

BLIND-TESTING EXPERIMENTS FOR INTERPRETING SPATIAL-PREDICTION PATTERNS OF LANDSLIDE HAZARD

ANDREA G. FABBRI¹ & CHANG-JO CHUNG²

¹DISAT, University of Milano-Bicocca, Milan, Italy.

²SpatialModels Inc., Ottawa, Canada.

ABSTRACT

In this contribution, we analyse three separate databases in case study areas each suggestive of particular strategies to better portray their predictive power. A database in north-eastern Spain is used to separate sub-areas, with hopefully more compatible geomorphologic settings. Another database in central Portugal offers the opportunity of representing the uncertainty of predicted hazard class membership via iterative cross-validation with systematically partitioned landslide occurrences. A third database in central Slovenia is used to interpret the predictive qualities of two dynamic types of landslides: one that is relatively well predicting and the other one poorly predicting. The diversity of the experiments and their results, point at strategies of blind-testing, still unexploited in spatial prediction modelling, that are not necessarily limited to the landslide hazard domain.

Keywords: blind testing, cross-validation, empirical likelihood ratios, landslide hazard, prediction patterns, prediction-rate curves, target pattern.

1 INTRODUCTION

When initially proposed, cross-validation of prediction patterns of landslide hazard had as a main target the assessment of their quality as temporal predictors [1,2]. It implied generating histograms of the distribution of validating occurrences, that is, occurrences not used in prediction modelling because of their subsequent temporal appearance, throughout equal-area ranked classes of a prediction pattern. Such a pattern is obtained by the application of a mathematical model to establish spatial relationships between a set of digitized landslide occurrences of a specific dynamic type that were mapped in an earlier time interval and a number of digitized maps hopefully representing the typical settings of those occurrences.

Each map is transformed into an array of prediction scores that are then integrated into a combined score. This is subsequently transformed into equal-area ranks after ordering of all values in decreasing order. The histograms obtained can also be analysed as cumulative curves in diagrams with the cumulative proportion of study area classified as hazardous on the horizontal axis versus the corresponding cumulative proportion of validation occurrence areas on the vertical axis.

The initial applications were on databases for which the occurrences were mapped using aerial photo-mosaics corresponding to particular years therefore allowing their time partitioning. Subsequent studies in a variety of hazardous regions led to the broadening of applications ranging from modelling temporal predictions to predictions using different spatial partitioning [3, 4] for studying the variability of prediction rates.

In this contribution, we analyse three separate case study areas whose databases were provided that are suggestive of different strategies to better portray their prediction power. The next section describes those databases while the subsequent one introduces favourability modelling terminology and cross-validation strategies by blind testing. It is followed by a description of the blind-testing experiments for the three study areas located respectively in northern Spain, in Portugal and in central Slovenia. The results, in terms of prediction patterns, tables of selected predicted ratio values for different mapping units and prediction-rate curves, point at strategies for blind testing still unexplored for database interpretation. Considerations are finally made on their promising value in further research not necessarily limited to landslides nor to hazards.

2 THREE DATABASES FOR LANDSLIDE HAZARD PREDICTION

The databases analysed in this contribution originated as case studies in joint European projects and were eventually shared between collaborating researchers.

The Deba Valley study area located in Northern Spain was constructed for experimenting on landslide hazard assessment by Remondo *et al.* (2003ab) [5,6] and later extended on risk assessment (2005) [7]. The Deba Valley spatial database that is used in this contribution has also been turned into a case study for training decision makers [8].

The part of the database considered here consists of digital images with the spatial distribution of 1,123 shallow translational landslides and associated flows, **stlaf**. They were mapped and dated through photointerpretation of aerial photo-mosaics and field observations. Of those landslides 906 occurred prior to 1997 and 217 during the period 1998–2001. Their average size being about 400 m², their distribution was digitally rendered as single picture elements or pixels of 10 m resolution. The same resolution was used for the remainder of the database, consisting of digitized maps to express the typical spatial settings of those landslide occurrences. They are 26 lithological units, 7 land-use classes, as categorical data, and aspect, curvature, elevation and slope, as continuous fields. They are listed in the upper part of Table 1, short-named as **luACES**, respectively. The study area is contained in a rectangular grid or raster of 1,886 pixels by 1,555 lines. It occupies 1,393,541 pixels of 10 m resolution and is shown in Fig. 1. Obviously, 906 pixels correspond to that many **pre-97** occurrences while 217 correspond to the **post-97** ones. To classify as hazardous the 1,393,541 pixels of the study area then, we will be using the 906 **pre-97** occurrence pixels.

The Fanhões-Trancão study area in Portugal is located just north of Lisbon. It was initially used as part of a course in applied geomorphology by Zêzere (1996) [9]. Later, a database was developed to experiment in quantitative landslide hazard prediction [10–12] and risk assessment [13]. The part of the database considered here contains the distribution of 92 trigger zones of shallow translational landslides, **stl**, divided into 43 that occurred before or during 1979 and 49 that occurred during or after 1980. The trigger zones were digitized as raster polygons with a resolution of 5 m into digital image of 760 pixels by 700 lines. Their distribution is shown in Fig. 4a. The same resolution was used for digitizing 6 lithological units, 5 land use classes, 5 superficial deposit units, as categorical data, aspect, elevation and slope, as continuous fields. They were short-named as **ludAES** as listed in Table 2. The study area contains 519,344 pixels. In it 946 pixels correspond to the 43 **pre-79** occurrences and 1,225 to the 49 **post-97** occurrences. The 946 pixels of the 43 **pre-97** occurrences will be used to classify as hazardous the 519,344 pixels of the study area.

The third database analysed is located in central Slovenia and comes from a study by Komac (2004) [14] made to generate a statistical landslide prediction map to be used in risk

analysis. It was later reconsidered for favourability modelling [15]. The part of the database analysed here includes the distribution of 41 instant slides, **is**, and 47 fossil/dormant landslides, **fdl**, digitized into images of 25 m resolution, shown in Figure 6a. No separation was available of the landslides into temporal groups. Their spatial settings were represented with images of the same resolution as the digitized categorical maps of 18 units of lithology, and 14 classes of land use. Their description is in the upper part of Table 3. In addition, continuous fields maps were aspect, curvature and slope, derived from the elevation contours, and three log-transformed distance maps from lithologic boundaries, from structural elements and from surficial waters. They were short-named as **luACS_DD_gD_w**. The part of the study area used here is contained within a rectangular raster of 1,144 pixels by 695 lines to occupy 795,080 pixels of 25 m resolution. The areas corresponding to the landslide occurrences are of 294 pixels for **is** and 284 pixels for **fdl**, respectively. The 294 **is** and the 284 **fdl** pixels will be used to classify as hazardous the 795,080 pixels of the study area.

Although extensively processed, the three databases still deserve being considered to assess and test the quality and reliability of the prediction patterns obtainable by new modelling strategies.

3 SPATIAL PREDICTION WITH FAVOURABILITY MODELLING

Favourability modelling is a broad term used to generalize the generation of spatial patterns representing a classification of a study area into ranks of spatial support in favour of the presence of either desired or unwanted occurrences. Desired occurrences can be resources sought and yet undiscovered while unwanted occurrences can be hazardous events some of which are not yet experienced.

The modelling stands on a proposition, that is, a mathematical statement that has to be proven true, on the presence of the occurrences given the presence of conditioning factors, otherwise also seen as indirect support of the proposition being true. The direct support being the known presence of the occurrences.

In the case of landslide hazard and having a spatial database, the proposition could be: “ P_i : a pixel i in the study area is affected by a future landslide | given the presence of the known conditioning factors.” Obviously the characteristics of the landslides must be sharp and homogeneous to imagine a consistent setting, and their presence should be as a distribution of trigger areas to represent direct supporting patterns of the proposition, or DSP. The conditioning factors represent evidence in favour of it, however, in an indirect way: they hopefully describe acceptably well at least in part the typical or common setting of those landslide occurrences. We have termed them indirect supporting patterns, ISP.

To model the quantitative support of the proposition for every pixel in the study area, its extension must be selected by expert knowledge. Various mathematical functions can be used, such as Bayesian probability, fuzzy sets, linear or logistic regression or the empirical likelihood ratio among several other ones. We will be using the latter, ELR, in the applications to follow.

Our spatial database covering the study area will preferentially consists of sets of co-registered digital images of the same resolution. It means to have one or more DSP and several ISPs. The latter will provide support as categorical and as continuous field images. The mathematical model function is used to establish and integrate spatial relationships between a DSP and the ISPs. For instance, the ELR function provides the ratios between the normalized frequencies of the categorical ISPs, and the density functions of the continuous field ISPs, in the presence of the DSP and that in its absence within the study area. An ELR value of 2, for example, means a frequency in the presence of the DSP twice that in its absence. We will be using this value as a kind of threshold later on to interpret the applications.

The combination rules and the assumptions implicit in ELR modelling permit to integrate the resulting values for different ISPs into one value for each pixel in the study area. ELR modelling has been amply discussed elsewhere by Chung (2006) [16]. The set of all ELR pixel values is termed a prediction image that is then transformed into a prediction pattern by converting the image values into ranks. For instance, a $1,000 \times 1,000$ pixel image study area will contain 1 million ELR values that can be conveniently ranked into 200 classes representing each 0.5% of the study area. No interpretable significance is recognized to the values except for their rank.

In order to assess the quality of the prediction pattern, several strategies can be applied depending on the characteristics of the database. For instance, if the DSP can be partitioned into time intervals, all occurrences for a given earlier period can be used to generate a prediction pattern and the ones from a later period are used to validate its ranked classes. Or else, if the study area is subdivided into sub-areas of similar geomorphology and physiography, one sub-area is used to generate a prediction pattern, the statistics of its spatial relationships are extended to obtain a prediction pattern in the other sub-area that is then validated using the occurrences available in it.

The critical point in cross-validating a prediction pattern is to be able to assess the quality or “goodness” of its classes as predictors, by the distribution of the occurrences we are pretending not to know and were not used in the generation of the pattern. Should we use all available occurrences in the study area as DSP, we would obtain the most detailed and probably reliable prediction pattern, whose quality as predictor, however, is still unknown. This is so unless we use some form of partitioning of the database to assess the prediction quality.

Some convenient strategy of cross-validation are iterative processes in which a specific number of occurrences are sequentially excluded, a prediction pattern is obtained with the remainder that is then cross-validated by the excluded occurrences. The process is repeated until all occurrences have been used to predict and to validate. Similarly, iterations could be obtained by selecting a given number of occurrences to predict and using the remainder to validate. Another useful strategy, particularly when the number of occurrences is large, is to select randomly a given number of occurrences to predict and use the remainder to validate. This process is then iterated a given number of times. All the iterative processes lead to a number of prediction patterns that are integrated statistically into what we have termed a target pattern. It combines by pixel value averaging all the prediction patterns obtained. The statistics of a target pattern can be analysed for measures of class membership, such as variance per pixel, obtaining as a representation of its uncertainty, the uncertainty pattern. We have used Spatial Target Mapping (STM) for spatial modelling with iterative cross-validation. The software is ancillary to a geographical information system [8, 17].

Only expert knowledge of the database and of the physical processes it portrays, in addition to favourability modelling, can complement the analyst’s imagination in constructing or following a useful strategy of cross-validation to assess the “goodness” of the database for prediction modelling. So far, in most applications, it was found that changing the mathematical model is mildly influential on the prediction pattern. More critical seems to be the database quality and contents [18].

The applications that follow in the different European study areas are to open a window on some potentially useful strategies for cross-validation that aim at extracting particular information from the data while attempting to answer specific questions that the database might or might not be fit to provide.

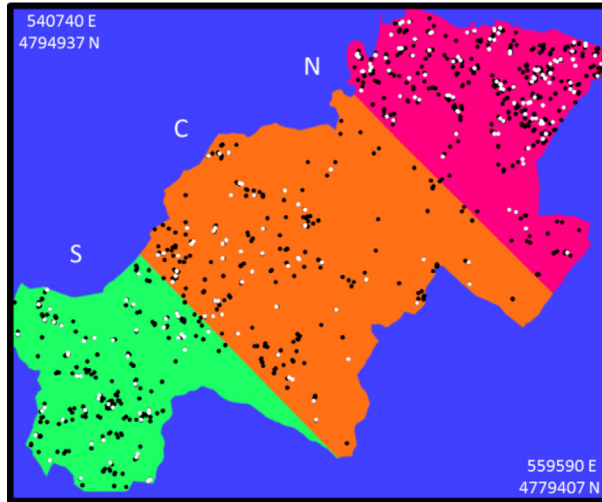


Figure 1: The Deba Valley study area in northern Spain, containing 906 **pre-97** (black dots) and 217 **post-97** (white dots) one-pixel landslide occurrences (dimensions exaggerated for visualization). It covers an area of 1,393,668 pixels. The three sub-areas, indicated by different colours cover 370,576 (North, N, red, with 460 **pre-97** and 112 **post-97**), 678,148 (Center, C, orange, with 225 and 54) and 342,278 (South, S, green, with 225 and 51).

4 EXPERIMENTS WITH BLIND-TESTING STRATEGIES

4.1 Partitioning of the study area

The entire Deba Valley database that covers the study area shown in Fig. 1 was used to generate the prediction pattern of Fig. 2a. The legend uses 13 groups of equal-area ranks of the predicted hazard classes of the 200 obtained as default. It is a convenient way to visualize the hazard values after having sequenced them in decreasing order. The groups of ranks are fixed and are assigned to a pseudo-colour look-up table with the hotter colours representing the higher ranks. The identical legend and colours will be used for all the patterns in the other applications. The prediction patterns used the 906 **pre-97** occurrences as DSP and **luACES** as ISPs. In order to assess the quality of the pattern as predictor, the distribution across the pattern of the 217 **post-97** occurrences was used to obtain the black cumulative prediction-rate curve in Fig. 3. It shows that the top 10% of the study area classified as hazardous contains 42% of the 217 **post-97** occurrences, and the top 20% contains 63% of them.

If we suspect that the prediction-rate curve reflects too great a variety of geomorphologic settings, we can subdivide the study area into more homogeneous sub-areas, as done in Fig. 1, to verify whether a local prediction provides better results. This is done with the three sub-areas shown in Fig. 1. Three prediction patterns for sub-areas N, C and S were obtained and mosaicked into one in Fig. 2b. In it, we can observe more local rankings that can be interpreted by their respective prediction-rate curves shown in Fig. 3. Sub-areas C and S provide initially steeper curves within the top 10% of their respective hazardous sub-areas.

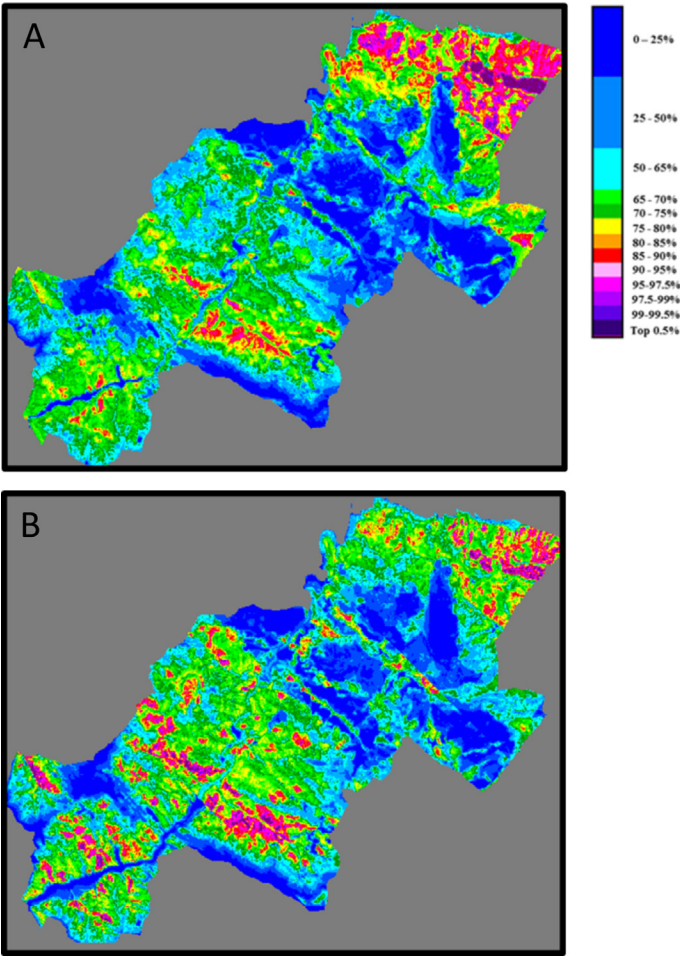


Figure 2: Prediction pattern obtained for the Deba Valley study area representing equal area ranks grouped as in the legend. In (A) the pattern obtained using the 906 **pre-97 stlaf** occurrences as DSP and **luACES** as ISPs. In (B) the mosaic of three prediction patterns for sub-areas N, C and S using the respective **pre-97** occurrences and ISPs.

Beyond that 10%, however, the curve from the entire study area prediction pattern remains steeper.

In terms of cost-benefit, within the top 10%, these considerations could be of guidance should hazard mitigation measures or avoidance be planned whose costs would increase exponentially with the proportion of study area considered so hazardous as to need such measures.

To interpret the prediction patterns, it becomes critical to consider how the different categorical units, classes and continuous field values contribute to supporting the proposition. Table 1 shows the ELR values for the different predictions in the entire study area and in the three sub-areas. ELR values range from 0 to infinity, and as a rule of thumb, we can use a threshold of 2.00 for simplifying the separation of those more supportive from the less or non-supportive ones. The table identifies the different characteristics of the prediction pat-

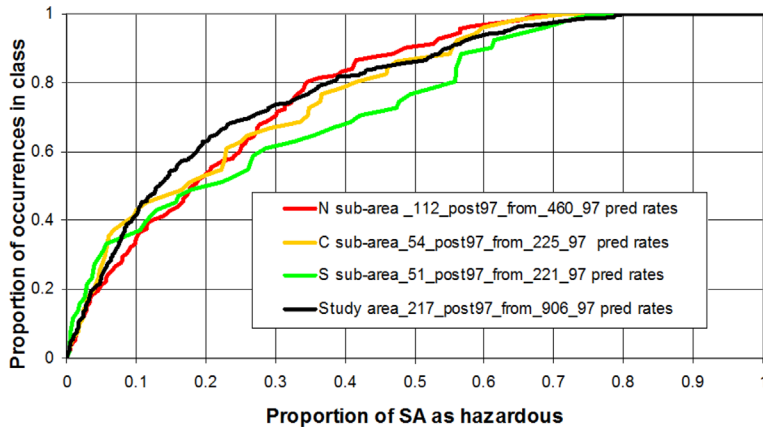


Figure 3: Cumulative prediction-rate curves for the 217 **post-97 stlaf** occurrences for the entire Deba Valley study area and for the three sub-areas N, C and S, black, red, yellow and green, respectively.

terns. For instance, land use class u_6 and lithology unit l_2 , l_4 and l_{12} , together with a curvature interval represent most of the support for the prediction pattern in the entire study area. As to the separate sub-areas, we can observe how the respective ratio values vary, including the absence in sub-area C, of lithology unit l_4 , l_{10} , l_{12} and l_{18} , so is for l_2 , l_4 and l_{12} in sub-area S. Also, unit l_{22} is of relatively greater support in sub-areas C and S than in sub-area N or in the entire study area.

4.2 Uncertainty of class membership

For the database of the Fanhões-Trancão study area, an ELR prediction was made by the 43 **pre-79** shallow translational landslide occurrences, **stl**, as DSP and **ludAES** as ISPs. The prediction pattern was to be validated with the distribution of the 49 **post-79** occurrences. Here, however, measures of uncertainty of the class membership of the occurrences were sought. For this three strategies were selected: (i) sequential exclusion of 3 occurrences from the 43 **stl pre-79** iterated 14 times ($43m3 \times 14$); (ii) random selection of 30 of the 43 iterated 14 times ($43r30 \times 14$); and (iii) the same random selection as in (ii) iterated 28 times ($43r30 \times 28$). Here, m stays for minus and r for random while x is the multiplier. Figure 4a shows the distribution of the two temporal groups of occurrences. Figure 4b shows the prediction pattern from the 43 **pre-79** DSP and **ludAES** as ISPs. In Fig. 4c we have the target pattern obtained by computing the median of the 14 ranks for each pixel of the corresponding 14 prediction patterns from the 14 iterations of prediction/cross-validation. Obviously, the two patterns are rather similar. However, the target pattern also provides a statistics about the range of the ranks about the median.

The rank of the range of ranks is shown in Fig. 4d. It represents a measure of uncertainty of class membership (of the 200 classes used as default). The wider is the range of ranks, the higher is the uncertainty rank in the pattern of Fig. 4d. To exemplify the possible use of such

Table 1: Categorical and continuous field ISPs and Empirical Likelihood Ratio values for predictions using the entire Deba Valley study area, SA, and the three sub-areas, N, C and S. Abbreviations are as follows: u_{1-7} the land use classes, l_{1-24} the lithology units, A, C, E and S, are aspect, curvature, elevation and slope, respectively. Abbreviations are bold if ELR > 2.00. In Italics is the corresponding range of classes with maximum ratio in brackets. Crossed units indicate their absence in the sub-area.

Land use classes, u	1 , Very dense broad-leaved forest; 2 , Dense broad-leaved forest; 3 , Semi-open broad-leaved forest; 4 , Very dense coniferous forest; 5 , Brush-land; 6 , Grassland, pastures y cultivated areas; 7 , Zones without vegetation.
Lithology classes, l	1 , Marly limestone; 2 , Muddy Flysh; 3 , Stratified limestone; 4 , Marly limestone; 5 , Marl; 6 , Sandstone and conglomerate; 7 , Sandy Flysh; 8 , Massive limestone; 9 , Calcareous lutite and sandy marl; 10 , Silicoclastic-calcareous Flysh; 11 , Calcareous marl and calcareous breccia; 12 , Calcareous Flysh; 13 , Pyroclastics; 14 , Lavas; 15 , Polygenic breccias; 16 , Siliceous breccias; 17 , Well graded gravel; 18 , Poorly graded gravel; 19 , Clayey gravel; 20 , Well graded sand; 21 , Poorly graded sand; 22 , Silty sand; 23 , Clayey sand; 24 , Silt and fine sand; 25 , Residual clay; 26 , Anthropogenic deposits.
SA stlaf 906 pre97	$u_6 = \mathbf{2.49}$; $l_2 = \mathbf{2.38}$, $l_4 = \mathbf{2.30}$, $l_5 = 1.38$, $l_{10} = 1.80$, $l_{12} = \mathbf{4.48}$, $l_{18} = 1.74$, $l_{22} = 1.72$; A < 1.25; C > 2.00 296.40–317.60 (<i>max 2.17</i>); E < 1.78; S < 1.78.
sub-a N stlaf 460 pre97	$u_6 = 1.83$; $l_2 = 1.46$, $l_4 = 1.20$, $l_7 = 1.23$, $l_{10} = 1.99$, $l_{12} = \mathbf{2.34}$, $l_{18} = 1.36$; $l_{22} = 0.65$; A < 1.52; C > 2.00 277.56–331.61 (<i>max 4.99</i>); E < 1.61; S < 1.59.
sub-a C stlaf 225 pre97	$u_6 = \mathbf{2.66}$; $l_2 = 0.49$, $l_3 = \mathbf{2.75}$, $l_4, l_5 = 1.60$; $l_7 = 1.34$, $l_{10}, l_{12}, l_{18}, l_{22} = \mathbf{2.53}$; $l_{24} = 1.16$; A < 1.50, C < 1.65; E < 1.57; S < 1.25.
sub-a S stlaf 221 pre97	$u_5 = 1.14$, $u_6 = \mathbf{2.29}$; $l_2, l_4, l_5 = 1.41$, $l_{10} = 1.14$, $l_{12}, l_{18} = \mathbf{2.07}$, $l_{22} = \mathbf{2.16}$; A < 1.50, C > 2.00 169.60–189.20 (<i>max 2.37</i>); E < 1.78; S < 1.42.

uncertainty, Fig. 4e shows the part of the target pattern in Fig. 4c, that corresponds to the lower 50% of the uncertainty pattern of Fig. 4d.

The target patterns were obtained also by the three strategies, 43m3 × 14, 43r30 × 14 and 43r30 × 28, and then cross-validated using not only the 43 **pre-79** occurrences but also the 49 **post-79** occurrences, as shown in Fig. 5a. There the prediction-rate curves are all very similar and the top 10% of the study area classified as hazardous contains 40% of the **post-79** occurrences while the top 20% contains the 55% and the top 30% the 75%. We can also verify together with the prediction pattern ranks of the occurrences the corresponding uncertainty pattern ranks.

This is done in the diagram in Fig. 5b for each of the 49 **post-79** occurrences in units of 1,000. In the diagram, the prediction rates for target pattern 43m3 × 14 are compared with the corresponding uncertainties and with those for target pattern 43r30 × 28. An uncertainty rank of 500 corresponds to the 50% uncertainty/target combination in Fig. 4e.

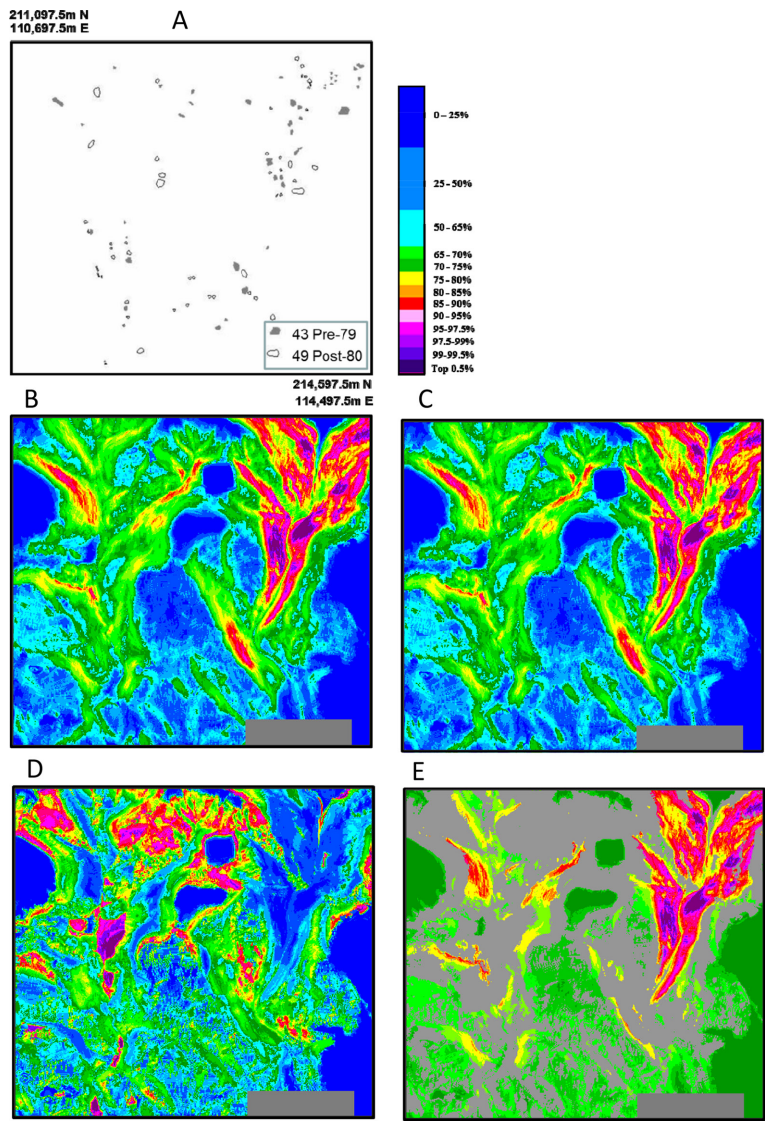


Figure 4: Distribution of shallow translational landslides in the Fanhões-Trancão study area in Portugal in (A). Prediction, target, uncertainty and 50% uncertainty/target combination patterns, respectively in (B) to (E).

Considerations on an acceptable uncertainty level could help in assessing the predicted occurrences not only on their prediction score but also on the uncertainty score of class membership.

Other considerations can be made on the predictability of the **post-79** occurrences by the **pre-79**, observing the prediction rates for the units, classes and values for the ISPs of the two groups, shown in Table 2. A prediction pattern using the **pre-79** occurrences shows higher ELR values for l_3 , u_2 and d_2 while the one obtained using the **post-79** (not shown here but source of the ratios) shows higher values for l_4 .

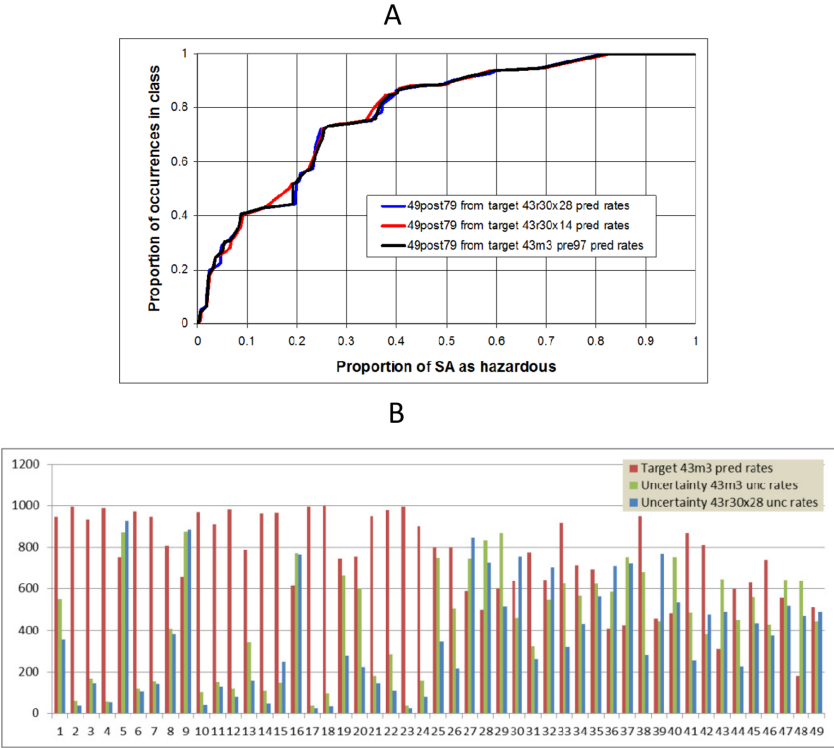


Figure 5: Target prediction rates in (A) for 49 **post-79 stl** occurrences for the three analyses using the 43 **pre-79** occurrences in the Fanhões-Trancão study area: 43r30 × 28, 43r30 × 14 and 43m3 × 14. In (B) the histogram of the 43m3 × 14 target rates for each of the 49 **post-79** occurrences, in red, and the corresponding uncertainty rates for analyses 43m3 × 14 in green and 43r30 × 28 in blue.

Table 2: Categorical and continuous field ISP and Empirical Likelihood Ratio values for predictions in the Fanhões-Trancão study area. Abbreviations are as follows: l_{1-6} the lithology units, u_{1-5} the land use classes, d_{1-3} the surficial deposit classes, A, E and S, are aspect, elevation and slope, respectively. Abbreviations are bold if ELR > 2.00. In *Italics* is the corresponding range of classes with maximum ratio in brackets.

Lithology classes, l	1 , Volcanics; 2 , Sandstones; 3 , Marls and marly limestones; 4 , Limestones; 5 , Lacustrine limestones; 6 , Conglomerates and sandstones.
Land use classes, u	1 , Cultivated areas; 2 , Dense shrubs; 3 , Herbaceous vegetation; 4 , Trees and tree cover; 5 , Urban and human intervention areas.
Surficial deposits, d	1 , slope deposits thinner than 0.5 m; 2 , slope deposits thicker than 0.5 m; 3 , fluvial deposits; 4 , Quaternary terrace; 5 , Cliff of cuestas.
stl 43 pre-79	$l_2 = \mathbf{2.69}$, $l_3 = \mathbf{2.50}$, $l_4 = 1.75$; $u_1 = 1.02$, $u_2 = \mathbf{2.24}$; $d_2 = \mathbf{2.21}$; A >2 237.90–102.52, 285.19–338.56 (<i>max 2.85</i>); E < 1.94; S >2 9.74–28.36 (<i>max 4.46</i>).
stl 49 post-79	$l_2 = \mathbf{2.40}$, $l_3 = 0.92$, $l_4 = \mathbf{2.14}$; $u_1 = 1.18$, $u_2 = 1.45$; $d_2 = 1.81$; A >2 244.75–307.57 (<i>max 3.09</i>); E < 1.89; S >2 13.14–29.00 (<i>max 4.28</i>).

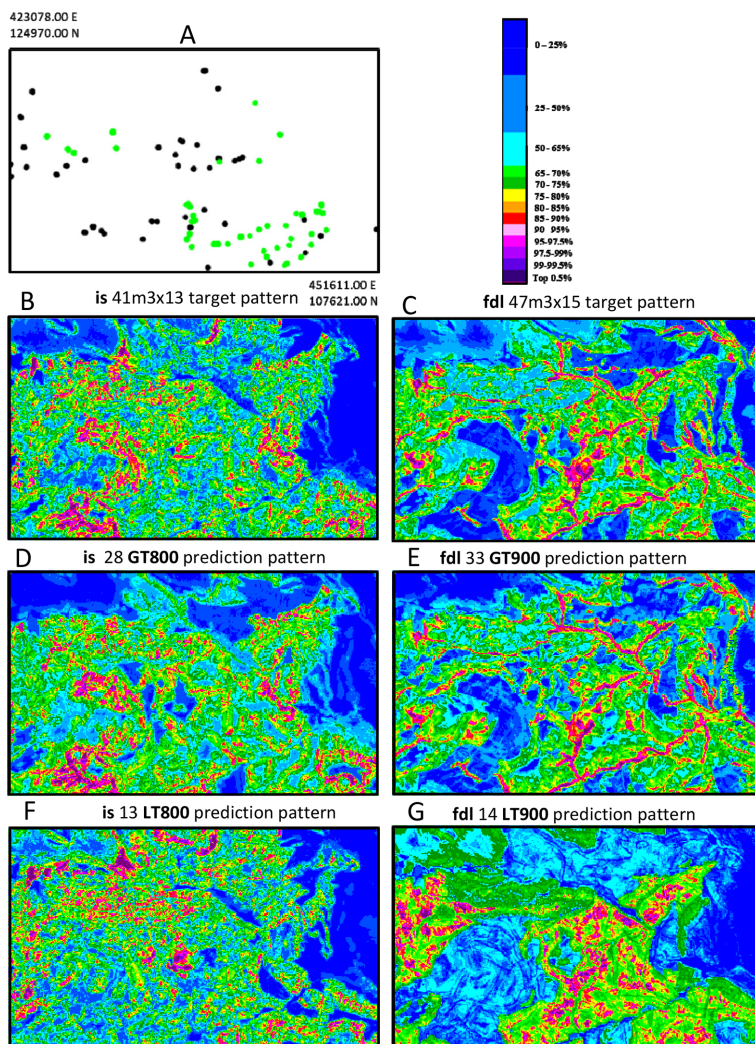


Figure 6: Distribution of 41 **is** occurrences in black, and 47 **fdl** in green, for the central Slovenia study area, in (A). Occurrence dimension exaggerated for visibility. Target patterns from 41m3 × 13 **is** in (B) and from 47m3 × 15 **fdl** in (C). In (D) and (E) the prediction patterns obtained for 28 **is** **GT800** and 33 **fdl** **GT900**. In (F) and (G) the prediction patterns for 13 **is** **LT800** and 14 **fdl** **LT900**. Explanations are in the text.

4.3 Separation of well predicting and well predicted occurrences

Strategies of iterative cross-validation can be used for other purposes such as for separating the occurrences of a DSP into groups of well predicting and well-predicted occurrences from poorly predicting and predicted ones.

This was done for the database of central Slovenia for two dynamic types of landslide occurrences. Figure 6a shows the distribution of 41 instant slides, **is**, in black, and of 47 fossil/dormant landslides, **fdl**, in green. In this case, no subdivision into time of occurrence

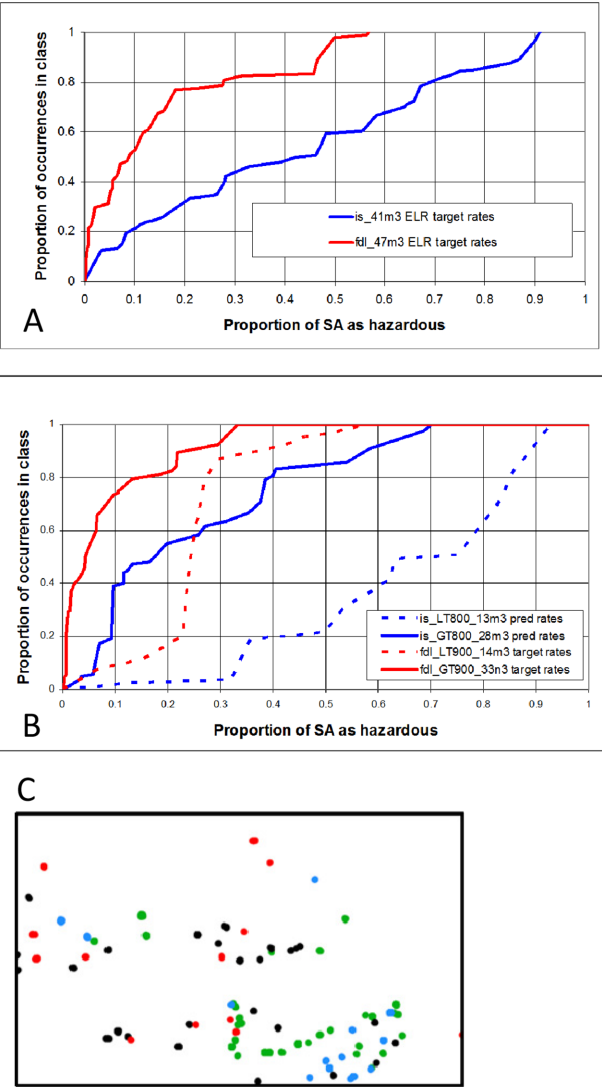


Figure 7: Prediction-rate curves of target patterns for **is** and **fdl** occurrences in the central Slovenia study area. In (A) for the 41 **is**, and the 47 **fdl**, and in (B) for the subgroups of 28 and 13 **is** and the 33 and 14 **fdl**. In (C) the distribution of occurrences of the 4 sub-groups: 28 and 13 **is** in black and red, 33 and 14 **fdl** in green and blue, respectively.

was available, so that the cross-validation strategies were used of $41m3 \times 14$ for **is** and $47m3 \times 15$ for **fdl**. Using $\mathbf{clACSD_gD_sD_w}$ as ISPs, they produced the respective prediction-rate curves shown in Fig. 7a: a rather good curve for **fdl** and a poor one for **is**. The corresponding target patterns are shown in Figs 6b and c.

Another strategy was then used to explore the good and bad target patterns. One occurrence was sequentially selected to predict all the remaining ones. This led to 41 iterations for

Table 3: Categorical and continuous field ISP and Empirical Likelihood Ratio values for predictions in the central Slovenia study area. Abbreviations are as follows: c_{1-14} the CORINE land use classes, l_{1-14} the lithology units, A, C and S, are aspect, curvature and slope, and D_g , D_s and D_w are the logarithmic distances from geological boundaries, structuring elements and surficial waters, respectively. Abbreviations are bold if $ELR > 2.00$. In Italics is the corresponding range of classes with maximum ratio in brackets.

CORINE land use classes, c	1 , Broad-leaved forest; 2 , Complex cultivated patterns; 3 , Mixed forest; 4 , Non-irrigated arable land; 5 , Coniferous forest; 6 , Industrial, commercial and service units; 7 , Transitional woodland scrub; 8 , Water courses; 9 , Mineral extracts; 10 , Natural grassland; 11 , Discontinuous urban fabric; 12 , Irrigated arable land; 13 , Pastures; 14 , Land principally occupied by agriculture, with significant areas of natural vegetation.
Lithology units, l	1 , Alluvium; 2 , Talus; 3 , Sandy and marly clay, clay, sandstone; 4 , Lake and marsh sediments; 5 , Morene, breccia; 6 , Conglomerate; 7 , Flish; 8 , Limestone, marly limestone, marl; 9 , Limestone; 10 , Shaly claystone and other rocks; 11 , Limestone and dolomite; 12 , Dolomite; 13 , Sandstone, shaly claystone, tuff; 14 , Sandstone, argillite, tuff; 15 , Pyroclastites; 16 , Pyroclastites, limestone; 17 , Sandstone, shaly claystone, alveolite, conglomerate, marl; 18 , Marly limestone, dolomite, sandy shale.
is 41	$c_2 = 2.22$; $l_3 = 2.43$, $l_{17} = 2.92$; A > 2.00 65.52–93.96 (<i>max</i> 2.47), C < 1.34 , S > 2.00 32.85–40.48 (<i>max</i> 3.31); $D_g < 1.65$, $D_s < 1.39$, $D_w > 2.00$ 1.562–3.701 (<i>max</i> 2.79).
is 28 GT800	$c_2 = 2.77$; $l_3 = 3.94$, $l_{17} = 0.00$; A > 2.00 62.28–84.60 (<i>max</i> 2.26); C < 1.54 ; S > 2.00 32.69–37.20 (<i>max</i> 2.70); $D_g < 1.83$, $D_s < 1.68$, $D_w > 2.00$ 0.000–3.764 (<i>max</i> 3.61).
is 13 LT800	$c_2 = 1.36$; $l_3 = 0.13$, $l_7 = 2.39$; $l_{17} = 7.38$; A > 2.00 67.32–105.84 (<i>max</i> 3.00); C < 1.07 ; S > 2.00 33.27–42.12 (<i>max</i> 5.83); $D_g < 1.58$; $D_s > 2.00$ 6.734–7.000 (<i>max</i> 2.24); $D_w > 2.00$ 6.574–7.064 (<i>max</i> 2.59).
fdl 47	$c_{13} = 4.80$, $c_{14} = 1.92$; $l_4 = 2.77$, $l_5 = 2.84$, $l_{15} = 1.43$; $l_{18} = 1.98$; A > 2.00 157.68–177.48 (<i>max</i> 2.09); C < 1.48 ; S < 1.09 ; $D_g > 2.00$ 1.606–4.158 (<i>max</i> 3.02), $D_s < 1.00$, $D_w > 2.00$ 3.615–4.570 (<i>max</i> 2.49).
fdl 33 GT900	$c_{13} = 5.99$, $c_{14} = 1.50$; $l_4 = 3.95$, $l_5 = 4.06$, $l_{15} = 0.97$, $l_{18} = 1.54$; A > 2.00 151.20–182.52 (<i>max</i> 2.05); C < 1.71 ; S < 1.09 ; $D_g > 2.00$ 1.523–4.365, (<i>max</i> 4.28); $D_s < 1.61$; $D_w > 2.00$ 3.449–4.570 (<i>max</i> 2.98).
fdl 14 LT900	$c_{13} = 2.00$, $c_{14} = 2.92$; $l_4 = 0.00$, $l_5 = 0.00$, $l_{15} = 2.51$, $l_{18} = 3.01$; A > 2.00 19.08–50.04, 159.48–174.96 (<i>max</i> 2.96); C < 1.19 ; S > 2.00 28.18–32.79 (<i>max</i> 2.49); $D_g < 1.75$, $D_s < 1.99$, $D_w < 1.99$.

is and 47 for **fdl**. The corresponding 41×41 and 47×47 arrays of prediction rates were used to separate the respective occurrences into a well predicted/predicting group and a group of poorly predicted/predicting ones.

A threshold of 800, that is, the top 20% of predicted values was arbitrarily selected for **is** occurrences, and one of 900, that is, the top 10% was selected for **fdl** occurrences. One group

contained all occurrences predicting more than 10 ones with value equal or above the threshold or predicted by at least 10 other ones with a value equal or above that threshold. Thus 28 **is** and 33 **fdl** occurrences were classified as well predicted/predicting and the remaining 13 and 14 were classified as poorly predicted/predicting **is** and **fdl**, respectively.

After the separation of the two groups for each type, new iterative cross-validations were obtained as $28m3 \times 9$, $33m3 \times 11$, $13m3 \times 4$ and $18m3 \times 6$. The respective prediction rates are shown in Fig. 7b, while the separate prediction patterns are in Fig. 6d–g. Note the similarity of the target patterns in Fig. 6b with the prediction pattern in Fig. 6d for **is** occurrences, and of those of Fig. 6c with that of Fig. 6e for **fdl** occurrences.

Table 3 shows the ELR values for the ISPs supporting the two sets of DSPs, **is** and **fdl** and for the respective sub-sets. We can observe how the subsets concentrate the support of the units and values of ISP ratios. In **is GT800**, we have ratios similar to those in **is 41** except for l_{17} with 0 ratio. In **is LT800**, the ratios are quite different with higher values for l_7 , l_{17} and S . As to **fdl GT900**, the ratios are close to those of **fdl 47** but with higher values for c_{13} , l_4 and l_5 . For **fdl LT900**, we have lower ratios for c_{13} and higher for c_{14} , ratios of 0 for l_4 and l_5 but higher for l_{15} and l_{18} . These differences are the reasons for the different rankings of the patterns in Fig. 6d and 6f as well for the ones in Fig. 6e and 6g.

5 QUESTIONS AND CONCLUDING REMARKS

Many aspects of spatial prediction modelling provide challenges that demand experimentations on representative databases. The three applications of blind-testing described offer the opportunity of questions for future research. How to identify optimal study area extensions to obtain acceptable prediction patterns and prediction-rate curves? The Deba Valley database provides an example of how to proceed. How to explore the uncertainty of class membership to interpret the target patterns? The Fanhões-Trancão database shows some strategies for that. How to identify the database signatures supporting a prediction pattern for a specific type of occurrence? What to do with poorly predicted/predicting occurrences? Should we ignore them as outliers? Or should they be part of the database signature with poorer prediction rates? Should we respect the signatures identified? Part of the central Slovenia database is used here to stimulate such questions. Furthermore, strategies can consider other types of spatial modelling for applications other than hazard prediction. Whenever a database targets relationships that are integral part of spatial processes we must be able to develop appropriate strategies of blind testing to assess the quality of the prediction patterns. The cross-validation strategies described here appear uncommon or unexploited to date.

ACKNOWLEDGMENTS

The Deba Valley database was provided by Dr. Juan Remondo Tejerina of the University of Cantabria, Santander, Spain. The Fanhões-Trancão database was provided by Dr. José Luís Zêzere of the University of Lisbon, Portugal. The central Slovenia database was provided by Dr. Marko Komac from the Geological Survey of Slovenia in Ljubljana.

REFERENCES

- [1] Chung, C.F., Fabbri, A.G. & Van Westen, C.J., Multivariate regression analysis for landslide hazard zonation. *Geographical Information Systems in Assessing Natural Hazards*, eds. A. Carrara & F. Guzzetti, Kluwer Academic Publishers: Dordrecht, pp. 107–133, 1995.

- http://dx.doi.org/10.1007/978-94-015-8404-3_7
- [2] Chung, C.F. & Fabbri, A.G., Probabilistic prediction models for landslide hazard mapping. *Photogrammetric Engineering & Remote Sensing PE&RS*, **65**(12), pp. 1389–1399, 1999.
- [3] Chung, C.F. & Fabbri, A.G., Validation of spatial prediction models for landslide hazard mapping. *Natural Hazards*, **30**(3), pp. 451–472, 2003.
<http://dx.doi.org/10.1023/B:NHAZ.00000007172.62651.2b>
- [4] Chung, C.F. & Fabbri, A.G., Predicting future landslides for risk analysis-spatial models and cross-validation of their results. *Geomorphology*, **94**(3–4), pp. 438–452, 2008.
<http://dx.doi.org/10.1016/j.geomorph.2006.12.036>
- [5] Remondo, J., González-Díez, A., Díaz de Terán, J.R. & Cendrero, A., Landslide susceptibility models utilising spatial data analysis techniques. A case study from the lower Deba valley, Guipúzcoa (Spain). *Natural Hazards*, **30**(3), pp. 267–279, 2003a.
<http://dx.doi.org/10.1023/B:NHAZ.00000007202.12543.3a>
- [6] Remondo, J., González-Díez, A., Díaz de Terán, J.R., Cendrero, A., Fabbri, A. & Chung, C.F., Validation of landslide susceptibility maps; examples and applications from a case study in northern Spain. *Natural Hazards*, **30**(3), pp. 437–449, 2003b.
<http://dx.doi.org/10.1023/B:NHAZ.00000007201.80743.fc>
- [7] Remondo, J., Bonachea, J. & 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.
<http://dx.doi.org/10.1007/s10346-005-0016-x>
- [8] Fabbri, A.G. & Chung, C.-J., Training decision-makers in hazard spatial prediction and risk assessment: ideas, tools, strategies and challenges. *Disaster Management and Human Health Risk*, eds. K. Duncan & C.A. Brebbia, WIT Press: Southampton, UK, pp. 285–296, 2009.
- [9] Zêzere, J.L., Landslides in the north of Lisbon region. *Fifth European Intensive Course on Applied Geomorphology – Mediterranean and Urban Areas*, eds. A.B. Ferreira & G.T. Vieira, Departamento de Geografia, Universidade de Lisboa, pp. 79–89, 1996.
- [10] Fabbri, A., Chung, C.F., Napolitano, P., Remondo, J. & Zêzere, J.L., Prediction rate functions of landslide susceptibility applied in the Iberian Peninsula. *Risk Analysis III. Series: Management Information Systems*, ed. C.A. Brebbia, WIT Press: Southampton, UK, 5, pp. 703–718, 2002.
- [11] Zêzere, J.L., Rodrigues, M.L., Reis, E., Garcia, R., Oliveira, R., Oliveira, S., Vieira, G. & Ferreira, A.B., Spatial and temporal data management for the probabilistic landslide hazard assessment considering landslide typology. *Landslides: Evaluation and Stabilization I*, eds. W.A. Lacerda, M. Ehrlich, S.A.B. Fontoura & A.S.F. Sayão, Taylor and Francis: London, pp. 117–123, 2004a.
<http://dx.doi.org/10.1201/b16816-15>
- [12] Zêzere, J.L., Reis, E., Garcia, R., Oliveira, S., Rodrigues, M.L., Vieira, G. & Ferreira, A.B., 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**(1), pp. 133–146, 2004b.
<http://dx.doi.org/10.5194/nhess-4-133-2004>
- [13] Zêzere, J.L., Garcia, R.A.C., Oliveira, S.C. & Reis E., Probabilistic landslide risk analysis considering direct costs in the area north of Lisbon (Portugal). *Geomorphology*, **94**(3–4), pp. 467–495, 2008.
<http://dx.doi.org/10.1016/j.geomorph.2006.10.040>

- [14] Komac, M., Statistical landslide prediction map as a basis for a risk map. *Risk Analysis IV*, ed. C.A. Brebbia, WIT Press: Southampton, UK, pp. 319–330, 2004.
- [15] Fabbri, A.G., Komac, M., Patera, A. & Chung, C.-J., Favourability modelling of landslide hazard with spatial uncertainty of class membership: a reapplication in central Slovenia. *Proceedings of IAMG 2015*, eds. H. Schaeben, R. Tolosana Delgado, K.G. van den Boogaart & R. van den Boogaart, Freiberg, Germany, pp. 848–856, 2015, ISBN 978-3-00-050337-5 (DVD).
- [16] Chung, C.F., Using likelihood ratio functions for modelling the conditional probability of occurrence of future landslides for risk assessment. *Computers & Geosciences*, **32**(8), pp. 1052–1065, 2006.
<http://dx.doi.org/10.1016/j.cageo.2006.02.003>
- [17] STM software, available at <http://www.spatialmodels.com>
- [18] Fabbri, A.G. & Chung, C.-J., On blind tests and spatial prediction models. *Natural Resources Research*, **17**(2), pp. 107–118, 2008.
<http://dx.doi.org/10.1007/s11053-008-9072-y>