Prediction rate functions of landslide susceptibility applied in the Iberian Peninsula

Fabbri A.G.¹, Chung C.F.², Napolitano P.³, Remondo J.⁴, and Zêzere J.L.⁵

¹ITC, 7500 AA Enschede, & SPINlab, Vrije Universiteit, 1081 HV Amsterdam, The Netherlands
²Geological Survey of Canada, Ottawa K1A 0E8, Canada
³ACTA, Via D. Fontana 40, 80128 Napoli, Italy
⁴DICITIMAC, Universidad de Cantabria, 39005 Santander, Spain
⁵Centro de Estudos Geográficos, Universidade de Lisboa, 1699 Lisboa, Portugal

Abstract

The prediction models of landslide susceptibility that we have developed, generate not only predicted hazard maps but also prediction-rate curves, which allow us to estimate the probabilities of the occurrences of future landslides from the hazard maps. For a risk-analysis estimating the “economic” values of the population, properties, economic activities, etc., the estimates of the probabilities are absolutely critical statistics. To use the hazard maps for risk analysis, we must be able to estimate the probability of occurrence of a future landslide at each hazard level in the maps. Without the estimates of such probabilities, the hazard maps can provide only indicators of landslide hazard, but they cannot be directly useful for a decision process. With those probabilities, however, decision makers can quantitatively assess the economic sterilization due to the possible damage under the assumptions of appropriate scenarios. Hence, they can take a learned and informed decision.

Predictions are based on “Favourability Functions” that integrate the spatial relationships between the distribution of trigger zones of specific dynamic types of landslides and the surrounding mapping units and contour intervals. The latter two represent the spatial support to estimate the likelihood of further failures.

The prediction-rate curves are obtained to interpret the predictions in spatial-temporal terms by partitioning the database of trigger-zone distribution and spatial-support layers (digital maps) in periods of activity, or in non-overlapping spatial subsets, needed for prediction and validation, respectively. The shapes of the prediction-rate curves empirically measure the quality of the predictions, their robustness, and provide the arguments for the cost-benefit analysis of the relationships between the spatial prediction patterns and the prediction-rate curves.

Two applications of Favourability Function models in the Iberian Peninsula are used to demonstrate the operational feasibility of generating the prediction-rate curves for the spatial-temporal validation needed for risk assessment.
1. Introduction

This contribution stems from the research activity of a project on “New Technologies for Landslide Hazard Assessment and Management in Europe” of the European Commission’s Environment Programme [1]. A major target of that project was the use of geographic information systems (GIS) to assess landslide susceptibility by generating hazard maps from systematically constructed spatial databases. For that prediction models were used based on an approach termed “Favourability Function” modeling, FF [2]. When we generate a hazard map for future landslides based on FF modeling, we also produce the corresponding prediction-rate curve, which allows us to estimate the probability of the occurrence of a future landslide in each hazard class in the map.

Traditionally, geomorphologists have constructed landslide hazard maps identifying areas likely to be affected by landslides. It has been achieved by geomorphological understanding of the area through aerial photographs and fieldwork [3]. Similarly, civil engineers have constructed slope stability maps based on deterministic models by studying and interpreting the physical processes of landslides using slope angles, soil cohesion, water saturation capacities, shearing resistance etc. Each point in the stability maps shows a level of “safety factor of slope failure” of the unit area surrounding the point [4]. While the hazard maps obtained from geomorphologic maps usually show three to five levels of hazard, the slope stability maps represent the levels of the safety factor in a continuous scale.

Landslide risk analysis estimates the “economic” values of each element vulnerable to landslide hazard. Vulnerable elements are the population, properties, economic activities, including public services, etc., at risk in a given study area. The specific risk is the expected degree of loss due to a specific damaging phenomenon, and it is computed for each vulnerable element. The total risk is the expected aggregated damage to human and economic activities due to landslides (e.g., number of persons injured, damage to properties or disruption of economic activities). Natural hazard susceptibility is the probability of occurrence of a potentially damaging phenomenon of given intensity during a given period. To use the hazard maps for risk analysis, we must be able to estimate the probability of occurrence of a future landslide at each hazard level in the maps.

Without the estimates of such probabilities, the hazard maps constructed by geomorphologists or civil engineers can provide only indicators of landslide hazard, but they cannot be directly useful for a decision process. However, if we have the estimates of those probabilities, then, by cost-benefit analyses, decision makers can quantitatively assess the economic sterilization due to the possible damage under the assumptions of appropriate scenarios. Hence, they can take a learned and informed decision, rather than opt for an emotional or a “gut-feeling” decision.

A brief background to predictive models based on FF modeling used is introduced here and the results of applications in two study areas in Portugal and Spain, respectively, are analyzed to further explore the use of prediction-rate curves to
derive probabilities assigned to hazardous area units that can be used for the representation of risk.

2. **FF modeling and prediction rate curve**

Spatial prediction models have been developed for landslide hazard mapping as a follow-up of research mainly in the area of mineral exploration [2,5]. To generate predictions, it is assumed that: (1) spatial databases are available or can be constructed that provide sufficient information to characterize the typical conditions in which individual types of landslides occur in a study area; and (2) that the conditions for the occurrence of the past landslides are the same or are very similar to the ones in which the future landslides will occur.

The term Favourability Function modeling is a framework that covers and connects a variety of approaches to spatial data integration. A procedural sequence of steps is provided in Table 1, where it is applied to decision problems that require the introduction of spatial components to predict the likely distribution of future events such as landslides. Various quantitative integration models have been discussed based on different interpretation of favourability as conditional probability, certainty, likelihood, belief, or fuzzy set membership function [2,6,7,8]. Applications of FF models have been made to many landslide-prone study areas, they have generated predicted hazard maps. The hazard maps show a relative hazard level for future landslides in a continuous scale at every point in the map. Often the hazard levels are divided into a number of hazard classes for visualization. These procedures are illustrated in Steps 1.1 and 1.2 in Table 1.

A critical issue in predictive modeling is the interpretation of the hazard levels in predicted maps. For this, prediction-rate curves have been generated. They require partitioning the distribution of the past landslides. One way in which the curves are obtained is the following. Using the FF modeling, we generate a predicted hazard map showing a relative hazard level in a continuous scale at every point in the map. This is similar to what is done for hazard maps by geomorphologists or for safety factor maps by civil engineers. The next step is to use the hazard map to compute a prediction-rate curve, as described in the partition, prediction, and cross-validation boxes (2.1, 2.2 and 2.3) of Table 1. To compute the prediction-rate curves, we must have partitioned the past landslides into either time periods, or into randomly selected sub-groups. Other partitioning criteria can also be used to generate the groups of the past landslides. One group has to be used to compute the predicted hazard maps and the other to compute the prediction-rate curves. An artificial example of a prediction-rate curve is shown in Figure 1.
Table 1. General strategy in Favorability Function (FF) modeling for landslide hazard mapping

<table>
<thead>
<tr>
<th>Step</th>
<th>Task</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Preparation</td>
<td>Prepare the past landslide layer containing the locations (or the scars) of the occurrences of the past landslides in study area A, and delineate the scarps (trigger areas) of the scars. Also prepare m map layers of causal (or correlated) factors of the occurrences. Each causal map layer consists of a number of mapping units and provides evidence for finding the future landslides. Co-register all the layers using the geo-reference and select a FF model for analysis.</td>
</tr>
<tr>
<td>1.2</td>
<td>Estimation</td>
<td>Construct FF models [2,6,7,8]. Compute a FF-value at each pixel: it shows the relative hazard level for that pixel being a part of the scarp of a future landslide. These computed FF-values are in a continuous scale and constitute the predicted hazard map. The values can be divided into a number of hazard classes for visualization.</td>
</tr>
<tr>
<td>2.1</td>
<td>Partition</td>
<td>To generate the prediction-rate curve for the hazard map, we divide the scarps of the past landslides into two groups, termed Estimation-group and Validation-group. Either time-partition or random-partition is generally obtained. Whenever possible, the time-partition, where the past landslides are divided into two time-periods, is strongly recommended.</td>
</tr>
<tr>
<td>2.2</td>
<td>Prediction</td>
<td>Using the Estimation step 1.2, generate a predicted hazard map using the Estimation-group landslides described in the Partition step 2.1.</td>
</tr>
<tr>
<td>2.3</td>
<td>Cross-validation</td>
<td>For any given hazard level in the predicted hazard map of the Prediction step 2.2, select all the pixels whose hazard levels are greater than the given level. Within the selected pixels, count the scarps in the Validation-group landslides. At each hazard level, compute two ratios, the first ratio is for the number of selected pixels and the total number of pixels in A, and the second ratio is for the counted scarps within the selected pixels and the total number of scarps in the Validation-group landslides. The sets of two ratios constitute the prediction-rate curve of the predicted hazard map. As the hazard level decreases, the number of selected pixels will increase, and the both the ratios will increase to 1. A hypothetical example of a prediction-rate curve is shown in Figure 1.</td>
</tr>
</tbody>
</table>
The prediction-rate curve for the hazard map in the Estimation step 1.2 is obtained by overlaying the predicted hazard map generated by using the distribution of the Estimation-group landslides from the Prediction step 2.2, with the distribution of the Validation-group landslides. The horizontal axis of the curve in Figure 1 is for the first ratio described in the Cross-validation step 2.3 and it indicates the proportion of the study area classified as hazardous. The vertical axis in Figure 1 is for the second ratio described in the Cross-validation step 2.3, and it indicates the proportion of the scarps in the Validation-group landslides within the selected hazardous pixels.

![Prediction-rate curve of landslide hazard](image)

Figure 1: Prediction-rate curve of landslide hazard. Points A, B, and C relate 10%, 20%, and 30% top values with prediction rates 66%, 82% and 88%, respectively.

In the diagram of Figure 1, the convex prediction rate curve indicates the “prediction power” of a predicted hazard map. If the map has no prediction power (i.e., if the predicted hazard map shows a random pattern with respect to the future landslides), then the expected proportion of the scarps in the Validation-group landslides within a set of selected hazardous pixels is equivalent to the proportion of selected pixels in the study area. For instance, point A in Figure 1 indicates (10%, 66%). It means that when we take a hazard level in the predicted map such that the proportional amount of pixels, whose level is greater than the chosen one, is 10% of the whole study area, then the proportion of the scarps in the Validation-group landslides within these 10% selected pixels is 66% (and not just 10% of the scarps as expected). In addition, the points B, and C in Figure 1 indicate the 20% and 30% top hazard classes of the study area, respectively. We can see that they correspond to the 82% and 88% prediction rates, so that the rate increases 16% going from A to B, and only 6% going from B to C. This may mean that the most hazardous 10% class predicted 66% of the Validation-group landslides, while the second most hazardous 10% class predicted 16% only. The subsequent hazardous 10% class predicted 6%,
i.e., less than the 10% expected. These three proportions, 66%, 16% and 6% are the estimated probabilities of the occurrences of future landslides within the three most hazardous predicted classes with relative sizes of 10% of the study area. Similarly we can estimate the probabilities of the occurrences of future landslides within any predicted hazard level in the predicted map.

In practice, we can study the prediction-rate curve and the corresponding predicted hazard maps in terms of cost-benefit analysis. Using the prediction rate curves, we can also measure the degree of support of combinations of causal factors, the effectiveness of different FF models, the time or space robustness of the predictions, etc. [7,9,10]. For instance, a decisional issue could be to identify areas of predicted highest hazard value that cover no more that 10% of the study area. Such areas could be a first priority for direct inspection on the field or for investing resources in prevention works. To exemplify how to use the prediction-rate curves in practice, we will use the spatial databases constructed for the two case studies in the Iberian Peninsula. They are described in the next section.

3. Two databases in Portugal and Spain

In Portugal, at the University of Lisbon, a spatial database was constructed for the Fanhões-Trancão area, north of Lisbon. The study area is 17.36 km² and is part of the dip-downstream slope of the Lousa-Bucelas cuesta, a sub-structural slope defined by a general concordance between topographic surface and south and southwest dip of the strata with angular values of 12°. Geologically the region is part of the Portuguese Meso-Cenozoic sedimentary basin and is located close to the contact between that morpho-structural unit and the Tagus River alluvional plain. The maximum elevation does not exceed 350 m a.s.l. The yearly average precipitation is only 700 mm, however, the area is characterized by a great irregularity of rainfall regime considered as a failure-triggering factor [11]. Detailed geologic-geomorphologic mapping at 1:2,000 identified 132 slope movements but 1:10,000 maps were compiled and digitized into a 5m x 5m resolution spatial database consisting of digital images of 703 x 761 pixels. The causal map layers used are: elevation, slope, aspect maps forming the digital elevation model (DEM), geology map, and land use map. The past landslides consisted of the 91 shallow translational slides. The analysis described in the next section will use the following causal factors: the DEM set, geology (6 units), surficial deposits (7 units) land use (5 units), and the 91 shallow translational landslides divided into two groups of randomly chosen slides. A detailed description of the database and of its statistical analysis is in Corominas et al. [1]. A morphologic synthesis of the study region has been provided by Zezere [11,12,13]. Figure 2a shows the distribution of the two groups of landslides used for prediction modeling.

In northern Spain, at the University of Cantabria in Santander and the University of Oviedo, a spatial database was constructed for the lower part of the Deba Valley, in the Basque Province of Guipuzcoa. The elevation varies between 0 and 700 m a.s.l.
and the terrain is moderately folded and faulted, consisting of a variety of limestones, marls, claystones, sandstones, flysch facies and volcanic rocks corresponding to the Cretaceous and Paleogene of the Basco-Cantabrico Pyrenees. The slopes are steep and there are surficial deposits, which represent different composition and thickness, determining the occurrence of many landslides triggered by intense rainfall episodes. The main annual rainfall is about 1,500 mm. Four dynamic types of landslides were distinguished: falls, translational, rotational, and flows. The study area covers over 500 km² and over 1500 separate landslides were mapped and described. Using field observations and photo-interpreting three different flight coverages of photographs (years 1954, 1991 and 1997), the translational landslides and associated flows could be classified into three temporal groups: before 1954, during 1954-1991 and during 1991-1997. Of the 1:5,000 mapped landslides, the main failure size is about 400 m², however, the resolution of the 1:25,000 maps compiled and digitized was 20m x 20m so that the distribution of each landslide corresponded with the closest 20m x 20m pixels. The database consisted of digital images of 2002 x 1801 pixels. A total of 1493 shallow translational slides and flows were used in the analysis. They were further split into two time periods, 532 that occurred during 1954-1991 and 907 during 1991-1997. The second group was further subdivided into two spatial groups containing each 50% of the randomly selected slides (i.e., 454 and 453). The partitioning of the landslides was to enable validating the predictions.

The causal factors are: lithologic map (30 units), vegetation/land use map (7 units), thickness of surficial deposits (3 classes) and the DEM set. A detailed description of the geology and geomorphology of the study area and of the database has been made by Corominas et al. [1] and Remondo [13]. Figure 4a shows the distribution of the two time-partitioned groups of landslides. The analytical results of the two databases are discussed in the next section.
1. Scars and scarps (in gray) of group 1 landslides
2. Scars and scarps (in black) of group 2 landslides
Figure 3
Figure 5: Two prediction-rate curves for the Spanish study area: 907 landslides for 1991-1997 were used to validate the temporal prediction obtained using 532 landslides for 1954-1991. The random partition used 454 landslides of the 907 to predict and 453 to validate.

4. Generation of prediction-rate curves from the two databases

The spatial database for the Fanhões-Trancão area in Portugal was used to generate the predicted hazard map shown in Figure 2b based on one of the FF models termed “Likelihood ratio function” [4]. The Estimation-group landslides consist of the 45 white polygons shown in Figure 2a to indicate the shallow translational landslides used to generate the predicted hazard map shown in Figure 2b. The illustration also shows the 46 black polygons that indicate the Validation-group landslides used to obtain the prediction-rate curve. The scarps locate the trigger areas of the landslides and the scars represent their larger bodies. The gray-level bar in Figure 2b indicates the decreasing predicted hazard level from white to black. The causal factors used are a combination of all six layers. The corresponding prediction-rate curve is shown in Figure 3, together with the curves obtained using single layers and all the six layers together. As it can be seen, slope is the single causal that provides most of the support so that it predicts much better than the other individual ones. An enlargement of the area in the black rectangular frame is shown below. In this
illustration, the prediction-rate curves help to understand the database effectiveness. For instance, the 15% of the study area with the highest predicted values contains 87% of the validation landslides when we use slope and aspect together. This is slightly better than using all six layers. Lower prediction rates are obtained using the other individual layers. The pattern in Figure 2b predicts only in space, owing to the fact that the database allowed only a space separation of the landslides. It is important to note that, although the prediction-rate curve was obtained by comparing the predicted hazard map, shown in Figure 2b, with the 46 Validation-group landslides, the prediction-rate curve should be used with the hazard map based on all 91 past landslides to estimate the probabilities of the occurrences of future landslides. Here the prediction refers to “the next 46 landslides” without a specific time period identified. The next application deals also with a prediction in time.

From the database over the lower Deba valley of northern Spain, the prediction in Figure 4b was obtained using the “Certainty factor” model [2,6]. The following causal map layers were used: elevation, slope and aspect classes, lithology and vegetation units. The landslide information consisted of 532 translational landslides that occurred during 1954-1991 and 907 for the period 1991-1997. Those mapped in the second period were used for validation. The two time partitions of landslides are shown in Figure 4a. In another experiment, the later 907 translational landslides were further subdivided into two groups of randomly selected landslides (454 and 453), similarly to what done in the Portuguese case study. The corresponding prediction-rate curves are shown in Figure 5. As it can be seen, the top 15% values predict only 53% of the validation landslides. In the previous example of Figure 3, the top 15% predicted 87% of the validation landslides. Furthermore, the two prediction-rate curves in Figure 5 for time and space predictions are very similar. The time prediction curve extends the prediction to “the next 7 years”, while the space prediction curve refers to “the next 453 landslides”. It also should be noted that, although the prediction rate curve was obtained by comparing the predicted hazard map, shown in Figure 4b, with the 453 Validation-group landslides, the prediction-rate curve should be used with the predicted hazard map based on all 907 past landslides to estimate the probabilities of the occurrences of future landslides.

5. Towards risk assessment

To illustrate how to use the prediction rate-curve to estimate the probabilities of the occurrences of future landslides for risk analysis, let us take Portuguese study area as an example. The first step is to generate a predicted hazard map using the same FF model (as in Figure 2) based on all six layers of the causal factors and the 91 past landslides. It should be somewhat similar to the predicted hazard map in Figure 2, where we have used all six layers and the only 45 Estimation-group landslides. Consider the most hazardous 0.2 km$^2$ (8,000 pixels of 5m x 5m) from the predicted hazard map based on all 91 landslides. 0.2 km$^2$ is approximately 1.5% of the whole study area. Let us look at the prediction-rate curve (black curve with triangles) for
“All six layers.” When the X-axis is 0.015 (1.5%), the corresponding Y-axis of the prediction rate curve is 0.346 (or 34.6%).

Suppose that we now assume that we expect 20 landslides over the next 10 years with an average size of the trigger areas of the future landslides that is approximately 10m x 10m. If we build 25 houses of the size of 10m x 15m within the most hazardous 0.2 km², the estimate of the probability that one of the houses will be part of the trigger areas of the 20 future landslides is approximately 40.5%. Obviously as we change the expected number of future landslides, the average size of future landslides, and the number of houses or their size, the estimate of the corresponding probability will also be changed accordingly.

Based on the estimates of such probabilities, once an acceptable high hazard area is identified as a priority for further analysis or for prevention follow-up, a second stage in the analysis can be initiated that requires an inventory of all human activities and assets that are within the its reach, or spatial domain, or zone of influence.

We are proposing a new way to express hazard spatially and statistically. We generate numerical predictions maps in which we try to estimate the probability of occurrence of each class. We produce a prediction-rate curve by using a validation technique. If our database allows temporal validation, we can predict within the time interval provided by the temporal partitioning.

This procedure would extend the general strategy described in Table 1, from a first stage in which we generate hazard predictions, to a second stage in which we estimate the probability for each hazard class and the corresponding spatial zone of influence. Further research work by the authors is directed to this target.

6. Acknowledgements

The initial development of this research has been supported by a project on “New Technologies for Landslide Hazard Assessment and Management in Europe, NEWTECH” (Contract ENV-CT96-0248) of the European Commission’s Environment Programme. The present work is now supported by another research project on “Assessment of Landslide Risk and Mitigation in Mountain Areas, ALARM” (Contract EVG1-CT-2001-00038) of the European Commission’s Fifth Framework Programme.

References


