Importance of conditional independence in choosing the best combination of predictive factors for mapping the susceptibility of the landslide in the Ksar El Kebir northern region_Morocco

El Hamdouni I, Ait Brahim L, Abdelouafi A
Université Mohammed V, Faculté des Sciences, Département sciences de la terre, Unité de recherche GEORISK, LGRN, 4 Avenue Ibn Battouta Rabat-Agdal, BP 1014, Rabat, Maroc.

Abstract: This article proposes the evaluation of landslides susceptibility in Ksar El Kebir region, by applying a bivariate indirect approach "weight of evidence". The approach consists in grasping the relation between the predictive factors (VP) and the spatial distribution of landslides (the variable to be modeled VM). The data modeling by GIS takes into account variables related to topography (slope, aspect, altitude), variables related to geology (lithology, fracturing), the land use and the distance from the drainage network. For each predictive factors V_P, two weights are calculated, one positive (W^+) and another one negative (W^-), which the values depend on the spatial relationship between the predictive variables and the variable to be modeled in the past [9]. Calculations of the values of W^+ and W^- for all V_P make it possible to calculate the a posteriori probability, which updates the a priori probability. Thus we obtained a map of susceptibility to landslides. The validity of the prediction model was verified by using the success rate curve (SRC) and the blind test (BT).

Keywords: Landslide; weight of evidence ; susceptibility to landslides ; conditional independence; success rate curve; blind test.

I. INTRODUCTION

Located west of the Rifan Range in Northern Morocco, the Ksar El Kebir region is one of the region's most affected by many landslide, including the impact on the natural environment and on the road infrastructure and buildings is certainly considerable. The studied area represents a crossroads connecting the main economic centers of the country; as a result, the mapping of susceptibility to land movements becomes a crucial need in regional planning. The landslides are responsible for much larger socio-economic losses than those commonly attributed to them [1]. However, their damage is mostly masked by associating them with triggering processes; but we find that some regions are much more likely than others even if they are exposed to the same triggers. As a result, the present work proposes a mapping of the susceptibility to slope movements, based on indirect methods. These methods rely essentially on the crossing of an inventory of gravity phenomena with local predispositions factors in order to identify the relationships between the presence of these factors and the onset of these phenomena. The contribution of GIS and probabilistic methods as well as comparisons of methods have largely fed the scientific literature. Of the large body of available methods, evidence theory is the most popular [2]-[3]-[10]. It is particularly adapted to the problems of slope movements in that it looks for relationships between a binary variable (presence / absence) of landslide and the different predictive variables in the cartographic form.

II. GEOMORPHOLOGIC AND STRUCTURAL FRAMEWORK

The studied area is located in the valley of the wadi Loukkos (Fig.1), limited to the West by a muddy alluvial plain on the edge of the Rharb plain; in the North-East the region extends to the massifs of Sidi Issef and El Kobba; in the East and South-East it is limited by Jbel Selloum in the region of Ouezzane (Fig.1). The eastern and north-eastern parts are dominated by the steep reliefs with a rugged topography whose altitudes peak at 600 m; the western portions of the alluvial plain are very low, reaching only 10 to 15 m high [4] or the slope is 7.10^{-4}.

The Ksar El Kebir region belongs to the external domain of the consequent Rifan chain of the alpine orogeny (Fig.1). The region is marked by a stack of several structural units separated from each other by abnormal contacts, the most important is the WSW extension of the Jebha-Chrafate accident [5]. It separates two groups of syn-orogenic turbidite complexes: Asilah-Larache sandstone (El Habt unit) and Sidi Mrayt basin in the North and Zoumi sandstone and Ouezzane domain in the South. [5] [6]. The tertiary series of Ksar El Kebir show several structurally distinct units whose stratigraphic distribution extends from the Middle Eocene to the Middle Miocene. Each of these units is made up of turbiditics set up by gravity collapses in a deep marine environment, the latter being the site of syn-sedimentary tectonics [6].

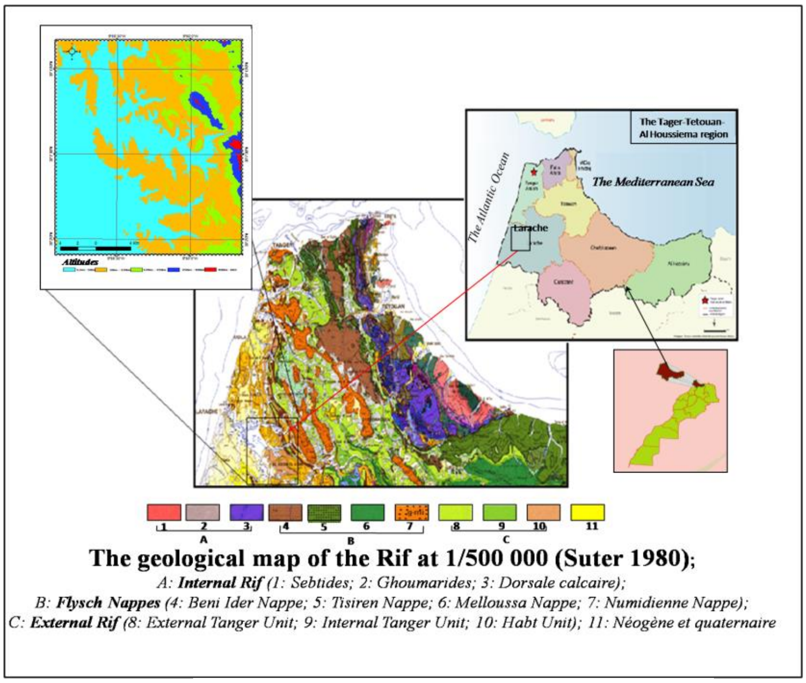


Figure.1: Situation of the study area.

III. METHODOLOGY AND TOOLS

An area is declared susceptible to ground movements when the ground conditions of that site are comparable to those of an area in which movement has already occurred [7]-[8]. (Fig.2)

The method used is that resulting from Bayes theorem called the weight of evidence theory "WofE" [9]-[3]. The principle of this method is to combine each class of predisposition factors with an inventory map of field instabilities in the form of points, in order to give it a weight to be attributed according to the density of landslides in each surface unit. The weights obtained for each factor class are summed one by one, the final calculation assigns several probabilities by combination of classes [2]-[10]. However, this hypothesis assumes that the different predisposition factors introduced into the model are independent of the probability of locating a ground motion [2]-[3]-[9]-[10], making the choice of independent parameters more difficult. The obtained map is evaluated statistically, then compared to an inventory map acquired by a geomorphologic approach (expert approach). Validation tests will be developed to evaluate the performance of the model. For this purpose, we realized a success rate curve (SRC) and blind test (BT) [10]-[11].

The principle of the model used in this paper is to define mathematical relationships between predisposing factors (predictors, Vp) and the spatial occurrence of Landslide Vm; This makes it possible to quantitatively evaluate the probability of rupture for regions not affected by landslides, and decreases the expert's subjectivity [2]- [3].

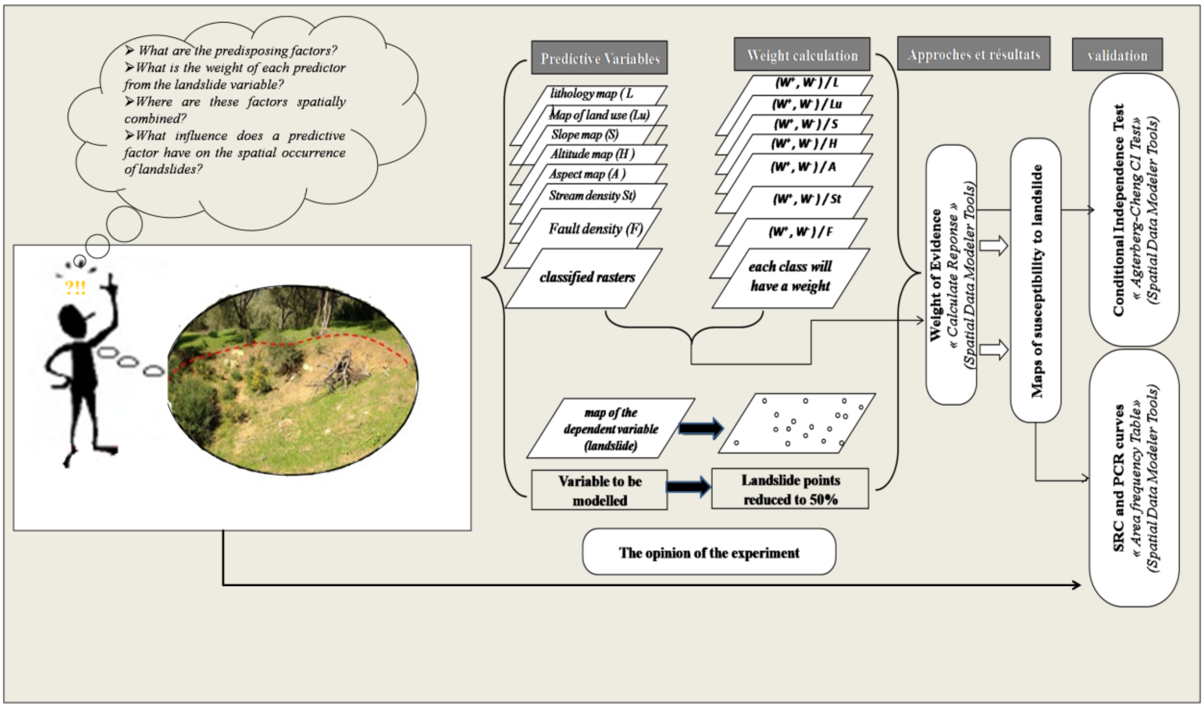


Figure.2: The modeling approach

A. *inventory of landslide*

In order to calculate the susceptibility of landslide, using the WofE method, an inventory of past instabilities is needed. For this, we have mapped the different types of instabilities (Fig.3), using satellite images, photo-interpretation, previous literature and field studies.

The ground instabilities thus mapped are introduced in the form of points in our model. The WofE approach assumes that there is only one point of variable that can be modeled per unit cell [9]; for this, we have subdivided the total number of landslides "N" into two groups A and B, so that each of them represents 50% of the known landslides. Arc-SDM has a tool designed for this purpose "Training Sites Reduction" with which we can reduce N at random to a desired percentage [13]-[15].

Figure.3: some examples of landslides

B. Data collection and formatting

The first step is to collect all the information and data needed for modeling. The importance of precision in the collection and storage of information has been widely described in many publications [2]-[9]-[10]. Two basic rules must be taken into consideration:

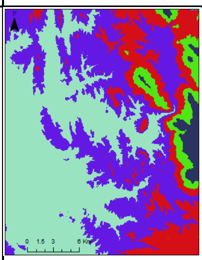
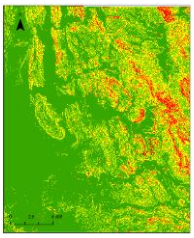
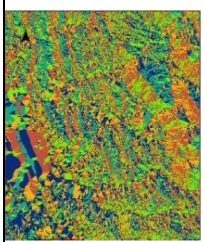
- ✓ The information must be homogeneous,
- ✓ The database must be organized from monothematic layers;

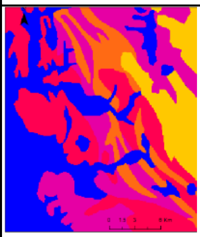
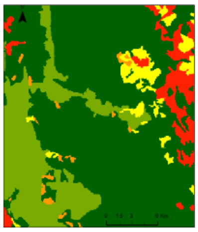
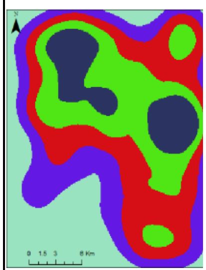
The essential steps in the zoning of the susceptibility to the movements of ground, by applying the weight of evidence approach, are defined in five steps [2]-[3]-[10]:

- ➢ Mapping of field instabilities and their distinction according to the type of activity, and based on information covering a broad period;
- ➢ Mapping and convert shapefile to raster and classified the parameters.
- ➢ The definition of the weights relative to each of the factors involved,
- ➢ Probabilistic modeling of mid-scale susceptibility; using the "*Spatial data modeller tools*" function on Arcgis 10.0.
- ➢ Validation of the models used by evaluation tests.

The data needed to obtain landslide susceptibility index values were acquired from cartographic materials, satellite images, and previous literature. These input data can be subdivided into five main groups: geomorphologic, topography, geology, hydro-geology, and land use (Table.I). For each factor, we will calculate a weight assigned to it during modeling.

Table.I. : Definition of prediction factors

Prediction factors	***Maps***	***classes***	***Legend***	***Data***	***Treatments***	***Cell Size***
Altitude		1 2 3 4 5	1 :0 – 20m 2 : 20 – 70m 3 : 70 – 150m 4 : 150 – 400m 5 :400 – 650m	DEM	-Spatial Analyst Tools« surface» - Reclassify	5m
Slope		1 2 3 4 5	1 :0 - 5° 2 : 5° - 10° 3 : 10° - 15° 4 : 15° - 25° 5 : > 25°	DEM	-Spatial Analyst Tools« surface» - Reclassify	5m
Aspect		1 2 3 4 5	1 : Nord 2 : Nord_Est 3 : Sud 4 : Sud_West 5 : Nord_West	DEM	-Spatial Analyst Tools« surface» -Reclassify	5m

Prediction factors	***Maps***	***classes***	***Legend***	***Data***	***Treatments***	***Cell Size***
Lithology		1 2 3 4 5	1 : sandstone 2 : clays 3 : sandstone marls 4 : limestone marl 5 : sandy loam	- Landsat TM image - geological map at 1/500000 ° - bibliography - Field study	- Image processing, - digitizing -Verification - convert to raster -Reclassify	92m
Land use		1 2 3 4 5	1 Forest 2 : Built 3 Scrub 4 : irrigated land 5 :Agricultural land	- Landsat TM image - bibliography - Field study	- Image processing, - digitizing -Verification - convert to raster -Reclassify	92m
Stream density		1 2 3 4 5	1 : 0.367 – 0.99 2 : 0.99 – 1.75 3 : 1.75 – 2.107 4 : 2.107 – 2.67 5 : 2.67 – 3.49	-Topographic map ; -The DEM	- Digitalization of the hydrographic network -Calculating density line on Arcgis 10 -Reclassify	92m
Fault density		1 2 3 4 5	1 :0 – 0.1 2 : 0.1 – 0.3 3 : 0.3 – 0.435 4 : 0.435 – 0.5 5 : 0.5 – 0.7	- Landsat TM image -structural board 1/500000° -bibliography -Study of the field	-Traitements d'images, -extraction des linéaments et verification -Calcul de la densité « Line density » - Reclassify	92m

C. Calculation of the weight of factors

The weights provide a measure of spatial association between the points of the terrain instabilities (variable to be modeled) and the predictive factors (Vp). Thus a weight is calculated for each class and at the level of each of the predictor variables. A positive value (1) of the weight W^+ indicates that there are more slips on this class than those due to chance; conversely, a negative value (2) W^- indicates that fewer slips occur than expected. The difference between the positive weight and the negative weight is known as the contrast ©. Thus, $C= W^+ - W^-$. Contrast is a global measure of the spatial association between the points of each factor and the variable to be modeled, combining the effects of the two weights. In general, absolute weight values between 0 and 0.5 are slightly predictive; values between 0.5 and 1 are moderately predictive; values between 1 and 2 are highly predictive, and values greater than 2 are highly predictive [2]-[10]-[3]. The weighting coefficients for the binary theorems are given by the ratio of the following conditional probabilities [10]-[13]]-[9].

$$(1) \quad W^{+} = \ln \frac{P\{B|D\}}{P\{B|\bar{D}\}} = \ln \frac{N\{B\cap D\}/N\{D\}}{N\{B\cap \bar{D}\}/\{\bar{D}\}}$$

$$(2) \quad W^{-} = \ln \frac{P\{\bar{B}|D\}}{P\{\bar{B}|\bar{D}\}} = \ln \frac{N\{\bar{B}\cap D\}/N\{D\}}{N\{\bar{B}\cap \bar{D}\}/N\{\bar{D}\}}$$

P: the probability;
N: Number of pixels
B: the potential presence of the predictive factor;
$\bar{B}$: the absence of the predictive factor;
D: the presence of the landslide;
$\bar{D}$: absence of landslide;

The ratio represents the probability of being present on the probability of absence of landslide within a class and for a given predictor. The weights are summed using the natural logarithm of ratios called "*logit*".

We have analyzed the weights W^{+}, W^{-} and the contrast attributed to each of the classes of the predictor variables; This made it possible to evaluate and interpret the respective role of each predisposition factor in the spatial occurrence of landslides. We have noted that land use and slope and lithology play a major role in susceptibility to terrain instabilities. Their standardized maximum contrast values are (*1.3913*), (*1.3064*), (*1.2111*) respectively. While the density of the hydrographical network and the exposure of the slopes are relatively less influential with a maximum contrast of (*0.5493*) and (*0.7521*). The density of faults and the hypsometry occupy an intermediate place with an absolute value of the contrast of (*0.9511*) and (*0.8135*) respectively. We have also noticed that for lithology, clays are characterized by a high susceptibility with a positive weight of *1.018*. (Table.II)

Tableau.II. calculation of the weight of the prediction factors (*C:Contrast)*

Factors	Classes	W+	W-	C
Fault density	0- 0.1	-0.342	0.0654	-0.407
	0.1- 0.3	0.023	-0.0036	0.027
	0.3- 0.435	-0.431	0.1195	-0.55
	0.43- 0.5	0.187	-0.0611	0.248
	0.5- 0.7	0.419	-0.13	0.549
Slope	0_5%	0.1157	-0.0661	0.182
	5%_10%	-1.162	0.1444	-1.306
	10% _15%	0.2362	-0.2195	0.456
	15%_ 25%	0	0	0
	>25%	-0.855	0.0269	-0.882
Stream density	0.36 – 0.99	-0.4307	0.0459	-0.476
	0.99 – 1.75	-0.1493	0.0173	-0.166
	1.75 – 2.11	-0.678	0.135	-0.813
	2.11 – 2.67	0.421	-0.225	0.646
	2.67 – 3.49	0.085	-0.0325	0.117
Altitude	0- 20m	0.108	-0.2079	0.316
	20 -70m	-0.649	0.1024	-0.752
	70- 150m	0.226	-0.0471	0.274
	150-400m	-0.690	0.0197	-0.710
	400-650m	0	0	0

Factors	Classes	W+	W-	C
Aspect	N	-0.145	0.0379	-0.1829
	N_E	-0.441	0.0751	-0.5166
	S	-0.319	0.0604	-0.3799
	S_W	-0.436	0.0739	-0.5099
	N_W	0.634	-0.3167	0.9511
Lithology	Sandstone	0	0	0
	Clays	1.018	-0.193	1.2111
	Sandstone & marls	-0.040	0.0122	-0.052
	limestone marl	-0.136	0.0398	-0.1757
	sandy loam	0.008	-0.0035	-0.1757
Land use	Forest	-1.335	0.0565	-1.391
	Built	0.389	-0.0063	0.395
	Scrub	-0.199	0.0088	-0.208
	irrigated land	-0.186	0.0439	-0.230
	Agricultural land	0.130	-0.3121	0.442

IV. MODELING RESULTS

A. *The susceptibility map*

In this step we have combined the predictive maps, in order to produce a constant posterior probability scale; this was possible thanks to the "*Calculate Response*" tool on ArcSDM, [11]-[15]. The susceptibility map obtained in this study is the result of a combination of the seven predictive factors and their weights. However, the *WofE* approach requires some independence; therefore, a quantification of factors and tests to decrease the degree of conditional dependence of these factors will be the goal of the next step "*CI test*". After a calibration procedure and validation of the model, we were able to make a map of the final susceptibility into five classes that interpret the respective weight of each of the factors used as explanatory variables. (Fig.4)

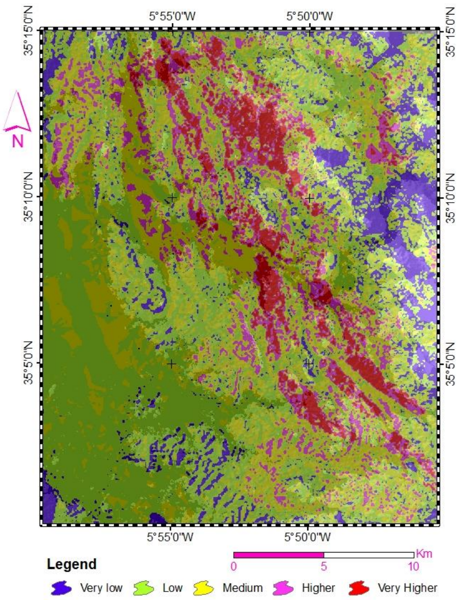


Figure.4: Map of the susceptibility to the instabilities of ground

B. *Conditional independence test*

WofE modeling assumes conditional independence (*CI*) between predictive factors [14]. However, in practice, and especially when working with geological and geomorphologic phenomena, the rule of conditional independence is often violated by certain measures [12]-[14]. As a result, the modeler must assess the seriousness of this violation [9]. Tests of the conditional independence between the different predictor variables must be applied [2]-[10].

In our work, we used the "*Agterberg-Cheng CI Test*" function on Arc-SDM to evaluate conditional independence. This tool offers a wide range of CI tests (the Overall test, the CI Ratio test n/T, the Agterberg-Cheng test (AC)). After several tests aimed at minimizing conditional independence, we have retained the following results [10]-[11]-[14]. (Tableau.III). These results have been accepted, although some tests show some dependence. Because when we try to reduce factor dependence, we risk losing valuable information [11]-[12]. However, too many conditionally independent variables can cause an adjustment of the model.

According to the results obtained, two out of three tests gave acceptable results (Table III); hence the possibility of keeping the model. In order to limit a possible loss of information, several methods have been developed in order to verify the degree of violation of conditional independence (*IC*) [16] developed a "*Conditional Dependence Adjusted WofE*" (CDAWE), the model significantly reduces the effects of conditional dependence by applying a correction to the posterior probability. Conditionally independent can cause a fit of the model.

Table.III : Independence test results

Number of instability points (n)	*Number of points expected (T)*	*T-n*	*σT*	*(T-n) / σT*	*Conditional Independence Tests*		
					Ratio (CI) n/T	*Test (AC)*	*Overall CI (OT)*
52	53.5	1.5	4.77	0.32	0.96	52%	74.6%

V. STATISTICAL VALIDATIONS OF MODELS

There are several ways to check the performance of the model by analyzing the posterior probability map, the most important being the *SRC* and the *PRC* efficiency curve, which we did using the Arc-SDM tool "*Area Frequency Table*".

A. *success rate curve (SRC)*

The success rate curve (*SRC*) or success curve is a test that ensures that the posterior probability corresponds to the slip points used to generate the prediction map [15]-[16] (Fig.5). The area under the curve makes it possible to obtain the efficiency of the model [12]-[13], which is of the order of 68.2%. This curve reflects how the model classifies the distribution of known landslides, but does not necessarily reflect how the model predicts unknown landslides, which is the goal of modeling. For this, we need to perform blind test (*BT*).

The success rate curve (*SRC*) or success curve is a test that ensures that the posterior probability corresponds to the slip points used to generate the prediction map [15]-[16] (fig.5). The area under the curve makes it possible to obtain the efficiency of the model [12]-[13], which is of the order of 68.2%. This curve reflects how the model classifies the distribution of known landslides, but does not necessarily reflect how the model predicts unknown landslides, which is the goal of modeling. For this, we need to perform blind test (*BT*).

B. *The blind test equations*

The blind test (*BT*) is the validation test used in our model (the *PRC* forecast curve). This test shows how our model predicts unused slips during modeling. As a result, a yield curve similar to the *SRC* is created, but using the slip points that were not used in the model. This curve shows us that it is the efficiency of the model to predict unknown slips. Fabbri and Chung (2008) [15] emphasize the importance of using BT in evaluating the predictive performance of spatial predictive models. Other authors [1]-[11] note that if a posterior probability map was generated randomly, the *PRC* would be a straight line with a slope of 1, so the prediction rate curve should be well above it.

To apply the test (*BT*), we used the group of instability points that was not introduced into the model. Thus, we realized a *PRC* curve which shows an efficiency of 65%. The prediction and success curves are quite similar, and the values of their "*area under the curve*" are very close (Fig.5).

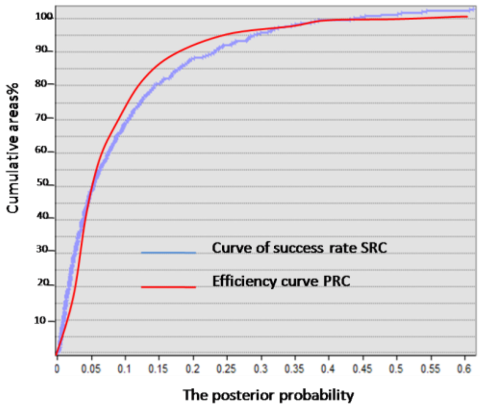


Figure.5: Comparison between the success rate curve and the forecast curve (blind test)

VI. CONCLUSION

The main objective of this study was to generate a map of susceptibility to landslides closest to reality. So, we used one of the probabilistic methods to reduce the degree of subjectivity. The WofE approach has been taken using the SDM tool on ArcGIS. The bulk of this project was dedicated to identifying appropriate data, improving data quality, and predicting model performance, which resulted in a relevant and adequate database.

The results obtained indicate that the combination of the seven predictor variables (slope gradient, lithology, land use, fault density, altitude, stream density, aspect) makes it possible to predict the spatial location, in terms of susceptibility, 68% of observed landslide areas. The model has been validated, using several tests; these have helped to define its reproducibility and the uncertainty associated with predictions. The bivariate analysis by the theory of evidence seems to be one of the most powerful methods, which confirms the results obtained by several authors: [2]-[3]-[11]-[15]. As a result, our model could be considered a success, as the success rate curve shows an efficiency of 68.2%.

ACKNOWLEDGMENT

I thank my friend "Meryem El Youssoufi" for helping to correct the language.

REFERENCES

[1] R. L. Schuster, W. J. Kockelman, “Principles of landslide hazard reduction, Landslides Investigation and Mitigation,” National Academy Press, 91-105,1996.
[2] Y. Thiery, S. Sterlacchini, J.P. Malet, A. Puissant, A. Remaitre, O. Maquaire, “Strategy to reduce subjectivity in landslide susceptibility zonation by GIS in complex mountainous environments,” Computers and Geosciences, 2007.
[3] C.J. Van Westen, M. Price, “GIS in landslide hazard zonation: a review, with examples from the Andes of Colombia,” Mountain Environments and Geographic Information Systems,1994.
[4] G. Maurer, “Les montagnes du Rif Central, étude géomorphologique,” travaux de l’Institut Scientifique Chérifien. Série Géologie ct Géographie Physique, vol. 14, 1968.
[5] G. Suter, “Carte géologique/structurale de la chaîne rifaine, échelle 1:500,000,” Serv. Carte géol. Maroc 245a/b,1980.
[6] A. Zakir , A. Chalouan, H. Feinberg, “Evolution tectono-sédimentaire d’un domaine d’avant-chaîne : exemple des bassins d’El Habt et Sidi Mrayt, Rif externe nord-occidental (Maroc), précisions stratigraphiques et modélisations tectonique,” Bulletin de la Société Géologique de France, vol. 175, 383-397,2004.
[7] D.J. Varnes, “Landslide Hazard Zonation a review of principles and practice,” IAEG Commission on Landslides, UNESCO, 1984.
[8] J.N. Hutchinson, “Landslide Hazard Assessment,” Landslides, Proceeding of 6th International Symposium on Landslides, Keynote Paper, Christchurch, vol.1, 1995.
[9] G.F. Bonham-Carter, “Geographic Information Systems for Geoscientists: Modeling with GIS, Computer methods in the geosciences,” Elsevier Science,vol. 13, 398, 1994.
[10] L. Aitbrahim and I. El hamdouni, “The Application of Logistic Regression for Mapping the Susceptibility of Versants to Landslides in the Region of El Quola "Southern RIF",” International Journal of Civil Engineering and Technology, vol. 9, pp. 1548–1558, Apr 2018. http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=9&IType=4
[11] Erich Schmitt, “Weights of Evidence Mineral Prospectivity Modelling with ArcGIS,” EOSC Directed Studies, 448, 2010.
[12] G. A. Partington, “Exploration targeting using GIS: more than a digital light table,” GeoComputing Conference, Brisbane,2010.
[13] G.F. Bonham-Carter, “Spatial Data Modeller (SDM) for ArcMAP 9.3 geoprocessing tools for spatial data modelling using weights of evidence, logistic regression, fuzzy logic and neural networks,” 2009.
[14] F.P. Agterberg, Q. Cheng, “Conditional independence test for weights of evidence modeling,” Natural Resources Research (2002).
[15] I.El Hamdouni, L.Aitbrahim, “Extraction of the predisposition parameters for mapping the susceptibility of slides lands, in the region of EL Quola, province of Larache (Nothern Rif),”: MATEC Web of Conferences 149:02045, DOI: 10.1051/matecconf/201814902045, LicenseCC BY 4.0, January 2018.
[16] C.J. Chung, A.G. Fabbri, “Validation of Spatial Prediction Models for Landslide Hazard Mapping, ” Natural Hazards, vol.30, 451–472 (2003).