

Occurrence neighbourhoods and risk assessment from landslide hazard in northern Spain

A. G. Fabbri^{1,3}, J. Remondo², C. Ballabio³, S. Poli³, C. F. Chung⁴
& H. J. Scholten¹

¹*SPINlab, Vrije Universiteit, Amsterdam, Netherlands*

²*DICITIMAC, Universidad de Cantabria, Santander, Spain*

³*DISAT, Università di Milano-Bicocca, Milan, Italy*

⁴*Geological Survey of Canada, Ottawa, Canada*

Abstract

This contribution analyzes the problem of selecting the desirable characteristics of a study area when using geo-information for natural risk assessment. Shape, boundary, density of detail of the study area and the distribution of hazardous occurrences can be fundamental in conditioning the estimation of values in a map of expected risk. A study area in the Basque Country of northern Spain is used in which previous studies produced maps of risks for linear infrastructures, land uses and buildings, from thousands of shallow translational landslides. The area is reconsidered here in terms of five telescopic sub-areas corresponding to different neighbourhoods of the landslide occurrences. The results of the corresponding hazard predictions are interpreted via prediction-rate tables and curves obtained from blind tests, i.e., prediction maps obtained using only part of the occurrences cross-validated with the distribution of the remaining occurrences. The subsequent introduction of socioeconomic thematic maps and scenarios enables the derivation of risk maps based on the prediction rates, the hazard maps and the socioeconomic indicator values. The comparison of the risk maps from the different study-area datasets is used to assess their impact on risk values and to provide guidance on how to perform the selection maintaining greater significance. A critical issue is the loss of significance when reducing study area neighbourhoods closer or further away from the hazardous locations. The application is an example of a general purpose spatial predictive modelling processing strategy for which dedicated software has been developed.

Keywords: study area selection, landslide hazard prediction, risk assessment, blind tests, spatial prediction modeling.



1 Introduction

This contribution explores one aspect usually ignored in studies of spatial prediction modelling: the statistical properties of a study area that make it worth using for the modelling. Today, many an application has been made to predict natural hazard and particular attention has been paid to landslide processes as causes of damage to human activities and assets. The use of geo-information technology has encouraged the construction of spatial databases from which the distribution of occurrence of hazardous events is inductively related (Alexander [1]) with that of contextual map data layers considered as relevant to, or typical of, or even causal factors to the occurrences.

Examples can be seen of studies of different area extensions with diverse densities of landslide occurrences (van Westen *et al.* [2], 21 km²; Zêzere *et al.* [3], 17 km²; Coe *et al.* [4], about 315 km², or Cardinali *et al.* [5], 79 places surrounding known landslides within a 8456 km² region; and Remondo *et al.* [6], a 500 km² area).

Generally, no particular statistical justification is used (put forward) for a study and mainly reasons of geomorphologic or physiographic uniformity, cartographic unit delimitation or sociologic relevance are the arguments used to justify the selection of study area and of its boundaries. Nevertheless, the geometrical statistical properties of a study area may be critical in the establishment of the spatial relationships supporting the derived ranking of hazard classes.

How could or should one proceed in selecting a representative or effective study area boundary in a specific case study? What is the zone of influence of the hazardous occurrences and the respective typical setting to be mapped from a spatial database? How confident are we that the database provides significant relationships between the occurrences and their spatial context? Does it contain all the variability to identify typical hazardous settings? What criteria have we used to select the study area, its shape and extension?

To help answering at least in part those questions, a study area database from earlier landslide hazard analyses has been used. It was artificially modified to represent different circular neighbourhoods of a fixed number of landslide locations. Following a description of the study area (previously analyzed for hazard prediction and risk assessment), five different sub-areas are generated of increasingly larger landslide neighbourhoods. The results of estimated probability of occurrence are obtained for each sub-area and compared, maintaining consistence between the experiments. Subsequently, using the same scenario of future processes and human impact, the corresponding risk assessments and risk maps are derived. Following a comparison of the risk values and rankings, a discussion is made of the results and the ensuing opportunities for further work in risk analysis.

2 Study area, sub-areas and the problem

The study area, located in Figure 1, is part of a spatial database for the lower Deba valley in the Basque province of Guipúzcoa in northern Spain. The



database has been the subject of a number of hazard studies (Remondo [7]; Remondo *et al.* [8], [9]). It has been later expanded with addition of socioeconomic information to enable risk assessment (Remondo *et al.* [6]; Bonachea [10]). It covers approximately 140 km², with elevations ranging between 0 m and 700 m a.s.l., and the main annual rainfall is 1500 mm.

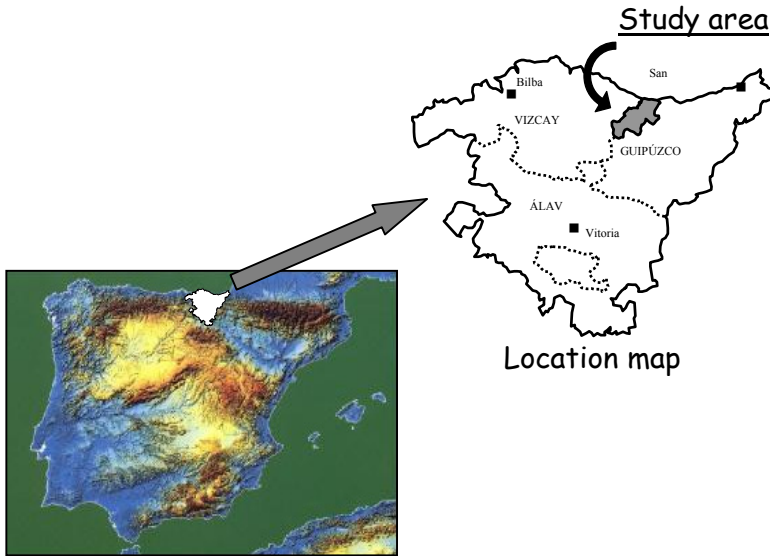


Figure 1: Location of the Deba Valley Study area in the Basque Country, northern Spain.

Geologically, the terrain consists of moderately folded and faulted limestones, marls, claystones, sandstones, flysch facies and volcanic rocks of the Cretaceous and Paleogene of the Basco-Cantabrico Pyrenees. Slopes are generally steep (average slope gradient is 22°) and there are surficial deposits of different composition and thickness that determine the occurrence of many hundreds of landslides triggered by intense rainfall episodes. Extensive field surveys and photo-interpretations of different flight coverages allowed mapping landslides of different dynamic types in temporal groups (Remondo *et al.* [9]).

For this study 1206 shallow translational landslides and associated flows were considered: 300 that occurred past 1997 and 906 pre 1977. Because some were reactivated after 1977, a subtotal of 1123 landslides was eventually used, with only 217 considered as post-1977.

The digital database consisted of rasterized images of 1886 columns x 1555 rows of 10 m pixel resolution. Within that image space the area of concern occupied 1,393,541 pixels. The average size of the landslide main failure is about 400 m², and the location of each was assigned to a single pixel of 10 m resolution. Besides the two digital images containing the location of the 1123 landslides (906 pre-1977, and 217 post-1977), shown in Figure 2, six “causal

factor” data layers were used: three continuous ones (elevation, slope and aspect) and three categorical ones (geology with 28 units, vegetation-land use with seven units, and thickness of surficial deposits with three units, between 0.5 and 3.0 m).

To obtain sub-area databases corresponding to landslide neighbourhoods of different radii, the image with the 1123 landslide pixels was spatially transformed to generate circular pixel patches around the pixel location of the landslides.

Different study areas were obtained sub-setting the images of the “causal factors” so that only the values corresponding to the patches were considered. The following five study areas were generated: named Mask 1 (the initial study area), Mask 2 (neighbourhoods of diameter 7 pixel), Mask 4 (of 11), Mask 6 (of 13), and Mask 9 (of 21). Their statistics is shown in Table 1. In this manner, study areas with different extensions and relative percentage of landslide pixels were obtained (0.08, 0.51, 1.10, 1.53, and 3.52%, respectively). Purpose of the transformation was to study their effect on the predicted hazard values, the ensuing estimates of probability of occurrence, and finally on the risk values assessed.

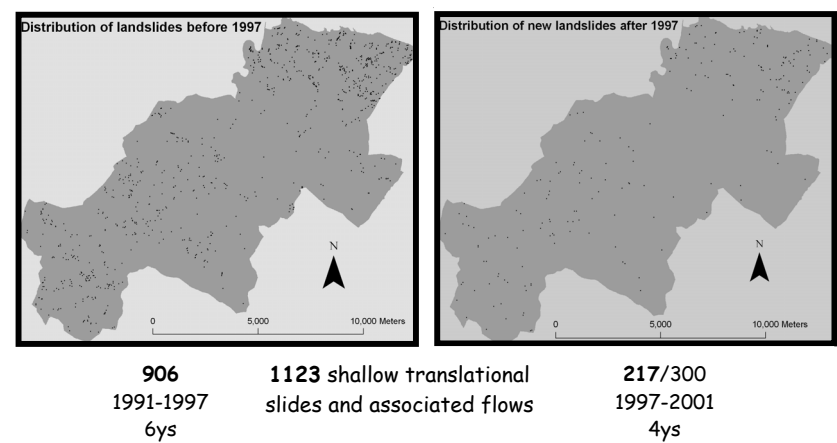


Figure 2: The distribution of hazardous events in the study area during two successive periods of six and four years, respectively. The distribution of the older period events was used for predicting and that of the later period was used for cross-validation.

Table 1: Masks, study areas and percentage relative of landslides.

Areas	Number of pixels	% of landslides
Mask 1	1,399,341	0.08
Mask 2	31,867	3.52
Mask 4	73,424	1.53
Mask 5	102,480	1.10
Mask 9	218,323	0.51

3 The analytical strategy

Prerequisite to spatial prediction modelling are several basic assumptions and related scenarios that justify its application. They relate to: (1) the significance of the database; (2) the distribution and type of hazardous events; (3) a scenario for the distribution of future events; (4) a scenario for the empirical validation of the prediction results; and (5) a scenario for the interpretation of the validation via cost-benefit analysis.

A spatial database must be assumed to be available that contains all or most or even a sufficient amount of the information about the past hazardous events and on their spatial setting such as causal factors or characterizing parameters.

The distribution of the past hazardous occurrences in time and in space, including their triggering factors, intensities and dynamic type is assumed to be similar to, or comparable with, that of future events for a same time interval of similar length.

The impact of the future hazardous events is similar to or comparable with that of the past events.

The prediction results, i.e., the hazard map, generated by using all available hazardous events can be evaluated by mean of an additional prediction obtained by subdividing the past hazardous events into two groups. One group is used for predicting, and the other for empirically cross-validating the second prediction results, a second hazard map. It means that the “validity” of the first prediction is assumed to correspond to that of the second prediction obtained by the “blind test.”

Once the empirical validation is performed, the statistics obtained for the several hazard classes should be evaluated in terms of cost-benefits to identify the hazard classes having greater reliability.

As a logical consequence of the assumptions and scenario just described, the following 3-stage strategy can be set up.

Stage 1. Select a prediction model out of a few that may be applied, use all available hazardous events of a given type and all causal data layers, and obtain a first prediction map. The map will have ordered levels of hazard for equal area classes as proportions of the study area (e.g., 0.5%, 1.0%, etc.).

Stage 2. Interpret the prediction results obtained in Stage 1 by subdividing the past hazardous events into two or more groups, to obtain, using the same prediction model, a new prediction map and its validation, respectively. The subdivision could be obtained by time partitions, if the hazardous events can be separated into an older and a more recent set of events. Should such a time partition not be possible, a random partition into two halves of the occurrence can be used, for instance using the first random half for predicting and the second half for validating. Another form of random partitioning the past occurrences is the take-one-out (or take-a-group-out) to see how well a prediction that uses the remainder fares. In such cases what is predicted is the next occurrence or the next group, without a specific time connotation. The results of the validation of the classes in the second prediction are in the form of tables, histograms and cumulative curves that express the relative prediction

power, assumed to be also the power of the first prediction. Once a scenario on the spatial impact of the future hazardous events is selected, it is possible to estimate the probability of occurrence of the future hazardous events (for instance assuming that in the future as many events will occur as the ones observed in the past). Such probabilities of occurrence for each predicted equal area hazard class can be scaled down to smaller unit areas such as pixels in the spatial database for subsequent risk analysis.

Stage 3. Use the additional socioeconomic indicator data layers about population density distribution, dwellings, infrastructures, activities and land use classes with their associated values, expected losses and vulnerabilities, to resolve for each one the risk equation,

$$R = H * V * E, \quad (1)$$

where R is the risk (€/year), H is the predicted hazard (0-1/year), V is the vulnerability (0-1) and E is the element value (€) exposed to the risk.

For each type of risk use the hazard map from the first prediction, assign to each pixel in each class the probability of occurrence obtained by the second prediction and then compute the risk values per pixel.

Various spatial prediction models can be used to establish a spatial relationship between the distribution of the hazardous events (points, lines or polygons) and that of the supporting or causal factors. Chung and Fabbri [11,12] discuss a mathematical framework for models based on fuzzy sets, likelihood ratios, linear and logistic regression and Bayesian probability. Here we have used the empirical likelihood ratio, ELR, as proposed by Chung [13]. The ELR highlights the difference between a function of the conditional probability of combinations of causal factors in the presence of hazardous events and a function of the conditional probability of combinations of causal factors in the absence of such events.

4 Predictions for study areas and impacts of neighbourhoods on risk maps

The initial database of study area Mask 1 was reduced in extension to generate the databases for study areas Mask 2, Mask 4, Mask 5 and Mask 9.

Landslide hazard (the term susceptibility is used as synonym) *Stage 1* prediction maps were generated for all the study areas, using first all the 1123 landslides, and successively using only the 906 pre-1997 landslides, shown in Figure 2. Figure 3 shows the *Stage 2* second prediction hazard map for the Mask 1 area, with the distribution of the 217 validation landslides. The cumulative prediction-rate curves for the *Stage 2* predictions and cross-validations are shown in the diagram of Figure 4. There the horizontal axis provides the relative proportions of the study areas classified as hazardous as classes in decreasing order of predicted values of likelihood ratios. On the vertical axis are the corresponding cumulative proportions of the 217 validation landslides. The diagram provides a measure of how “good” the predictions are. While a

distribution of values on the diagonal line connecting points (0, 0) and (1,1) indicate a random distribution, one away from it indicates that the validation landslides tend to cluster within classes with high predicted values.

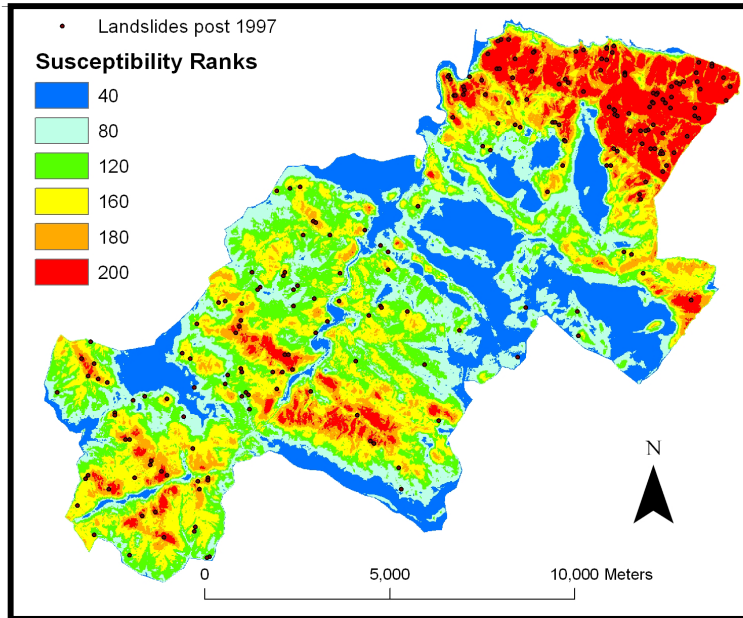


Figure 3: The second prediction map obtained using the distribution of the 906 landslides during 1991-1997, and the distribution of the 217 landslides that occurred during 1997-2001. The rectangular inset will be used to visually compare prediction maps of the other study areas.

Clearly, the relative significance of the prediction-rate curves is much greater for Mask 1 to decrease drastically from Mask 9 to mask 2. A measure of effectiveness of the prediction-rate curve in terms of cost-benefit analysis was used by Chung and Fabbri [14] to isolate the parts of a prediction-rate curve with inclination acceptably steeper than that of the diagonal line. In this case and only for the Mask 1 prediction, the effective classes correspond to the most hazardous 10% of the study area (it predicts 47% of the “future” landslides), and the least hazardous 30%, as shown by the two black horizontal bars in Figure 4. As a consequence, we can see that reducing the size of the landslide neighbourhoods, the distribution of predicted landslides in the classes approaches randomness.

We can fit a continuous line to the prediction-rate curve and convert the prediction rates into probabilities of occurrence, as shown in Figure 5, by the following expression that computes values for a pixel x :

$$\hat{P}_x = 1 - [1 - p_y]^{na/ny} \quad (2)$$

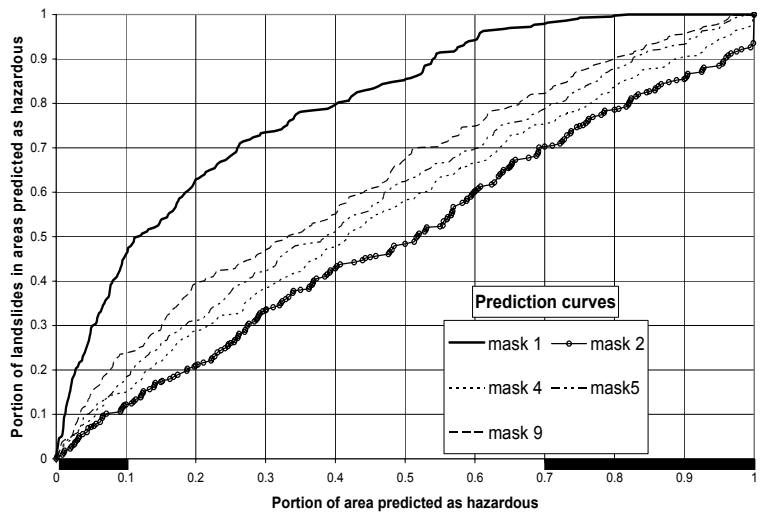


Figure 4: The cumulative prediction-rate curves for the five study areas corresponding to the different landslide neighbourhoods. The black bars on the horizontal axis identify the effective portion of the curve for the Mask 1 study area.

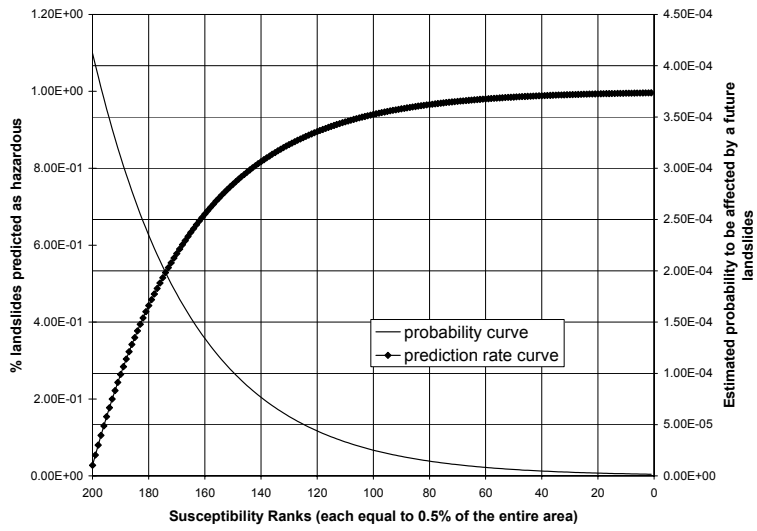


Figure 5: The fitted prediction rate curve for the Mask 1 study area and the respective curve for the probability of occurrence.



where, n_γ is the cumulative number of pixels in hazard classes whose levels are \geq to the class; n_α is the number of pixels in the expected future landslides; and p_γ is the cumulative portion of the landslides in the validation group of the hazard classes whose levels are \geq to the class (from the prediction-rate table).

Table 2: Probabilities of occurrence for study areas Mask 1 and Mask 2.

Mask 1			
Classes	% area	% landslides	Probabilities
5	0.05	0.2949	6.825E-04
10	0.05	0.1649	5.063E-04
15	0.05	0.0761	3.759E-04
20	0.05	0.0921	2.792E-04
25	0.05	0.0530	2.075E-04
30	0.05	0.0530	1.542E-04
35	0.05	0.0461	1.147E-04
40	0.05	0.0161	8.526E-05
45	0.05	0.0346	6.340E-05
50	0.05	0.0230	4.715E-05
70	0.2	0.1245	1.444E-05
100	0.3	0.0207	2.519E-06
Mask 2			
Classes	% area	% landslides	Probabilities
5	0.05	0.0714	1.210E-02
10	0.05	0.0507	1.096E-02
15	0.05	0.0507	9.937E-03
20	0.05	0.0369	9.006E-03
25	0.05	0.0507	8.163E-03
30	0.05	0.0737	7.399E-03
35	0.05	0.0461	6.706E-03
40	0.05	0.0484	6.077E-03
45	0.05	0.0253	5.508E-03
50	0.05	0.030	4.992E-03
70	0.2	0.2189	3.368E-03
100	0.3	0.2972	1.885E-03

Assumptions that allow the computations are: (i) a four year time partition is acceptable; (ii) there is uniformity of physical settings in time; and (iii) there is a similarity of triggering factors between the two sets of landslides. Table 2 shows the probability of occurrence for the two extreme study areas Mask 1 and Mask 2. There, for corresponding equal area classes we can see the distribution of validation landslides, and the related probability of occurrence. Clearly, for the Mask 2 area the probabilities of occurrence are two orders of magnitude larger than those for Mask 1.

Table 3: Land use class values and vulnerabilities. Values in €/100m2 of land, 1 pixel.

Code	Land use class	Value	Vulnerability
1	Very dense forest	40	0.1
2	Dense deciduous forest	30	0.2
3	Half open deciduous forest	20	0.35
4	Very dense coniferous	64	0.32
5	Shrubs and bushes	10	0.1
6	Grasslands and cultivation	100	0.6
7	Areas without vegetation	0	0

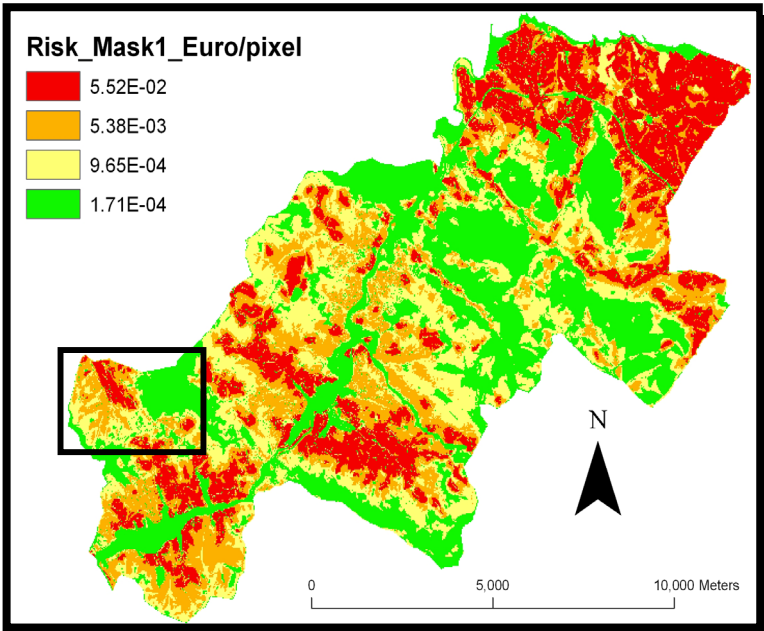


Figure 6: The land-use risk map obtained for study area Mask 1.

In this application the probabilities of occurrence obtained for the five study areas, and the *Stage 1* prediction maps were used to assess the risk to land uses. For seven classes of land use a land-use map (not shown here) monetary values and vulnerabilities were obtained as shown in Table 3. Expression (1) was used to obtain risk values per pixel and then to generate risk maps like the one in Figure 6 for the Mask 1 study area. Figure 7 provides a visual comparison for a sub-area, marked in Figure 6, of risk maps for Mask 1, Mask 9 and Mask 5. Clearly, the risk values obtained tend to increase when reducing the study areas, from about 0.06 to 0.37 €/pixel, in this case.

These experiments indicate that to obtain a meaningful hazard prediction and risk assessment, it is critical to have a spatial database that captures a representative sample of landscape variability to identify classes in a study area that express the typical and recognizable setting of the hazardous events. Measures of the relative “goodness” of a spatial prediction can be easily obtained by cross-validation via blind tests, i.e., pretending that part of the known events are unknown and then using them as proxies for the future events.

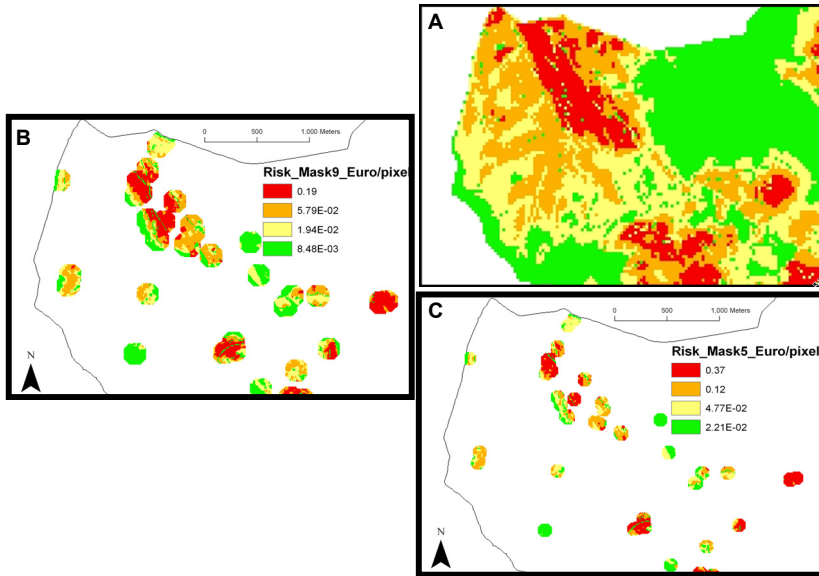


Figure 7: Visual comparison of part of the land-use risk maps for study areas Mask 1 in (A), Mask 9 in (B) and Mask 5 in (C).

5 Concluding remarks

In this contribution a study area previously used for landslide hazard prediction was artificially transformed into several study areas corresponding to landslide neighbourhoods of different radii. It was done to study how the loss of significance varies with the decrease in study area size. It could be seen that the estimated probability of occurrence increases two orders of magnitude when the study area is reduced to 7-pixel circular neighbourhoods of the known hazardous events.

This type of analysis, based on cross-validation strategies is considered important in selecting the study area characteristics for a reliable risk assessment. Furthermore, uncertainty analysis via cross-validation can be performed to demonstrate that uncertainty also varies in a similar manner. For this contribution SPM-SRA software systems for Spatial Prediction Modeling and Spatial Risk Analysis were used that are based on iterated cross-validations (SPM-SRA, [15]; Chung and Fabbri, [16]).

Acknowledgements

This contribution was partly supported by the European Commission's Project ALARM, on Assessment of Landslide Risk and Mitigation in mountain areas (EVG1-CT-2001-00038, 2001-2004), and more recently by Project Mountain Risks: from prediction to management and governance (MRTN-CT-2006-035978, 2007-2010).

References

- [1] Alexander, D., *Principles of Emergency Planning and Management*. Oxford, Oxford University Press, 340 p., 2002.
- [2] van Westen, C.J., Rengers, N. & Soeters, R., Use of geomorphological information in indirect landslide susceptibility assessment. *Natural Hazards*, **30**, pp. 399–419, 2003.
- [3] Zêzere J.L., Reis E., Garcia R., Oliveira S., Rodriques M.L., Vieira G., and Ferreira A.B., 2004, Integration of spatial and temporal data for the definition of different landslide hazard scenarios in the area north of Lisbon (Portugal). *Natural Hazards and Earth system Sciences*, **4**, pp. 133–146, 2004.
- [4] Coe J.A., Michael J.A., Crovelli R.A., Savage W.Z., Laprade W.T. and Nashem W.D., Probabilistic assessment of precipitation-triggered landslides using historical records of landslide occurrence, Seattle, Washington. *Environmental Engineering Geoscience*, **10(2)**, pp. 103–122, 2004.
- [5] Cardinali, M., Reichenbach, P., Guzzetti, F., Ardizzone, F., Antonini, G., Galli, M., Cacciano, M., Castellano, M., & Salvati, P., A geomorphological approach to the estimation of landslide hazards and risks in Umbria, central Italy. *Natural Hazards and Earth System Sciences*, **2**, pp. 57–72, 2002.
- [6] Remondo, J., Bonachea, J. And Cendrero, A., A statistical approach to landslide risk modelling at basin scale: from landslide susceptibility to quantitative risk assessment. *Landslides*, **2(4)**, pp. 321–328, 2005.
- [7] Remondo, J., *Elaboración y validación de mapas de susceptibilidad de deslizamientos mediante técnicas de análisis espacial*. Ph.D. Thesis, Universidad de Oviedo, Spain, 2001.
- [8] Remondo, J., González-Díez, A., Díaz de Terán J.R. and Cendrero, A., Landslide susceptibility models utilising spatial data análisis techniques. A case study from the lower Deba valley, Guipúzcoa (Spain). *Natural Hazards*, **30**, pp. 267–279, 2003a.
- [9] Remondo, J., González-Díez, A., Díaz de Terán J.R., Cendrero, A., Fabbri, A. and Chung, C.F., Validation of landslide susceptibility maps; examples and applications from a case study in northern Spain. *Natural Hazards*, **30**, pp. 437–449, 2003b.
- [10] Bonachea, J., *Desarrollo, aplicación y validación de procedimientos y modelos para la evaluación de amenazas, vulnerabilidad y riesgo debidos a procesos geomorfológicos*. Ph.D. thesis, Universidad de Cantabria,

- Santander, Spain, 356 p., July 2006 (http://www.tesisenred.net/TDR-1124106-134112/index_cs.html).
- [11] Chung, C.F., & Fabbri, A. G., The representation of geoscience information for data integration. *Nonrenewable Resources*, **2(2)**, pp. 122–139, 1993.
 - [12] Chung, C.F., & Fabbri, A. G., Probabilistic prediction models for landslide hazard zonation. *Photogrammetric Engineering & Remote Sensing*, **65(12)**, pp. 1389–1399, 1999.
 - [13] Chung, C. J., Using likelihood ratio functions for modelling the conditional probability of occurrence of future landslides for risk assessment. *Computers & Geosciences*, **32(8)**, pp. 1052–1068, 2006.
 - [14] Chung, C.F., and Fabbri, A. G., Validation of spatial prediction models for landslide hazard mapping. *Natural Hazards*, **30**, pp. 451–472, 2003.
 - [15] SPM-SRA, *Spatial Prediction Modeling System User Guide*, 146 p., and *Spatial Risk Analysis User's Guide*, 40 p. <http://www.spatialmodels.com>, 2006.
 - [16] Chung, C.F., and Fabbri, A. G., Systematic procedures of landslide hazard mapping for risk assessment using spatial prediction models. *Landslide Hazard and Risk*, eds., T. Glade, T., M. G. Anderson & M. J. Crozier, M.J., Wiley, Chichester, pp. 139–174, 2005.

