformación en variables continuas (Zhu y Huang, 2006; Grozavu *et al.*, 2013; Trigila *et al.*, 2015). Esta metodología

evita la creación de un número excesivo de variables dummy, pero requiere el paso anterior del cálculo de las densidades de deslizamientos. Siguiendo a Bai *et al.* (2010), se ha aplicado la siguiente fórmula:

$$LD_i = \frac{LA_i/A_i}{LA/A} \quad (3)$$

donde LD_i es el valor de la densidad de deslizamientos para la clase i , LA_i y A_i son el área de deslizamientos de tierra en la clase i y el área total de la clase i , respectivamente; y LA y A representan la superficie total de deslizamientos en el área de estudio y superficie total del área de estudio (Grozavu *et al.*, 2013).

3.4.4. Modelo IV

El modelo IV ha sido calculado siguiendo el mismo tratamiento de las variables del modelo III. Sin embargo, en este caso, el conjunto de variables introducido en la RL ha sido modificado por criterio experto tomando como referencia los resultados estadísticos de la fase de selección de las variables, así como los resultados de los modelos I, II y III.

3.5. Validación de modelos

Para la evaluación de los resultados de cada modelo se ha utilizado una muestra de deslizamientos diferente a la utilizada para su elaboración, denominada muestra de validación (Figura 3). Mediante el cruce de esta muestra con los resultados de cada mapa de susceptibilidad, se han calculado las curvas de porcentaje acumulado (Duman *et al.*, 2006) y las curvas de tasa de predicción (Chung y Fabbri, 2003) con sus correspondientes valores del área bajo la curva (AUC), como indicador numérico para la comparación entre los modelos.

Tal y como señalaron Duman *et al.* (2006), una de las etapas más importantes de la cartografía de susceptibilidad de deslizamientos es la evaluación de la efectividad del mapa

producido. Así mismo, Can *et al.* (2005) aseguran que un modelo de susceptibilidad correcto debería cumplir los siguientes dos requisitos:

- *En el mapa, la mayor parte de los deslizamientos reservados para la validación deberían estar ubicados en los píxeles incluidos en las clases de mayor susceptibilidad.*
- *Al mismo tiempo, estas clases de alta susceptibilidad deberían cubrir la menor área posible, ya que, si estas cubren grandes extensiones del territorio, gran parte de los deslizamientos descritos se incluirían en las clases de alta susceptibilidad sin que ello signifique que el modelo sea lo suficientemente exacto.*

Por lo tanto, se han realizado dos curvas de porcentaje acumulado para cada modelo (ver ejemplo en Figura 4). La primera (curva - a) relaciona la clase de susceptibilidad de deslizamientos con el porcentaje acumulado del número de deslizamientos que pertenecen a cada clase. Así, para satisfacer la primera condición, esta curva debe mantenerse lo más cercana posible del eje X en las clases inferiores de susceptibilidad, y luego aumentar a medida que las clases de susceptibilidad se hacen más altas. La segunda (curva - b), relaciona las clases de susceptibilidad con el porcentaje acumulado del terreno que ocupa cada una de ellas. En este caso, para satisfacer la segunda condición, la curva debe aumentar tanto como sea posible en las clases más bajas de susceptibilidad y luego volverse cada vez más estable a medida que las clases de susceptibilidad se vuelven más altas. Con el objetivo de trazar las curvas de porcentaje acumulado, los valores de susceptibilidad de deslizamientos resultantes de cada modelo (de 0 a 1) han sido reclasificados en 10 clases de intervalos iguales.

Por otro lado, con el fin de obtener un valor objetivo y comparable para evaluar los modelos, se han realizado las curvas de tasa de predicción tal y como se sugiere en Chung y Fab-

bri (2003). Esta vez, el porcentaje acumulado de deslizamientos ocurridos en cada clase, ordenado de mayor a menor, se relaciona con el porcentaje acumulado del área de cada clase de susceptibilidad (Bonachea, 2006; Amorim, 2012; Trigila *et al.*, 2015). Al respecto, además de la ya significativa forma de la curva, se ha considerado el área bajo la curva (AUC) como un valor numérico adecuado para la comparación objetiva entre los diferentes modelos, de manera que un valor de 0,5 indicaría un poder de predicción prácticamente aleatorio, mientras que valores cercanos a 1 indicarían una capacidad de predicción óptima (Yilmaz, 2009).

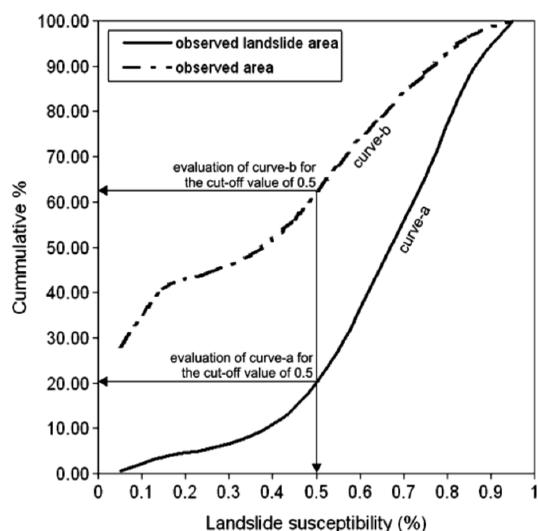


Figura 4: Ejemplo de curvas de porcentaje acumulado.
Fuente: Duman *et al.* (2006).

Figure 4: Example of accumulated percentage curves.
Source: Duman *et al.* (2006).

4. Resultados y comparación

4.1. Selección de variables explicativas

Todas las variables continuas han mostrado un resultado negativo en las pruebas de Kolmogorov-Smirnov (la variable no está distribuida de manera normal) por lo que las pruebas de significancia estadística se han llevado a cabo con el estadístico de Man Whitney y χ^2 . Entre las 22 variables independientes de inicio, 6 han sido excluidas debido a que han

superado el valor umbral de significancia definido en 0,05 (señalizados con un asterisco en la Tabla 1). Esto significa que *Permeabilidad*, *Orientación EO*, *Concavidad-Convexidad* y todas las variables derivadas de la *Curvatura* no han presentado una diferencia estadísticamente significativa entre los puntos estables e inestables.

Entre las 16 variables restantes, la prueba de correlación ha mostrado una alta colinealidad entre algunas variables (en negrita en la Tabla 1). Por ello, para asegurar la independencia de las variables de entrada, las variables continuas *Pendiente*, *Pendiente senoidal* y *Rugosidad* se han incluido en la RL por separado. Del mismo modo, al tratarse de diferentes clasificaciones de una misma variable, tanto *Uso del suelo 1* como *Uso del suelo 2* representan la misma información, por lo que no han sido incluidas juntas en la RL. Además, también se han considerado la *Distancia al cauce* y la *Distancia a la red de transportes* de manera independiente a la *Distancia al cauce CAT* y la *Distancia a la red de transportes CAT* respectivamente, ya que estas últimas son el resultado de la reclasificación de las primeras.

Utilizando el método “Backward Wald” (Pardo y Ruiz, 2002), en todas las combinaciones probadas, el *Espesor del regolito*, *Orientación*, *Orientación NS*, *TWI*, *Distancia al cauce CAT* y *Distancia al cauce* han sido eliminados de la ecuación del modelo final.

En consecuencia, una vez eliminadas las variables no significativas e identificadas las correlaciones, se han llevado a cabo diferentes combinaciones de variables para la regresión logística. En las mismas el mayor índice de acierto general ha sido alcanzado considerando solo la *Litología*, el *Uso del suelo 1*, la *Distancia a la red de transportes CAT*, la *Altitud*, la *Pendiente senoidal* y el *NDVI* (Tabla 2). Por lo tanto, éstas han sido las variables seleccionadas para el desarrollo de los primeros tres modelos de susceptibilidad.

Tabla 2: Coeficientes estimadores (β) de cada modelo de regresión logística y los valores de densidad de deslizamientos (LD) para cada clase.

Table 2: Estimate coefficients (β) for each logistic regression model and the landslide density (LD) values for each class.

		Modelo I	Modelo II	Modelo III		Modelo IV
		coef. β	coef. β	LD	coef. β	coef. β
Litología	Rocas detríticas de grano grueso	0	0	1,346	0,615	0,827
	(Areniscas). Dominante					
	Detríticos alternantes	-1,334	-1,371	0,5994		
	Calizas	-4,039	-3,906	0,0503		
	Margas	-0,884	-0,961	20,422		
	Margas descarboxatadas	-2,771	-4,429	0,2845		
	Depósitos superficiales	-3,721	-3,207	0,5356		
	Calizas impuras y calcarenitas	-1,516	-1,54	10,961		
	Alternancia de margocalizas, margas calizas y calcarenitas	-1,552	-1,611	13,561		
	Pizarras	-3,008	-3,671	0,0674		
	Arcillas con yesos y otras sales	-3,221	-2,849	11,025		
	Ofitas	-1,669	-1,224	26,592		
	Rocas detríticas de grano fino (lutitas). Dominante	-4,142	-3,803	0,4593		
Uso del suelo 1	Urbano	0	0	0,32	0,559	0,782
	Vías de comunicación	0,002	0,576	19,193		
	Bosque	-0,332	-1,047	0,4361		
	Bosque de Plantación	0,247	-0,334	0,5323		
	Herbazal Pastizal	-17,491	-15,376	0		
	Cultivos	-2,373	-2,594	0,4198		
	Bosque de Galería	23,014	22,302	0,9212		
	Arbustados	1,528	0,96	14,334		
	Pastizal Matorral	1,471	1,591	0,1995		
	Prados	1,374	1,038	28,446		
	Prados con setos	-20,967	-20,61	0		
	Distancia a la red de transportes CAT	0 - 20 m	0	0	5,745	0,499
20 - 50 m		-0,52	-0,752	46,428		
50 - 100 m		-1,712	-1,759	30,149		
100 - 150 m		-1,471	-1,491	25,151		
150 - 200 m		-2,395	-2,536	12,551		
200 - 250 m		-3,391	-3,365	0,6732		
250 - 300 m		-2,112	-1,991	0,6295		
>300 m		-3,329	-3,569	0,2682		
NDVI	-0.6 y -0.2	-5,342	0	16,226	-4,574	-4,587
	-0.2 y 0.2		-1,124	0,9534		
	0.2 y 0.4		-0,216	10,395		

		Modelo I	Modelo II	Modelo III		Modelo IV
		coef. β	coef. β	LD	coef. β	coef. β
Pendiente senoidal	0 - 0.2	5,975	0	0,0595	5,312	4,853
	0.2 - 0.4		2,526	0,2702		
	0.4 - 0.6		2,516	0,3784		
	0.6 - 0.8		4,559	12,982		
	0.8 - 1		5,265	14,433		
Altitud	0 - 300 m	-0,002	0	16,888	-0,002	--
	300 - 600 m		-0,226	0,8119		
	600 - 900 m		-1,934	0,0979		
	900 - 1200 m		-20,5	0		
	1200 - 1550 m		-18,328	0		
Intercepto		-0,829	0,431		-5,85	-5,75

4.2. Comparación de modelos I, II y III

La tabla 2 muestra los coeficientes asignados a cada variable explicativa mediante los cuales se define el modelo de susceptibilidad de deslizamientos. Los signos y valores absolutos de los coeficientes indican cuáles de los factores ambientales desempeñan un papel decisivo en la aparición de deslizamientos. En las variables categóricas, un valor positivo y alto indica un papel importante a favor del deslizamiento de esa clase, mientras que un valor negativo significa un papel importante a favor de la estabilización de la ladera. Sin embargo, cabe señalar que debido a la reclasificación en variables dummy utilizada en los modelos I y II, los coeficientes de las clases categóricas se hacen en referencia a la primera clase de cada variable, por lo que se trata de valores de magnitud relativa; y en consecuencia no son comparables con valores de otros modelos. Por su parte, en las variables continuas, un coeficiente alto y positivo significa que cuanto mayor es el valor de píxel en esa variable, favorecerá en mayor medida a la probabilidad de ocurrencia de un deslizamiento de tierra. Por contra, si resultan coeficientes negativos, la presencia

de valores absolutos altos representaría un efecto contrario.

Atendiendo a los valores de la tabla 2, se puede observar que la *Distancia a la red de transportes CAT*, la *Pendiente senoidal* y la *Altitud* presentan resultados relativos similares para los tres modelos. Y aunque este último no implica casi relevancia alguna (coeficiente cercano a 0), las otras dos variables han demostrado ser realmente significativas para los modelos de susceptibilidad de deslizamientos en éste área de estudio y con este inventario de deslizamientos. Así, la *Distancia a la red de transporte CAT* y la *Pendiente senoidal* se consideran los principales factores que explican la distribución de la variable dependiente. Así mismo, se podría inferir que, valores de pendiente senoidal cercanos a 1 (alrededor de 45° de pendiente), altitudes entre 0 y 300 m, un NDVI entre 0,2 y 0,4 (vegetación deficiente), distancias de la red de transportes entre 0 y 20 m, con suelos de matorrales, pastos o prados y rocas detríticas de grano grueso o margas son las características más comunes en las laderas afectadas por deslizamientos de tierra en la cuenca del río Oria (Tabla 2).

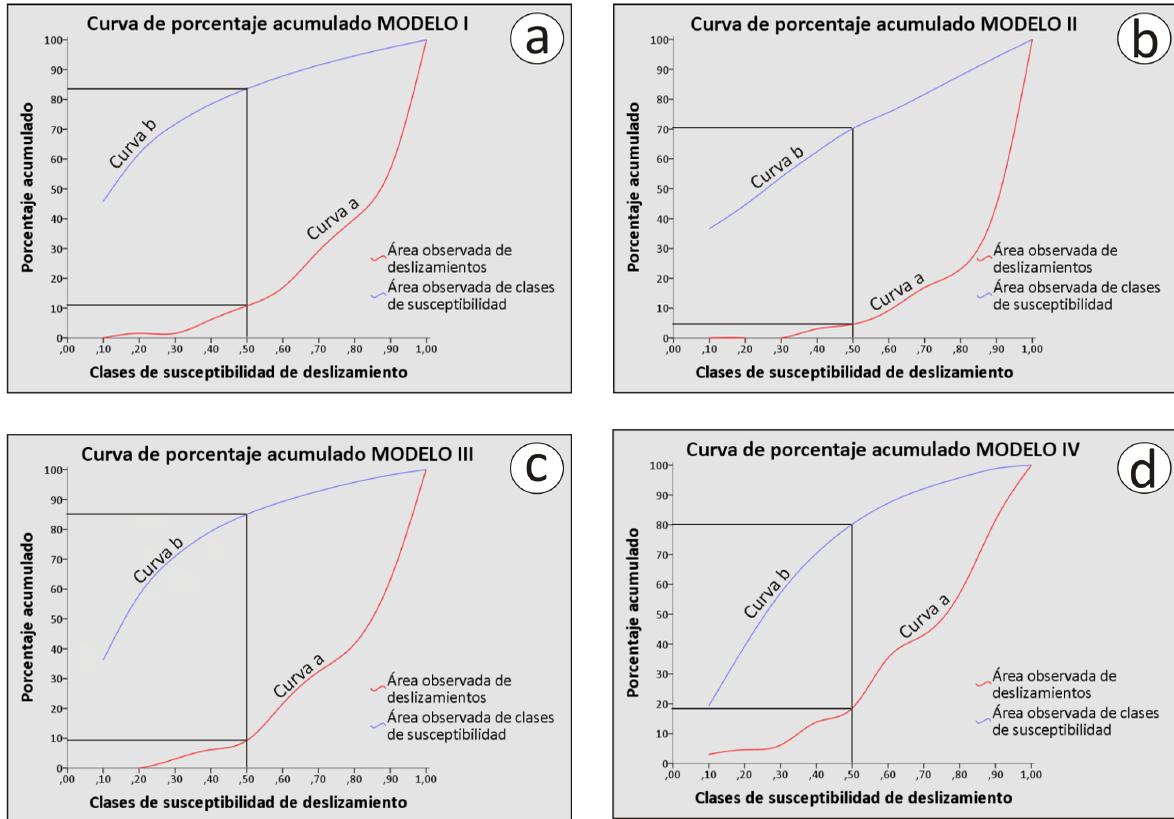


Figura 5: Las curvas de porcentaje acumulado para los modelos I, II y IV respectivamente. Las curvas están suavizadas.

Figure 5: The cumulative percentage curves for models I, II and III respectively. Smooth curves.

Las curvas de porcentaje acumulado (Figuras 5 a, b y c) muestran una alta coincidencia con respecto a los puntos de inestabilidad reservados para la validación. En los tres casos la gran mayoría de la muestra de validación (entre el 90 y el 95%) se ha localizado en clases de susceptibilidad de deslizamientos superiores al 0.5 de probabilidad, tal y como se puede apreciar en las curvas-a de cada modelo. Al respecto, destaca el modelo II (Figura 5 b), cuya curva-a muestra que menos del 5% de los deslizamientos coinciden con las clases de menor susceptibilidad, mientras que en los modelos I y III (Figuras 5 a y 5 c) este valor aumenta hasta cerca del 10%. Sin embargo, las curvas-b indican que no todos los modelos poseen la misma capacidad de discriminación. En el modelo II, las clases de alta probabilidad cubren un mayor porcentaje del área de estudio (30% aproximadamente) que en los modelos I y III (15%).

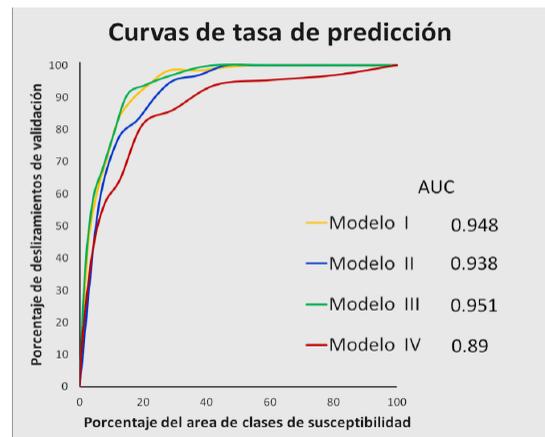


Figura 6: Curvas de tasa de predicción para los modelos I, II, III y IV, y su correspondiente área bajo la curva (AUC). Las curvas están suavizadas.

Figure 6: Prediction rate curves for models I, II, III and IV, and their corresponding area under the curve (AUC). Smooth curves.

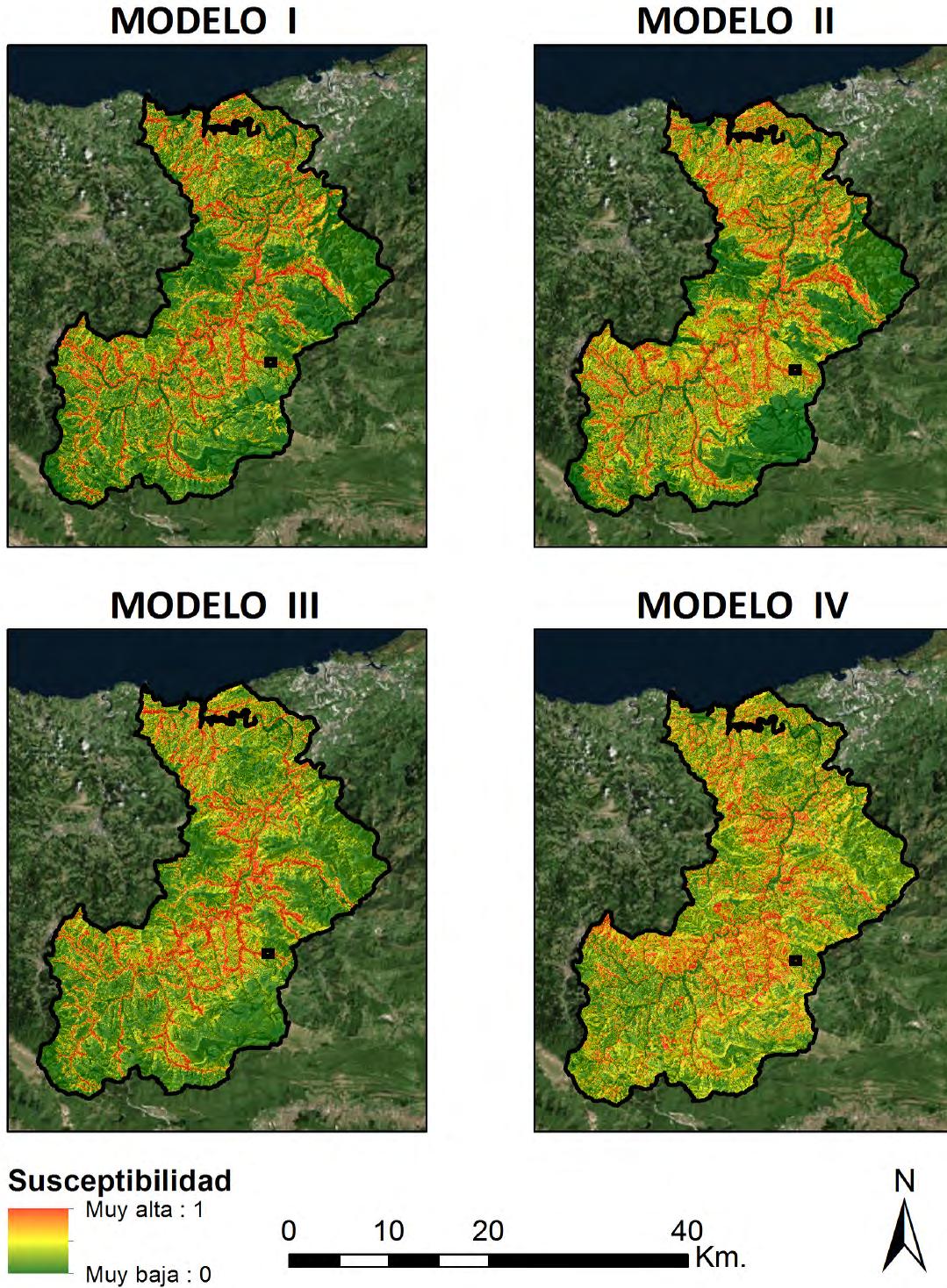


Figura 7: Mapas de susceptibilidad de deslizamientos correspondientes a los modelos I, II, III y IV.

Figure 7: Landslide susceptibility maps for models I, II, III and IV.

Respecto a las curvas de tasa de predicción (Figura 6), sus formas revelan que las diferentes estrategias de tratamiento de las variables explicativas para la RL resultan muy satisfactorias, aunque se pueden apreciar pequeñas diferencias entre los tres modelos de susceptibilidad. Numéricamente, con un AUC de 0,951 el modelo III ofrece el mejor rendimiento seguido de cerca por el modelo I (AUC=0,948), mientras que el modelo II presentaría una capacidad de acierto ligeramente inferior (AUC=0,938) (Figura 6).

A partir de los modelos resultantes de la regresión logística se han representado las probabilidades de ocurrencia de deslizamientos superficiales en los siguientes mapas de susceptibilidad de deslizamientos (Figuras 7 a, b y c). Entre ellos, la mayor diferencia se puede observar en el modelo II (Figura 7 b) donde, a partir de una inspección visual, resalta la uniformidad de las zonas no susceptibles en la parte este de la cuenca (zonas montañosas); frente a la mayor difuminación que presentan los modelos I y III (Figuras 7 a y 7 c).

En todo caso, destaca la amplia similitud que presentan los tres mapas en los que se distribuyen las clases de alta susceptibilidad siguiendo claramente un patrón lineal parecido a la red de drenaje. De hecho, una inspección con mayor detalle (Figuras 8 a, b y c) de los mapas ha revelado la casi completa dependencia de la distribución de la susceptibilidad a la variable *Distancia a la red de transportes CAT*. Esto podría manifestar cierta influencia que pudiera tener el inventario de deslizamientos de manera intrínseca con respecto a esta variable, ya que la información de las localizaciones de los deslizamientos ha sido obtenida en trabajo de campo.

4.3. Resultados del modelo IV

Atendiendo a los resultados de los tres primeros modelos, se ha constatado que, (i) la variable *Altitud* no ejerce prácticamente ningún efecto en la presencia o ausencia de los deslizamientos inventariados (coeficiente β muy próximo a 0), y (ii) la distribución espacial de

los mapas de susceptibilidad se encuentra fuertemente condicionada por la variable *Distancia a la red de transportes CAT*, lo que sugiere que, al menos parte del inventario, podría estar sesgado por esta variable.

Por esa razón, se ha elaborado un modelo de susceptibilidad adicional en el que se ha aplicado el mismo tratamiento de variables que en el modelo III, pero eliminando la *Altitud* y la *Distancia a la red de transportes CAT* como variables explicativas y se han introducido únicamente las variables *Litología*, *Uso del suelo 1*, *NDVI* y *Pendiente senoidal*.

El Mapa de Susceptibilidad de Deslizamientos del modelo IV (Figura 7 d) muestra notables diferencias respecto a los modelos anteriores. Los valores de probabilidad de ocurrencia de deslizamientos no se encuentran distribuidos siguiendo los patrones lineales presentes en los modelos I, II y III, y observando la porción ampliada (Figura 8 d) estas diferencias se manifiestan en mayor medida.

Respecto a las pruebas de validación, las curvas de porcentaje acumulado curva-a y curva-b (Figura 5 d) muestran que menos del 20% de los deslizamientos de validación coinciden con clases de susceptibilidad inferiores al 0,5. Mientras, alrededor del 80% del territorio estudiado ha sido clasificado como zona de baja susceptibilidad. Esta misma información representada en la curva de tasa de predicción presenta un elevado valor de AUC (0,89) (Figura 6).

5. Discusión

Existe una amplia variedad de metodologías disponibles para la confección de mapas de susceptibilidad de deslizamientos y cada una de ellas es altamente dependiente de ciertas decisiones que han de tomarse a lo largo de su ejecución. Qué tipo de variable dependiente utilizar (puntos, polígonos o mixtos) y la fuente de la misma, qué unidad de cartografiado aplicar al análisis (píxeles, unidades de pendiente, ...), qué variables considerar como relevantes o incluso qué tipo de trata-

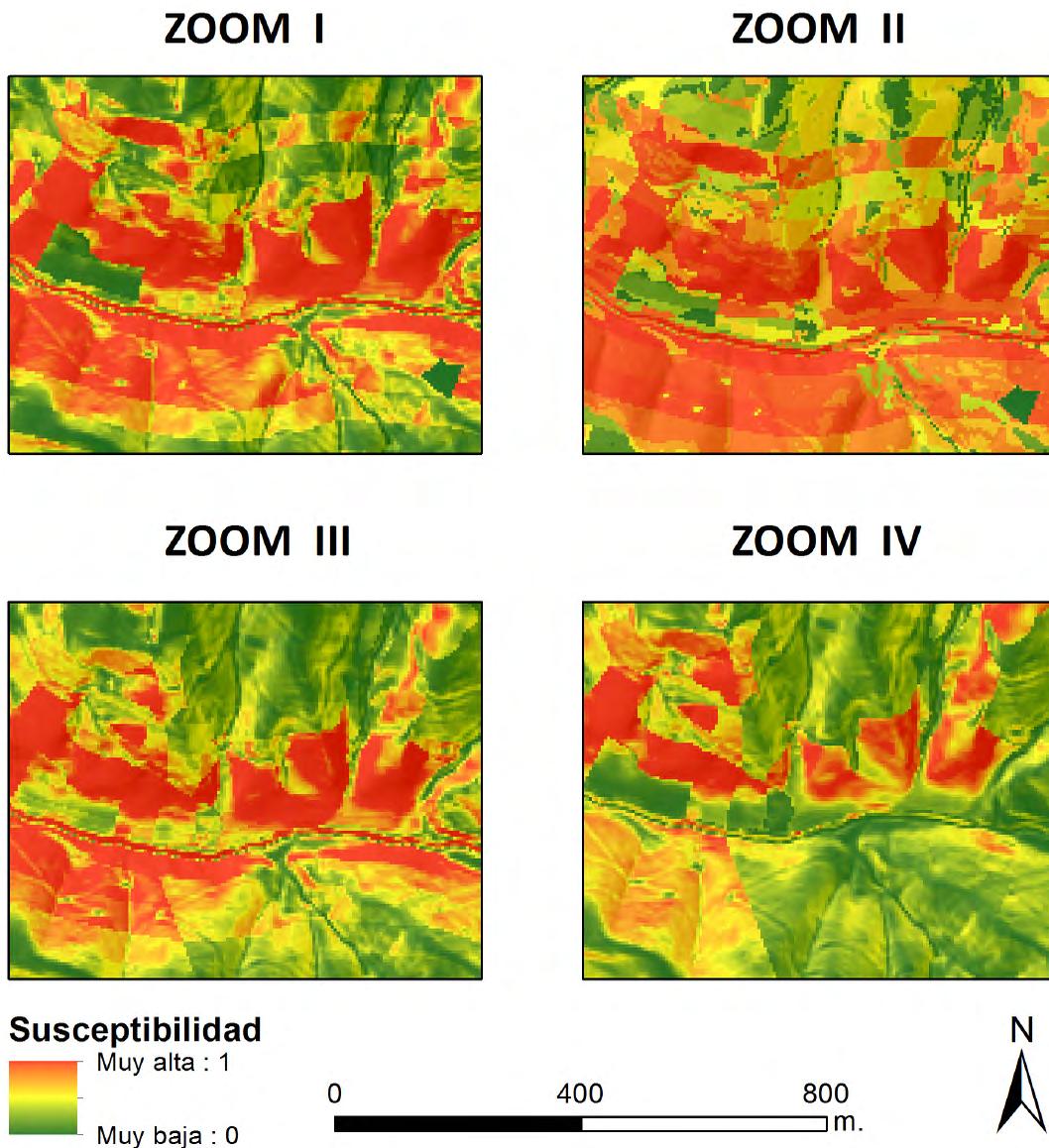


Figura 8: Ampliación en detalle de los mapas de susceptibilidad I, II, III y IV.

Figure 8: Detailed zoom of susceptibility maps I, II, III and IV.

miento deben recibir, son solo algunas de las cuestiones que deben plantearse a la hora de llevar a cabo un análisis de susceptibilidad de deslizamientos. En este artículo se ha abordado el aspecto de la selección de variables explicativas y el tipo de tratamiento más adecuado que éstas deberían recibir dependiendo de su condición continua o categórica. Se ha llevado a cabo un procedimiento

de selección de variables experimental y la comparación entre varios modelos de susceptibilidad en los que las variables explicativas han seguido diferentes estrategias de tratamiento (ver apartados 3.2 y 3.4). Ello ha permitido detectar pequeñas diferencias que han servido para determinar la mejor manera de introducir las variables en la regresión logística.

Atendiendo a la capacidad de predicción de los primeros 3 modelos, el modelo I presenta buenos resultados (Figuras 5 y 6), pero debido al tamaño reducido de algunas categorías, la aparición de valores extremos de los coeficientes podría condicionar fuertemente los resultados. Lo mismo sucede en el caso del modelo II, que presenta resultados ligeramente inferiores, quizás por la división subjetiva que se aplica a las variables continuas para su conversión en categóricas. Como ejemplo de este efecto, destacan los altos valores absolutos de los coeficientes asignados a las clases de mayor elevación, que enmascaran cualquier posible influencia de las demás variables. No obstante, a pesar de su alto rendimiento de predicción, hay que tener en cuenta la gran extensión que toman las clases de alta susceptibilidad, lo que disminuye fuertemente la capacidad de discriminación de este modelo. Además, la codificación de todas las variables en variables de confusión binarias supone un aumento considerable del tiempo de cálculo del modelo y la interpretación de los coeficientes de estimación del modelo se hace más laboriosa. Por su parte, el modelo III presenta los mejores resultados, con excelente equilibrio entre la capacidad de predicción y el poder de discriminación de modo que la transformación de las variables categóricas mediante la asignación de valores de densidad de deslizamientos a las diferentes clases permite mitigar el efecto de las categorías más pequeñas, ofreciendo un modelo de susceptibilidad más robusto, tal y como sugieren Grozavu *et al.* (2013) y Trigila *et al.* (2015).

El procedimiento de selección de variables explicativas cuyos resultados se exponen en el apartado 4.1 ha demostrado ser válido para la elección objetiva de variables que se encuentren relacionadas de forma directa con la distribución espacial de deslizamientos y la identificación de variables altamente relacionadas entre sí. La eliminación de las primeras 6 variables con el test de significancia estadística se trata de un resultado que cabía esperar, considerando que todos estos factores ambientales presentan una distribución

espacial muy uniforme en el área de estudio. Además, el hecho de que el propio software haya eliminado de manera sistemática otras 6 variables concuerda con los resultados de otras investigaciones (Carrara, 1983; Guzzetti *et al.*, 1999; Remondo *et al.*, 2008; Bonachea, 2006; Yilmaz, 2009). En el caso de la *Distancia al cauce*, su eliminación podría deberse a la alta densidad de la red de drenaje, lo que facilita que la mayoría de los puntos, sean estables o inestables, caigan cerca de un río, y por lo tanto la variabilidad se reduce de manera acentuada. Hay que tener en cuenta que algunas de las variables son simples reclasificaciones de otras, que han sido transformadas para abarcar el mayor abanico de posibilidades. Pero lógicamente, en estos casos la correlación es prácticamente perfecta y en consecuencia este tipo de variables no serán introducidas conjuntamente.

Así, se han podido seleccionar únicamente seis variables de un conjunto inicial de 22. Sin embargo, esta selección no ha resultado del todo satisfactoria, ya que una vez definidos los modelos de RL y los mapas de susceptibilidad, el comportamiento de algunas variables ha delatado su inadecuada aptitud para el desarrollo de un modelo de susceptibilidad con garantías para este caso de estudio. Por un lado, la *Altitud*, aunque el estadístico de Man Whitney sugiere una relación estadísticamente significativa con la presencia o ausencia de deslizamientos, una vez introducido en la RL ha mostrado un efecto prácticamente nulo dentro del modelo, con un coeficiente muy cercano al valor de 0. De hecho, la justificación geomorfológica de esta variable no es del todo clara, ya que la altitud, por sí sola, no supone un efecto directo en la desestabilización de laderas. Sin embargo, sí que puede llegar a representar la asociación de cualquier otra variable, como por ejemplo la litología o el uso del suelo. Por otro lado, las pruebas estadísticas han avalado la relación de la *Distancia a la red de transportes CAT* con la presencia o ausencia de inestabilidades de ladera. El hecho de haber obtenido el inventario de deslizamientos a partir de trabajo de campo, probablemente esconde una

influencia intrínseca de la variable explicativa respecto a la variable de respuesta. De esta manera se podría considerar que al menos parte del inventario de deslizamientos esté constituido por inestabilidades directamente relacionadas con las modificaciones artificiales de los taludes de carreteras u otras infraestructuras lineales, tales como las vías del tren. Ello sugiere que, para poder utilizar la variable *Distancia a la red de transportes CAT* como variable explicativa, sería necesario dividir el inventario entre deslizamientos de origen “natural” y deslizamientos de origen “antrópico”, una información que lamentablemente no se encuentra disponible. De hecho, el resultado sesgado que esta variable ejerce en los modelos de susceptibilidad I, II y III ha sido detectado a través de sus representaciones gráficas (Figuras 7 a, b y c), donde la concentración prácticamente completa de las clases de susceptibilidad más altas se encuentra alrededor de las vías de comunicación.

Siguiendo este razonamiento, se deduce que los mapas de susceptibilidad I, II y III no son operacionalmente aceptables, aunque la comparación de los resultados de validación obtenidos aplicando diferentes estrategias de tratamiento de las variables explicativas sí que ha permitido identificar al procedimiento seguido por el modelo III como el más apropiado. Esto ha hecho necesario el desarrollo de un cuarto mapa de susceptibilidad (apartado 4.3), siguiendo la estrategia de tratamiento de variables aplicado en el modelo III, pero esta vez eliminando el sesgo que la variable *Distancia a la red de transportes CAT* pudiera producir suprimiendo su uso en la ecuación. A su vez, la variable *Altitud* también ha sido eliminada a consecuencia de su nula contribución en las ecuaciones de los modelos I, II y III.

Como resultado, el modelo IV presenta una capacidad de predicción y discriminación muy buena con un valor de AUC de la curva de tasa de predicción de 0,89 (Figura 6). Además, la representación gráfica del modelo (Figuras 7 d y 8 d) muestra una distribución espacial de las zonas de alta suscep-

tibilidad ligeramente marcada por los tipos litológicos, pero mucho más difuminada por el resto de las variables explicativas, lo que ofrece un mapa de susceptibilidad de deslizamientos satisfactorio.

En definitiva, las labores desempeñadas durante este estudio han planteado nuevas cuestiones a considerar en futuras investigaciones. La más importante podría ser el estudio de la incertidumbre que un inventario de deslizamientos basado exclusivamente en trabajo de campo puede introducir en un modelo estadístico de susceptibilidad. Este artículo sugiere que los inventarios realizados por trabajo de campo pueden condicionar de manera directa la elección de ciertas variables explicativas. Asimismo, el trabajo de campo no permite abarcar la totalidad de un área de estudio de manera completa y homogénea, ya que siempre dependerá de la accesibilidad y la visibilidad de cada lugar y por tanto la incertidumbre introducida, sobre todo por los puntos considerados como “estable”, podría ser muy importante.

6. Conclusiones finales

El presente trabajo ofrece diversos resultados que se consideran objetivos y que están relacionados con las decisiones que se han de tomar al inicio de todo proceso de confección de mapas de susceptibilidad de deslizamientos en los que se aplica la RL.

La comparación de tres modelos de susceptibilidad desarrollados a partir de diferentes estrategias de tratamiento de las variables explicativas ha concluido en que la transformación de las variables categóricas en continuas mediante la asignación de valores relativos a la densidad de deslizamientos resulta más ventajoso para la adecuada ejecución de modelos de susceptibilidad de deslizamientos mediante RL, ya que mediante este procedimiento se evita la creación de una gran cantidad de variables de confusión (dummy variables) (Pardo y Ruiz, 2002) y se mantiene la máxima capacidad de predicción.

El procedimiento de selección de variables para un modelo de RL ha ofrecido una válida opción a la hora de decidir qué variables introducir en el análisis estadístico, aunque los indicadores estadísticos no ofrecen la certeza absoluta de que las variables seleccionadas sean las más adecuadas. En este caso, la existencia de una variable, como la *Distancia a la red de transportes CAT*, que presenta una dependencia constatable respecto a la variable dependiente, indica la necesidad de una justificación geomorfológica frente a los resultados para asegurar unos mapas de susceptibilidad fiables y ajustados a la realidad.

Finalmente, mediante esta investigación se espera haber podido avanzar hacia una mejora de la metodología para el análisis de susceptibilidad de deslizamientos y a su vez, haber aportado un documento cartográfico de utilidad para la planificación territorial del área de estudio.

Agradecimientos

Este trabajo ha contado con la aportación de la Cátedra UNESCO de Paisaje y Patrimonio de la Universidad del País Vasco (UPV/EHU). Nuestro agradecimiento a Patxi Tamés y José Antonio Navarro (Diputación Foral de Gipuzkoa) por proporcionar los documentos oficiales inéditos relacionados, a Irantzu Barrios e Ismael Barbero por su asesoramiento sobre el procedimiento estadístico y a Alfredo Ollero y Estela Nadal por sus comentarios críticos que han contribuido a mejorar este manuscrito.

Bibliografía

Amorim, S. F. (2012). *Estudio comparativo de métodos para la evaluación de la susceptibilidad del terreno a la formación de deslizamientos superficiales: Aplicación al Pirineo Oriental*. Tesis doctoral, Universidad Politécnica de Catalunya, Barcelona.

Atkinson, P. M. y Massari, R. (1998). Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy. *Computers*

& *Geosciences*, 24(4): 373-385. [https://doi.org/10.1016/S0098-3004\(97\)00117-9](https://doi.org/10.1016/S0098-3004(97)00117-9)

Ayala-Carcedo, F.J. y Olcina, J. (2002). *Riesgos Naturales*. Ariel Ciencia, 1512 pp.

Bai, S.B.; Wang, J.; Lü, G.N.; Zhou, P.G.; Hou, S.S. y Xu, S.N. (2010). Gis-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China. *Geomorphology*, 115(1): 23-31. <https://doi.org/10.1016/j.geomorph.2009.09.025>

Blais Stevens, A.; Behnia, P.; Kremer, M.; Page, A.; Kung, R. y Bonham-Carter, G. (2012). Landslide susceptibility mapping of the Sea to Sky transportation corridor, British Columbia, Canada: comparison of two methods. *Bulletin of Engineering Geology and the Environment*, 71(3): 447-466. <https://doi.org/10.1007/s10064-012-0421-z>

Bonachea, J. (2006). *Desarrollo, aplicación y validación de procedimientos y modelos para la evaluación de amenazas, vulnerabilidad y riesgo debidos a procesos geomorfológicos*. Tesis doctoral, Universidad de Cantabria.

Bonachea, J.; Remondo, J.; Terán, D.; Díaz, J. R.; González Díez, A. y Cendrero, A. (2009). Landslide risk models for decision making. *Risk analysis*, 29(11): 1629-1643. <https://doi.org/10.1111/j.1539-6924.2009.01283.x>

Budimir, M.; Atkinson, P. y Lewis, H. (2015). A systematic review of landslide probability mapping using logistic regression. *Landslides*, 12(3): 419-436. <https://doi.org/10.1007/s10346-014-0550-5>

Campos, J. y García-Dueñas, V. (1972). *Mapa Geológico de España escala 1:50.000. 2ª Serie (MAGNA), Hoja de San Sebastián*. IGME. Servicio de Publicaciones Ministerio de Industria, Madrid.

Campos, J.; Olivé, A.; Ramírez, J.I.; Solé, J. y Villalobos, L. (1983). *Mapa Geológico de España escala 1:50.000. 2ª Serie (MAGNA), Hoja de Tolosa*. IGME. Servicio de Publicaciones Ministerio de Industria, Madrid.

Can, T.; Nefeslioglu, H. A.; Gokceoglu, C.; Sonmez, H. y Duman, T. Y. (2005). Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses. *Geomorphology*, 72(1): 250-271. <https://doi.org/10.1016/j.geomorph.2005.05.011>

Carrara, A. (1983). Multivariate models for landslide hazard evaluation. *Mathematical geology*, 15(3): 403-426. <https://doi.org/10.1007/BF01031290>

- Cascini, L. (2008). Applicability of landslide susceptibility and hazard zoning at different scales. *Engineering Geology*, 102(3): 164-177. <https://doi.org/10.1016/j.enggeo.2008.03.016>
- Chung, C.J. F. y Fabbri, A. G. (2003). Validation of spatial prediction models for landslide hazard mapping. *Natural Hazards*, 30(3): 451-472. <https://doi.org/10.1023/B:NHAZ.0000007172.62651.2b>
- Chung, C. y Fabbri, A. G. (2005). Systematic procedures of landslide hazard mapping for risk assessment using spatial prediction models. *Landslide hazard and risk*. Wiley, New York, 139-177. <https://doi.org/10.1002/9780470012659.ch4>
- Corominas, J. y Mavrouli, J. (2011). Living with landslide risk in Europe: Assessment, effects of global change, and risk management strategies. Documento técnico, SafeLand. *7th Framework Programme Cooperation Theme 6 Environment (including climate change) Sub-Activity 6.1.3 Natural Hazards*.
- Costanzo, D.; Chacón, J.; Conoscenti, C.; Irigaray, C. y Rotigliano, E. (2014). Forward logistic regression for earth-flow landslide susceptibility assessment in the Platani river basin (southern Sicily, Italy). *Landslides*, 11(4): 639-653. <https://doi.org/10.1007/s10346-013-0415-3>
- CRED (2014). *Atlas of mortality and economic losses from weather, climate and water extremes (1970-2012)*. Documento técnico, World Meteorological Organization, Geneva.
- Dai, F. y Lee, C. (2002). Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology*, 42(3): 213-228. [https://doi.org/10.1016/S0169-555X\(01\)00087-3](https://doi.org/10.1016/S0169-555X(01)00087-3)
- Das, I.; Sahoo, S.; Van Westen, C.; Stein, A. y Hack, R. (2010). Landslide susceptibility assessment using logistic regression and its comparison with a rock mass classification system, along a road section in the northern Himalayas (India). *Geomorphology*, 114(4): 627-637. <https://doi.org/10.1016/j.geomorph.2009.09.023>
- De Winter, J. C. y Dodou, D. (2010). Five point likert items: t test versus mann-whitney-wilcoxon. *Practical Assessment, Research & Evaluation*, 15(11): 1-12.
- Devkota, K. C.; Regmi, A. D.; Pourghasemi, H. R.; Yoshida, K.; Pradhan, B.; Ryu, I. C.; Dhital, M. R. y Althuwaynee, O. F. (2013). Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling-Narayanghat road section in Nepal Himalaya. *Natural Hazards*, 65(1): 135-165. <https://doi.org/10.1007/s11069-012-0347-6>
- Diario Vasco (18 de marzo, 2013). *Las lluvias provocan daños de casi 8 millones en la red viaria de Gipuzkoa*. Diario Vasco.
- Diputación Foral de Gipuzkoa (1987). *Estudio de riesgos naturales del Territorio Histórico de Gipuzkoa*. Dirección General de Medio Ambiente. Informe inédito.
- Diputación Foral de Gipuzkoa (2007). *Elaboración de modelos de predicción de riesgos naturales de incendios, deslizamientos e inundaciones en el Territorio Histórico de Gipuzkoa*. Dirección General de Medio Ambiente. Informe inédito.
- Diputación Foral de Gipuzkoa (2013). *Evaluación y gestión integrada de riesgos geotécnicos en la red de carreteras de la Diputación Foral de Gipuzkoa*. Dpto. de Movilidad e Infraestructuras Viarias. Informe inédito.
- Duman, T. Y.; Can, T.; Gokceoglu, C.; Nefeslioglu, H. A. y Sonmez, H. (2006). Application of logistic regression for landslide susceptibility zoning of cekmece area, Istanbul, Turkey. *Environmental Geology*, 51(2): 241-256. <https://doi.org/10.1007/s00254-006-0322-1>
- EM-DAT. *The International Disaster Database*. www.emdat.be
- EVE (2010). *Mapa Geológico del País Vasco 1:100.000*. Ente Vasco de la Energía.
- Fernández-Arroyabe, P. y Martín-Vide, J. (2012). Regionalization of the probability of wet spells and rainfall persistence in the Basque Country (Northern Spain). *International Journal of Climatology*, 32(12): 1909-1920. <https://doi.org/10.1002/joc.2405>
- Felicitísimo, Á. M.; Cuartero, A.; Remondo, J. y Quirós, E. (2013). Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. *Landslides*, 10(2): 175-189. <https://doi.org/10.1007/s10346-012-0320-1>
- González-Hidalgo, J. C.; Brunetti, M. y de Luis, M. (2011). A new tool for monthly precipitation analysis in Spain: MOPREDAS database (monthly precipitation trends December 1945-November 2005). *International Journal of Climatology*, 31(5): 715-731. <https://doi.org/10.1002/joc.2115>
- Grozavu, A.; Pleşcan, S.; Patriche, C. V.; Mărgărint, M. C. y Roşca, B. (2013). Landslide susceptibility assessment: GIS application to a complex mountainous environment. En: *The*

- Carpathians: Integrating Nature and Society Towards Sustainability*, 31-44. Springer. https://doi.org/10.1007/978-3-642-12725-0_4
- Günther, A.; Van Den Eeckhaut, M.; Malet, J.P.; Reichenbach, P. y Hervás, J. (2014). Climate physiographically differentiated Pan-European landslide susceptibility assessment using spatial multi-criteria evaluation and transnational landslide information. *Geomorphology*, 224: 69-85. <https://doi.org/10.1016/j.geomorph.2014.07.011>
- Guzzetti, F.; Carrara, A.; Cardinali, M. y Reichenbach, P. (1999). Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology*, 31(1): 181-216. [https://doi.org/10.1016/S0169-555X\(99\)00078-1](https://doi.org/10.1016/S0169-555X(99)00078-1)
- Guzzetti, F.; Reichenbach, P.; Cardinali, M.; Galli, M.; Ardizzone, F. (2005). Probabilistic landslide hazard assessment at the basin scale. *Geomorphology*, 72(1):272-299. <https://doi.org/10.1016/j.geomorph.2005.06.002>
- IDE de Euskadi. *Infraestructura de Datos Espaciales de Euskadi*. www.geo.euskadi.eus
- INGEMISA (1996). *Inventario y Análisis de las Áreas sometidas a Riesgo de Inestabilidades del Terreno de la C.A.P.V.* Documento técnico, Gobierno Vasco.
- Kumar, R. y Anbalagan, R. (2015). Landslide susceptibility zonation of Tehri reservoir rim region using binary logistic regression model. *Current Science*, 108(9): 1662-1672.
- Lee, S. (2005). Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *International Journal of Remote Sensing*, 26(7): 1477-1491. <https://doi.org/10.1080/01431160412331331012>
- Nefeslioglu, H.; Gokceoglu, C. y Sonmez, H. (2008). An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Engineering Geology*, 97(3): 171-191. <https://doi.org/10.1016/j.enggeo.2008.01.004>
- Nefeslioglu, H.; Gokceoglu, C.; Sonmez, H. y Gorum, T. (2011). Medium-scale hazard mapping for shallow landslide initiation: the Buyukoy catchment area (Cayeli, Rize, Turkey). *Landslides*, 8(4), 459-483. <https://doi.org/10.1007/s10346-011-0267-7>
- Pardo, A. y Ruiz, M. A. (2002). *SPSS 11. Guía para el análisis de datos*. McGraw-Hill/Interamericana de España, Madrid.
- Petley, D. (2012). Global patterns of loss of life from landslides. *Geology*, 40: 927-930. <https://doi.org/10.1130/G33217.1>
- Pourghasemi, H.; Moradi, H. y Aghda, S. F. (2013). Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. *Natural Hazards*, 69(1): 749-779. <https://doi.org/10.1007/s11069-013-0728-5>
- Ramírez, P. E.; Fito, J. M. E. y Rojo, A. B. (2005). Evaluación espacial de los peligros naturales en el valle de Oiartzun (Guipuzkoa). *Munibe Ciencias Naturales. Natur zientziak*, 1(56): 5-20.
- Rana, R. y Singhal, R. (2015). Chi square test and its application in hypothesis testing. *Journal of the Practice of Cardiovascular Sciences*, 1(1): 69-71. <https://doi.org/10.4103/2395-5414.157577>
- Remondo, J.; Bonachea, J. y Cendrero, A. (2008). Quantitative landslide risk assessment and mapping on the basis of recent occurrences. *Geomorphology*, 94(3): 496-507. <https://doi.org/10.1016/j.geomorph.2006.10.041>
- Rouse Jr, J.; Haas, R.; Schell, J. y Deering, D. (1974). *Monitoring vegetation systems in the great plains with erts*. NASA special publication, 351: 309.
- Schicker, R. D. (2010). *Quantitative landslide susceptibility assessment of the Waikato region using GIS*. Tesis doctoral, The University of Waikato.
- Soeters, R. y Van Westen, C. (1996). Slope stability recognition analysis and zonation. En: *Landslides: Investigation and Mitigation*, 129-177.
- Trigila, A.; Iadanza, C.; Esposito, C. y Scarascia-Mugnozza, G. (2015). Comparison of logistic regression and random forests techniques for shallow landslide susceptibility assessment in Giampileri (NE Sicily, Italy). *Geomorphology*, 249: 119-136. <https://doi.org/10.1016/j.geomorph.2015.06.001>
- Van Den Eeckhaut, M.; Hervás, J.; Jaedicke, C.; Malet, J.P.; Montanarella, L. y Nadim, F. (2012). Statistical modelling of Europe-wide landslide susceptibility using limited landslide inventory data. *Landslides*, 9(3): 357-369. <https://doi.org/10.1007/s10346-011-0299-z>
- Van Den Eeckhaut, M.; Vanwalleggem, T.; Poesen, J.; Govers, G.; Verstraeten, G. y Vandekerckhove, L. (2006). Prediction of landslide susceptibility using rare events logistic regression: a case study in the Flemish Ardennes (Belgium). *Geomorphology*, 76(3): 392-410. <https://doi.org/10.1016/j.geomorph.2005.12.003>

- Van Westen, C. J. (1993). *Application of geographic information systems to landslide hazard zonation*. Tesis doctoral, TU Delft, Delft University of Technology.
- Varnes, D. J. (1958). Landslide types and processes. *Landslides and Engineering Practice*, 24: 20-47.
- Wang, Y.T.; Seijmonsbergen, A. C.; Bouten, W. y Chen, Q.T. (2015). Using statistical learning algorithms in regional landslide susceptibility zonation with limited landslide field data. *Journal of Mountain Science*, 12(2): 268-288. <https://doi.org/10.1007/s11629-014-3134-x>
- Yilmaz, I. (2009). Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat Turkey). *Computers & Geosciences*, 35(6): 1125-1138. <https://doi.org/10.1016/j.cageo.2008.08.007>
- Zhu, L. y Huang, J.F. (2006). GIS-based logistic regression method for landslide susceptibility mapping in regional scale. *Journal of Zhejiang University-SCIENCE A*, 7(12): 2007-2017. <https://doi.org/10.1631/jzus.2006.A2007>

