18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

The weight of evidence statistical method in landslide susceptibility mapping of the Rio Pardu Valley (Sardinia, Italy)

Barbieri G. and **P. Cambuli**

Dipartimento di Ingegneria del Territorio, Università degli Studi di Cagliari,
P.za D'armi, 09123 Cagliari, Italia
Email: barbieri@unica.it

Abstract: The landslide susceptibility of hillsides has been assessed using the statistical methodology known as "weight of evidence", in order to estimate its potentiality, its difficulty of application and its suitability to specific geomorphological settings. The Valley of the Rio Pardu River, eastern-central part of Sardinia, Italy, has been chosen for applying the method. On the basis of the IFFI Project (Inventory of Landslides Phenomena in Italy), the Rio Pardu Valley turns out to be one of more hazardous zone in Sardinia for the presence of ancient and current landslides, whose causes have to be searched in the structural and lithological characteristics, the extreme meteorological conditions and the anthropic factors.

The weight of evidence method has been implemented by means of the ArcView 3.2 software and the ArcSDM extension (Spatial Data Modeller). Like all the statistical methodologies, this method requires to identify and to locate on a map all the instability phenomena which affected the studied area. On the basis of the territorial distribution of the past and present landslides the method calculates the weights to be assigned to the single classes of every considered parameter. The landslide map of the Pardu Valley has been provided by the IFFI Project. The map has been digitized, georeferenced and supplemented by the observation of the 2006 colour orthophotos.

For evaluating the landslide hazard of the studied area, three basic parameters for landslide susceptibility have been taken into consideration: the Lithology, the Land Use and the Slope.

In order to check the reliability and the forecasting capability of the method, the influence of each parameter on the instability phenomena has been calculated analyzing a sampled area, obtained by extracting at random a little percentage of the studied territory. In the sampled area the same proportion between the area affected by landslides and the total area calculated on the whole territory, equal to about the 0.063%, has been kept.

Two subsets have been so created: the Training set and the Validation set. In the extracting process, the morphometric, lithological and land use characteristics have been considered, in order to obtain a sample which was actually representative of the studied area, so that the distributions of the various classes of every single considered parameter were almost the same between the sampled area (training set) and the total area (validation set). The Training set has been divided into two subgroups: the subgroup of the areas which do not present landslide phenomena, and the subgroup of the areas which present landslide phenomena.

Then a statistical analysis of different parameters has been carried out, using the weight of evidence method, and the weights to be attributed to each class of every single parameter have been then obtained.

Keywords: *Landslide susceptibility zoning, Weight of evidence statistical method, Rio Pardu Italy.*

1. INTRODUCTION

The catchment basin of the Rio Pardu river, in the Ogliastra region, eastern-central part of Sardinia (Italy), was surveyed for an area of about 29 km^2, including the little village of Gairo S. Elena, which was re-founded after the old village of Gairo was destroyed in 1952 by a landslide.

The Rio Pardu valley is extensively covered by Palaeozoic metasedimentary rocks, that are widely present in almost all the south-eastern part of Sardinia. The present geomorphological features derived from the tectonic phases of Alpine Orogeny. The major tectonic frameworks are oriented NW-SE and NE-SW (Vardabasso, 1956; Cocozza, 1974).

The Rio Pardu valley developed along one of these tectonic alignments, oriented NW-SE, that crosses the central part of Sardinia. To this main tectonic alignment more differently oriented fault systems are associated. The geomorphological processes responsible for the present landform started in Pliocene, at the end of the tectonic activity. The river erosion significantly deepened the valley, producing talus deposit, especially on the left side of the valley, at the toe of limestone relieves, in the form of heterogeneous and chaotic blocks.

This paper aims at assessing the landslide susceptibility of this area, using the statistical methodology known as "Weight of Evidence" (Bonham-Carter et al., 1988; 1989; Agterberg et al., 1990).

2. THE WEIGHT OF EVIDENCE STATISTICAL METHOD

The weight of evidence method is based on a statistical bayesian bivariate approach. Originally, this method was developed for gold mineralization researches. After several years, the interest for this method extended to several researchers working in the field of landslide hazard assessment (Lee et al., 2004; Rezaei Moghaddam et al., 2007).

The weight of evidence method was implemented using the ArcView 3.2 software and the ArcSDM extension (Spatial Data Modeller), developed by Kemp et al. (2001). This extension implements some of the most commonly used statistical spatial models. The weight of evidence method is based on the Bayes theorem and on the concepts of prior and posterior probability, for assessing the relations between the spatial distribution of the areas affected by landslides and the spatial distribution of the analyzed landslide susceptibility factors (or parameters). It is therefore possible to calculate the degree of influence that each factor had, and will have in the future, on the development of landslide events. If a part A_f of the studied area is affected by landslide phenomena, the prior probability of finding a landslide within A_t (total studied area) is:

$$P_f = \frac{A_f}{A_t} \qquad 1$$

This initial estimation can be then increased or decreased depending on the relations between analyzed factors and landslides. The probability of finding one of the factors examined in the study area is given by:

$$P_v = \frac{T_v}{A_t} \qquad 2$$

where T_v = total area occupied by a certain class of a certain factor (e.g. slope class 20-35%).

For the whole territory surveyed, the probability of finding a landslide in the areas occupied by the *n* class of the *j* parameter is the ratio between the probability of finding a landslide inside the territory occupied by the *n* class of the *j* parameter and the probability of finding an area occupied by the *n* class of the *j* parameter in the whole territory:

$$P\{A_f \,|T_v\} = \frac{P(T_v \cap A_f)}{P_v} = \frac{P\{T_v \,|A_f\}}{P_v} \times P_f \qquad 3$$

Similarly, the posterior probability of finding a landslide in the areas not occupied by the *n* class of the *j* parameter is:

$$P\{A_f \,|\overline{T_v}\} = \frac{P(\overline{T_v} \cap A_f)}{\overline{P_v}} = \frac{P\{\overline{T_v} \,|A_f\}}{\overline{P_v}} \times P_f \qquad 4$$

For mathematical reasons, probabilities can be expressed more conveniently as odds:

$$O = \frac{P}{1-P} \qquad 5$$

Then, equations 3 and 4 can be expressed as:

$$O\{A_f \mid T_v\} = \frac{O\{T_v \mid A_f\}}{O_v} \times O_f \qquad 6$$

$$O\{A_f \mid \overline{T_v}\} = \frac{O\{\overline{T_v} \mid A_f\}}{\overline{O_v}} \times O_f \qquad 7$$

In the weight of evidence method, the natural logarithms are applied to both equations 6 and 7 to obtain:

$$\ln O_T\{A_f \mid T_v\} = W_v^+ + \ln O_f \qquad 8$$

$$\ln O_T\{A_f \mid \overline{T_v}\} = W_v^- + \ln O_f \qquad 9$$

where W+ is the positive weight to be assigned when the class *n* of the *j* parameter is present, and W- is the negative weight to be assigned when the class *n* of the *j* parameter is absent.

The weights are calculated by the following equations:

$$W^+ = \ln \frac{P\{T_v \mid A_f\}}{P\{T_v \mid \overline{A_f}\}} = \ln\left(\frac{\frac{T_v \cap A_f}{A_f}}{\frac{T_v \cap \overline{A_f}}{\overline{A_f}}}\right) = \ln\left(\frac{\frac{Landslide\ area\ in\ the\ considered\ class}{Total\ lanslide\ area}}{\frac{Stable\ area\ in\ the\ considered\ class}{Total\ stable\ area}}\right) \qquad 10$$

$$W_v^- = \ln \frac{P\{\overline{T_v} \mid A_f\}}{P\{\overline{T_v} \mid \overline{A_f}\}} = \ln\left(\frac{\frac{\overline{T_v} \cap A_f}{A_f}}{\frac{\overline{T_v} \cap \overline{A_f}}{\overline{A_f}}}\right) = \ln\left(\frac{\frac{Total\ lanslides\ area\ in\ the\ other\ classes}{Total\ lanslide\ area}}{\frac{Total\ stable\ area\ in\ the\ other\ classes}{Total\ stable\ area}}\right) \qquad 11$$

The W+ positive weight is directly proportional to the influence that the *n* class of the *j* parameter has on landslides development.

To analyze the influence of several parameters on the distribution of landslides in the area, the weights of each parameter are summarized, as these parameters are mutually statistically independent:

$$\ln O_T\{A_f \mid T_{v1}^k \cap T_{v2}^k \cap T_{v3}^k \ldots\ldots \cap T_{vn}^k\} = \sum_{i=1}^{n} W_{vi}^K + \ln O_f \qquad 12$$

where k can take the sign + or - depending on the presence or absence of this parameter.

The difference between the positive and negative weights, as computed for each class of each parameter analyzed, is a good indicator of its relation with landslides:

$$C = W^+ - W^- \qquad 13$$

The value of C is typically between 0 and 2; when the value of C tends to zero, the presence of the considered parameter does not affect the distribution of landslides in the area; conversely, when C is approximately 2 or more, the correlation is very significant.

T is the sum of the product between posterior probability and the area, extended to all elementary cells in which the territory was divided: ideally T should then be approximately equal to the total landslide area. In the practical applications, it frequently happens that T is greater than the landslide area, because there are generally some conditional dependencies between the variables; if T does not exceed landslide area for more than 15%, conditional independence between the considered variables exists. This test is called "omnibus test" (Agterberg and Cheng, 2002).

3. ANALYSIS

In order to construct a landslide susceptibility map, three basic susceptibility landslide parameters were utilized: Lithology, Land Use and Slope.

In order to test the reliability and the predictive capabilities of the weight of evidence method, the influence of each parameter on the landslide susceptibility is calculated in a data sub-set obtained randomly extracting around one percent of the whole study area (training set). These samples have the same proportion between the landslide area and the total area than the whole studied area (validation set), that is about 0,063%. The extraction process took into account the morphological, lithological and land use characteristics of the area, in order to obtain a sample that is truly representative of the whole studied area.

The various thematic maps of the training set were overlapped with the landslides map. On the basis of these intersections, by using ArcView 3.2 software and its extension ArcSDM (Spatial Data Modeller) positive and negative weights for each class of each parameter were calculated (Table 1):

Table 1. The weights calculated for the Lithology parameter

Classes	Areas (km^2)	Landslides (n.)	Weight +
Alluvial deposits	0.0076	0	
Limestones and dolomites	0.0109	1	-0.5935
Metasandstones	0.2494	27	-0.4242
Limestones talus	0.0326	7	0.2950
Porfiric dikes	0.0004	1	4.4843
Argillaceous talus	0.0988	29	0.6327

A quick analysis of these results highlighted some aspects of this methodology:

- the absence of landslide areas in the training set occupied by Alluvial deposits makes impossible to calculate their weight;
- the presence of a single landslide in the training set occupied by Porfiric dikes means that the weight assigned to them is too high, because there are some statistical problems due to the very small area occupied by this formation.

These two anomalies are attributable to the fact that these two lithologies occupy a very small area, compared to the extension of the basin. Looking at the weights assigned to other formations, the best correlation between the Argillaceous talus and the landslides may be observed. Limestones and dolomites received the lowest weights, and this was expected as these rocks have excellent mechanical properties. Metasandstones obtained only a slightly higher weight than limestones and dolomites: they have very different geomechanical features from the point of view of fracturing and alteration, that are certainly greater in the metasandstones than in limestones and dolomites. Statistical analyses do not highlight these patterns and underestimate the weight assigned to metasandstones. The Table 2 sets out the main results of the analysis with regard to the Land Use parameter.

Table 2. The weights calculated for the Land Use parameter

Classes	Areas (km^2)	Landslides (n.)	Weight +
Shrubby areas	0.1656	29	0.0780
Crops areas	0.0187	0	
Heterogeneous agricultural areas	0.0160	2	-0.2776
Areas with rock outcrop	0.0338	11	0.7465
Forest	0.1587	22	-0.1672
Urbanization	0.0069	1	-0.1142

The calculation of the weights for the "Slope" parameter presents more difficulties than the other factors, since the continuous nature of this parameter does not allow an immediate processing. To solve this problem, the discreetization in a finite number of classes is necessary; the final classes were found after numerous attempts, when the best correlation between the parameter "Slope" and landslides were found (Table 3).

Table 3. The weights calculated for the Slope parameter

Classes	Areas (km^2)	Landslides (n.)	Weight +
0-10°	0.0218	1	-1.3003
10-20°	0.1220	10	-0.7105
20-40°	0.2307	44	0.1674
40-60°	0.0226	9	0.9713
60-90°	0.0024	1	1.0328

4. THE LANDSLIDE SUSCEPTIBILITY MAP

The weight calculated for each class of each parameter is assigned to each elementary grid cell.

For each cell the final probability is the sum of the weights of each parameter and the prior probability. The final probability was obtained using the expression 12:

$$\ln P.Finale = \sum W^{+} + \ln O_{f} \Rightarrow P.Finale = Exp(\sum W^{+} + \ln O_{f}) = Exp(2.012 + \ln(0.052)) = 0.283$$

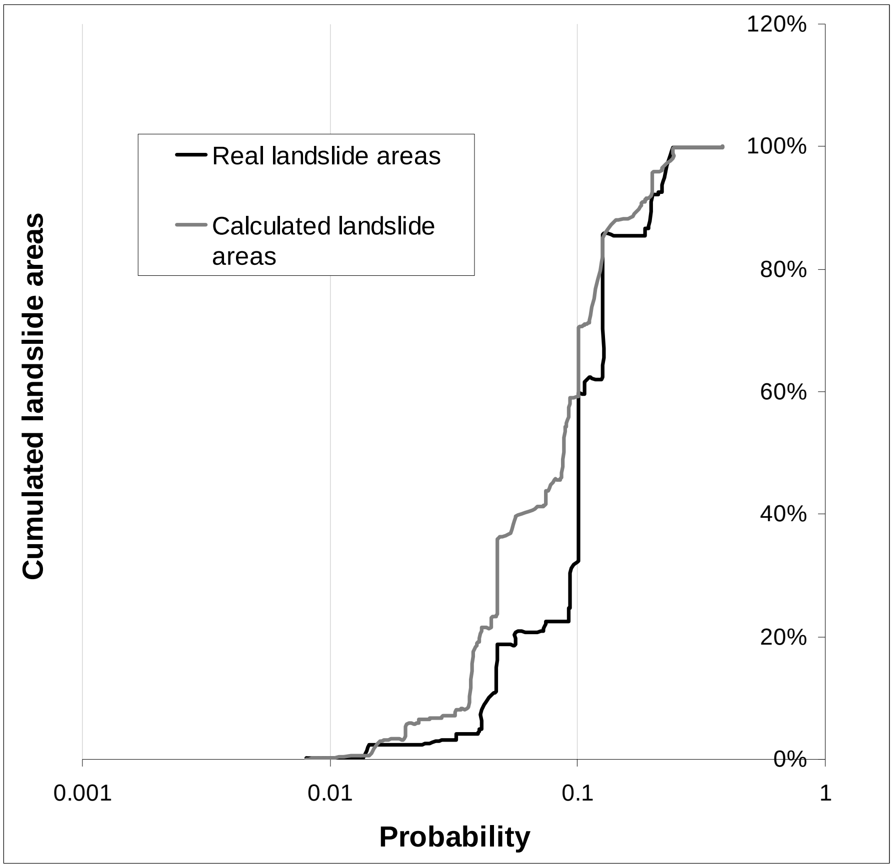
In the Table 4 an example of how to calculate the weights for the cells is shown.

Table 4. Weights assigned to the cells 1 and 2

Id	Slope	Slope W+	Land use	Land use W+	Lithology	Lithology W+	Sum of Weights	Prior probability	Final probability
1	10-20°	-0.7105	Forest	-0.1672	Metasand-stones	-0.4242	-1.3019	0.052	0.014
2	40-60°	0.9713	Areas with rock outcrop	0.7465	Limestones talus	0.295	2.012	0.052	0.283

After the calculation, the weights are applied to the validation set to obtain the landslide susceptibility map. The probability was split in four classes: the range of each class was not chosen arbitrarily, but on the basis of the graph shown in Figure 1.

Figure 1. Ranges of the probability classes



Looking at this graph, two main observations can be made:

- the two curves, related to calculated and real landslide areas, seem to be very similar;
- in both curves, some abrupt changes of the slope may be observed, where the limits of the probability classes were placed.

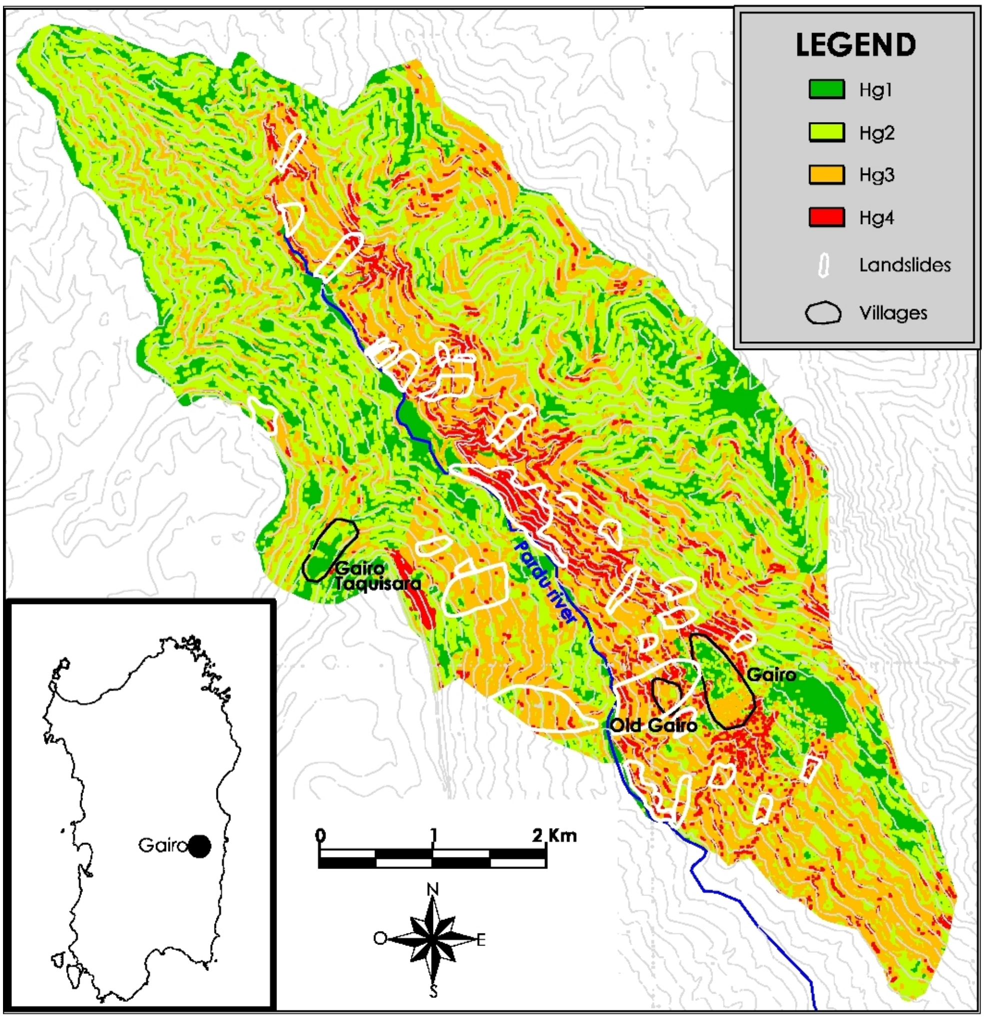
On the ground of these observations, four classes of landslide susceptibility were singled out (Table 5).

Table 5. Classes of landslide susceptibility

Classes	Value
Hg1	0 – 0.04
Hg2	0.04 – 0.09
Hg3	0.09 – 0.18
Hg4	0.18 – 1

On the basis of all these data, the landslide susceptibility map of the studied area was drawn (Figure 2).

Figure 2. Landslide susceptibility map of the Rio Pardu Valley near Gairo S. Elena



Checking this map, the stable areas appear those with a slope less than 20° and, partly, the areas where the metasandstones, limestones and dolomites outcrop. The areas classified as Hg2 are occupied by forests and are concentrated in the north-east part of the Pardu Valley and partly at north of the Gairo village. Finally the areas classified as Hg3 and Hg4 are concentrated where the Limestone and Argillaceous talus outcrop and the slope is greater then 40°.

5. CONCLUSIONS

The analysis performed using the weight of evidence method showed the following advantages:

- it is an objective system that can discriminate between the various parameters, in order to understand which are the most important parameters in the development of landslide phenomena;
- the weights are calculated separately for each study area, allowing the method to choose different weights, for the same parameters, for different geomorphological settings;
- the process of assigning weights is objective and almost independent by choices of the observer.

The analysis showed also the following disadvantages:
- the results of different analyzed areas are comparable only if the classification ranking is standardized;
- the method is not suitable for areas where different types of landslide movements occur;
- the method need accurate and reliable information on past landslide movements;
- the method either overvalues or undervalues the weights of the parameters related to very small portions of the study area.

REFERENCES

Agterberg, F.P., G.F. Bonham-Carter, and D.F. Wright, (1990), Statistical Pattern Integration for Mineral Exploration. *Geological Survey of Canada Contribution,* 24088.

Agterberg, F. P. and Q. Cheng, (2002), Conditional Independence Test for Weights-of-Evidence Modeling. *Natural Resources Research*, 11(4), 249-255.

Bonham-Carter, G.F., F.P. Agterberg, and D.F. Wright, (1988), Integration of Geological Datasets for Gold Exploration in Nova Scotia. *American Society for Photogrammetry and Remote Sensing*.

Bonham-Carter, G.F., F.P. Agterberg, and D.F. Wright, (1989), Weights of evidence modelling: a new approach to mapping mineral potential. *Statistical Application in the Earth Sciences*, Agterberg, F.P. and Bonham-Carter, G.F. (Eds.), Geological Survey of Canada, 89(9), 171-183.

Cocozza, T. (1974), Schema stratigrafico-strutturale del massiccio sardo-corso e minerogenesi della Sardegna. *Memorie della società geologica italiana*, 13.

Kemp, L.D., G.F. Bonham-Carter, G.L. Raines, and C.G. Looney, (2001), Arc-SDM: Arcview extension for spatial data modelling using weights of evidence, logistic regression, fuzzy logic and neural network analysis.

Lee, S., J. Choi, and K. Min, (2004), Probabilistic landslide hazard mapping using GIS and remote sensing data at Boun, Korea. *International Journal of Remote Sensing*, 25(11), 2037-2052.

Rezaei Moghaddam, M.H., M. Khayyam, M. Ahmadi, and M. Farajzadeh, (2007), Mapping susceptibility landslide by using the weight-of-evidence model: a case study in Merek Valley, Iran. *Journal of Applied Science*, 7(22), 3342-3355.

Vardabasso, S. (1956), La fase Sarda dell'orogenesi Caledonica in Sardegna. *Zeitschrift der Deutschen Geologischen Gesellschaft, Band Symp*, 120-127.