The use of predictive modeling techniques for optimal exploitation of spatial databases: a case study about landslide hazard mapping in the Northern Apennines (Italy)[1]

Alberto Pistocchi(*) · Lucia Luzi(**) · Paola Napolitano(***)

(*) GECOsistema srl , viale G. Carducci, 15, 47023 Cesena, Italy
alberto.pistocchi@gecosistema.it

(**) INGV Milano, Italy

(***) ACTA Studio Associato, Naples, Italy

Abstract. A case study is presented in which different probabilistic prediction models (Bayesian probability, fuzzy logic "and", "or", "sum", "product", "gamma" operations, and certainty factors) are used to produce landslide hazard maps for a hilly and mountainous region in the northern Apennines, Italy. Seven data layers are exploited to detect the most vulnerable areas: lithology, distance from the geological lineaments, annual rainfall amount, land cover type, topographic slope and aspect, and the distance from hydrographic network segments. The results of the different predictions are compared using the prediction rate index and critically discussed, to evaluate the possibility of using readily available databases for land planning.

Keywords. Favorability functions - Integration modeling - Landslide hazard - Spatial database

General issues

In the last few years, a great development has been seen in the set up of spatial databases by regional land planning authorities all over Europe. However, it seems that many databases are still ineffective for decision support, and the use of available data is very often quite naive. In particular, final data users and decision makers tend to have little consciousness about the modeling capabilities of geographic information systems (GIS; Campbell and Masser *199*5).

Predictive models, based on causal relationships between underlying physical phenomena, are in widespread use among hydrologists, geoscientists, environmental analysts, and engineers, for applications in the field of natural risk assessment, natural resource management, pollution prevention and reclamation, and environmental impact assessment. In the case of natural hazards such as landslides, the traditional approach relies on the expertise of geologists and geomorphologists to detect the features that account for the occurrence of landslides. This brings an appropriate recognition of past occurrences, but cannot support any prediction apart from the expert's subjective and qualitative judgement.

Therefore a prediction made through "objective" replicable models, with limited arbitrary choices by the analyst, might be of relevant interest.

This leads to the investigation of the possibility of using probabilistic approaches to prediction, where prior knowledge of landslide occurrences is used to the full extent to make probabilistic predictions via well-established parametric, fuzzy, or stochastic map overlaying methods.

Over the last years, many approaches of this kind have been explored (Carrara *198*3; Carrara and others *199*5; Luzi and others *199*8; Massari and Atkinson *199*8). All these methods have been extensively compared using sensitivity analysis or the performance of different methods on the same case study (van Westen *199*3). A major difficulty with these applications is the comparison of different maps (Guzzetti and others *200*0).

Chung and Fabbri *(199*3) proposed a framework approach to the problem, referred to as the favorability function mapping. In their work, the authors show how a wide range of probabilistic, fuzzy, and evidential functions can be used to detect the most favorable areas for the occurrence of phenomena, such as landslides or mineral deposits. In this way, a unique criterion for the comparison of different prediction maps is easily found in what Chung and Fabbri *(199*9) call the prediction rate, i.e., a measure of the goodness of the model's validation.

---

[1] The full version of this paper is published on Environmental Geology (2002): 41 765-775

The aim of this paper is to discuss the applicability of the favorability function modeling in producing a hazard map for landslide occurrences using standard currently available databases, and to check how this method might improve the use of information in an existing database, compared with other techniques.
It is explained how favorability function modeling can be used as a conceptual scheme for the structuring of databases: data collection is strictly dependent on the theoretical framework in which information will be used.

Application

The area used for the case study is the Savio River catchment, a province of Forlì-Cesena, Emilia Romagna in northern Italy. A geological outline of the area indicates a basically sedimentary basin with a dominance of marls and sandstones (the "Marnoso Arenacea Romagnola" formation). The area is covered by a large number of landslides, occurring in most cases as slide type movements or debris flows, in different geological units.
Before describing the application of the favorability function modeling to landslide hazard zoning, it must be recalled that the aim of this study is to evaluate the predictive capability of a real world database, rather than producing reliable hazard maps, and thus the best available information has been used without any further field investigation and data capturing.
The data used in the research were made available by the Regione Emilia Romagna Geological Survey (Regione Emilia Romagna *1991*). The database used for the case study consists of thematic layers concerning:

1. structural lineaments (faults, synclines, and anticlines), scale 1:50,000;
2. lithological units, scale 1:50,000;
3. land cover obtained from TM Landsat imagery according to the CORINE European Project Guidelines (Briggs and Martin *198*8), scale 1: 50,000;
4. digital terrain model (DTM) obtained from 50-m equidistance contour lines available from the cartographic database of the Regione Emilia Romagna through standard linear interpolation;
5. rainfall measurements at seven gauge stations over the area (data published by the Regione Emilia Romagna *199*6);
6. digitized hydrographic network, scale 1: 10,000.

The detailed data description and transformation for the analysis is treated in the paper by Pistocchi et al. (2002).
All thematic data considered have, in principle, the possibility of mutual association, that would lead to a computational effort that might give useless results because of redundant information. Before the favorability modeling calculation, an association test has been performed over the seven themes used for the analysis, classified into discrete legends and tested with respect to the map of active landslides.
Four association indices were calculated for each map pair (Press and others *198*6; Bonham Carter *199*4): a) the chi-square index; b) the Cramers index and c) the contingency index.
As comes from the analysis, it must be noticed that landslides show some association with lithology, and a spatial trend associated with elevation/rainfall and land cover.
If one looks for association among causal factors, it must be noticed that lithology is associated with elevation/rainfall and land cover, whereas the association is weaker with slope, and even less with the other themes. The inadequate DTM available for the study seems to be a prime cause of this. All other associations can be considered irrelevant.
During each run of the favorability modelling calculation, only one half of the known landslides (chosen by random sampling) was used to make the prediction map, whereas the remaining ones were considered as a validation data set. To predict the landslide hazard pattern, potential causal factors were initially used, and in a second attempt only the three most relevant factors were used.

Discussion of the results

The computation of favorability functions was performed for different modeling hypotheses (Pistocchi et al., 2002). The test on the predictive power of each computed favorability map was done using a prediction rate curve (Chung and Fabbri *199*9). This curve was obtained by plotting the cumulative percentage of the study area, sorted by decreasing value of the favorability value, as the abscissa against the cumulative percentage of the landslide area as the ordinate. In a wider sense, the closer the curve approaches the ordinate axis, the

more the prediction fits. In contrast, the more the curve approaches the 45° straight line, the less it is useful to combine factors because the prediction is close to a random distribution of the favorability values.
From the comparison of the prediction rates one can assess that:

1. There seems to be no significant improvement when using six causal factors instead of the three more strictly related to landslides (lithology, land cover, and rainfall).
2. A further smoothing effect can be observed when all landslide bodies are used as evidence, instead of the trigger points only.
3. From the prediction rate diagrams it can be seen that the causal factors that have higher predictive capability are in any case lithology (thus confirming the choice of the Regione Emilia Romagna Geological Survey to use this layer alone for hazard mapping), and then land cover and rainfall. All other themes are not relevant for prediction.
4. The seven predictors used in this case study behave very similarly in prediction, except for the case of Bayesian probability - which is very sensitive to the actual dispersion of the data, and approaches a random prediction when all landslide bodies are used as evidences - and fuzzy "or", "and" in some cases. In general, it seems that the certainty factor predictor is the most useful method in this specific case study (see Figure 1 and Figure 2).

This analysis has led to the recognition that the existing database is not fit, however, for predictive modeling because of inadequate topographic data. This constitutes an input for addressing future survey and data capturing to define a better digital terrain model.

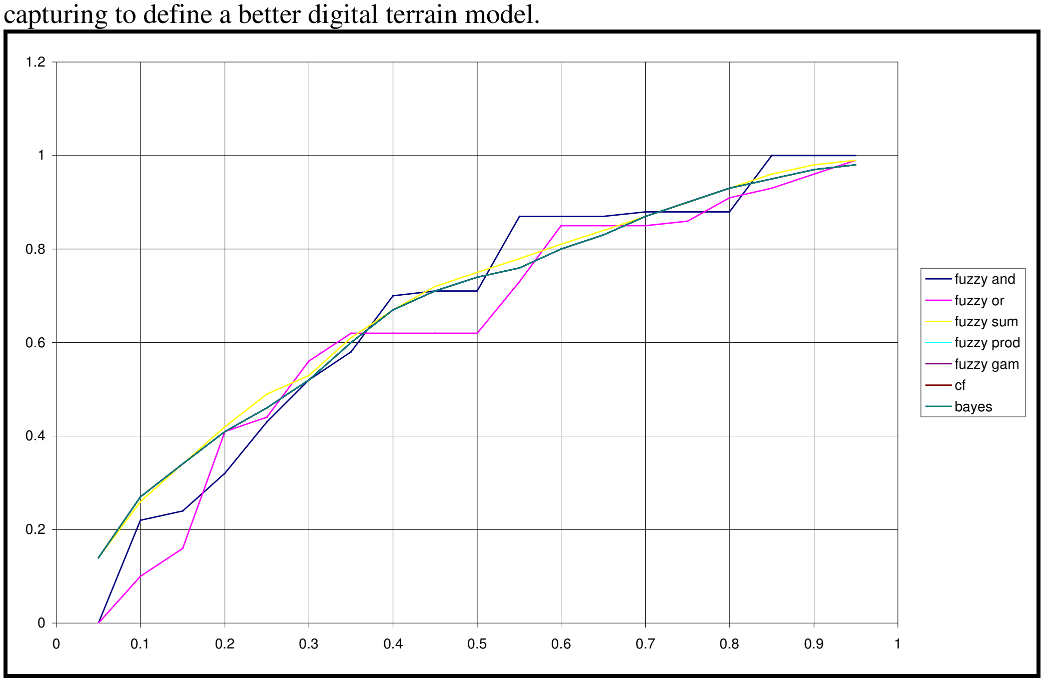

Figure 1 - prediction rate of the seven predictors using the six relevant causal factors and the trigger points of landslides.

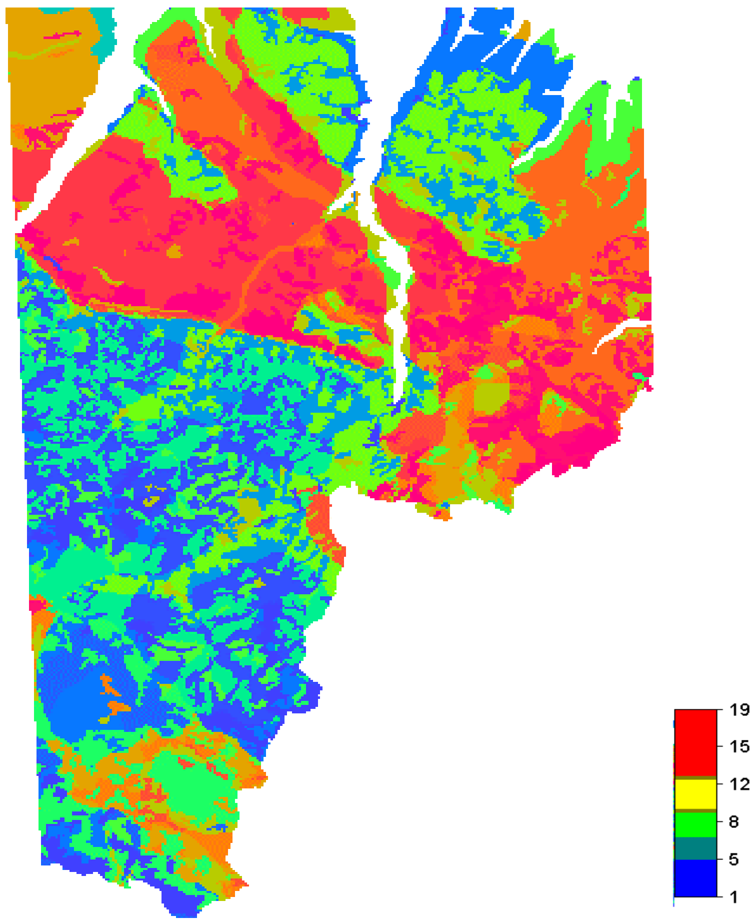

Figure 2 - prediction maps according to the certainty factors predictor, using rainfall, lithology and land cover themes and the trigger points as evidences. The region is subdivided into equal area classes in increasing order of favourability (class 20=max favourability) according to the legend.

As soon as the improved causal factor map is produced, or a new causal factor is supposed to be relevant for the phenomenon, the calculations can be repeated and a new prediction map can be produced. The validation using prediction rates allows a check for actual and effective improvements, and can be used to orient further efforts in data acquisition and geotechnical monitoring.

References

Bonham-Carter G (1994) GIS for geoscientists - modeling with GIS. Elsevier, New York

Briggs DJ, Martin DM (1988) CORINE: an environmental information system for the European Community. Eur Environ Rev 2(1):29-34

Campbell H, Masser I (1995) GIS and organizations: how effective are GIS in practice? Taylor and Francis, Bristol

Carrara A (1983) Multivariate models for landslide hazard evaluation. Math Geol 15(3):403-427

Carrara A, Cardinali M, Guzzetti F, Reichenbach P (1995) GIS based techniques for mapping landslide hazard. In: Carrara A, Guzzetti F (eds) Geographical information systems in assessing natural hazards. Kluwer, Dordrecht, pp 135-175

Chung CF, Fabbri AG (1993) The representation of geoscience information for data integration. Non Renew Resour 2(2):122-139

Chung CF, Fabbri AG (1999) Probabilistic prediction models for landslide hazard mapping. Photogrammet Eng Remote Sensing 65(12):1389-1399

Guzzetti F, Cardinali M, Reichenbach P, Carrara A (2000) Comparing landslide maps: a case study in the upper Tiber river basin, central Italy. Environ Manage 25(3):247-263

Luzi L, Pergalani F, Terlien MTJ (1998) A probabilistic approach to assess uncertainty of landslide vulnerability to earthquakes. In: Buccianti A, Nardi G, Potenza R (eds) Proceedings of the fourth annual conference of the International Association for Mathematical Geology. De Frede, Naples, pp 235-240

Massari R, Atkinson PM (1998) Using the Gibbs sampler in mapping susceptibility to landsliding. In: Buccianti A, Nardi G, Potenza R (eds) Proceedings of the fourth annual conference of the International Association for Mathematical Geology. De Frede, Naples, pp 247-252

Pistocchi A, Luzi L, Napolitano P. (2002) The use of predictive modeling techniques for optimal exploitation of spatial databases: a case study in landslide hazard mapping with expert system-like methods. Environmental Geology 41, pp. 765-775

Press WH, Flannery BP, Teukolsky SA, Vetterling VT (1986) Numerical recipes: the art of scientific computing. Cambridge University Press, Cambridge

Regione Emilia Romagna (1991) Piano Territoriale Paesistico Regionale, Bologna

Regione Emilia Romagna-Servizio Meteorologico Regionale (1996) Tavole climatologiche dell'Emilia Romagna 1951-94, Bologna

Van Westen CJ (1993) Application of geographical information systems to landslide hazard zonation. PhD Thesis, Technical University of Delft, ITC publication no 15