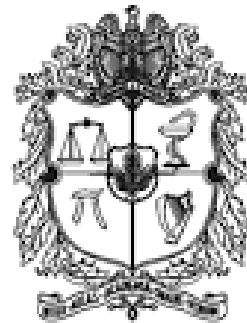


CARTOGRAFÍA GEOTÉCNICA

Métodos Estadísticos Multivariados

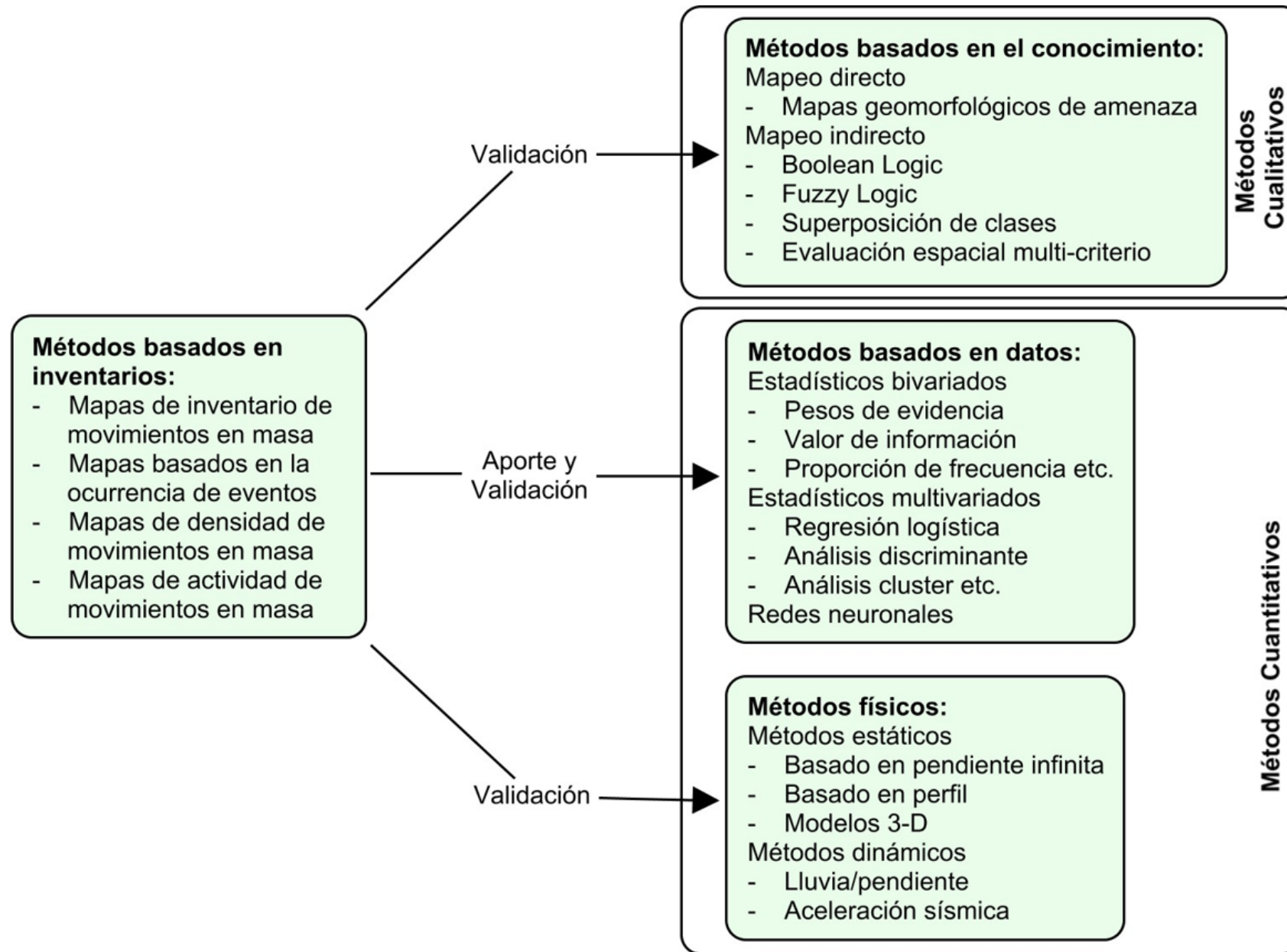
Prof. Edier Aristizábal



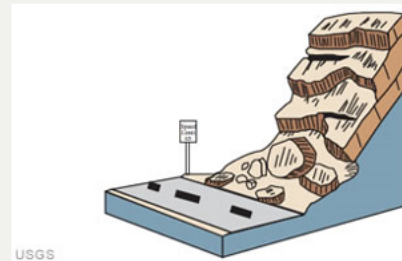
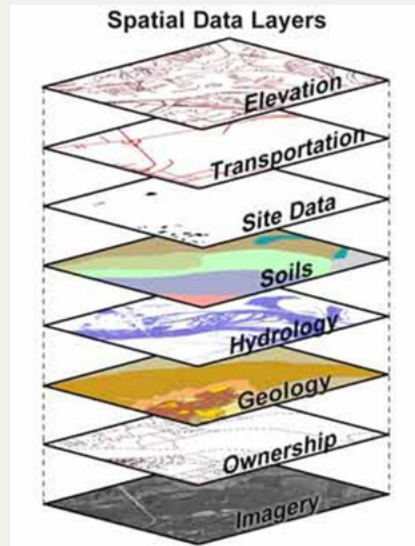
UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN

Oct, 16 / 2020



Métodos Estadístico (data-driven)

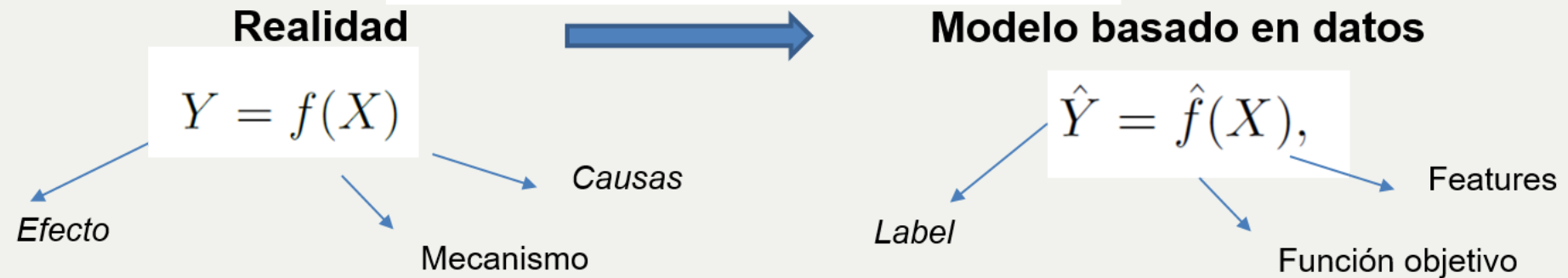


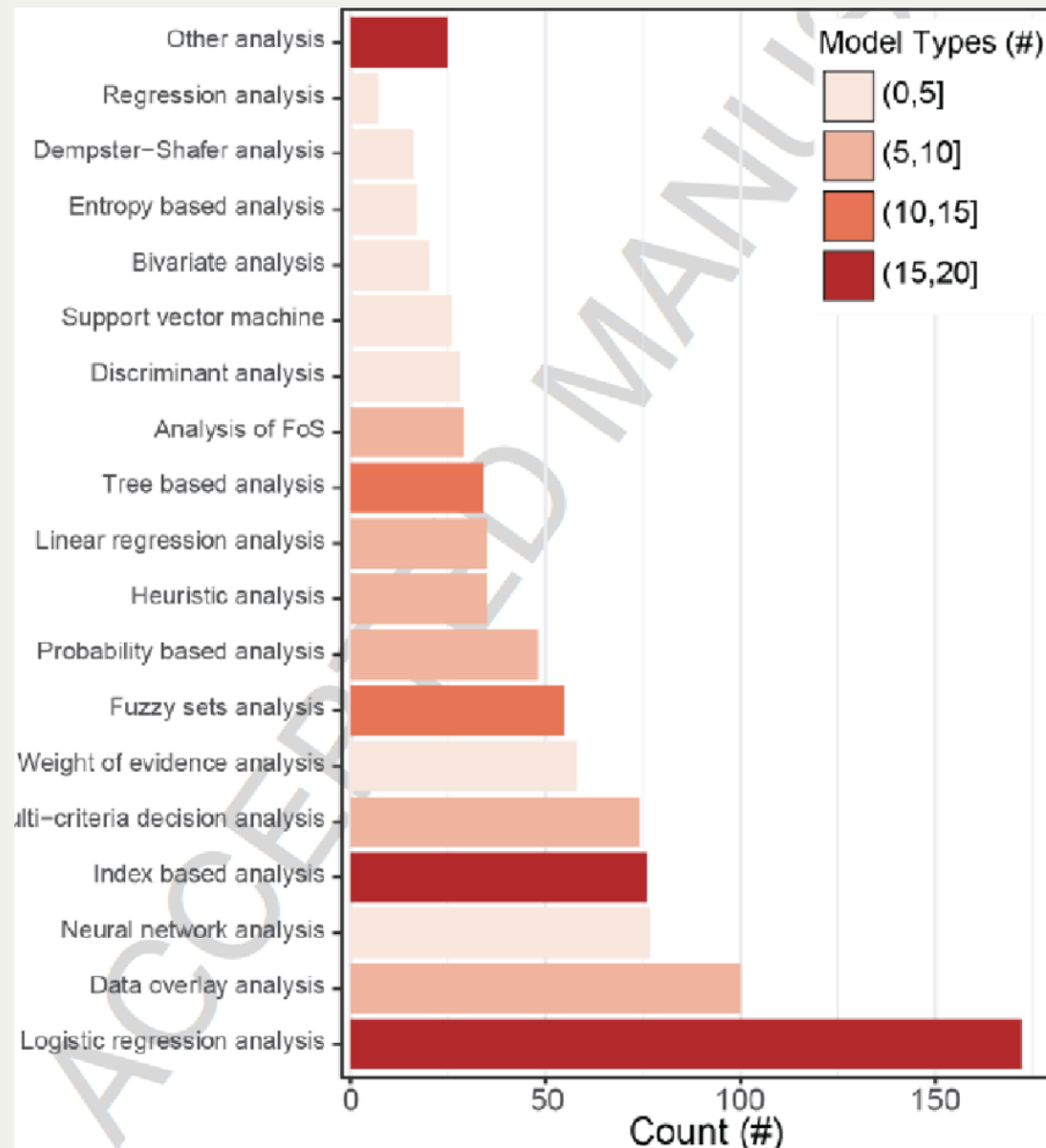
D features (attributes)

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

(b)

$$OutputVariable = f(InputVariables)$$





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 traslacional).

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), Luzi (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 Multivariado

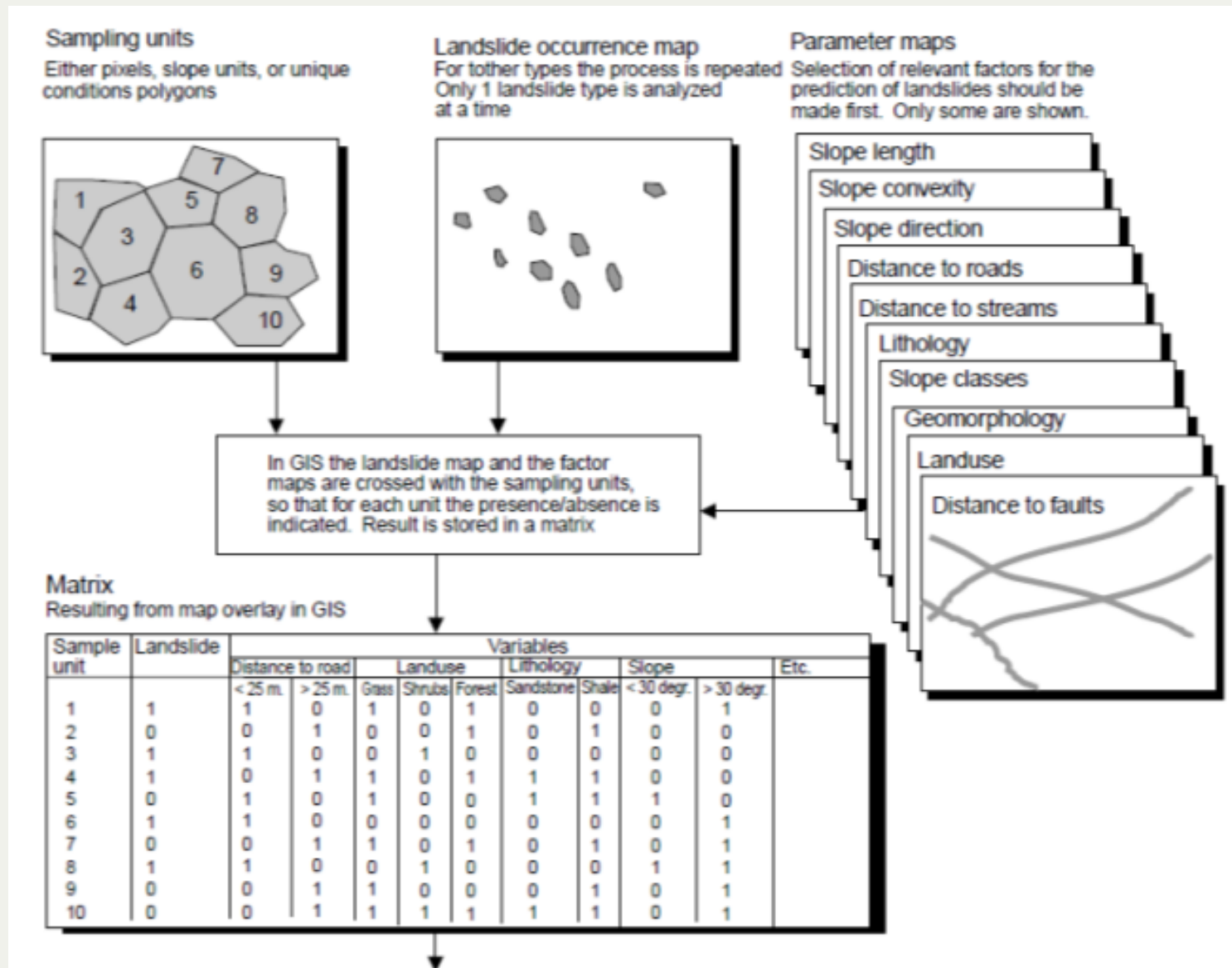
Evalúan la relación combinada entre una variable dependiente (ocurrencia de deslizamientos) y una serie de variables independientes (factores que controlan la ocurrencia de deslizamientos). Convirtiéndose en el método estándar para escalas regionales de evaluación de la susceptibilidad.

Limitaciones

- La hipótesis general que los deslizamientos ocurren por la misma combinación de factores a lo largo del área de estudio.
- Ignorar el hecho que la ocurrencia de ciertos tipos de deslizamientos están controlados por ciertos factores de causa que deben ser analizados o investigados individualmente.
- Extender el control de algunos factores espaciales que pueden variar ampliamente en áreas con condiciones geológicas complejas y ambientes culturales.
- La falta de un criterio de experto adecuado en diferentes tipos de deslizamientos, procesos y factores de causas.

Fuente: AGS (2007) Guidelines for landslide susceptibility, hazard and risk assessment

Procedimiento (i)



Fuente: van Westen et al, (1997)

Procedimiento (ii)

Example 1: negative relation between parameter and landslides



Result of sampling:
positive relation

Unit	Landslide	Parameter
1	0	0
2	0	0
3	1	1
4	1	1
5	0	0
6	1	1
7	1	1
8	0	0
9	0	0
10	1	1

Sampling:
1 = present
0 = absent

Example 2: positive relation between parameter and landslides

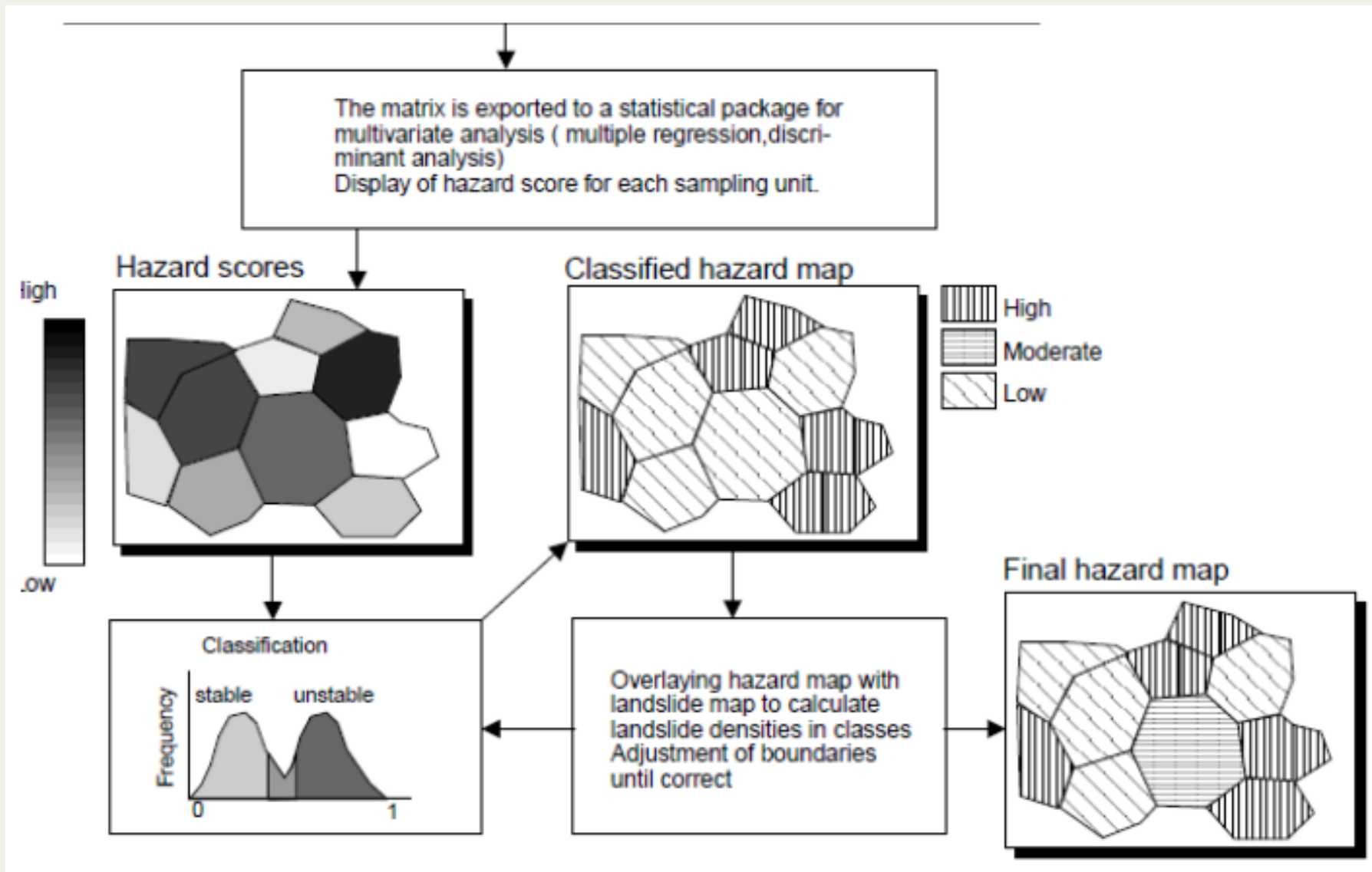


Result of sampling:
no relation

Unit	Landslide	Parameter
1	0	1
2	0	1
3	1	1
4	1	1
5	0	1
6	1	1
7	1	1
8	0	1
9	0	1
10	1	1

Fuente: van Westen et al, (1997)

Procedimiento (ii)



Fuente: van Westen et al, (1997)

Matrix Assessment

También denominado **Método de Análisis Condicional** (Clerici et al., 2002; 2006). No requiere asunciones de aleatoriedad del fenómeno bajo análisis.

Metodología

1. Mapa de inventario de deslizamientos
2. Mapa de factores explicativos
3. Definir UCU
4. Cruzar UCU vs inventario de deslizamientos
5. Definir para cada UCU % de área con deslizamientos [0-1]
6. Determine categorías relativas de susceptibilidad a deslizamientos (k-means cluster analysis).

La densidad de deslizamientos es equivalente a la probabilidad futura de la ocurrencia de deslizamientos. L: probabilidad de ocurrencia de deslizamiento. UCU: Unidad de condiciones Únicas.

$$P(L/UCU) = (\text{landslide area \& UCU area}) / \text{UCU area}$$

Fuente: Constanzo et al., (2012), Clerici et al. (2006) ; De Graff et al. 1991; 2012)

A. Landslide Matrix

Litología	Exposición			
	N	NE	SE	E
A	71	16	0	0
B	63	121	0	0
C	90	31	0	15
D	0	18	0	12
E	0	22	4	9
Pendiente				
M N O P				

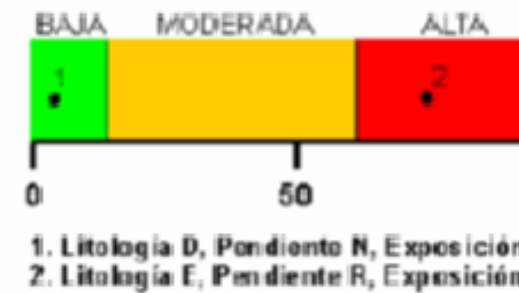
B. Management Unit Matrix

Litología	Exposición			
	N	NE	SE	E
A	210	90	6	66
B	230	631	29	110
C	710	122	38	98
D	135	310	125	23
E	90	89	10	12
Pendiente				
M N O P				

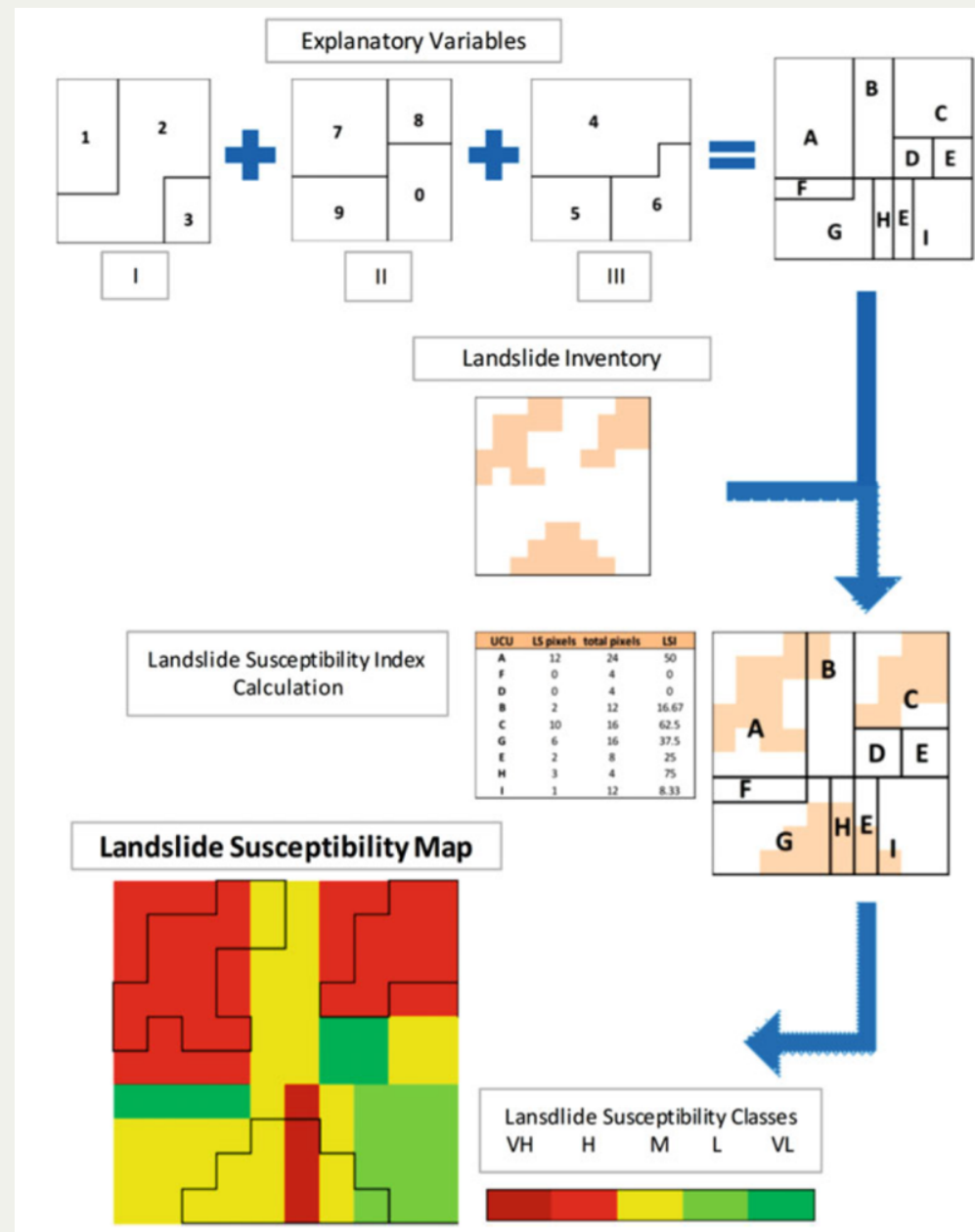
C. Matrix

Litología	Exposición			
	N	NE	SE	E
A	34	18	0	0
B	27	19	0	0
C	13	25	0	15
D	0	6 ¹	0	52
E	0	25	40	75 ²
Pendiente				
M N O P				

D. Susceptibility class






Fuente: De Graff et al. 1991; 2012)



Fuente: De Graff et al. 1991; 2012)

Table of UCUs characteristics in the Parma River basin^a

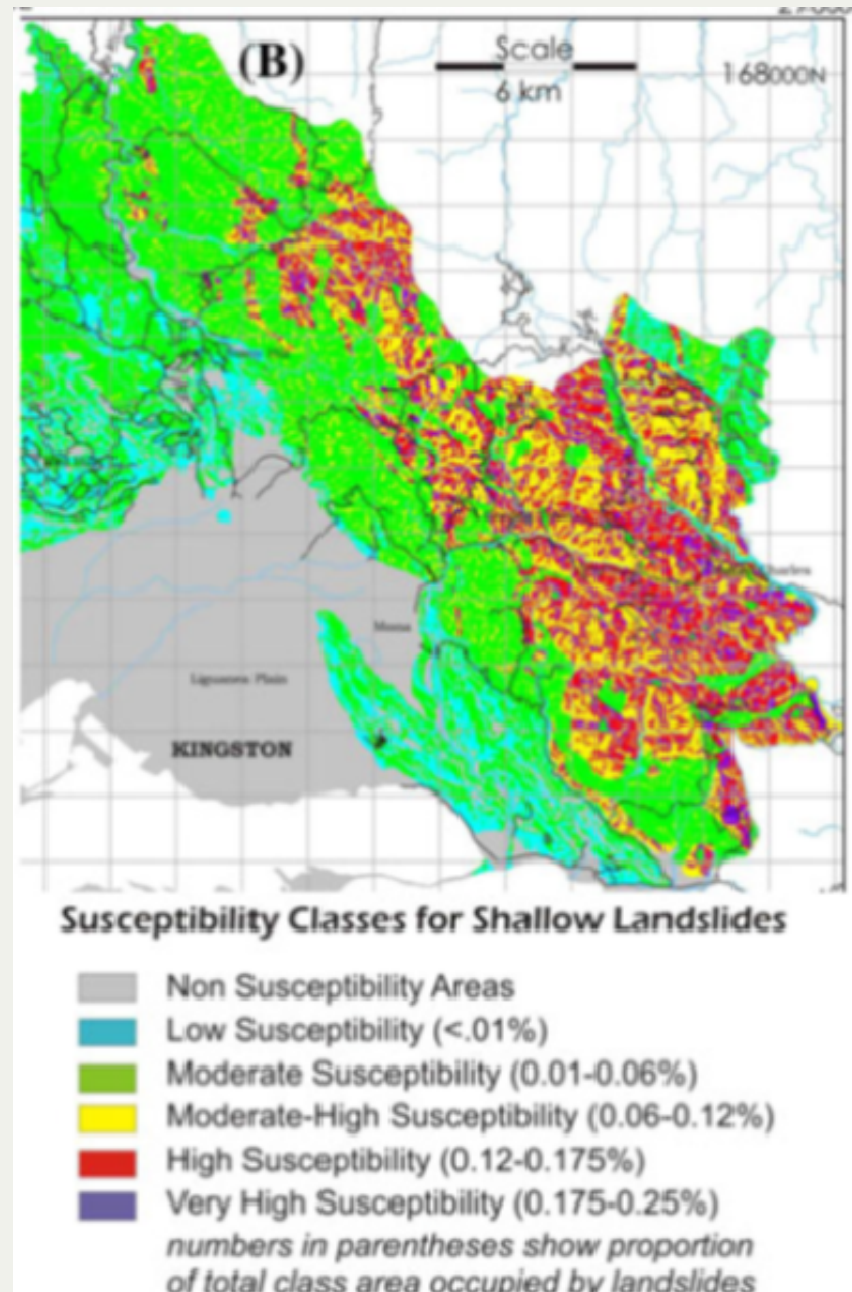
1	1	2	1	1	1	15	68	22
2	1	2	2	1	1	48	117	41
3	1	2	3	1	1	19	151	13
4	1	2	4	1	1	0	85	0
5	1	2	5	1	1	0	3	0
6	1	4	1	1	1	831	3385	25
7	1	4	2	1	1	924	5735	16
8	1	4	3	1	1	213	4187	5
9	1	4	4	1	1	8	871	1
10	1	4	5	1	1	0	14	0
11	1	6	2	1	1	0	1	0
12	1	6	3	1	1	0	7	0
13	1	6	4	1	1	0	7	0
14	2	1	1	1	1	160	259	62
15	2	1	1	2	1	1634	5903	28
2117	12	6	1	2	1	0	1572	0
2118	12	6	1	3	1	0	2183	0
2119	12	6	1	4	1	0	56	0
2120	12	6	1	5	1	0	40	0
2121	12	6	2	1	1	0	28	0
2122	12	6	2	2	1	0	98	0

-  # celdas con deslizamientos en la UCU
-  # celdas totales de la UCU
-  % densidad deslizamientos de la UCU

Class limits and extension

Class number	Class name	Class value
1	VERY LOW	0–13%
2	LOW	14–27%
3	MEDIUM	28–41%
4	HIGH	42–55%
5	VERY HIGH	> 55%
LANDSLIDES		
UNDEFINED AREAS		
TOTAL		

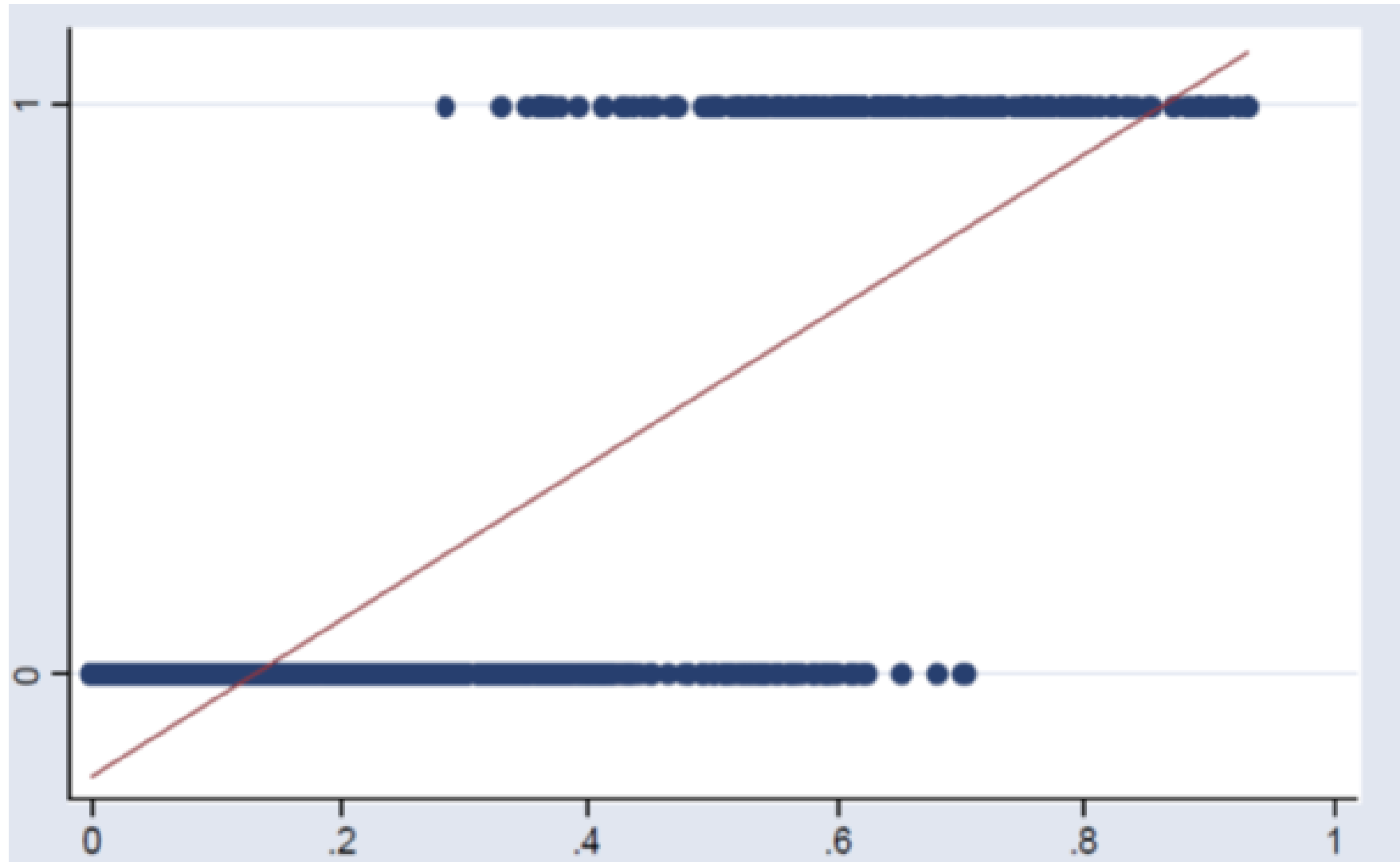
Fuente: Clerici et al. (2002)



Fuente: De Graff et al. 1991; 2012)

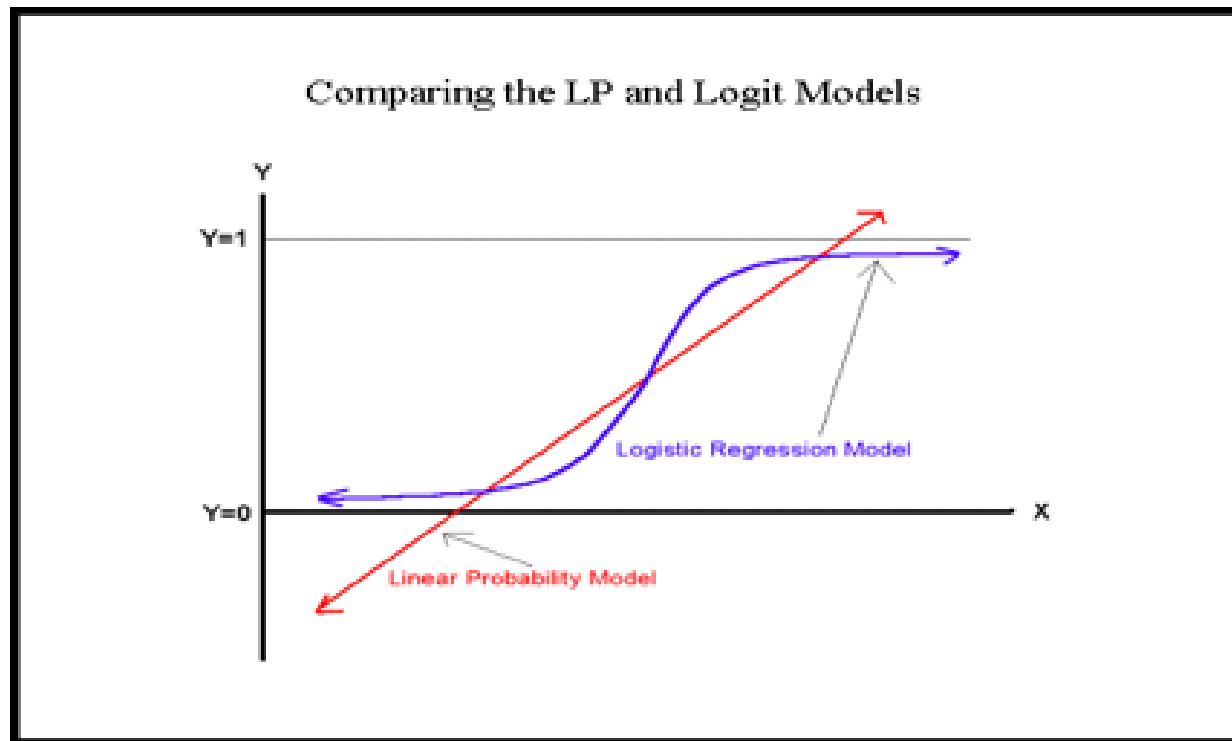
Función *LOGIT*

Variable dependiente= ocurrencia de MenM



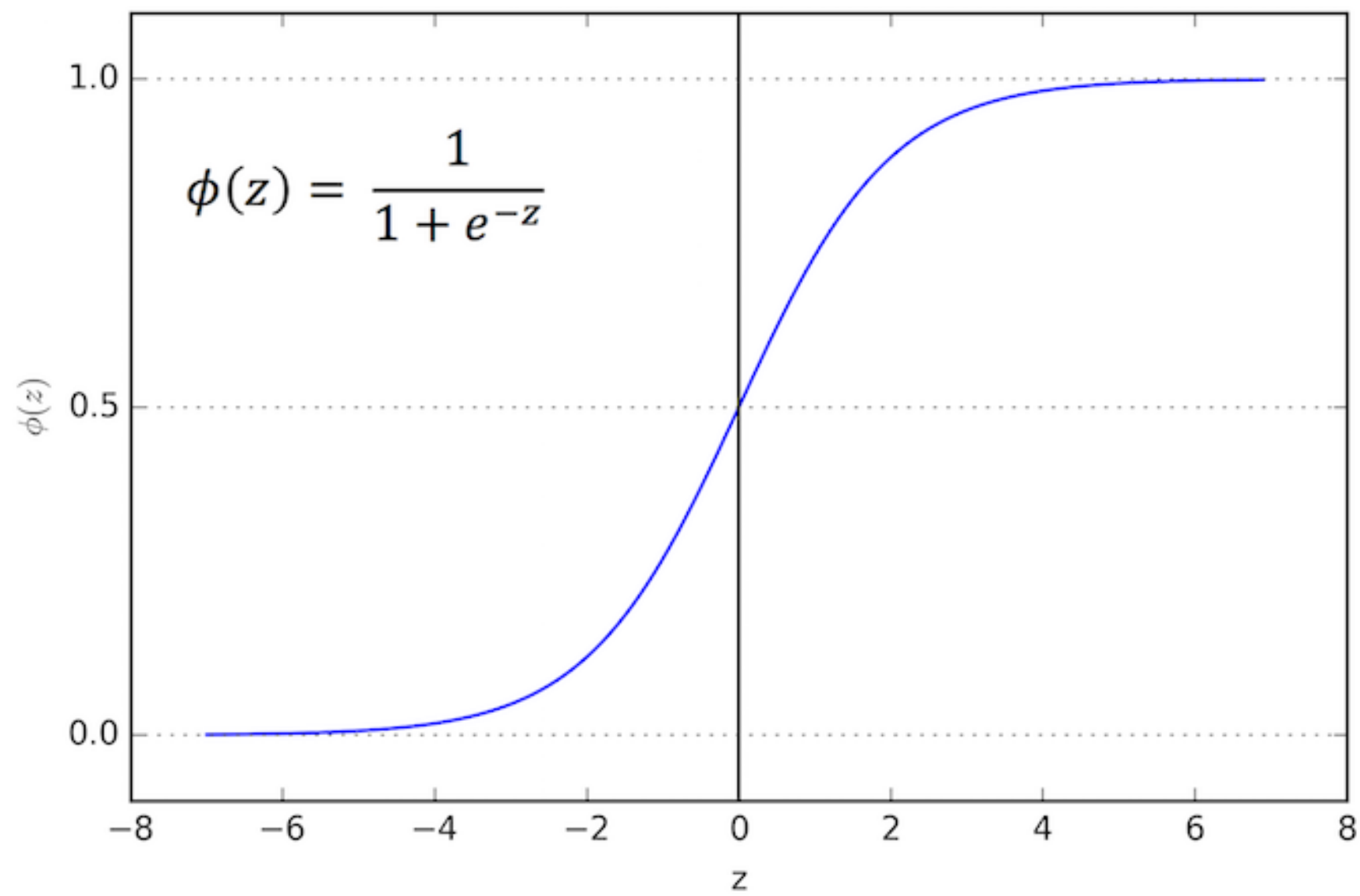
Variables independientes = Geología, pendiente, uso del suelo, etc

Probabilidad de la variable dependiente



$$P(y|x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

Relación lineal en función de las variables independientes



Regresión Logística

La Regresión Logística es una combinación lineal de variables independientes (factores explicativos) para explicar la varianza en una variable dependiente (inventario de deslizamientos) tipo dummy [0 – 1].

Ventajas

1. Las variables predictoras pueden ser continuas, discretas, dicótomas, o cualquier combinación de ellas.
2. La variable dependiente es dicotoma (binaria)
3. A pesar de que el modelo transformado es lineal en las variables, las probabilidades no son lineales

Desventajas

1. Los pesos de las variables terminan siendo un promedio para toda el área de estudio, los cuales en realidad pueden diferir en diferentes partes del área de estudio.
2. La función objetivo es una combinación lineal de las variables independientes

Sea $p(x)$ la probabilidad de éxito cuando el valor de la variable predictora es x , entonces:

$$p(x) = \frac{e^{a+\sum bx}}{1+e^{a+\sum bx}} = \frac{1}{1+e^{-(a+\sum bx)}}$$

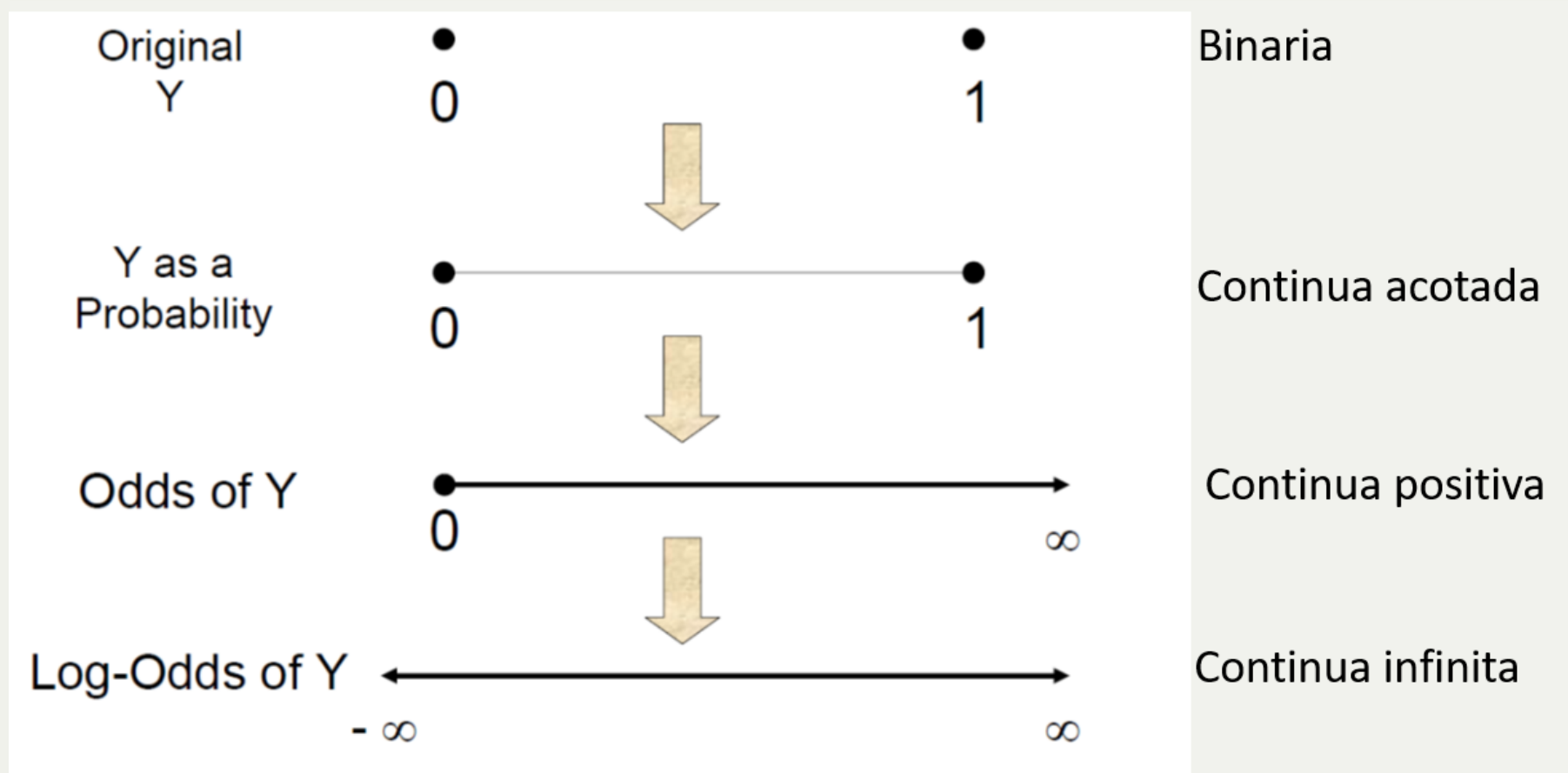
$$\frac{p(x)}{1-p(x)} = e^{a+\sum bx}$$

$$\text{Ln}\left(\frac{p(x)}{1-p(x)}\right) = a + \sum bx$$

Donde a es el intercepto del modelo, b son los coeficientes del modelo de regresión logística, y x son las variables independientes (predictoras).

$$P(y = 1) = \frac{1}{1+e^{-(a+\sum bx)}}$$

Donde, P es la probabilidad de Bernoulli que una unidad de terreno pertenece al grupo de no deslizamientos o al grupo de si deslizamiento. P varía de 0 a 1 en forma de curva "S" (logística).



Count		Pendiente		Total
		Pendiente < 20	Pendiente > 20	
No MenM	.00	508	590	1098
MenM	1.00	45	33	78
Total		553	623	1176

$$\frac{p(x)}{1 - p(x)} \rightarrow \frac{8\%}{92\%} = 0,089 \rightarrow 92\% \quad 8\% \quad 95\% \quad 5\% \rightarrow \frac{5\%}{95\%} = 0,056$$

$$\text{Odds en MenM con pendiente} > 20 = \frac{33/623}{590/623} = \frac{33}{590} = 0,056$$

$$\text{Odds en MenM con pendiente} < 20 = \frac{45/553}{508/553} = \frac{45}{508} = 0,089$$

$$\text{cambio en los Odds debido a HS} = \frac{0,056}{0,089} = 0,63$$

$$\text{Coeficiente de regresión logística} = \ln(0,63) = -0,46$$

Count

		Pendiente		Total
		Pendiente < 20	Pendiente > 20	
No MenM	.00	508	590	1098
MenM	1.00	45	33	78
Total		553	623	1176

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	Pendiente	-.460	.237	3.763	1	.052	.631	.397	1.005
	Constant	-2.424	.156	242.857	1	.000	.089		

$$\ln \left[\frac{p(x)}{1-p(x)} \right] = \beta_0 + \beta_1 X = -2.424 - 0.46X$$

$$P(x) = \frac{1}{1 + e^{-(-2.424 - 0.46x)}}$$

Cual es la probabilidad de un MenM con o sin pendiente >20?

$$\text{Cambio en odds} = e^{-.46} = .63$$

Cambio en los Odds: que en una ladera con pendiente > 20 ocurra un MenM es 0.63 comparado con otras laderas. Es decir es menos probable.

Causative factors	Categories	B^a	S.E. ^b	Wald ^c	Df ^d	Sig. ^e	Exp(B) ^f
Lithology (L_C)	Mudstone, shale, and Quaternary deposits	−0.020	0.496	0.002	1	0.967	0.980
	Sandstones and thinly bedded limestones	−0.822	0.120	47.175	1	0.000	0.439
	Limestones and massive sandstones	0.861	0.064	182.968	1	0.000	2.365
Bedding structure (BS_C)	Over-dip slopes	−0.680	0.453	2.255	1	0.133	0.507
	Under-dip slopes	0.029	0.184	0.024	1	0.876	1.029
	Dip-oblique slopes	0.410	0.107	14.595	1	0.000	1.507
	Transverse slopes	0.425	0.108	15.345	1	0.000	1.529
	Anaclinal-oblique slopes	0.316	0.098	10.412	1	0.001	1.371
	Anaclinal slopes	0.339	0.111	9.298	1	0.002	1.404
Slope (S)		−4.127	0.285	209.197	1	0.000	0.016
Aspect (A_C)	Flat	−3.166	0.624	25.720	1	0.000	0.042
	North	−3.416	1.332	6.576	1	0.010	0.033
	Northeast	0.350	0.103	11.653	1	0.001	1.419
	$z = L_c + BS_c + A_c - 4.127 \times S - 16.316 \times E - 0.247 \times PrC + 0.226 \times PlC - 1.875 \times FVC + 3.786$					0.461	1.081
0.383						0.906	
0.000						0.600	
Southwest						0.397	0.113
	West	−0.923	0.124	55.487	1	0.000	0.397
	Northwest	−0.597	0.116	26.729	1	0.000	0.550
Elevation (E)		−16.316	0.405	1627.088	1	0.000	0.000
Profile curvature (PrC)		−0.247	1.658	0.022	1	0.882	0.781
Plan curvature (PlC)		0.226	2.088	0.012	1	0.914	1.253
Fractional vegetation cover (FVC)		−1.875	0.229	66.924	1	0.000	0.153
Constant		3.786	1.273	8.846	1	0.003	44.097

Fuente: Chen et al (2016), río Yangtze (China)

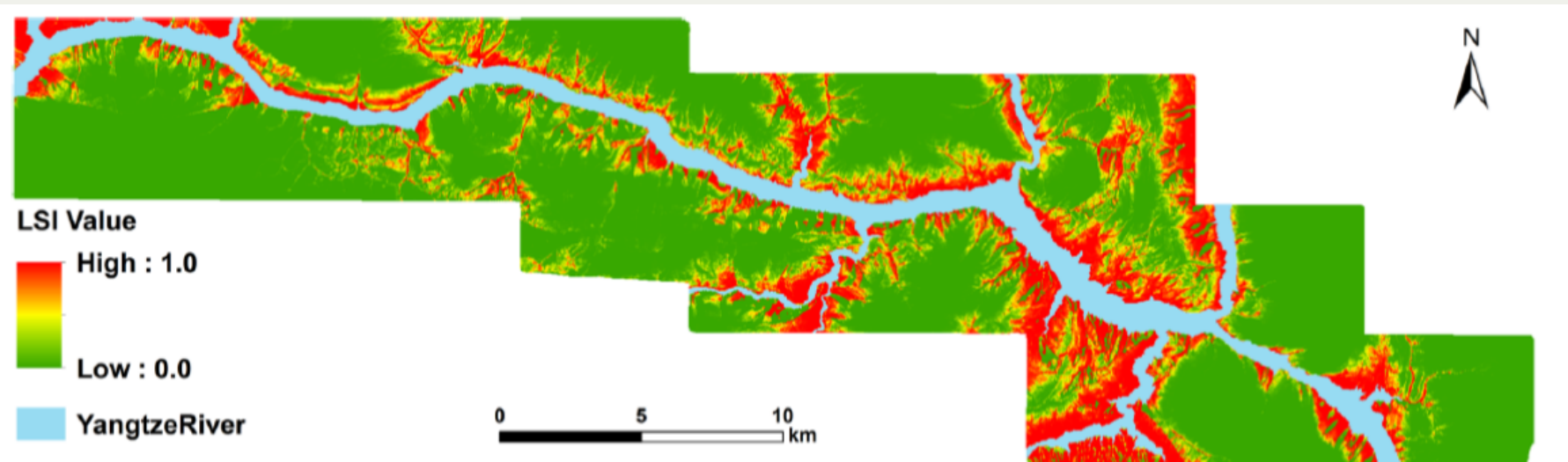


Fig. 12 Landslide susceptibility maps obtained from LR model

Fuente: Chen et al (2016), río Yangtze (China)