On Spatial Prediction Models or the Unbearable Lightness Of GIS

Andrea G. Fabbri
SPIN*lab*, IVM, VU, NL

Introduction

Rector of the Vrije Universiteit, ladies and gentlemen:

On this occasion I am going to expose my personal views on physical and social modeling issues of concern related with the endowed chair of Geo-information and Risk Management that I have the honor to occupy at the SPIN*lab* since January of last year. The subject of my lecture, as it can be guessed from the title, is meant to be provocative! But what do I mean by "provocative"?

Using the term "provocative" reminds me of another, less momentous, occasion in which I was asked to stimulate an audience of mathematically oriented earth scientists. It was during a European Science Foundation Workshop on the future of space-time modeling in the geo-sciences. I recall it took place in the beautiful hills of Tuscany in Italy, in a lovely resort named "Il Ciocco." My provocation consisted of describing the surveyor geologist as a guy carrying a portable computer to the field and who would systematically capture digitally all observations. He/she would store them on floppies and construct spatial databases to generate all kinds of geological maps "on demand" for a multitude of users and uses.

My portray generated havoc and the 90 people or so in the audience started shouting against or in favor of that vision even down to heavy expressions to one another! The chairman was staring at me with visible perplexity. Had I been too successful in my stimulation? That was over 15 years ago and I realize now that I was simply pointing at the *obvious* for some and simultaneously at the *tragic* for others. I was predicting that geo-information would penetrate surveying practices to the point that newly constructed databases of field observations would allow focused analyses during and after data capture providing support and transparency to thematic maps for exploration and risk assessment as well. In short, they would become the basis of predictions. Here is an illustration of what M.Sc. students at ITC, my former Institute, were doing a few years ago in the field to produce digital geological maps. Figures 1 and 2 show some of the students at work and one of their products. This is rather normal now among those who wish to exploit new technologies (Schetselaar *et al.*, 2006; Clegg *et al.*, 2006), although most maps to date have **an inventory rather than a prediction theme**!

I also realize now that geo-information is all a **matter of responsibility** that we are prepared to accept… here we need models to predict and risk maps to show what and where … statistically. That is a responsibility! Going back to the spatial prediction provocation now, risk is a human condition due to either natural processes or human activities. Risk is something that we either want to avoid or at least minimize. It contains a probability and a potential damage. And this is provocative! Predictions are extremely hard to come by… Am I a visionary to expect a future in which predictions are going to be common? Are they going to be a duty, or a right or a necessity, or a prerequisite, or even compulsory by law?

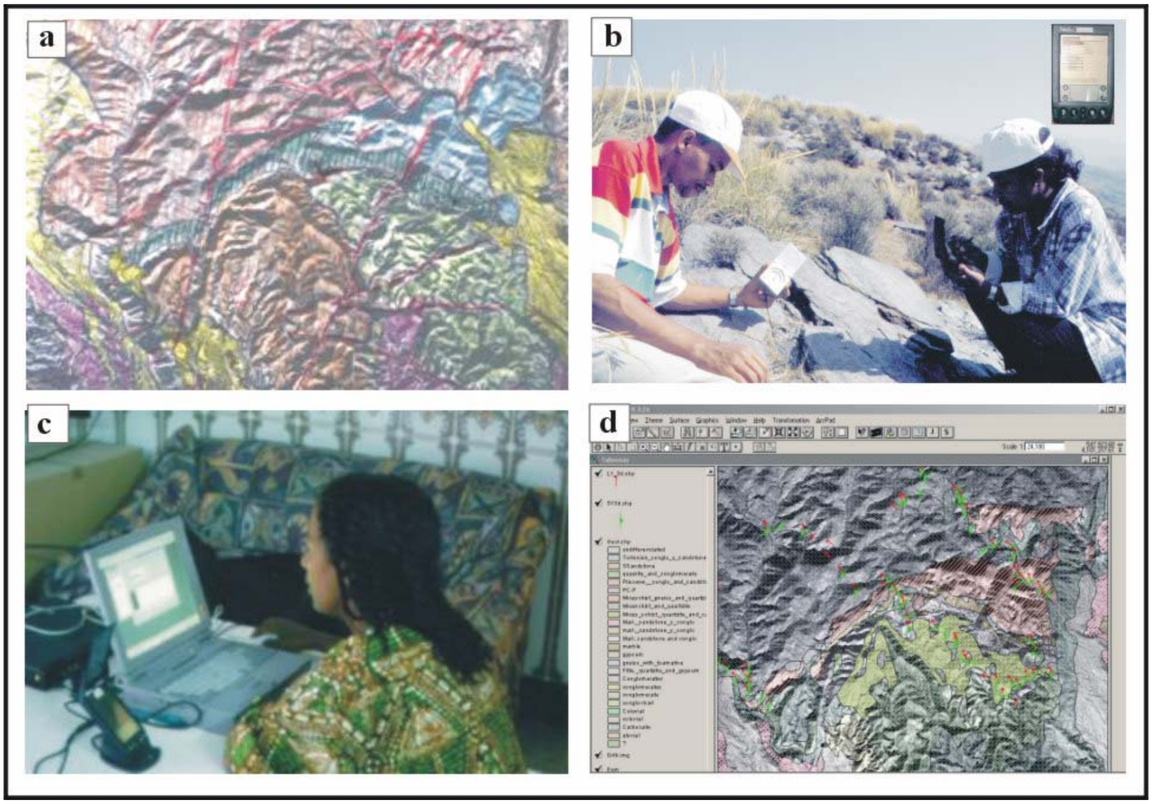


Figure 1: Geology students of ITC at work in the field using PDAs and PCs. (Ernst M. Schetselaar, personal communication).

Here are some questions to consider about risk:

- Do we want to predict hazards and assess risks to minimize the latter? Hazard is the probability of occurrence of hazardous events, while risk is what potentially happens when they occur.
- Should the modeling the future space-occupancy condition be a light or a heavy task? Is this a dilemma in risk research or just in risk communication?
- What to do to avoid risks or to manage acceptable risks?
- How well are we performing when anticipating risk level distribution in time and in space?

These are today very provocative questions that made me propose a title suggestive of Milan Kundera's novel "The Unbearable Lightness of Being", where he writes (Kundera, 1984, p. 2-3):

> *"… the idea of eternal return implies a perspective from which things appear other than as we know them: they appear without the mitigating circumstances of their transitory nature. … If eternal return is the heaviest of burdens, then our lives can stand out against it in all their splendid lightness. But is heaviness truly deplorable and lightness splendid?"*

These words, taken out of their initial broad existential context, when placed into the spatial prediction context can have a strong significance in the specific instance of generating and using geo-information for risk or for disaster management. Why to model physical and social processes? What are the models anyway? Who needs the results of such models, and furthermore, how good are the results? Or alternatively, should we just wait and see whether a disaster really happens where or when predicted to occur? Or even, should we simply make an account of what has happened and pretend that it will help in the future, thus avoiding responsibility? Is that convenient in view of the critical hazardous processes today … going from the avian influenza spread, to widespread land sliding (not in Holland of course…), or from human induced subsidence, earthquakes, tsunamis, wildfires, to environmental impacts and terrorism?

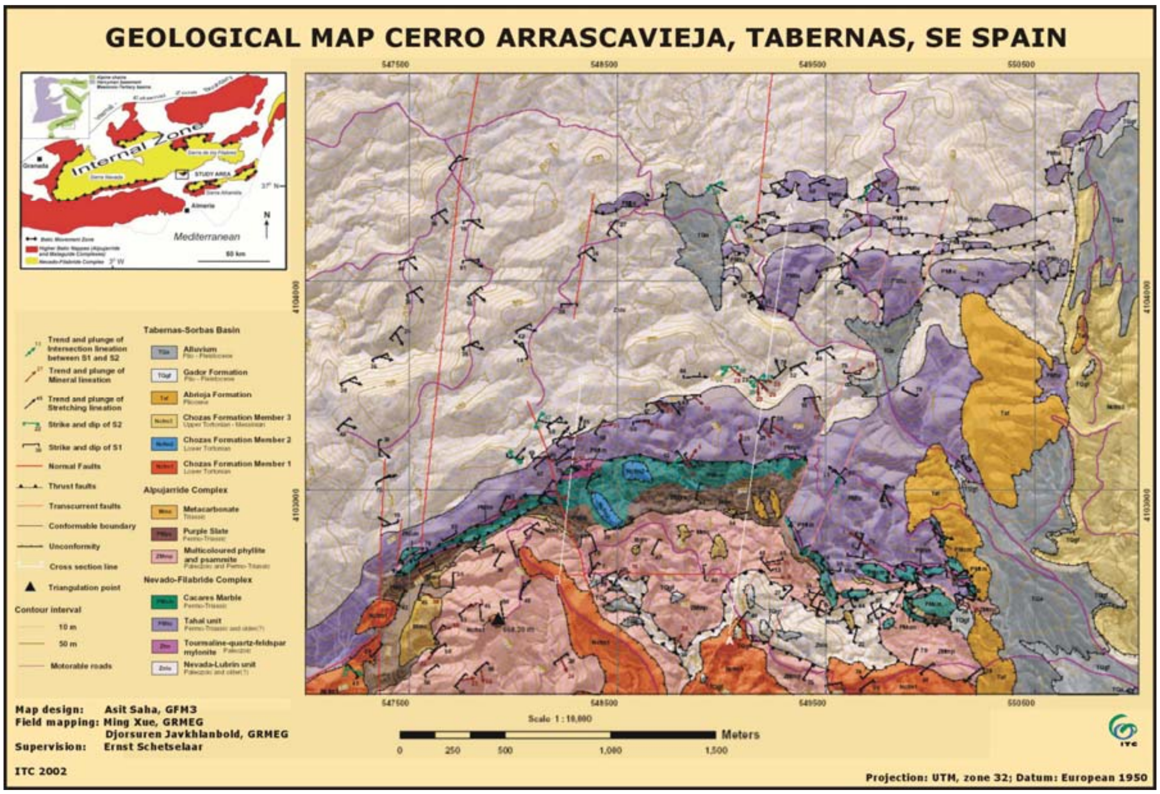


Figure 2: A geological map of the Arrascavieja area, Tabernas, SE Spain, compiled by a team of students. The map compilation Takes around two weeks of office work, teaching the database management and cartographic aspects of digital geologic mapping (after Schetselaar *et al*., 2006).

Modeling means to systematically use expert's knowledge observations to recognize the typical conditions (settings, trends or signatures) of hazardous events and to construct scenarios for their prediction in space or possibly in time. Such predictions over relatively large areas are of necessity probabilistic. That leads to establishing who the experts are, and when of necessity they use statistics, i.e., asking the question: what do they really have to know in order to be effective and accept responsibilities for their results? What I have just put as a casual bunch of questions is definitely provocative for all those who do not dare exposing what they know about risk and space and also what they do not know or perhaps cannot know.

To respond to that sort of questions, let me refer to some research and training activities that led me to the present personal view. I have myself been dealing with spatial prediction models for a while, actually since 1969, when I joined the Geological Survey of Canada as a young researcher. At that time, natural resource exploration was the front runner program in Canadian Government's research. Our task in "geo-mathematics" was to reduce … **the risk of failing to discover new resources** by drilling in the wrong places. For this we would digitize geological, geophysical and geochemical anomaly maps and the distribution of mineral occurrences to establish relationships indicating the typical settings to be mapped for larger regions and to guide exploration. That was in the early 1970s and we were using early geo-information and GIS (termed *image processing systems* at that time!) and we waited some years to see whether the new discoveries occurred there where we statistically predicted high mineral potential with statistical methods. Eventually we were to construct special validation techniques and strategies to assess the "goodness" of our predictions. I will be coming back to that.

When later on I left Canada as a senior geo-information scientist to come and join ITC in the Netherlands, natural resources were much less of a priority than was predicting natural hazards and assessing risks. As a matter of fact, besides becoming involved in lecturing on environmental geosciences, I found myself naturally trapped in a sequence of EC Research Network Projects and that still continues today. Here they are with self explanatory titles:

- "**Geomorphology and Environmental Impact Assessment**: a network of researchers in the European Community (1993-1996) HC&M, ERBCXRXCT 930311;
- "**New Technologies for Landslide Hazard Assessment and Management** in Europe", NEWTECH (1996-1998) CEC Environmental Programme, ENV-CT96-024;
- "A European Research Network for the **Application of Geomorphology and Environmental Impact Assessment to Transportation Systems** (1998-2001) TMR, ERBFMRXCT970162;
- "**Quantitative Indicators and Indices of Environmental Quality**: a Euro-Latin American Network for Environmental Assessment and Monitoring", ELANEM (1999-2002) INCO-DC, ERBIC18CT980290;
- "**Assessment of Landslide Risks and Mitigation in mountain Areas**", ALARM (2001-2004) EVG1-2001-00018;
- "**Mountain Risks: from Prediction to Management to Governance**", MOUNTAIN RISKS (2007-2010) 6th Framework's, MRTN-CT-2006-035978, a Marie Curie Research Training Network.

Remarkable about those projects are the increasing relevance of geo-information and eventually the critical contribution of economics and the social sciences in cross-disciplinary approaches to estimate, evaluate and express risk. For most teams in those projects, however, the idea of modeling using digital geo-information was not easily accepted and much happened that avoided or slowed down predictive modeling to restrict research to **representing what had happened** and not to deal with what will happen. This indeed became a provocative issue! Nevertheless some teams in the networks, namely the ones from the Iberian peninsula, made considerable progress not only in the use of geo-information but also in hazard prediction and risk analysis (Remondo *et al*., 2003, 2005; Zêzere *et al.,* 2004). Another peculiar aspect of European research networking attitude was to propose a different study area for each participating team, thus not truly facilitating too much the cross-breeding of methods, techniques and approaches. This I consider still a tendency hard to correct that does not lead to extensive use of geo-information and spatial prediction modeling. One more provocation here…

These were simply research activities in which I was involved either as a principal investigator or as coordinator, with a focus on predicting, representing and communicating hazard and risk related mainly to mass movements or to environmental quality deterioration. Those activities, nevertheless, stimulated greatly the development of validation procedures to assess the **"goodness" of spatial predictions** of hazards and risks as we will see later on. As such, I consider them very precious to me.

What are the plans of the EU in the risk area today then? Looking at the actual EC's 7th Framework on: Information Communication Technology, ICT, for Sustainable Growth we can see emphasis on: Disater Risk Reduction, DRR, Emergency Management, EM, and the 7th Framework R&TD, with an indicative budget of 53 M €. In addition, the Natural Hazard Programme in the 2007 calls has an indicative budget of 13 M €. Their target and the respective indicative budget, point at the level of concern for risk R&D in Europe. It becomes apparent that things like unsustainable trends, risk exposures to pollution and disasters, and energy sources increasing green house gases are critical targets. Terms like generic solutions, validation, risk reduction, emergency management, interoperability of ICT-based solutions and vulnerability, all rely heavily on modeling the future with geo-information. Where are we likely then to find representations of future hazards and risks? Are the attitudes that I have encountered still giving preference to the technology for describing the past disregarding the technology for predicting the future and for disaster preparedness? These questions, I am afraid, are still in need of satisfactory answers! It seems that even **training initiatives are mostly limited to describing the past**. If so, is that a light choice instead of a heavy one? Where are research endeavors that anticipate the future? That should be our ultimate target, would it not? We have one major provocation here!

Before providing examples of our risk maps, I would like to consider some authoritative views (certainly more authoritative than my humble ones) that motivated the use of models to predict hazardous events and consequent risks. A major economic consultant, Peter L. Bernstein (1996) traced the "story of risk" and wrote that: "The revolutionary idea that defines the boundary between modern times and the past is the mastery of risk: the notion that the future is more than a whim of the gods and that men and women are not passive before nature." Then, he added "… the ability to define what may happen in the future and to choose among alternatives lies at the heart of contemporary societies. Risk management guides us over a vast range of decision-making …" Furthermore, he noted that: "In 1703, Gottfried von Leibniz commented to the Swiss scientist and mathematician Jacob Bernoulli that 'Nature has established patterns originating in the return of events, **but only for the most part**,". That was a remarkable and becoming comment, indeed! It led to the application of probability theory to natural risk, for instance.

Another incentive to predictive modeling is provided by Monmonnier's (1997, p. 293) careful analysis of "cartographies of danger" when he points at **hazard-zone**

mapping as a *recent phenomenon* that seems to focus on forecasting and monitoring while prior cartographies used to be mainly descriptive and explanatory of past hazardous events. That means that "most risk maps involve statistical models of some sort for estimating the likelihood of rare events such as volcanic eruptions or disastrous floods … and forecasting requires a representative record of the hazard's magnitude and variability … comparatively rare hazards, like volcanic eruptions are inherently uncertain … we cannot generate a future that uniformly replicates the past."

He continues to describe how GIS can be used to simulate hypothetical earthquakes. Simulations of the intensity of ground shaking are obtained by coverages of shaking vulnerability and seismic energy. Map overlays can aggregate risk scores from point and linear hazard to yield composite risk maps that can be overlaid with coverages for vulnerable subpopulations to identify areas most in need of vigilance, mitigation and preparedness. Of course, many references are made to extensive documentation material available from the U.S. Geological Survey, probably the most active large federal institution in natural risk representation and communication. I will refer to them later on in this presentation.

Another related work is David Alexander's book (2002) on the principles of emergency planning and management, which dedicates a chapter to methodology for analytical techniques. In it, modeling, simulations and scenarios are discussed for hazard mitigation and zonation, and emergency decision-making, however, no theory of spatial prediction is discussed. Risk estimation is simplified by relating the distribution of physical hazard to that of human vulnerability giving the resulting pattern a temporal framework. Risk turns into the likelihood that casualties, damages or losses, will occur in a given interval of time or within a given return period. Again, GIS overlays of geo-information can be used to build predictive models of hazard. But how are they obtained?

Richard Posner (2004) in his book on "Catastrophe, Risk and Response" is not so hopeful that risk or major disasters can be managed, understood or predicted (e.g., meteorite impacts, epidemics or terrorist attacks). Risk analysis is still a scientific endeavor and scientists are considered unreliable when evaluating issues in which they have vested interests. Furthermore, in the US judges (like himself) often are not in the position to evaluate how to decide on risk matters or controversies, mostly because of their lack of exposure to science so that they are not prepared to understand technical information. One would wonder then, who is to take action on such risky issues in the U.S., such as global change! That seems a likely role for the new Department of Homeland Security, DHS. For instance, Plum and Beitel (2006) at the Idaho National Laboratory, an institute that collaborates with the DHS, are concerned with risk communication and propose the development of a World Risk Index, WRI, by the Department of Homeland Security (DHS was established just after the September 11 2001 disaster). An example of classical calculation of risk that is used is: *Risk = Probability (of Attack) * Consequence (of Attack).* Similarly to what is done in seismology, the levels of risk are best expressed using a logarithmic scale. They provide an alternative to using threat thresholds, which imply potential consequences and require the people **to predict the probability of an event** and to calculate the risk given their personal situation. They propose scales that must communicate risk as a function of event probability and the probability of consequence. Tables 1 and 2 provide World Risk Levels for ranges of loss of life as absolute numbers and as percentages of loss of population, respectively. The tables can be adapted to either large or small countries or communities or even to the entire world. We could wonder if and when such indexes would **trigger compulsory action**. However, even in this work, no particular attention is given to methods of spatial prediction and to the reliability of their application. Here I am personally provoked now! Because I continue feeling uncomfortable about this, I want now to discuss how to define and visualize risk and how risk maps can be constructed.

Table 1: World's Risk Index by number of loss of life (after Plum and Beitel, 2006).

	Range of lost life		
World Risk Level	**Low range**	**% median life lost**	**High range**
WRL 10	100,000,000	1,000,000,000	10,000,000,000
WRL 9	10,000,000	100,000,000	1,000,000,000
WRL 8	1,000,000	10,000,000	100,000,000
WRL 7	100,000	1,000,000	10,000,000
WRL 6	10,000	100,000	1,000,000
WRL 5	1,000	10,000	100,000
WRL 4	100	1,000	10,000
WRL 3	10	100	1,000
WRL 2	1	10	100
WRL 1	0.	1	10

Table 2: World's Risk Index by number of loss of life (after Plum and Beitel, 2006).

	Range of lost life		
World Risk Level	**Low range**	**% median life lost**	**High range**
WRL 10	1.00%	31.6%	1000.0%
WRL 9	0.10%	3.2%	100.0%
WRL 8	0.01%	0.32%	10.0%
WRL 7	0.001%	0.032%	1.0%
WRL 6	0.0001%	0.0032%	0.1%
WRL 5	0.00001%	0.00032%	0.01%
WRL 4	0.000001%	0.000032%	0.001%
WRL 3	0.0000001%	0.0000032%	0.0001%
WRL 2	0.00000001%	0.00000032%	0.00001%
WRL 1	0.000000001%	0.000000032%	0.000001%
WRL 0	0.0000000001%	0.0000000032%	0.0000001%

Risk visualization

First, let us review what risk is intended to be and how it is normally visualized. Risk visualization for risk management combines risk analysis and risk evaluation (see Plattner, 2004, and http://www.sra.org/resources_glossary.php, for a discussion of various risk terms). In essence, risk is a human condition related to the probability that one or more natural or technological processes take place affecting negatively our daily life, there where we are more exposed to the damage. In practice it is the spatial distribution of the natural and technological processes and the exposed socioeconomic activities that are critical to the risk management. Such distribution is conveniently communicated in map form.

However, the technical difficulties in generating risk maps are numerous and multidisciplinary, ranging from the poor availability of consistent data, the need to model the hazardous processes in space and in time, the complexity of valuating human life, assets and activities and the co-occurrence of more than one risk. I am intentionally avoiding the political difficulties here! I do not want a provocation flood right now! Clearly, the risk mapping task involves objective and subjective aspects and representations that have to be directed not only to specialists in the risk areas, but also to non-specialist decision-makers and to the general public whose perception of risks can be an important factor in risk management. As a result, the generation of a risk map implies a strong responsibility for the producer and for the local administration that eventually distributes it and explains its usability. What is a risk map then?

Some current examples of risk maps

A slightly modified version of this and the following section are taken from Fabbri *et al.* (2005). Recently I have made a naïve search on the Internet to describe the present general understanding of risk maps. Using the two keyword "risk maps" has immediately lead to over 10 million hits! Clearly the topic happens to be trendy and lively, however there must be a large variety of interpretations of how those maps should look like, on their meaning and use.

For instance, **Audinet** (http://www.audinet.org/docs/riskas.doc), a website hit for risk maps, provides a risk assessment survey and risk mapping tool to prioritize business risks. It consists of a checklist to rate (1 to 10) the significance of various risks, initially without regard to their likelihood of occurrence, and then to rate the latter (1 to 5) to obtain a diagram with significance on the vertical axis and likelihood of occurrence on the horizontal one. The corresponding high and low values allow identifying and prioritizing four quadrants for: (I) Prevention of risks at source, high priority, (II) Detection and monitoring, significant risks, (III) Monitoring, less significant risks, and (IV) Low control of risk that are non significant. Obviously the map is simply intended to represent only the decision space and not the physical space. Another example are the risk maps made available by Risk Management Solutions (http://www.rms.com//Publications?Maps.asp/) that offers Natural Hazard Risk, Terrorism Risk, Water risk and Enterprise Risk Services and a variety of RMS catastrophe maps of the US, Latin America, Europe and Japan. They are small scale maps for posters intended to assist catastrophe managers and the like at conferences etc. Contoured values for entire continents or countries show a common measure of combined relative risk for the most typical insured hazards (termed aggregate Average Annual Loss or AAL), a so-called Risk Thermometer for selected cities, and the footprints and industrial losses for historical disasters. Again, such products are not meant for a close analytical scrutiny for risk management. It could be another provocation here but the authors state that themselves…

An extensive risk map production is through the U.S. Geological Survey's National Service Hazard Project (http://www.eqhamaps.usgs.gov/), the Multi-Hazard

Mapping Institution (http://www.hazardmaps.usgs.gov/atlas.php), and the National Landslide Hazard Program:

(hppt://landslides.usgs.gov/html_files/landslides/program.html/).

In particular, the USGS **geo-hazards** research work provides a list of research projects and staff where articles can be downloaded for instance on **landslide recurrence intervals and probabilities** in the Seattle area, Washington State (Coe *et al.*, 2004), (http://geohazards.cz.usgs.gov/ research.html/). Maps are provided by those authors of landslide densities, mean recurrence intervals and "exceedence" probabilities for different probability models applied in that study area. However, as the authors warn, they are to be used as **a general guide to landslide occurrences** and not as a prediction of landslide hazard at specific sites! Oh boy! That is honest, alas a rare attitude!

Clearly, as seen in those few examples, we can go from general and broad representations of risk to detailed risk maps for specific areas of concern, so that even the characterization of all types of risk maps on the World Wide Web would become a research endeavor in itself (e.g., http://www.territorio.t-6.it/**armonia_**overview.htm)! Nevertheless, one of the problems encountered to date is that **none of the risk maps analyzed seem to contain measures of *credibility, uncertainty* and *robustness*** of the spatial prediction representations. In particular, it is not clear whether the risk is represented as an aggregation of past events or as a prediction of future ones. Because of this, I and my collaborators (Fabbri *et al*., 2004; Chung *et al*., 2005) have introduced an analytical strategy to provide such measures for spatial predictions of hazard and risk maps via **empirical validation techniques**. That approach I wish to discuss next. It aims at avoiding a number of common pitfalls in existing hazard/risk mapping models applied in GIS analysis (Chung and Fabbri, 2004), such as:

(a) the absence of statements on the assumptions made in the prediction models,
(b) the lack of validation of the prediction results, and
(c) the absence of estimations of the conditional probabilities of future events given the characterizations of an area within a study area.

Overcoming these deficiencies is a necessary but not sufficient condition. The following (provocative) points are still major challenges:

(1) need of *spatial database* that captures the distribution of hazardous processes, their settings and of the socioeconomic elements exposed to risk;
(2) necessity to use *models for estimating* the hazard probabilities;
(3) requirement of techniques for *estimating the uncertainties* associated with the models and for estimating the uncertainties associated with the database;
(4) development of *scenarios* necessary to compute the risks; and,
(5) different techniques needed for *representing the risk maps* so that the risk levels and the associated uncertainties can be understood.

Let us now look at some results based on spatial validation strategies for resolving those problems (Fabbri *et al*., 2005).

Application examples of risk and hazard mapping

Risk is a condition that is evaluated by combining the presence of exposed vulnerable elements and the probability of occurrence of hazardous processes. Without the former no risk condition can occur. It is represented commonly either as monetary loss, i.e., $

values, or as a number of human casualties expected. An example of a risk map can be seen in Figure 3 that shows a 5-class population risk in South Korea for the Boeun study area that is affected by landslides of surficial debris flow type. Naturally, most of the map has no risk values, which is due to the absence of urban settlements at those locations.

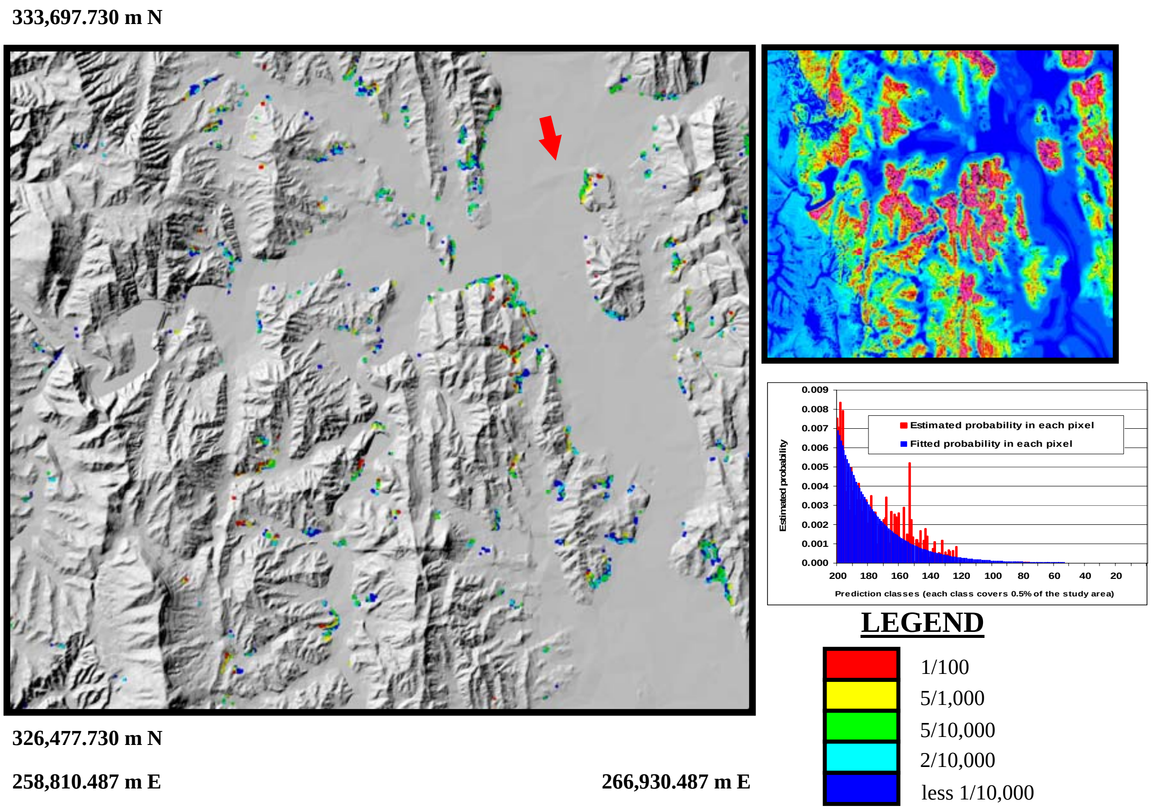


Fig. 3: A 5-class population risk map of the Boeun study area located in South Korea, affected by landslide processes. The classes have been warped on a shaded relief enhanced elevation image. On the right side are the landslide-hazard prediction image and the histogram of probability of occurrence necessary to compute the values for the risk map (after Chung *et al*., 2004).

The classes indicate the proportions of casualties expected per 5 m pixel. To understand the significance of the risk maps, it is necessary to know how it has been constructed using a spatial database, a specific mathematical model with its assumptions and the analytical strategy used for the prediction of the hazard. The Boeun study area is 58.4 km^2 and has about 45,600 inhabitants living in 15,000 households. The spatial database (Fabbri *et al*., 2004) is a set of digital images of 1624 x 1444 pixels of 5m x 5m resolution: the DEM, surficial geology, forest coverage, land use, drainage and the distribution of 420 past surficial debris flow landslides that occurred prior to 1997. In addition, several socioeconomic “indicator” images were compiled to represent the vulnerable elements: the distribution of population density, of road networks, buildings of several types and of the drainage features and embankments. For these, values in $ for 5m pixels and the corresponding vulnerability levels (values between 0, no damage, and 1 total destructions) were also compiled. Furthermore, information became available on 44 new landslides in the area that occurred in 1998 and occupied 2,000 pixels. They caused about $ 200,000 of damage to man-made properties and three injuries to persons.

The information on the number of pixels affected in 1998 allowed estimating the risk level distribution in the study area.

What was done, then, was to apply a **Three-Stage** analytical strategy of risk assessment keeping in mind the actual damages and casualties due to the 1998 landslides but using only the numbers of pixels affected in 1998 to set up a computational scenario and the distribution of the 420 pre-1997 landslides. In the **First Stage** the distribution of the 420 pre-1997 landslides was used with a Fuzzy Set prediction model (Chung and Fabbri, 2001) to classify the study area from the spatial relationships between the landslide distribution and the digital images of DEM, surficial geology, forest cover, land use and drainage patterns. The prediction is represented as a 200-value hazard image (using a pseudo-color look-up table) shown on the upper right in Figure 3.

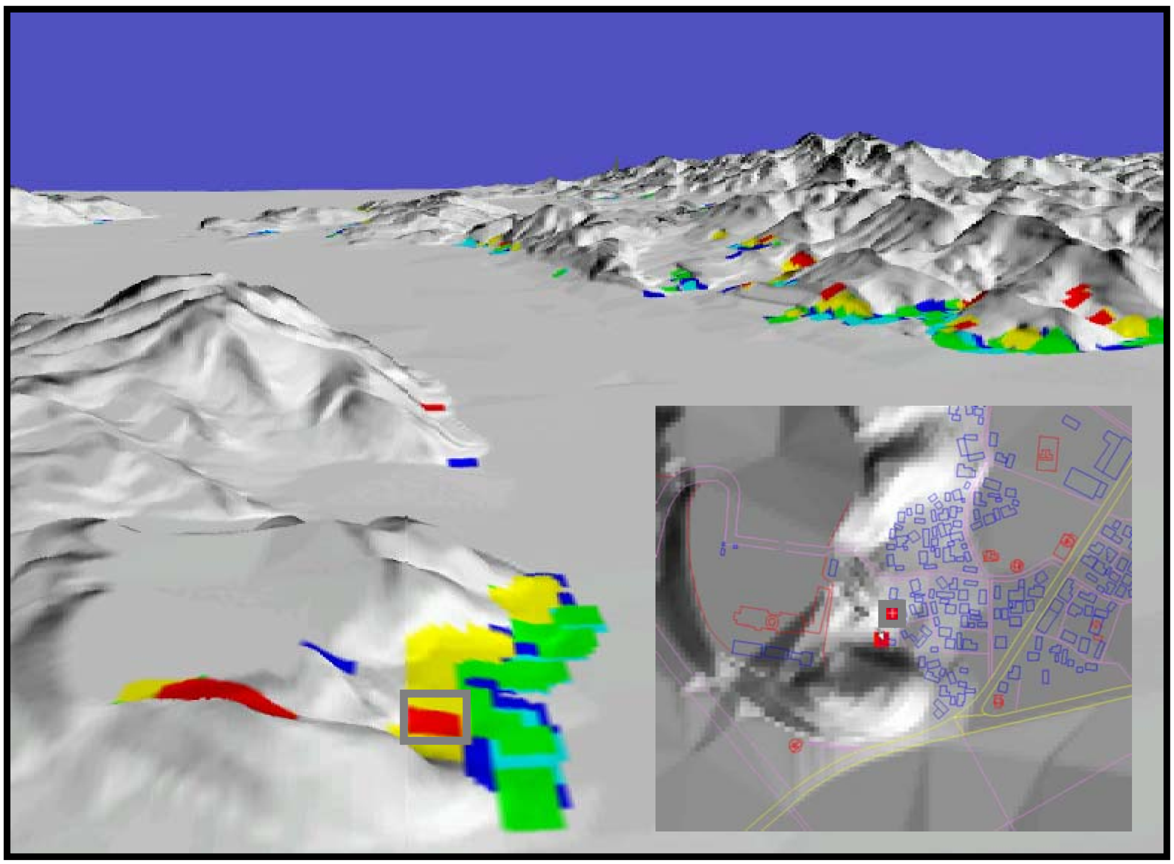


Fig. 4: A fly-through 3-D visualization was obtained of a portion of the risk map in Figure 3 where the flight direction is indicated by a red arrow. The grey box indicates the location of a house where a casualty took place. The inset on the lower right shows a top view of the population density database.

In the **Second Stage**, a second hazard prediction was obtained by the same model but using only the distribution of a random half of the 420 landslides. That of the remaining 210 landslides was compared with the 200 hazard classes obtained in the second prediction to see whether the high hazard classes contain high proportion of the "validation" landslides. This was to obtain a prediction-rate table, also visualized as a prediction-rate curve, expressing the predictability of the events given the database and the 200 classes of hazard (200 used as default). Cost-benefit analysis can be applied to the characterization of the curve into meaningful sections.

Finally, in the **Third Stage** a realistic scenario assumed that 2,000 pixels would be affected by landslides in 1998 so that the probability of occurrence of future landslides could be estimated at each pixel of each class. The estimated probability histogram is shown to the lower right of Figure 3. Those probability values have to be used to combine the first prediction map from the **First Stage** with the socioeconomic data and images using the risk expression, **R = E · V · H**, where **E** indicate the element exposed, **V** its vulnerability, and **H** the probability of occurrence of the hazardous event. The combination of digital images of probability values and vulnerability/dollar values allows computing the risk map in Figure 3. To better communicate the risk visually, a fly-through risk map is shown in Figure 4 where a partial view of the image in Figure 3 is shown.

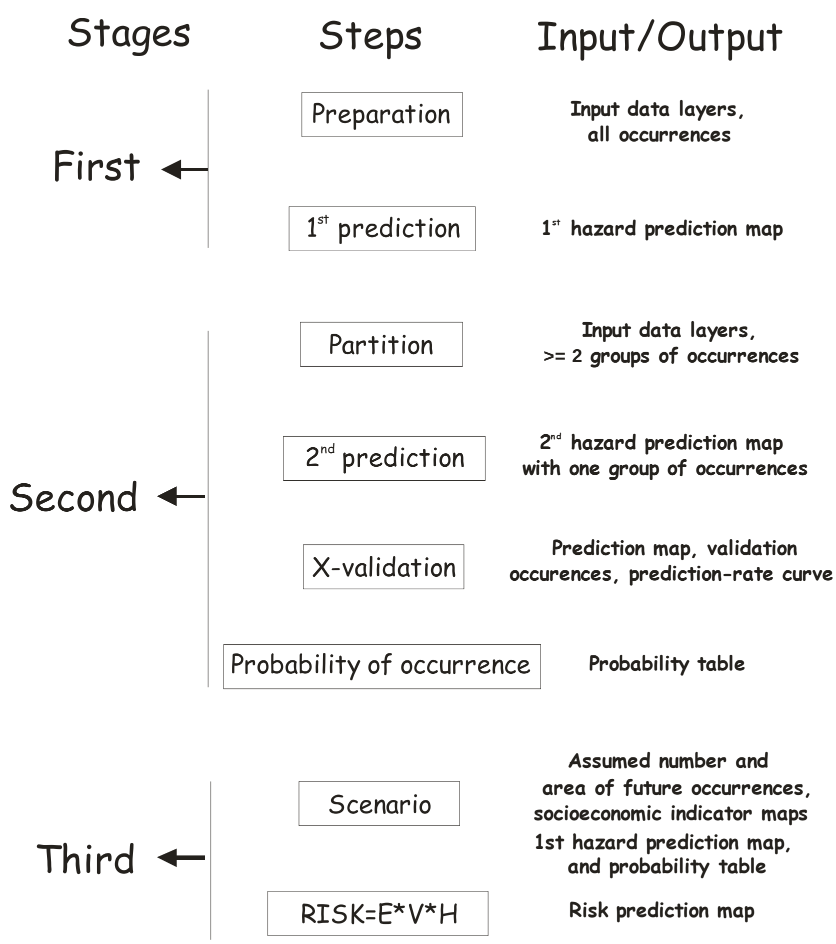


Figure 5: The three stages of the analytical strategy for predicting hazard and risk. The second stage is the critical one for the many purposes of empirical validation: “goodness”, sensitivity, robustness, comparisons, probability of occurrence, etc.

The risk map is evidently a complex construct whose understanding is not trivial due to the analytical steps and the necessary assumptions. One critical issue then is how credible and reliable a risk map is. The **Three-Stage** strategy proposed and shown in Figure 5, is indeed transparent and repeatable, however, in the above application it only provides the empirical validation of the predicted hazard map using a random half of the events. Thus it does not tell us when to expect the events and it only tells us that, given the data in the database, the expected casualties in the study area are 3.14, almost coinciding in this case with the 3 casualties observed in 1998. More considerations on this case study can be found in Chung *et al.* (2005), where also risks to infrastructures and to vegetation are discussed. In this example we have used empirical validation techniques not only to demonstrate and measure the spatial support to the predicted hazard map, but also to estimate the **probability of occurrence** through a scenario that exploited the notion of the 2,000 pixels affected in 1998.

To estimate the risk uncertainty in time and in space, however, we will need more information in the spatial database with the distribution of the hazardous events time intervals and in space subdivisions and a number of different validation experiments. If a time division is not possible because there is no information on the time of occurrence of the past events, they can be randomly subdivided (repeatedly) into two or more groups to obtain other validations.

All such experiments will generate prediction-rate tables and curves that can be compared to assess the uncertainty of the prediction results in the hazard map that is to be used to generate the risk map from the estimation of the probability of occurrence. The programming of a new spatial prediction modeling software intended as complementary to conventional GIS has been a target of mine and of my of my collaborators (Fabbri *et al*., 2004; http://www.spatialmodels.com). I believe that the process of generating credible and convincing risk maps can take advantage of the strategy described here to communicate risk with the public at large. To this end we have dedicated some efforts in training young researchers.

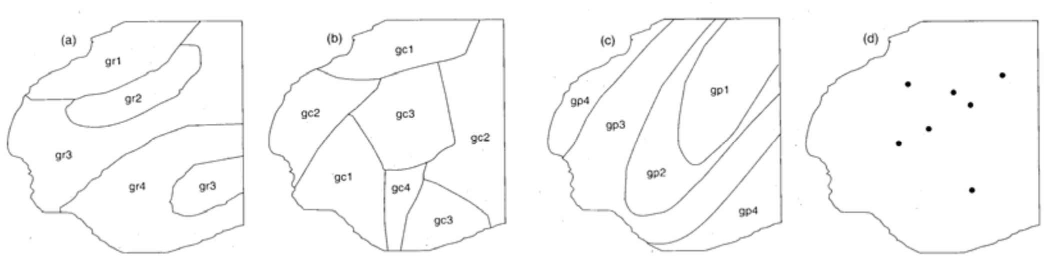
A training application of predictive modeling

To guide students in the application of spatial prediction models a simplified database after Chung and Fabbri (1993) continues being used in international workshops and university lectures. The database for a hypothetical study area consists of three digital maps of four thematic classes supposed to partly reflect the typical setting (circumstances, causal factors or environment) of the occurrence of several point-like events. Figure 6 shows the maps and the respective spatial statistics of the map units and the events. Figure 7 shows different study areas, located elsewhere, with similar map units and event distributions.

Students are asked to “make up” the meaning of the maps and of the events to establish the spatial relationships using at least two different spatial model interpretations (e.g., Bayesian probability, Certainty factor, CF, Fuzzy set membership function) and then to integrate the three maps to obtain the ranking of the polygons resulting from the overlays. Those interpretations allow us to express relative frequencies of event occurrence within recognizable spatial settings. It is the interpretations of their settings that can be integrated into predictive indices by following correctly the corresponding statistical rules. Comparisons are requested of the results of the rankings from the models. Their interpretation and discussion are the scope of the exercise.

Obviously this requires an "in-depth" knowledge of the underlying processes that led to the occurrences. Such knowledge is available with the students owing to their complete and consistent imaginary construct (e.g., forest fires, mass movements, flooding, environmental impacts).

Mathematical expressions for different interpretations of spatial models are provided as well as the database statistics from a GIS, shown in Figures 6 and 7 and in Table 2. The students are asked to use a spreadsheet for computing the values of general spatial relationship functions termed "favorability functions" for the thematic units of each map and to combine them into an overall favorability index according to the respective rules of each model interpretation (e.g., probability, certainty, possibility, referring to the spatial model interpretation mentioned earlier).



Unit	# pixels	Occ.	Prob.
gr1	7,777	1	.000129
gr2	5,236	1	.000191
gr3	21,445	5	.000233
gr4	15,255	0	.000000

Unit	# pixels	Occ.	Prob.
gc1	15,304	1	.000065
gc2	18,654	2	.000107
gc3	13,219	4	.000303
gc4	2,546	0	.000000

Unit	# pixels	Occ.	Prob.
gp1	10,733	3	.000279
gp2	9,579	2	.000208
gp3	18,828	2	.000106
gp4	10,583	0	.000000

Symb.	Occ.	Ratio	Prob.
•	7	7/49723	.000141

Figure 6: Three maps of causal factors and one with event occurrence distribution, with respective statistics.

For instance, the range of Bayesian probability values is 0/1, while for the CF it is -1/+1. It means that for each point pair in the database the favorability values can be compared to rank the corresponding class (sets of polygons). Bayesian probability relates (1) the conditional probability of event occurrence conditional to the presence of each map unit (evidence), (2) that of event occurrence and (3) that of the map unit presence in the entire study area, with (4) that of the presence of the map unit conditional upon the presence of the event occurrence. The integration of the favorability values for each map is obtained using the combination rules of probability.

The CF uses the same type of initial statistics from a GIS but considered as the differences between *a priori* probabilities of event occurrence and of map unit presence and as *a posteriori* probabilities of event occurrence given the presence of the units and of the map units given the event occurrence (corresponding to the Bayesian conditional probabilities). Particular normalization rules are used as well as combination rules that maintain the -1/+1 ranges in the integration. The CF happens to be different in meaning as well as in values from the Bayesian probability so that a comparison of the relative ranks can only be made in terms of equal area classes (as in any environmental index). Such relative ranks need careful interpretation and discussion.

The spatial prediction problem can be seen in various ways, from (1) assuming that no event occurrence is located in the study area (in this case ranking the units only by expert's knowledge), to (2) assuming that some occurrences are in the study area, to (3) assuming that some occurrences are available only or also in other similar areas outside the study area, etc.

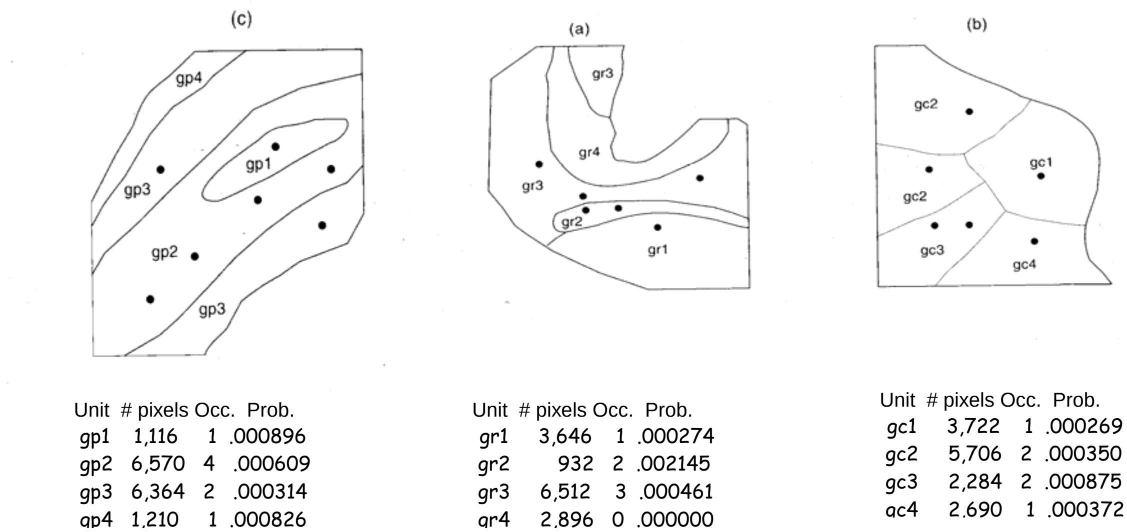


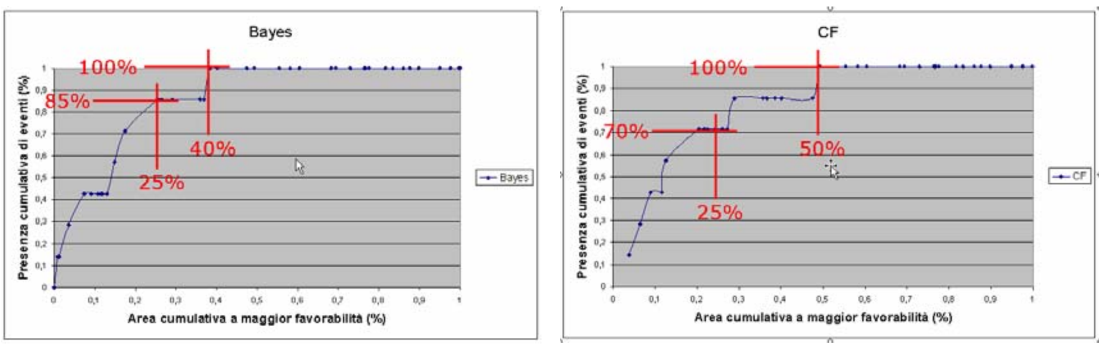
Figure 7: Three additional maps of similar type in different study areas with event occurrences and respective statistics.

Table 2: Pixel counts and events for the unique-condition map units obtained by overlaying maps gr, gc and gp and the events in Figure 6 (a), (b), (c) and (d).

gr∩gc∩gp	# pixels /events	gr∩gc∩gp	# pixels /events	gr∩gc∩gp	# pixels /events
1 1 1	480	3 1 1	802	4 2 2	394
1 1 2	818	3 1 2	845 / 1	4 2 3	2050
1 1 3	3090	3 1 3	3646	4 2 4	878
1 1 4	1343	3 2 1	3946 / 1	4 3 1	1291
1 2 3	731 / 1	3 2 2	902	4 3 2	809
1 2 4	1137	3 2 3	3360	4 3 3	1054
1 3 2	6	3 2 4	3893	4 3 4	2561
1 3 3	88	3 3 1	1885 / 1	4 4 2	1319
2 1 1	539	3 3 2	1238 / 1	4 4 3	1058
2 1 3	138	3 3 3	438 / 1	4 4 4	169
2 2 1	388	3 3 4	500		
2 2 3	906	4 1 2	1831		
2 3 1	1327 / 1	4 1 3	1672	**Total area**	49723
2 3 2	1336	4 1 4	100		
2 3 3	602	4 2 1	69		

The exercise is difficult although rewarding, because it compels the analyst to reflect on the meaning of the spatial relationship, on the assumption of the model (e.g., conditional independence of the evidences) and of the scenario used for predicting the distribution of the future events. Normally students are asked to work in teams of 3 or 4 and the exercise can take from 6 to 30 hours. Figure 80 shows part of a student report. The exercise leads to provocative questions like the following:

(1) What are the assumptions of the scenario to be constructed to estimate the probability of occurrence?
(2) What are the assumptions of the prediction model used?
(3) What are the assumptions of the database used for predicting?
(4) Who do you expect should be the modeler and who the decision maker?
(5) How would you validate the result of your spatial prediction to be convinced or to convince of its "goodness"?
(6) What would you do with a later (or a near by) set of event occurrences?



The Bayesian probability model appears to group in the 40% of the cumulative area all the events, while the CF requires 50% of the study area to contain all the events.
Even at the 25% cumulative area, the CF contains 70% of the events, while 85% are contained in the prediction by the Bayesian probability model . . .

Figure 8: Particular of a student's report where fitting curves were generated to compare the two predictions.

In conclusion, it seems a fact that spatial prediction of hazard and risk is critically desirable and feasible, however, most material for training and communication lacks instances of modeling procedures.

For example, a UNDP, United Nations Development Programme's Disaster Management Training Program has been established since 1992 at the Disaster Management Centre at the University of Wisconsin-Madison, U.S.A. (http://dmc.engr.wisc.edu/courses/ and http://www.undmtp.org/modules.htm). The training material covers many topics in a general manner, however, no particular mention is made of any spatial prediction methods even in modules dealing with hazard and vulnerability.

In a very recent volume by the Committee on Planning for Catastrophe (NRC, 2007), entitled "Successful Response Starts with a Map" a short section is dedicated to tools for data exploitation. Geospatial tools for visualization, analysis and decision support for emergency management and disaster research are considered briefly to emphasize the existing impediments for managers to take advantage of them. Examples are mentioned of modeling ground shaking, volcano ash fall, evacuation simulations and the use of geospatial models in forest fire fighting. Three main issues, however, are brought up as instances of serious reasons of concern:

1. only very general estimates at regional scales are commonly made with damages expressed as the probability of the buildings being in one of four damage states: slight, moderate, extensive and total;
2. even if a "common denominator" set of tools were to be available, needy users generally fail to take advantage of their capabilities;
3. geospatial information need to ensure that the underlying assumptions, data quality and usefulness are [understood and] conveyed properly.

In that volume, recommendation to academia is made to offer training in geospatial tools, to FEMA (the Federal Emergency Management Agency) is made to employ geospatial professionals, and to DHS is made to link up with qualified geospatial professionals. Clearly, even if reference is made to an earlier NRC (2006) report on technologies and methods for disaster research, in that report no space is reserved for spatial prediction modeling! The tools, as such in any case, have hardly enough intellectual content for training. One has to wonder then, who, where and how such research can be performed that would lead to risk maps that show the likelihood of what is to happen in a reasonably near future. Isn't that provocative now?

Concluding remarks

I have exposed my views and experience on geo-information in mapping and the consequent responsibility of predicting hazard and risk. I have quoted some reputable views on (i) the societal benefits of mastering risk, (ii) on the probability theory to estimate the likelihood of hazardous events for a "cartography of danger." Also, (iii) simulations and scenarios for hazard mitigation and (iv) emergency decision making are strategic steps to compose geo-information overlays. In addition there are the views that (v) judges and people at large unfortunately have serious difficulties in comprehending risk and risk maps so that (vi) they require education in risk matters (that appear too technical to them). Even a World Risk Index (vii) implies potential consequences and requires people to predict the probability of events.

Unfortunately, most "risk maps" (viii) do not seem to expose how the spatial data have been related by mathematical and conceptual models, assumptions, and the uncertainty levels. Examples of a landslide risk map (ix) and of a spatial database for training in predictive modeling (x) were provided.

Where are we now? We have observed a serious neglect of the analytical theory at the base of predictive models. Most of the emphasis that we notice is on technology and tools as such. But what is innovation, really? That is certainly not "having new or more tools", it is rather mastering ways to problem solving and spatial-temporal understanding and anticipating. I continue my provocation also in concluding!

But I wish to mention a number of new initiatives of the Spatial Information Laboratory, SPIN*lab*, with which I have been involved since its foundation in 2002. The

UNIGIS Program has initiated a course module on the "Ongoing Debate on the Definition of GIS. The basic issue is whether we should speak of **GI-System**, a technology tool, of **GI-Science**, a scientific discipline [or endeavor], or **GI-Services**, an information provider. A series of short papers on a variety of relevant applications and problem areas are available on diverse topics such as: the value of GIS for science, GIS and archeology, marine biology and epidemiology (http://www.spinlab.vu.nl/). Furthermore, an International Workshop is planned for next September 25-26, in Amsterdam, on "**Geo-ICT and Location-based Science.**" There the key goals are to analyze the significance, effectiveness and methodological obstacles to optimal use of Geo-ICT in location-based approaches (http://gisinscience.blogspot.com/)

In addition, the following three activities are worth mentioning: (1) a set of new lectures in Earth Sciences and Economics on the topic of Natural Hazards, Risks and Vulnerability; (2) a research project on Geographical Data Infrastructure for Disaster Management, GDI4DM, that has developed an open-standard generic platform GDI for administration and analysis to predict the development of crisis events and to advise on consequent change in plans. This is to be tested next September in collaboration with police, fire department and medical support (http://www.gdi4dm.nl/); and (3) the EDUGIS Educational Portal for the young that offers a GIS course module to explore spatial data relevance in the Netherlands (http://www.edugis.nl). I am confident that the predictive modeling challenges just discussed will be raised and developed further within those timely initiatives.

Our training of young scientists and our research focus indeed must be on: constructing solutions to risk management. Our task is not light when we are facing the "unbearable lightness" of GIS! Can we train young researchers on spatial prediction models and not just on using some off-the-shelf GIS products? Who is the modeler? Who is the decision-maker? Do they relate? What is the role of a professor of Geo-information and Risk Management then? Shouldn't he/she cover the entire disaster management cycle from prevention/mitigation, preparedness, alert, response, recovery and post disaster measures? Wouldn't we be better off when prioritizing prevention and preparedness? In other words, concentrating first on prediction and planning? A **risk manager** avoids and mitigates; a **disaster manager** monitors and rescues. Who of the two would you rather be?

Personally, I have been involved in risk related activities at the SPIN*lab*, contributing to the organization of a series of symposia on Geo-information for Disaster Management, Gi4DM, two of which that have taken place in Delft (2005), and in Goa, India (2006), and two more that are to take place in Toronto (2007) and finally in Beijing (2008). I had the pleasure of lecturing for UNIGIS workshops, of supervising a Ph.D. candidate on risk analysis and management in India, and of contributing to the SPIN*lab*'s activities mentioned earlier. A research agenda is to ensure that methods of spatial predictive modeling of hazard and risk are applied properly to obtain credible patterns with associated usable uncertainty.

In conclusion, the problem is that all-the-tech-in-the-world cannot substitute for process insight, model conceptualization and application understanding. The representation of the distribution of past events is hardly of guidance in risk management. We need also representations of future events.

Ah! …The unbearable lightness of GIS!

I would like to thank the Rector of the VU for the opportunity of this "*inaugurele rede*" and the audience as a whole for still being with me and waiting for the reception that is to follow next.

References

Alexander D., 2002, *Principles of Emergency Planning and Management*. Oxford, Oxford University Press, 340 p.

Bernstein P.L., 1996, Against the Gods. The remarkable story of risk. New York, John Wiley and Sons. Ltd., 383 p.

Chung, C.F. and Fabbri, A.G., 1993, Representation of geoscience data for information integration. Jour of Non-renewable Resources, v.2, n. 2, p. 122-139.

Chung C.F. and Fabbri A.G., 2001: Prediction models for landslide hazard using a fuzzy set approach. In, M. Marchetti and V. Rivas, eds., Geomorphology and Environmental impact Assessment, Balkema, Rotterdam, 31-47.

Chung C.F. and Fabbri A.G., 2004: Systematic procedures of landslide hazard mapping for risk assessment using spatial prediction models. In, Glade T., Anderson M.G., and Crozier M.J., eds., *Landslide Hazard and Risk*. New York, John Wiley & Sons, 139-174.

Chung, C.F., Fabbri, A. G., Jang, D. H., and Scholten, H. J., 2005: Risk assessment using spatial prediction model for natural disaster preparedness. In, van Oosterom P., Zlatanova S., and Fendel E.M., Geo-informastion for Disaster Management. Berlin, Springer, 619-640. Procs. of Gi4DM, The First Symposium on Geo-information for Disaster Management, Delft, Netherlands, March 21-23, 2005.

Clegg P., Bruciatelli L., Domingos F., Jones R.R., De Donatis M., and Wilson R.W., 2006, Digital geological mapping with tablet PC and PDA: a comparison. Computers & Geosciences, v. 32, n. 10, p. 1682-1698.

Coe J.A., Michael J.A., Crovelli R.A., Savage W.Z., Laprade W.T. and Nashem W.D., 2004, Probabilistic assessment of precipitation-triggered landslides using historical records of landslide occurrence, Seattle, Washington. Environmental Engineering Geoscience, 10/2, 103-122.

Fabbri A.G., Chung C.F. and Jang D.H., 2005, Spatial visualization of risk: a challenge for its management. In, Q. Cheng and G. Bonham-Carter, eds., GIS and Spatial Analysis, Vol. 1, Proceedings of IAMG '05, Toronto, Canada, Aug. 21-26,2005, p. 346-351.

Fabbri A.G., Chung C.F. and Jang D.H., 2004, A software approach to spatial predictions if natural hazards and consequent risks. In, Brebbia C.A., ed., *Risk Analysis IV*, Southampton, Boston, WIT Press, 289-305.

Kundera M., 1984, *The Unbearable Lightness of Being* (Nesmesitna lehkost byti), 1995 Edition, Faber and Faber Limited, London, 305 p. Harper & Row Publishers Inc., New York. (Translated from the Czech by Michael Henry Heim).

Monmonnier M. S., 1997: *Cartographies of Danger: Mapping Hazards in America*. Chicago, University of Chicago Press, 363 pp.

NRC (National Research Council), 2006, Facing Hazards and Disaters: Understanding Human Dimensions. Washington, D.C.: The National Academic Press.

NRC (National Research Council), 2007, Successful Response Starts wit a Map: Improving Geospatial Support for Disaster Management. Washington D.C., The National Academies Press, 184 p. (http://www.nap.edu/catalog/11793.html

Plattner, T., 2004: An integrative model of natural hazard risk evaluation. In Brebbia C. A., ed., *Risk Analysis IV*, Southampton, Boston, WIT Press, 649-658.

Plum M.M. and Beitel G.A., 2006, Rationale and development of a security assurance index with application toward the development of a World Risk Index. In, Risk Analysis V, WIT Press, Southampton, UK, p. 325-340 (http://www.witpress.com/).

Posner Richard A., 2004, *Catastrophe, Risk and Response*. New York, Oxford University Press Inc., 334 p.

Remondo J, Bonachea J., and Cendrero A., 2005, A statistical approach to landslide risk modeling at basin scale: from landslide susceptibility to quantitative risk assessment. Landslides, v. 2, n. 4, p. 321-328.

Remondo, J., Gonzalez, A., Diaz de Teran, J.R., Cendrero, A, Fabbri, A.G., and Chung, C.F., 2003, Validation of landslide susceptibility maps: examples and applications from a case study in Northern Spain. Natural Hazards,v. 30, p. 437-449.

Schetselaar E.M., Harrap R., and Brodaric B., 2006, Digital capture of geological field data: bridging the gap between a mapping heritage and the integrated dissemination of geoscience information. In, GIS for the Earth Sciences, Geological Association of Canada, Special Publication 44, Edited by Jeff Harris, Chapter 21, p.469-480.

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, v. 4, p. 133-146.