

CARTOGRAFÍA GEOTÉCNICA

Métodos Estadísticos Bivariados

Prof. Edier Aristizábal



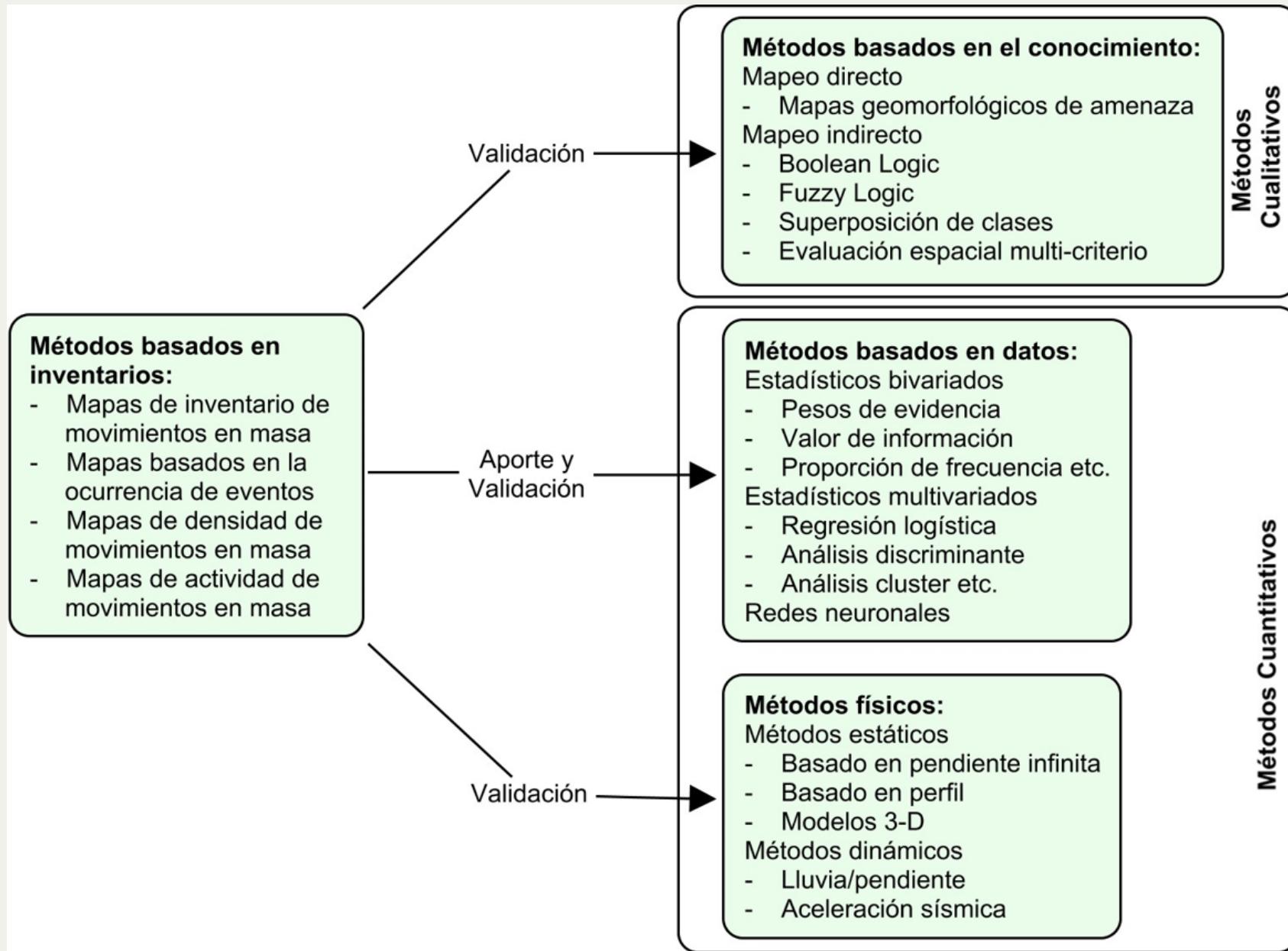
UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN

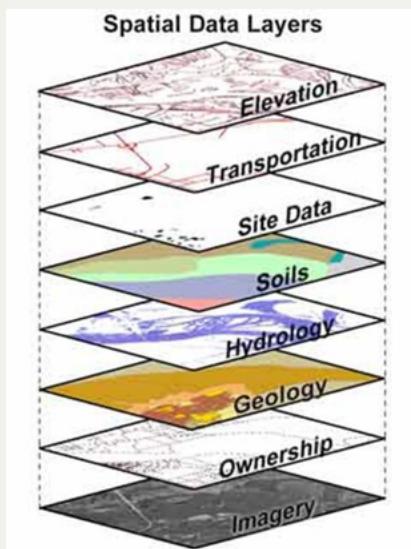
Oct, 1 / 2020

Contenido

- Métodos
- Modelos basados en datos
- Modelos bivariados
- Likelihood
- Statistical Index model
- Weights of Evidence
- Evidential belief function
- Certainty factor



Métodos Estadístico (data-driven)



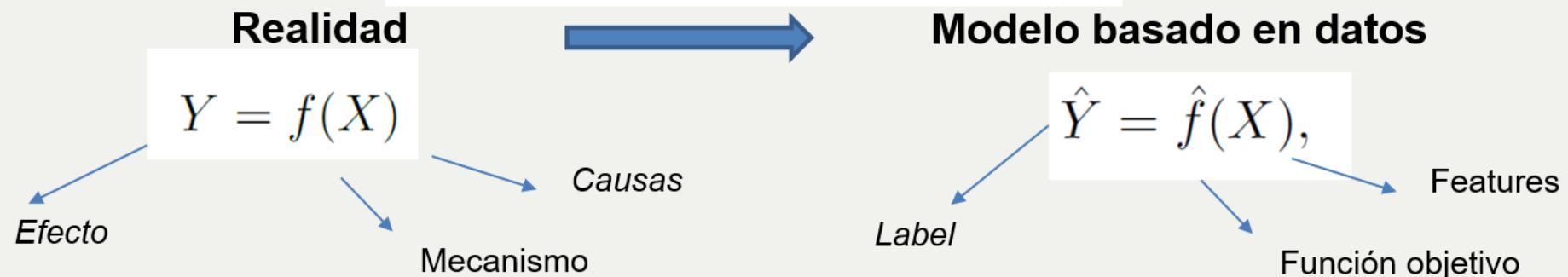
(b)

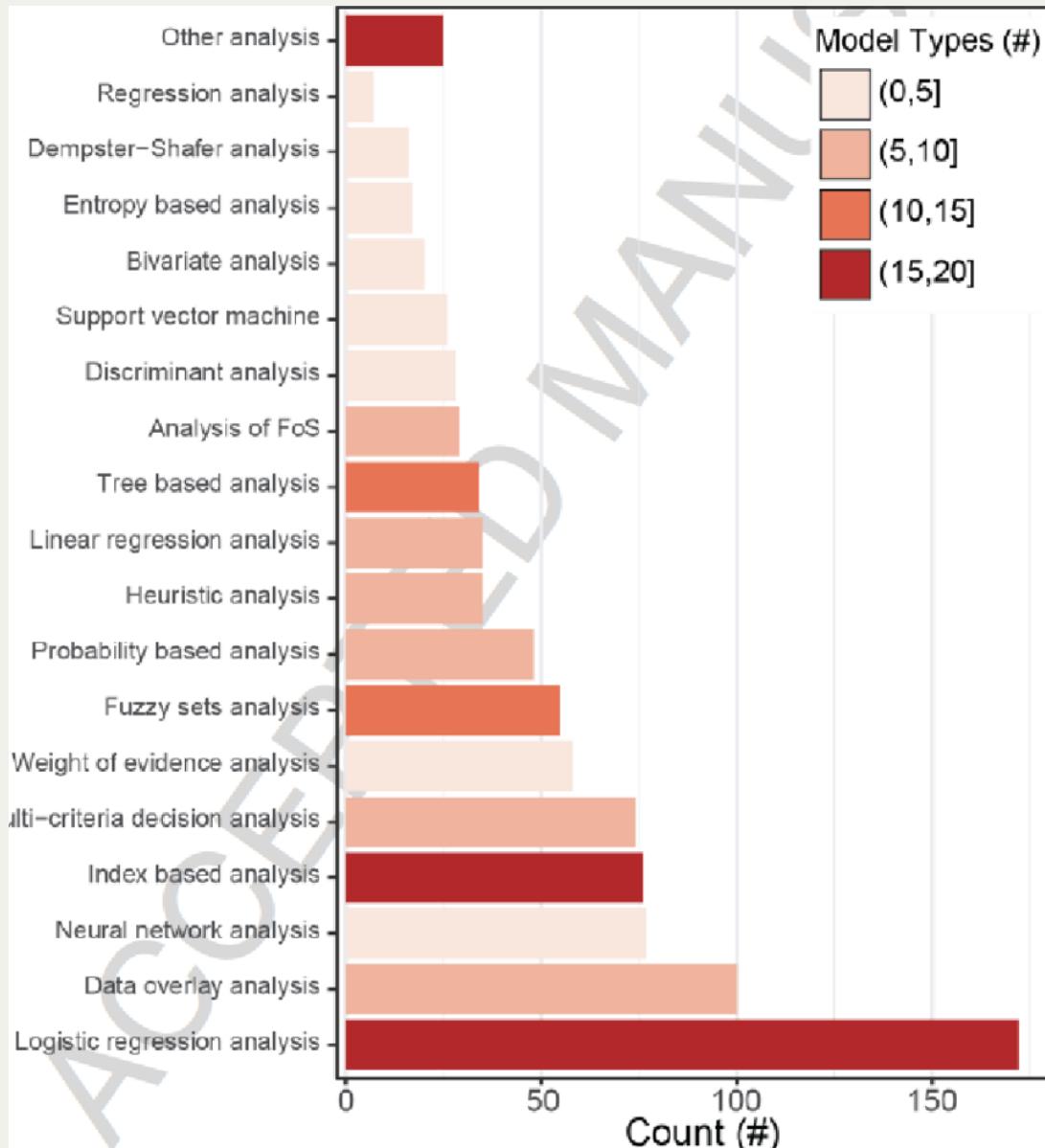
D features (attributes)

N cases

Color	Shape	Size (cm)	Label
Blue	Square	10	1
Red	Ellipse	2.4	1
Red	Ellipse	20.7	0

$$OutputVariable = f(InputVariables)$$





Fuente: Reichenbach et al. (2018)

Los métodos estadísticos se ajustan mejor para movimientos en masa:

- Que no se mueven mucho de su área fuente
- Que no cambian significativamente su tamaño y geometría durante el movimiento (rotacional o translacional).

Y menos aplicables a movimientos en masa:

- Movimientos en masa que viajan largas distancias (cientos de miles de metros)
- Que cambian significativamente su volumen y geometría durante su movimiento desde el área fuente a la zona depositacional.
- Para debris flow se recomienda generalmente modelos con base física.

La falta de algunos movimientos en masa en el inventario puede no representar un problema para el modelo de susceptibilidad, en tanto que el inventario tenga una razonable representación de la abundancia y distribución de los movimientos en la zona de estudio. Consistencia es mas importante que este completo.

Fuente: Reichenbach et al. (2018)

Table 5 Recommended methods for data-driven landslide susceptibility assessment

	Method	References
Bivariate statistical methods	Likelihood ratio model (LRM)	Lee (2005)
	Information value method	Yin and Yan (1988)
	Weights of evidence modelling	van Westen (1993), Bonham-Carter (1994), Suzen and Doyuran (2004)
	Favourability functions	Chung and Fabbri (1993), Luzzi (1995)
Multivariate statistical method	Discriminant analysis	Carrara (1983), Gorsevski et al. (2000)
	Logistic regression	Ohlmacher and Davis (2003), Gorsevski et al. (2006a)
ANN	Artificial neural networks	Lee et al. (2004), Ermini et al. (2005), Kanungo et al. (2006)

Fuente: Corominas et al. (2014)

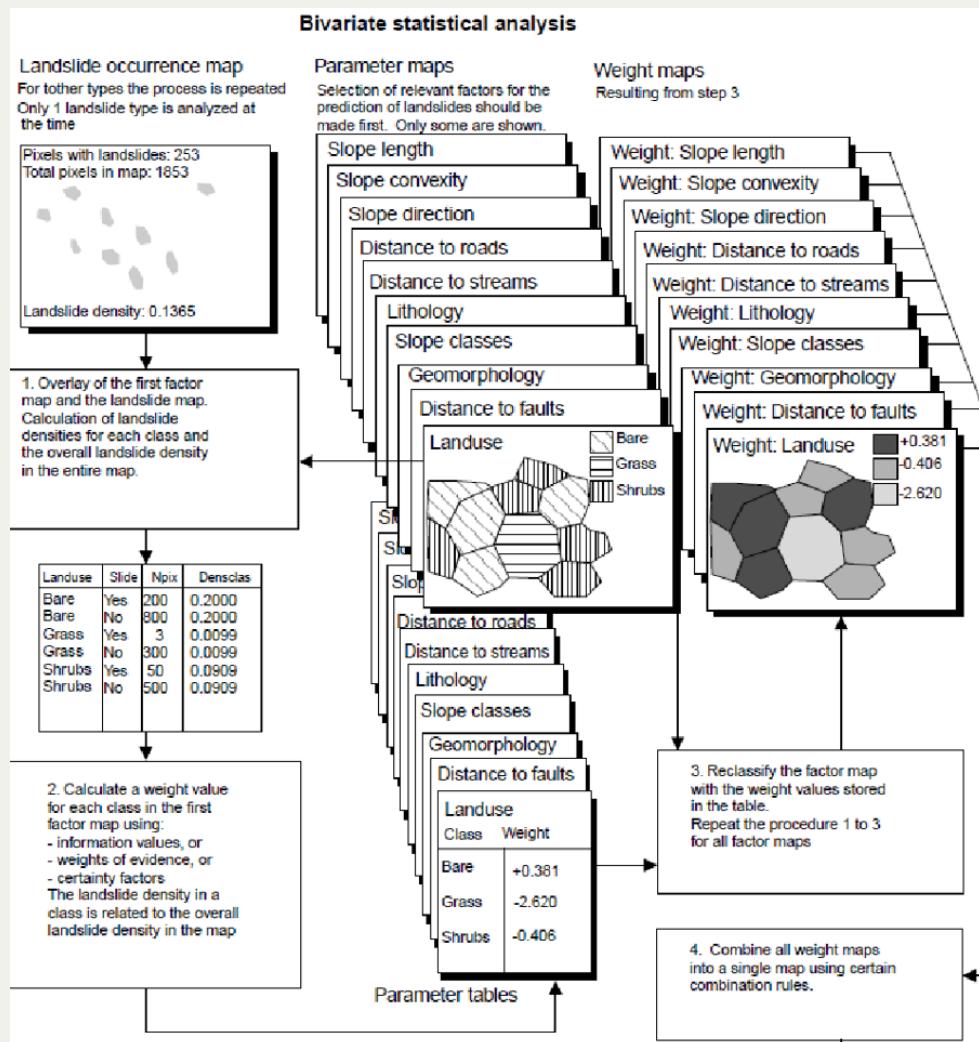
Método Estadístico Bivariado

Cada mapa de factor (pendientes, geología, uso del suelo, etc,) es combinado con el mapa de distribución de movimientos en masa, y los valores de los pesos, basados generalmente en densidad de deslizamientos, son calculados para cada clase de la variable (rango de pendiente, unidad litológica, tipo de uso del suelo, etc,).

- Los métodos estadísticos bivariados son una buena herramienta de aprendizaje para el analista encontrar qué factores o combinación de factores juegan un papel importante en la ocurrencia de movimientos en masa.
- Pueden ser combinados con métodos heurísticos
- Están basados en la asociación observada entre la distribución de los movimientos en masa y cada clase.
- Todos los parámetros tienen una importancia igual sobre la ocurrencia de deslizamientos.

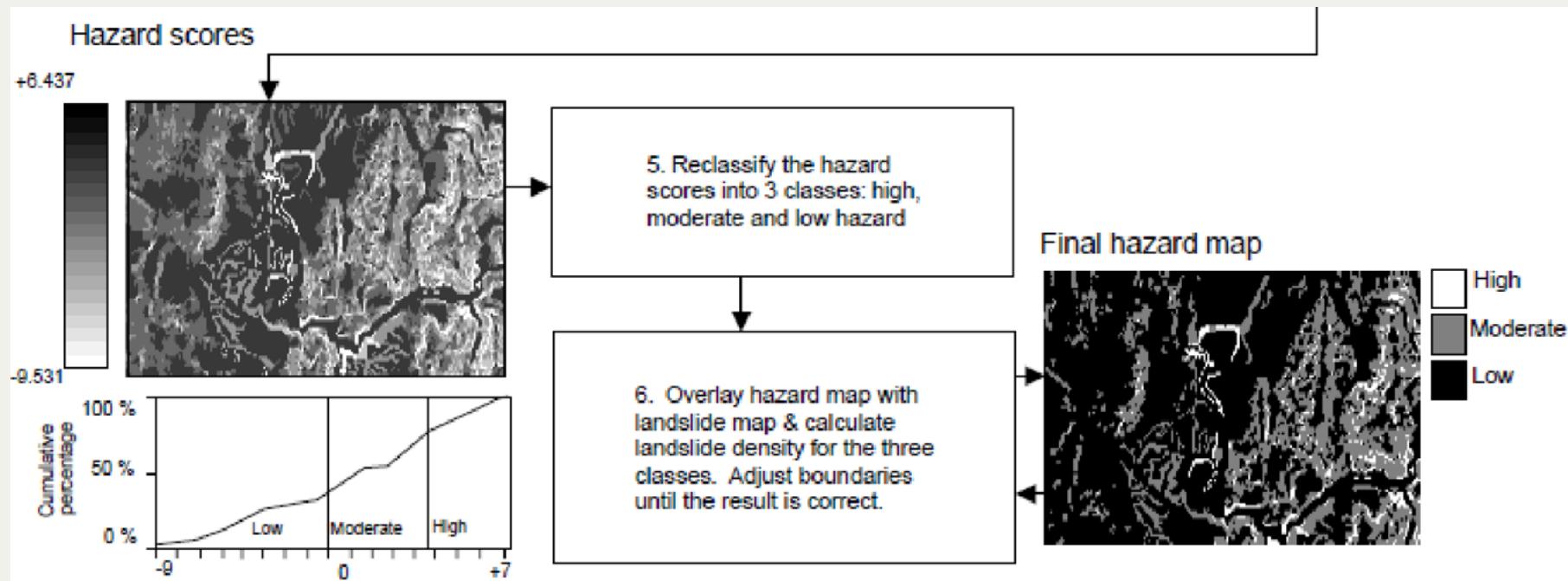
Fuente: AGS (2007)

Procedimiento (i)



Fuente: van Westen et al, (1997)

Procedimiento (ii)



Landslide Susceptibility Index:

$$LSI = \sum FR = Fr_1 + Fr_2 + Fr_3 + \dots + Fr_n$$

Fuente: van Westen et al, (1997)

Likelihood - Frequency Ratio Model

$$AR_i \equiv \frac{C_i}{C_t} = \frac{C_i}{\sum_{i=1}^n C_i},$$

C_i es el número de celdas con valor i, C_t es el total de número de celdas de la zona de studio, n es el total de valores enteros en X.

$$LR_i \equiv \frac{L_i}{L_t} = \frac{L_i}{\sum_{i=1}^n L_i}.$$

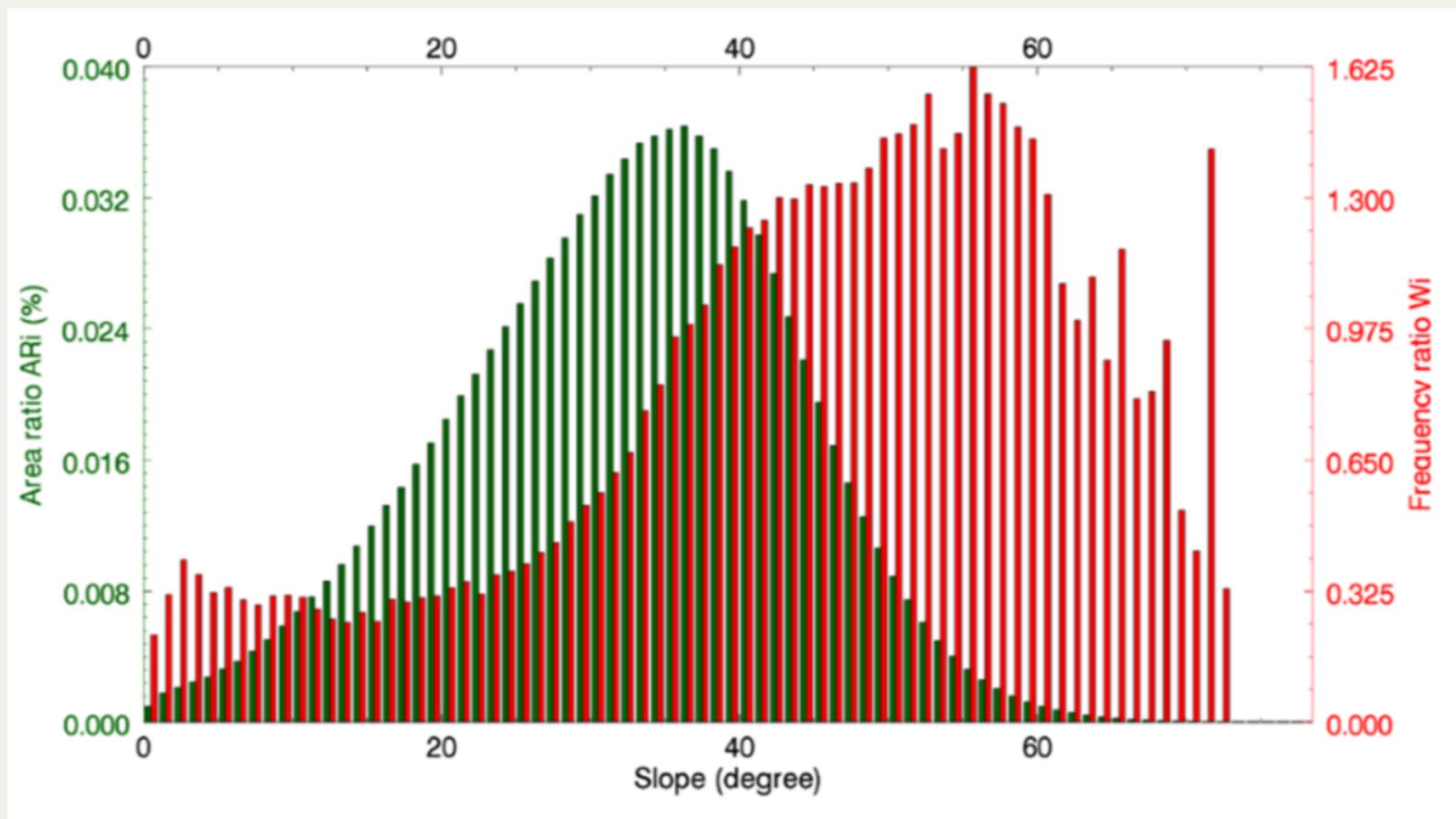
L_i es la cuenca de factor X con valor i que caen dentro del area de deslizamientos, L_t es el total de area con deslizamientos, n es el número de valores enteros de X el número de factores X.

$$W_i^{WF} \equiv \frac{LR_i}{AR_i}.$$

La relación de frecuencia (FR) de la clase de cada factor condicionante es calculada dividiendo la relación de la ocurrencia de movimientos en masa por la relación del área de la clase.

Parameter	Classes	% of total area (a)	% of landslide area (b)	Frequency ratio (b/a)
Geology ^a	Alv	3.86	1.54	0.40
	Pl	2.39	2.23	0.94
	Ev	36.63	13.17	0.36
	Kru	54.65	80.02	1.46
	Jkr	0.11	0.00	0.00
	Jlh	0.67	0.00	0.00
	γ_2	1.63	3.04	1.87
	Kk	0.07	0.00	0.00
Slope (%)	0–10	13.72	2.41	0.18
	10–20	6.94	5.71	0.82
	20–30	11.07	11.79	1.07
	30–50	26.82	30.83	1.15
	>50	41.45	49.26	1.19

Fuente: Chen et al, (2016); Yalcin et al. (2011)



Fuente: Liu et al. (2016)

Table 2 Likelihood ratio values of the landslide-conditioning parameters

Parameter	Class	Pixels with landslide	Landslide density ^a (%)	Numbers of pixels in domain	Percentage of domain ^b (%)	Likelihood ratio ^c
Slope gradient	0–10	1,025	18.88	124,261	18.02	
	10–20	1,901	35.02	267,775	38.83	
	20–30	2,053	37.82	219,276	31.80	?
	30–40	399	7.35	64,470	9.35	
	40–50	50	0.92	7,075	1.03	
	>50	0	0.00	6,668	0.97	
Lithology	Slope debris	815	15.02	187,257	27.16	
	Alluvium	0	0	84,915	12.31	
	Microgabro–Gabro	33	0.61	7,925	1.15	?
	Granodiorite–quartz microdiorite	17	0.31	15,497	2.25	
	Andesite and pyroclast	2,598	47.89	265,717	38.54	
	Dacite and pyroclast	1,962	36.17	128,214	18.60	

^a Ratio of landslides occurred

^b Ratio of landslides not occurred

^c Ratio of landslides occurred divided by ratio of landslides not occurred

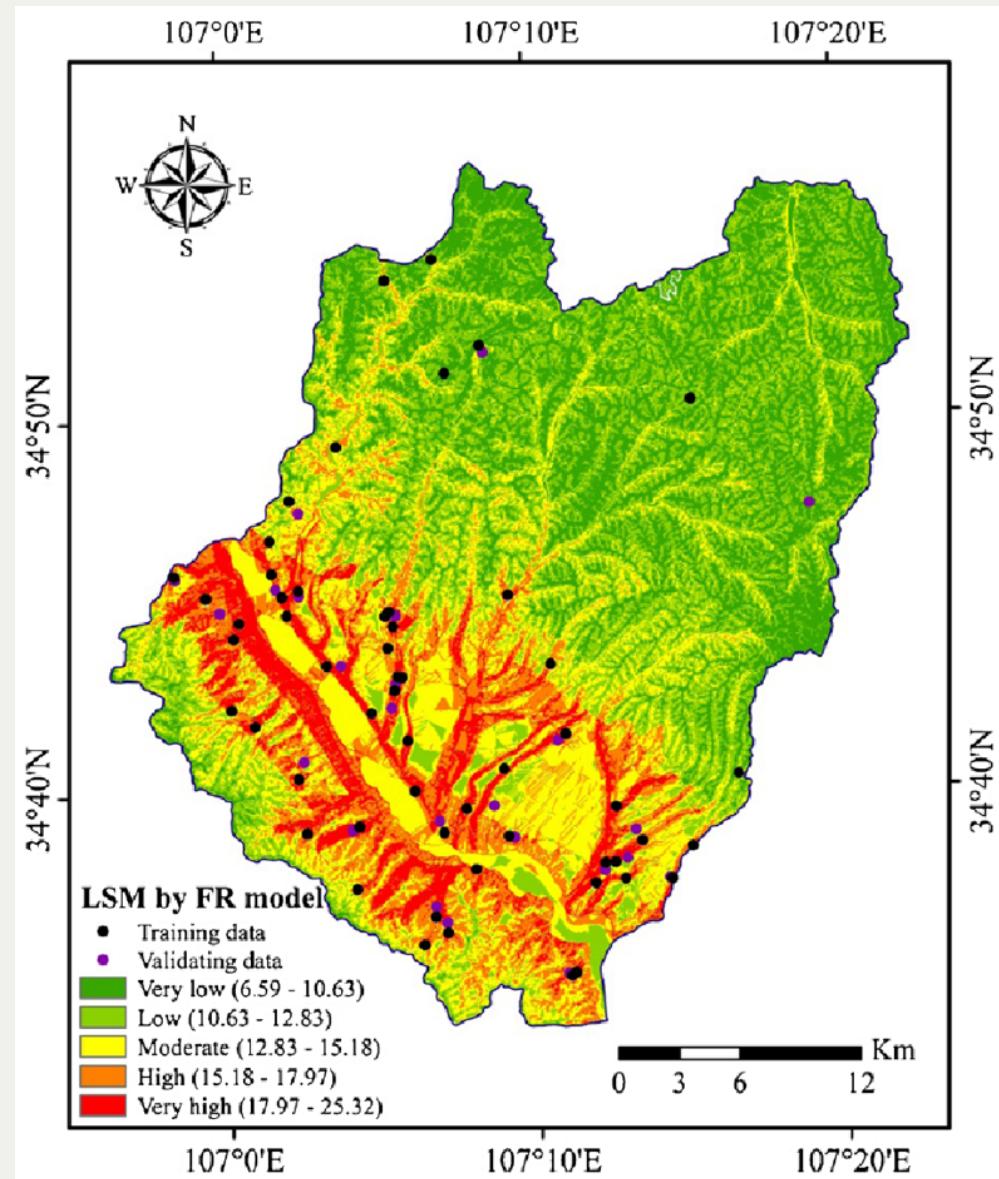
Fuente: Akgun et al (2008)

Table 2 Likelihood ratio values of the landslide-conditioning parameters

Parameter	Class	Pixels with landslide	Landslide density ^a (%)	Numbers of pixels in domain	Percentage of domain ^b (%)	Likelihood ratio ^c
Slope gradient	0–10	1,025	18.88	124,261	18.02	1.05
	10–20	1,901	35.02	267,775	38.83	0.90
	20–30	2,053	37.82	219,276	31.80	1.19
	30–40	399	7.35	64,470	9.35	0.79
	40–50	50	0.92	7,075	1.03	0.90
	>50	0	0.00	6,668	0.97	0.00
Lithology	Slope debris	815	15.02	187,257	27.16	0.55
	Alluvium	0	0	84,915	12.31	0.00
	Microgabro–Gabro	33	0.61	7,925	1.15	0.53
	Granodiorite–quartz microdiorite	17	0.31	15,497	2.25	0.14
	Andesite and pyroclast	2,598	47.89	265,717	38.54	1.24
	Dacite and pyroclast	1,962	36.17	128,214	18.60	1.95

^a Ratio of landslides occurred^b Ratio of landslides not occurred^c Ratio of landslides occurred divided by ratio of landslides not occurred

Fuente: Akgun et al (2008)



Fuente: Chen et al. (2016)

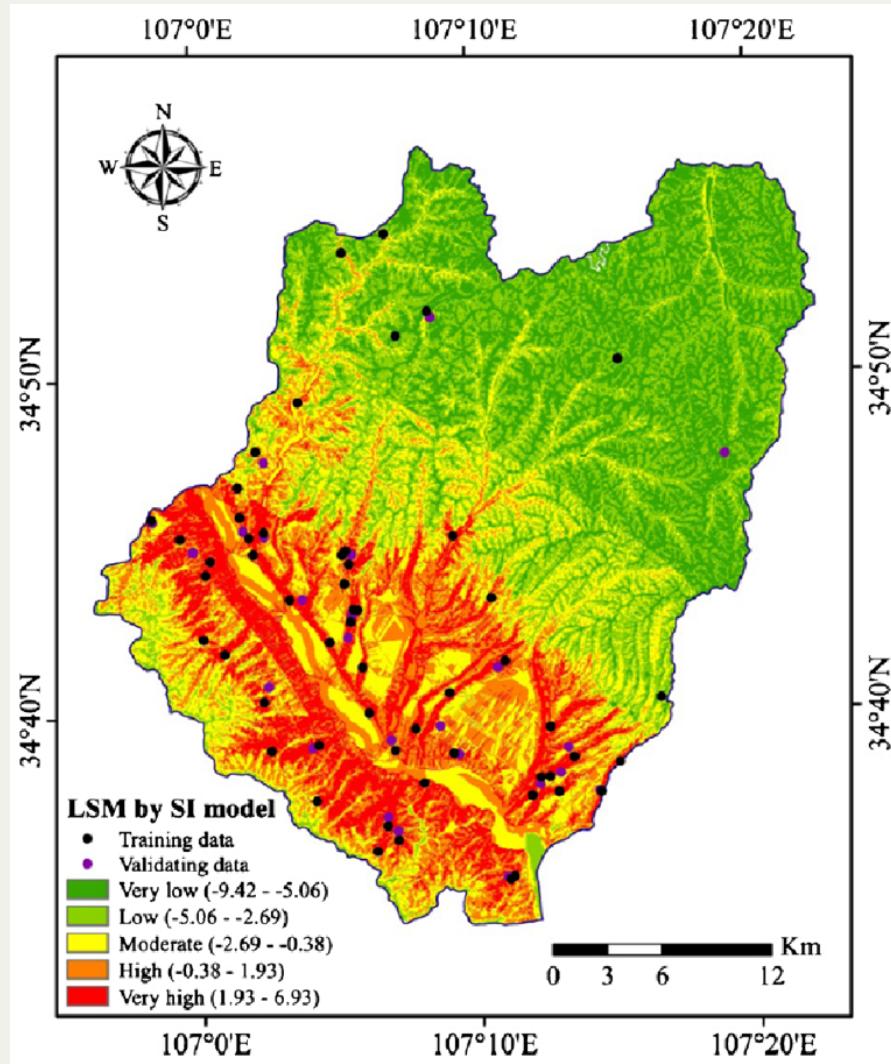
Statistical Index Model

El valor del peso para cada unidad categórica es definido como el logaritmo natural de la densidad de deslizamientos en una clase dividido la densidad de deslizamientos en toda la cuenca, de acuerdo con la siguiente ecuación:

$$W_{ij} = \ln\left(\frac{D_{ij}}{D}\right) = \ln\left[\frac{N_{ij}}{S_{ij}} / \frac{N}{S}\right] = \ln\left[\frac{\text{DensClass}}{\text{DensMap}}\right]$$

Donde W_{ij} es el peso dado a una cierta clase i de un parámetro j . D_{ij} es la densidad de deslizamientos de la clase i del parámetro j . D es la densidad total de deslizamientos de todo el mapa. N_{ij} es el número de píxeles en una cierta clase i del parámetro j . N es el total de celdas que corresponden a deslizamientos en todo el mapa. S es el total de píxeles del mapa.

Fuente: Chen et al, (2016); van Westen (1997)



Fuente: Barredo et al. (2000)

Weights-of-Evidence (WoE)

Ej. 1. Si tomamos $p(50)=2/3$, entonces:

$$Odds = \frac{p(50)}{1 - p(50)} = \frac{2/3}{1 - 2/3} = 2/1$$

Cuando $x=50$, un éxito es 2 veces tan probable como una falla. Por lo que se puede decir simplemente que los **Odds** son 2 a 1.

- $p = 0 \rightarrow O(p) = 0$
- $p = 1/4 \rightarrow O(p) = 1/3$ ("Odds are 1-to-3 against")
- $p = 1/2 \rightarrow O(p) = 1$ ("Even odds")
- $p = 3/4 \rightarrow O(p) = 3$ ("Odds are 3-to-1 in favor")
- $p = 1 \rightarrow O(p) = \infty$

Ej. 2. Los **Odds** que un caballo gane la carrera es 1 a 3, lo que significa que de cada 4 carreras que corra ganará 1 y perderá 3. Por lo tanto me deben pagar \$3 por cada peso que apueste.

La probabilidad priori $P(c)$ es usualmente estimada empíricamente con el conocimiento acerca de la ocurrencia de un evento c en el pasado bajo iguales condiciones. Cuando la probabilidad de la evidencia $P(x)$ es integrada en el cálculo de la probabilidad, se conoce como probabilidad condicionada o posterior $P(c|x)$, es decir la probabilidad que un evento x ocurrirá bajo la presencia de una evidencia x . El teorema de Bayes integra ambas probabilidades.

The diagram shows the formula for posterior probability $P(c|x)$ with arrows pointing from its components to their respective labels:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

- Likelihood: $P(x|c)$
- Class Prior Probability: $P(c)$
- Posterior Probability: $P(c|x)$
- Predictor Prior Probability: $P(x)$

Below the formula, the joint probability expression is shown:

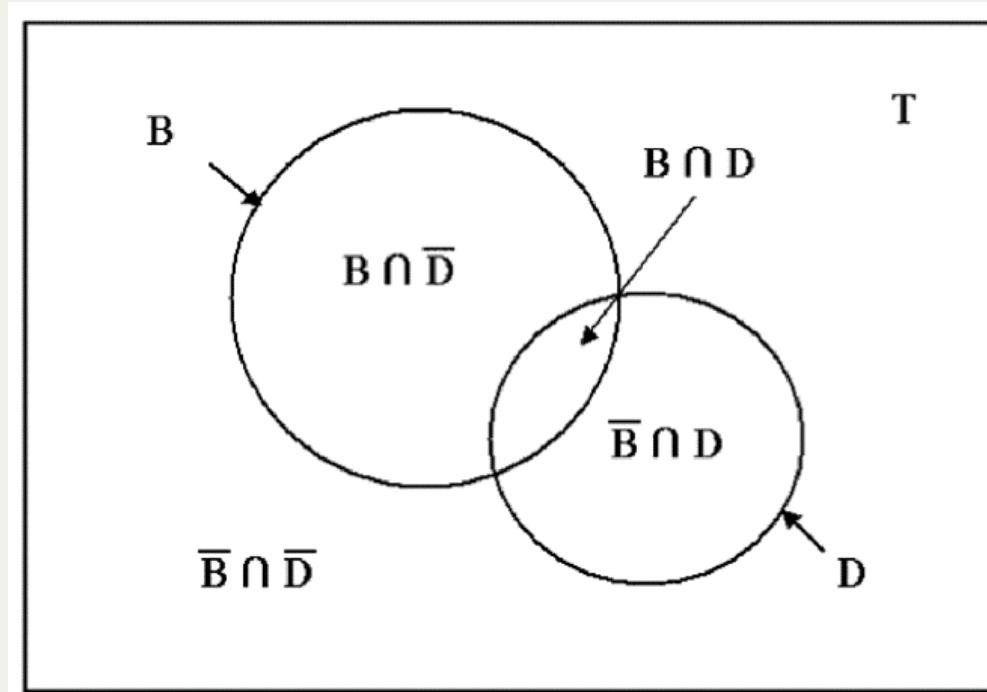
$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \cdots \times P(x_n|c) \times P(c)$$

Este método está basado en la teoría de probabilidad Bayesiana. En el cual para determinar la relativa importancia de los datos se utiliza los conceptos de probabilidad priori y posterior.

$$C = W_j^+ - W_j^- = \ln \left(\frac{P\{D | B_i\}}{P\{D | \bar{B}_i\}} \right) / \left(\frac{P\{\bar{D} | B_i\}}{P\{\bar{D} | \bar{B}_i\}} \right)$$

Fuente: Lee et al., (2002); Vahidnia et al (2009)

Un peso positivo de W_+ indica que la presencia del factor es favorable para la ocurrencia de deslizamientos, y la magnitud indica una correlación negativa. Un valor positivo de W_- señala que la ausencia del factor de causa es favorable para la ocurrencia de deslizamientos, **Contraste (C)**, y refleja la asociación espacial total entre una variable predictora y la ocurrencia de deslizamientos. Un valor de 0 indica que la clase considerada del factor de causa no es significativa en el análisis.



$$P\{D\} = \frac{N\{D\}}{N\{T\}}$$

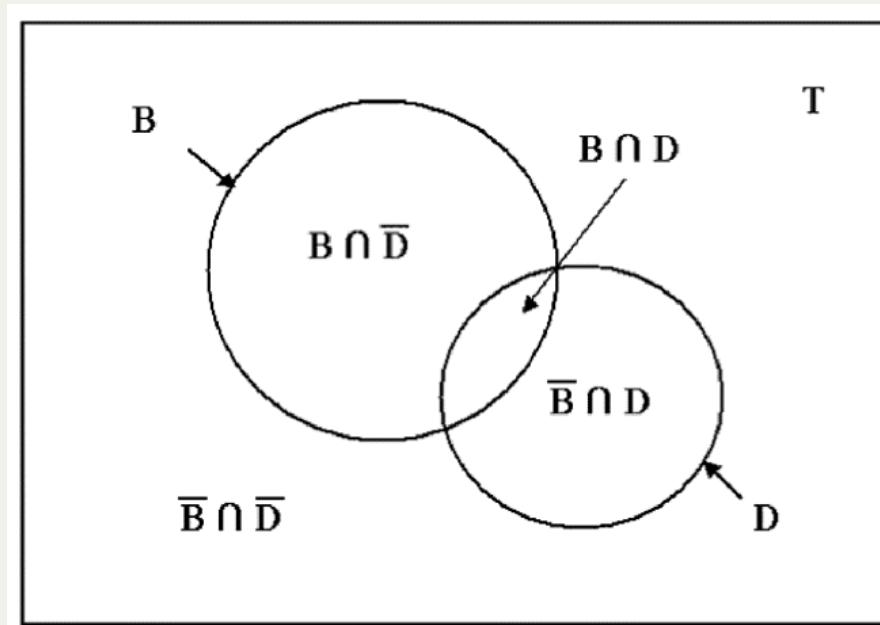
Probabilidad posterior

$$P\{D|B\} = \frac{P\{D \cap B\}}{P\{B\}} = P\{D\} \frac{P\{B|D\}}{P\{B\}}$$

Likehood

Probabilidad priori

Evidencia



$$O\{D|B\} = O\{D\} \frac{P\{B|D\}}{P\{B|\bar{D}\}} \rightarrow \text{Presencia}$$



$$W^+ = \log_e \frac{P\{B|D\}}{P\{B|\bar{D}\}}$$

$$\log_e O\{D|B\} = \log_e O\{D\} + W^+$$

$$O\{D|\bar{B}\} = O\{D\} \frac{P\{\bar{B}|D\}}{P\{\bar{B}|\bar{D}\}} \rightarrow \text{No presencia}$$



$$W^- = \log_e \frac{P\{\bar{B}|D\}}{P\{\bar{B}|\bar{D}\}}$$

$$\log_e O\{D|\bar{B}\} = \log_e O\{D\} + W^-$$

N_{pix} = number of pixels	B_i : Potential landslide conditioning factor		
	(Present)	(Absent)	
S: Landslides	Present	N_{pix_1}	N_{pix_2}
	Absent	N_{pix_3}	N_{pix_4}

$$W_I^+ = \log_e \frac{\frac{N_{pix_1}}{N_{pix_1} + N_{pix_2}}}{\frac{N_{pix_3}}{N_{pix_3} + N_{pix_4}}}$$

$$W_I^- = \log_e \frac{\frac{N_{pix_2}}{N_{pix_1} + N_{pix_2}}}{\frac{N_{pix_4}}{N_{pix_3} + N_{pix_4}}}$$

La probabilidad final (P) de cada celda es la suma de los pesos de cada parámetro y la probabilidad priori ($Pp(s)$)

$$P = \exp \left\{ \sum W^+ + \ln Pp(s) \right\}$$

$$Pp(s) = \frac{\text{Number of landslide pixels}}{\text{Number of total study area pixels}}$$

Fuente: Lee et al. (2002); van Westen et al (2003)

$$W^+ = \ln \left[\frac{\frac{A1}{A1+A2}}{\frac{A3}{A3+A4}} \right]$$

$$W^- = \ln \left[\frac{\frac{A2}{A1+A2}}{\frac{A4}{A3+A4}} \right]$$

- **A1** es el número de píxeles con deslizamientos en una clase
- **A2** es el número de píxeles con deslizamientos que no están presentes en dicha clase
- **A3** es el número de píxeles en la clase en los cuales no hay deslizamientos
- **A4** es el número de píxeles en la clase donde no hay deslizamientos y que tampoco el factor es presente.

Fuente: Ozdemir & Altural (2013)



Pertenecen a la clase B pero no son MenM

Pertenecen a la clase y son MenM

No pertenecen a la clase B y no son MenM

No pertenecen a la clase B y son MenM

$$A_1 = 1$$

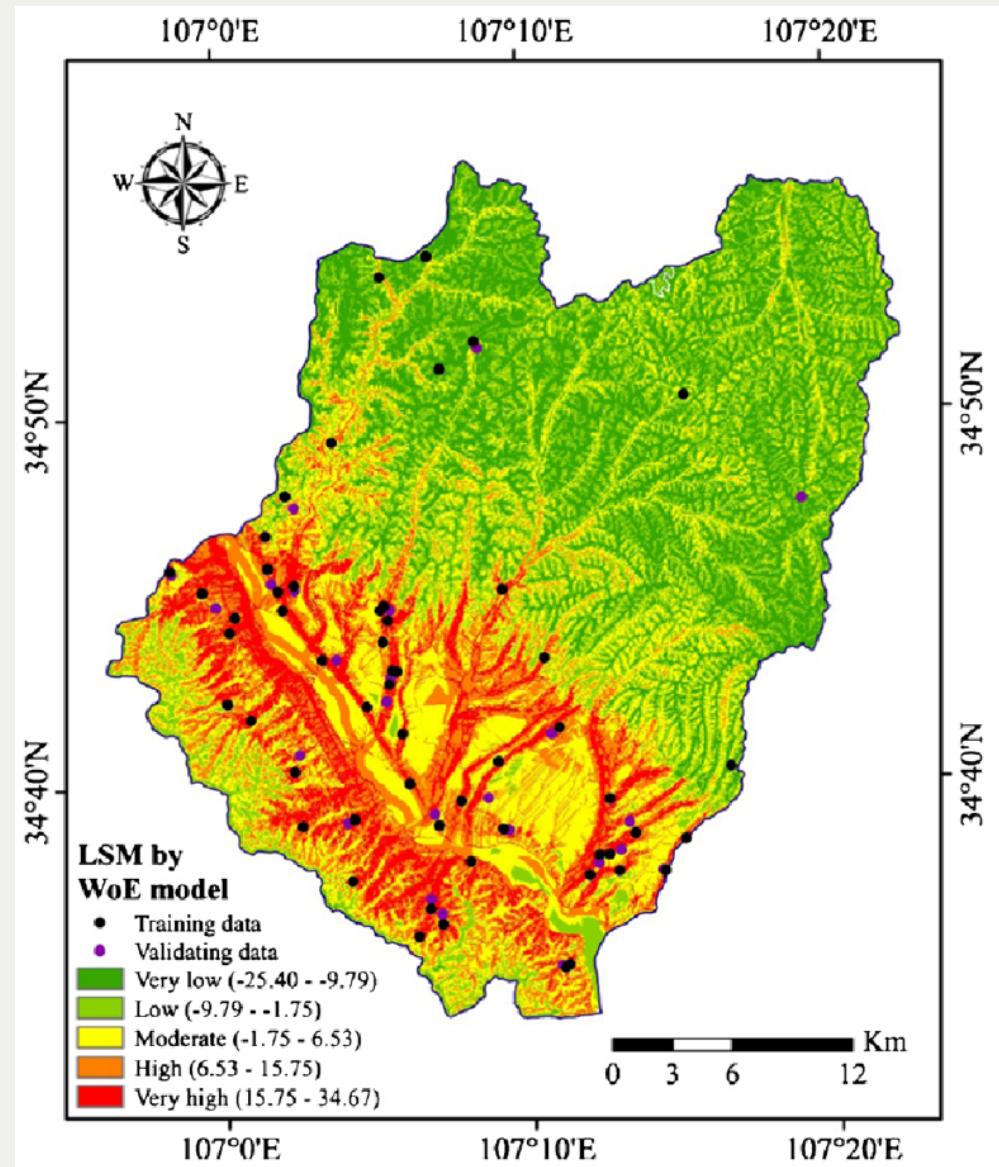
$$A_2 = 2$$

$$A_3 = 17$$

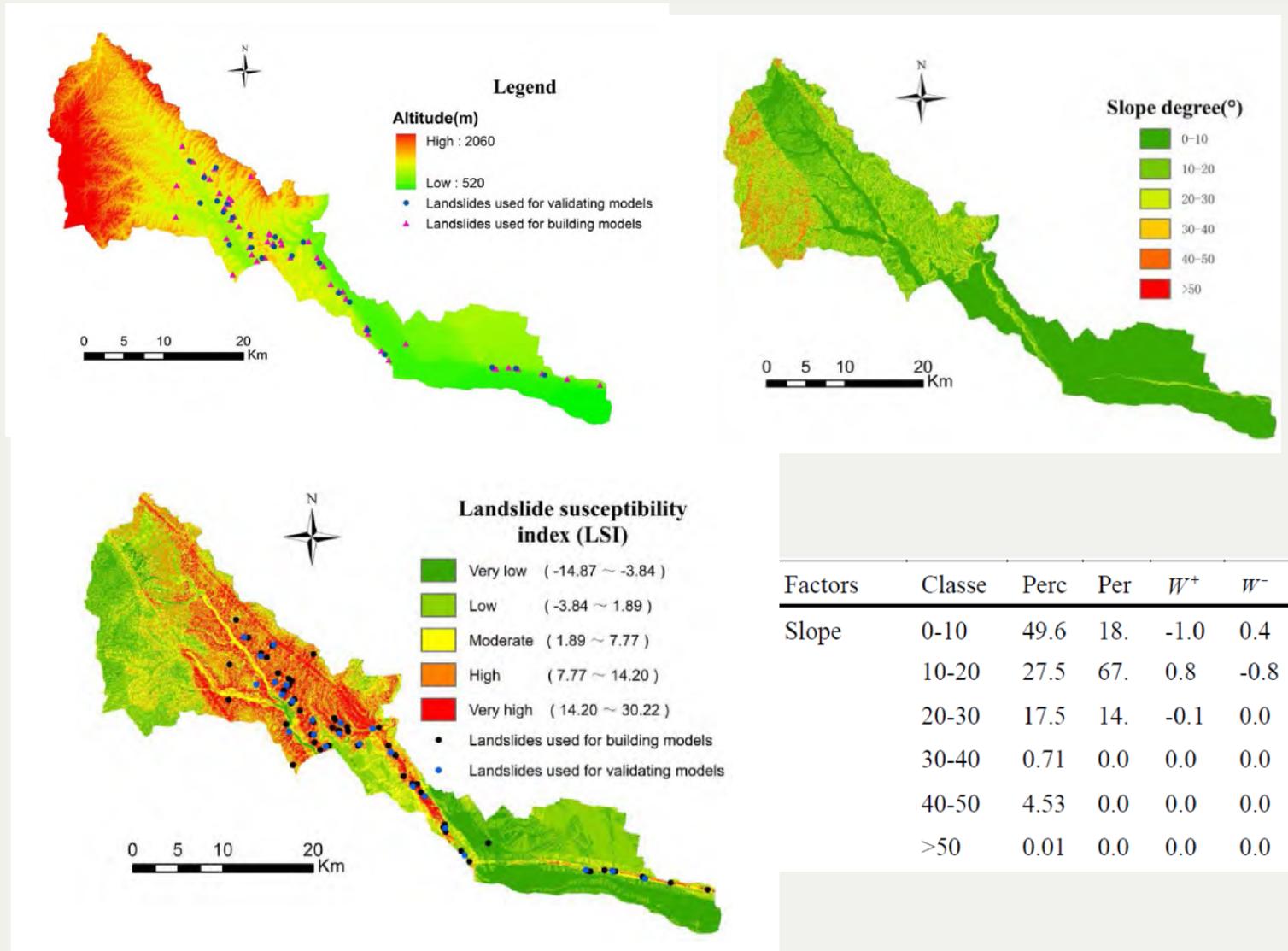
$$A_4 = 46$$

$$W^+ = ?$$

$$W^- = ?$$



Fuente: Chen et al. (2016)



Factors	Classe	Perc	Per	W^+	W^-	C
Slope	0-10	49.6	18.	-1.0	0.4	-1.4
	10-20	27.5	67.	0.8	-0.8	1.6
	20-30	17.5	14.	-0.1	0.0	-0.2
	30-40	0.71	0.0	0.0	0.0	0.0
	40-50	4.53	0.0	0.0	0.0	0.0
	>50	0.01	0.0	0.0	0.0	0.0

Fuente: Chen & Li (2014)

Evidential Belief Function (EBF)

El modelo EBF (Dempster - Shafer) consiste en el grado de Belief (Bel), Disbelief (Dis), Uncertainty (Unc) y Plausibility (Pls), con valores entre 0 y 1. Bel y Pls representan el límite de probabilidad bayesiana inferior y superior. Mientras que, Pls es mayor o igual que Bel, y Unc es la diferencia entre ellos, indicando la incertidumbre del resultado (Ghosh & Carranza, 2010; Chen et al., 2016).

$$Bel_{C_{ij}} = \frac{Wc_{ij}L}{\sum_{j=1}^m Wc_{ij}L}$$

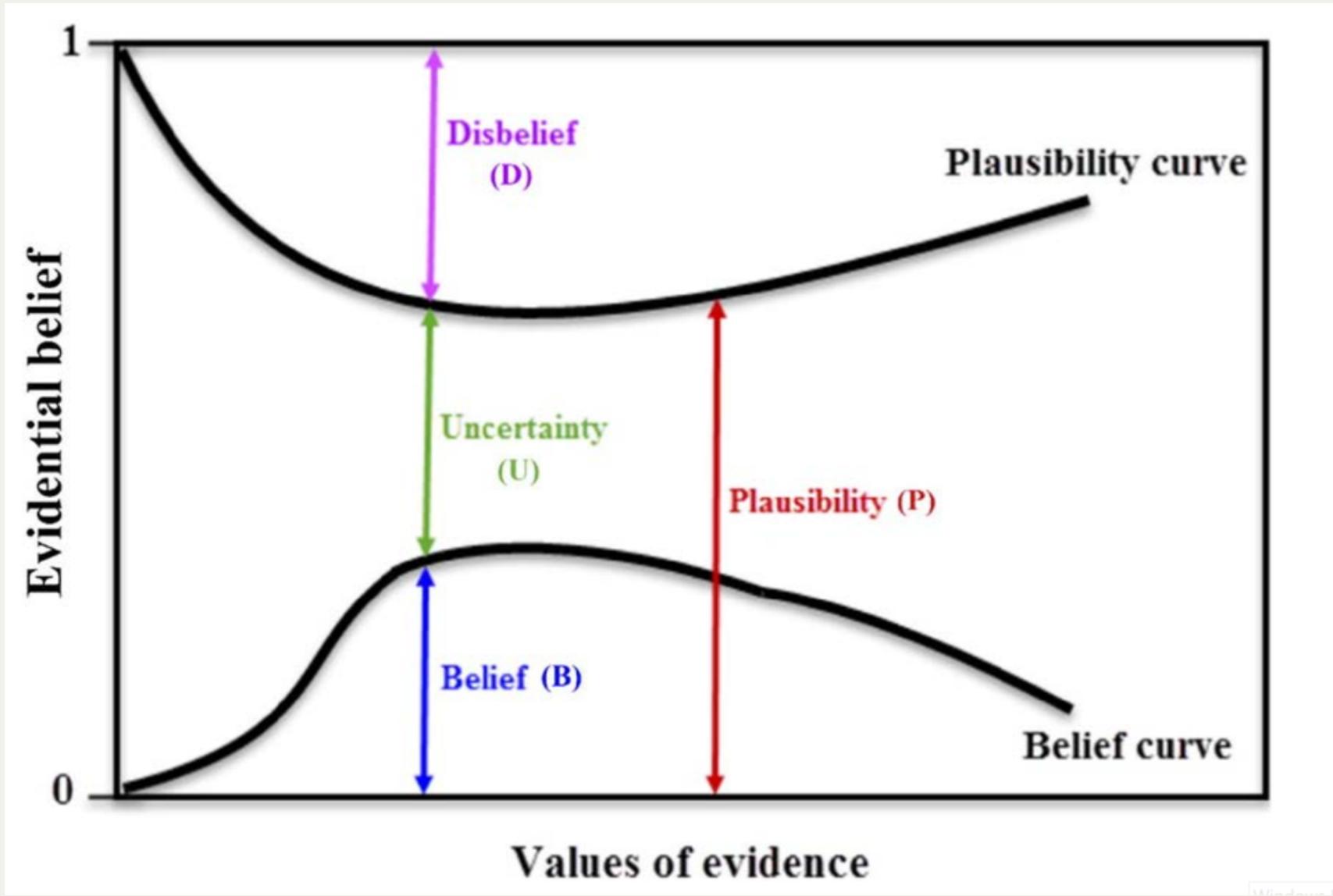
$$Wc_{ij}L = \frac{\frac{N(C_{ij} \cap L)}{N(C_{ij})}}{\frac{N(L) - N(C_{ij} \cap L)}{N(S) - N(C_{ij})}}$$

$$Disc_{ij} = \frac{Wc_{ij}\bar{L}}{\sum_{j=1}^m Wc_{ij}\bar{L}}$$

$$Wc_{ij}\bar{L} = \frac{\frac{N(C_{ij}) - N(C_{ij} \cap L)}{N(C_{ij})}}{\frac{N(S) - N(L) - [N(C_{ij}) - N(C_{ij} \cap L)]}{N(S) - N(C_{ij})}}$$

$$Unc_{C_{ji}} = 1 - \left(W_{C_{ji}R} / \sum_{j=1}^m , W_{C_{ji}R} \right) - \left(W_{C_{ji}\bar{R}} / \sum_{j=1}^m , W_{C_{ji}\bar{R}} \right) = 1 - Bel_{C_{ji}} - Dis_{C_{ji}}.$$

$N(L)$ es el número total de pixeles con deslizamientos. $N(S)$ es el número total de pixeles. Donde j son las variables, i son las clases, por lo que $N(C_{ij})$ es el número de pixeles en la clase C_{ij} .



Fuente: Nampak et al. (2014)

Rockslide locations (■) and spatial evidence X_1

$$N(T) = 792$$

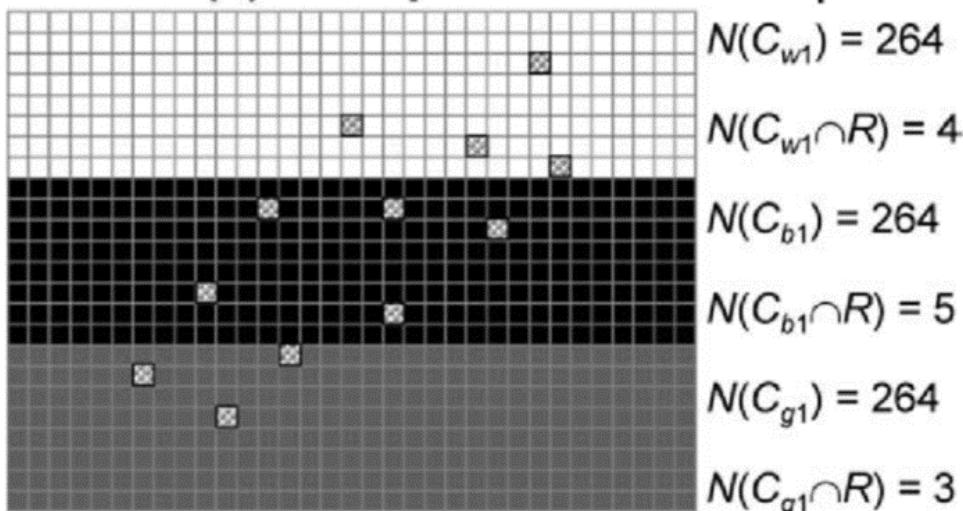
$$N(R) = 12$$

Classes (j) of evidence (C_j):

w = white

b = black

g = grey



$$N(C_{w1}) = 264$$

$$N(C_{w1} \cap R) = 4$$

$$N(C_{b1}) = 264$$

$$N(C_{b1} \cap R) = 5$$

$$N(C_{g1}) = 264$$

$$N(C_{g1} \cap R) = 3$$

Data-driven estimates of Bel_{Cj1} [Eq. (1a)]:

$$Bel_{C_{w1}} = 0.322 \quad Bel_{C_{b1}} = 0.467 \quad Bel_{C_{g1}} = 0.210$$

Data-driven estimates of Dis_{Cj1} [Eq. (2a)]:

$$Dis_{C_{w1}} = 0.333 \quad Dis_{C_{b1}} = 0.332 \quad Dis_{C_{g1}} = 0.334$$

Data-driven estimates of Unc_{Cj1} [Eq. (3)]:

$$Unc_{C_{w1}} = 0.345 \quad Unc_{C_{b1}} = 0.201 \quad Unc_{C_{g1}} = 0.456$$

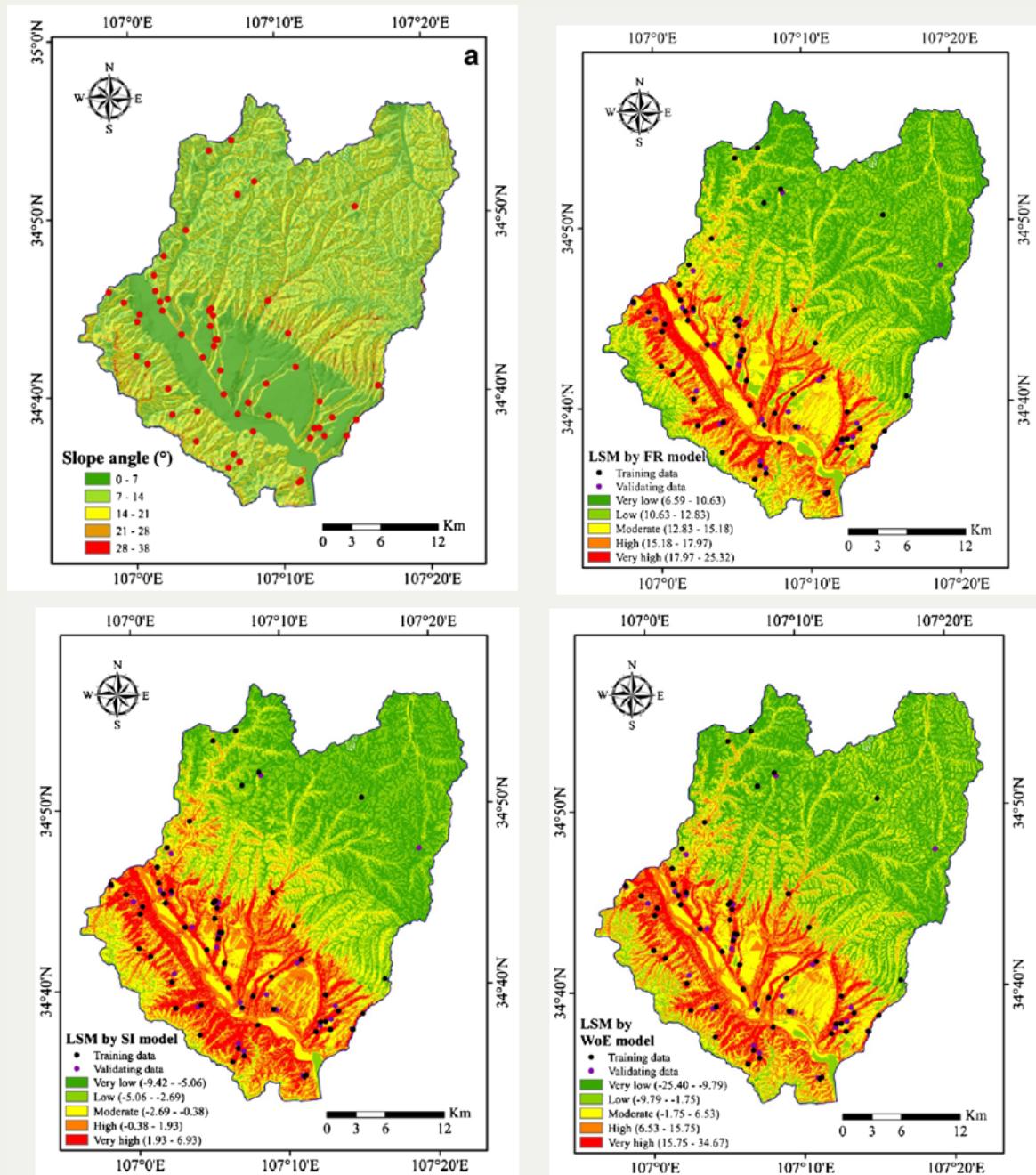
Fuente: Ghosh & Carranza (2010)

Certainty Factor (CF)

El valor de CF varía entre -1 y +1. Un valor positivo significa un incremento en la certidumbre de la ocurrencia de deslizamientos, mientras que un valor negativo corresponde a una reducción en dicha certidumbre. Un valor cercano a 0 significa que la probabilidad priori es muy similar a la probabilidad condicional (Pourghsemi et al. 2012).

$$CF = \begin{cases} \frac{PP_a - PP_s}{PP_a(1 - PP_s)}, & PP_a \geq PP_s \\ \frac{PP_a - PP_s}{PP_s(1 - PP_a)}, & PP_a < PP_s \end{cases}$$

Donde PPa es la probabilidad condicional de la ocurrencia de un deslizamiento en la clase a. PPs es la probabilidad priori del número total de deslizamientos en el área (Kanungo et al. 2011).



Fuente: Chen et al. (2016)