

# Combining seismic and geologic data for reproducible Seismic source models



**B. Le Goff & D.D. Fitzenz**

*Centro de Geofísica de Évora, University of Évora, Évora, Portugal*

## **SUMMARY:**

An important step in the Probabilistic Seismic Hazard Analysis (PSHA) consists in defining the seismic source model. Given that most of the faults, in low seismic regions, are not characterized well enough, the source models are defined as areal zones, delimited with finite boundary polygons, within which the seismicity and the geological features are deemed homogeneous, leading to different problems.

We investigate an alternative approach, using Bayesian methods, to model the seismotectonic zoning, with two main objectives: 1) obtain a reproducible method that 2) preserves the information on the sources and extent of the uncertainties. We start with two zones, characterized by two different surface activity rates. The inference of this model allows us to recover the geographical limit between this zones. Future work will strive to incorporate all available data, such as structural data, and to extend this model to  $n$  zones with a unspecified shape.

*Keywords: Seismotectonic zoning, Probabilistic Seismic Hazard Assessment, Bayesian methods*

## **1. INTRODUCTION**

Several studies have shown the determinant impact of seismotectonic zoning in Probabilistic Seismic Hazard Assessment (PSHA) (Beauval, 2003; Beauval and Scotti, 2004; Bender, 1986; Le Goff et al., 2009; Woo, 1996). The seismotectonic zoning allows to link the seismicity with the tectonically-active geological structures, in order to define sources for use in PSHA computation. Usually, and because faults are often not characterized well-enough, source zones are defined as surfaces and modeled as polygons. In that case, they are delimited with fixed, infinitely thin boundaries. In each zone, the geological expression of active tectonics and the seismicity are deemed homogeneous (e.g., focal depths and mechanisms, seismicity rate, and maximum magnitude), and each point of a zone is considered an equally likely source of earthquake.

Besides the lack of data (e.g., short period of instrumental observation of small events, short catalog of large events, blind faults), the establishment of a traditional seismotectonic zoning generates different shortcomings. The finite boundaries of the different zones are set by expert decisions, leading to different problems : 1) the superposition of a resulting hazard map with areal source zoning model underlines the large sensitivity of the results to this method (Beauval, 2003); 2) it is not reproducible: different experts come up with different zonings using the same input data. 3) the final seismotectonic zoning is not provided with error maps reflecting the original density of information used for both the assessment of the common characteristics and the calculation of seismicity rates of each zone.

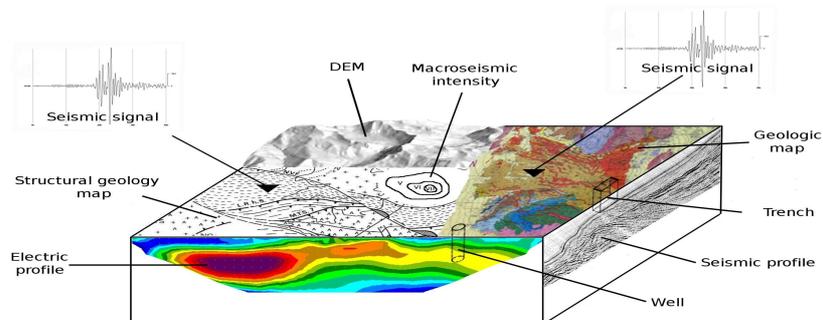
Some approaches, such as the K-means method for region partitioning (Weatherhill and Burton, 2009), strive to resolve these shortcomings but are still not satisfying, because of their lack of robustness (problems determining the optimal number of clusters and the initial cluster centers and

may lead to a local optimum instead of a global optimum partition) and the fact that, at best, only the seismic catalogs and the focal mechanisms are used. In order to avoid the abrupt change in the seismicity rate at source boundaries, observable in the resulting hazard map, Bender (1986) proposes to use standard "hard" seismotectonic zoning but to provide smoothing using the epicenter location uncertainty, considered normally distributed. A point source will contribute to the seismicity rate calculation of several zones, according to its epicentral location uncertainty. In that paper, the assumption on the location uncertainty characterization, arbitrary defined by 0, 10, 25 and 50 kilometers, is questionable. Each event has its own location uncertainty, depending on its position related to the seismic network and its date of occurrence (improvement of seismic station accuracy and network density).

This paper aims to propose an exploratory study for alternative procedures in area source modeling. We search to obtain a method which will be robust and reproducible. In this way, we strive to combine different data using Bayesian methods. We start to develop a generative model with two zones, characterized by two different surface activity rates, creating synthetic catalogs drawn from a Poisson distribution as occurrence model, a truncated Gutenberg-Richter law as magnitude-frequency relationship and a uniform spatial distribution. The inference of this model allows us to assess the minimum number of data,  $n_{min}$ , required in an earthquake catalog to recover the activity rates of both zones and the geographical limit between them, with some confidence. In the first sections, we see the data used to develop a seismotectonic zoning (section 2) and its use in the Probabilistic Seismic Hazard Assessment (PSHA) (section 3). Then, we describe the method developed to obtain the limit between the two zones (section 4) and present some conclusion. Future improvements of this method will strive to reduce the minimum number of data,  $n_{min}$  and to extend the model to  $n$  zones of polygonal shape (section 5).

## 2. INPUT DATA

Different data sets are traditionally used in order to establish a seismotectonic zoning. These data are of different kinds and are characterized by different dimensions. The structural geology map contains much information about faults (e.g., fault orientations, fault geometry, fault type, tectonic ensembles) and has an important role in the development of seismotectonic zoning. Two different seismicity catalogs are used in seismotectonic zoning computations : historical and instrumental. The idea is 1) to locate the hypocenters and 2) to compute the focal mechanisms and magnitudes to try and image active faults and characterize the type of faulting (strike-slip, or dip-slip faulting). These catalogs contain uncertainties, notably in hypocenter locations and magnitude calculations, and the systematic errors (round-off). Because the evolution of seismic networks is recent the period of instrumental observation of small events is short. Other data are useful in the conception of seismotectonic zoning in order to have a better constrained model. Indirectly, geological maps, expressing the age of the formations, trenches, geophysical data and digital elevation model (DEM) allow to improve the location of faults. An example of such composite data sets is shown in figure 1.



**Figure 1.** 3D diagram representing the useful data for using in PSHA. In this figure, the segmentation is done to show the kinds of data and do not express a seismotectonic zoning

In conclusion, we have 2D maps with features such as 1D fault lines, or dip angles; 3D points (hypocenters resulting from the inversion of seismograms), points on a surface with intensity data; borehole or seismic transects intersecting faults.

### **3. PSHA**

As reminded by an INGV open letter, the best approach to protect population and building from collapsing is not through earthquake prediction but through the application of appropriate safety measures. The development of seismic hazard maps provides the specifications required by building codes to avoid collapse of buildings and the resulting fatalities, and the information to convey to the population the basic concepts of earthquake hazard, awareness, preparedness and response.

There are several approaches to assess seismic hazard in a probabilistic sense. The most widely used is the Cornell-McGuire approach (Cornell, 1968; McGuire, 1976). It consists in estimating the probability of exceedance of a ground-motion target level at a site, over a given time window. Usually, this target level is described by the Peak Ground Acceleration (PGA) or the Peak Ground Velocity (PGV). Considering a Poissonian model for earthquake occurrence, it is common to refer to the return periods (inverse of the annual rate) instead of the annual rates. According to the application domain, these return periods range from 100 to  $10^7$  years.

A number of steps need to be completed on the seismicity data before the actual PSHA can take place (Le Goff and al., 2009). First, the seismic catalogs have to be homogenized in a common magnitude scale. Second, the seismicity is usually modeled as a succession of independent events (with a Poissonian distribution of inter-event times), meaning that there is neither foreshocks, aftershocks, nor triggered events. The catalogs have to be filtered. Third, from all available seismic events and seismotectonic data, a seismotectonic zoning is achieved in order to identify and characterize the source of seismicity. A ground motion prediction model is used to estimate exceedance probability of a target level and for a given magnitude/distance. Finally, the exceedance probability contributions of all couples magnitude/distance are summed up.

Beyond random (aleatory) uncertainties, intrinsic to data (e.g., catalog uncertainties), epistemic uncertainties are generated when the choice of different parameters or models is done. Beauval (2003, 2004) has demonstrated that both the truncation of the predicted ground-motion model and the choice of the magnitude-intensity correlation are dominant with respect to the level of hazard inferred. In terms of spatial distribution of hazard, the impact of the location of seismotectonic zoning boundaries is what matters most. Usually, these epistemic uncertainties are accounted for using a logic tree approach where each branch is weighted by an expert or a panel of experts.

The choice to focus this study on the seismotectonic zoning is due to the impact of its source zone boundaries into the hazard map. Moreover, the current seismotectonic zoning model does not allow to represent any variation in faulting mechanism. The assessment of seismotectonic zoning is not reproducible. Indeed, different experts, using same data, provide different zoning, based on their differing interpretation.

## **4. 2-ZONE INFERENCE**

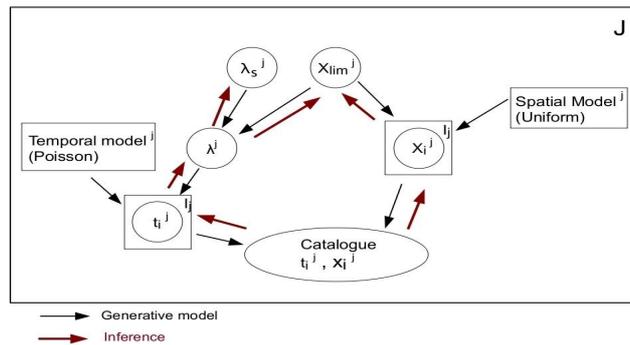
### **4.1. Introduction**

The seismotectonic zoning is generally used in regions where the seismicity is low to moderate or where most of the seismogenic structures are not known. However this approach presents some limitations. For example, the expert decisions used to achieve this zoning are not traceable back to the data that dominated the decision-making. Moreover, no information on the uncertainties are provided. The objective of our model is to use a suite of data sets, from different domains, including tectonic and

geologic information. We choose the Bayesian methods in order to define the generative model, using, first, the earthquake locations and time of occurrence. This model can be inverted (inference) to obtain the posterior probability of the target variables, as the geographic limit between zones or the characteristic of each zone. This posterior probability may be updated when new data is available.

## 4.2. Graphical model

The graphical model constitutes a guide to define the joint probability. In such a generative model, each node represents a random variable, parameter of our study. An arrow pointing from a node to another expresses the dependency between these two nodes, expressed as a conditional probability density function. The black arrows show the generative model or direct model while the red ones express the inference model (Figure 2).



**Figure 2.** Graphical model for generating a synthetic catalog. Black arrows represent the causal relationships between the different parameters of the forward model. The red arrow represents the inference, which allows to estimate the limit between geographical zones and their seismic activity rates.

For the  $j^{\text{th}}$  zone, the seismicity rate is generated from the surface seismicity rate and the size of the zone, which depends on the limit between the two zones. Then, this seismicity rate is coupled to a Poisson model to determine the different times of events. The seismicity rate and the observational period allow to define the number of earthquakes,  $I$ , in the zone  $j$ . The limit between the two zones is also used to obtain the different locations of the  $I^j$  events, drawn in a uniform spatial distribution. Both times and locations are used to generate the synthetic catalog. The inference provides the joint probability of the model and, after marginalization, provides the optimal value for the location of the limit between the two zones and the surface seismicity rate in each zone.

## 4.3. Synthetic catalog

A synthetic catalog is generated in order to control the input parameters of the model and to have an idea on the expected results. A series of synthetic catalogs was generated in order to test the resolving power of the method, in a controlled case, where the input parameters, the temporal occurrence and spatial occurrence models are known.

### 4.3.1. Temporal repartition of the events

Even though the Poisson model is often discussed for individual faults (e.g., Zöller et al., 2007; Kuehn et al., 2008; Fitzenz et al., 2010), this distribution is generally considered to model the distribution of inter-event times for earthquakes in a large region, regardless of their magnitude. This model implies that the probability of occurrence of an earthquake does not depend on the elapsed time since the last earthquake. The events are then considered independent of one another, i.e., that there is neither foreshocks, aftershocks, nor triggered events.

The cumulative distribution function of the Poisson law, expressing the probability to have at least one event in the time interval  $t$ , is defined as follow:

$$F(\Delta t) = 1 - e^{-\lambda \Delta t} \quad (4.1)$$

$t$  is the time interval between two events;  $\lambda$  is the seismic activity rate.

The seismic activity rate,  $\lambda$ , is generated from the surface seismicity rate of the  $j^{\text{th}}$  zone and the limit between the 2 zones. For the  $j^{\text{th}}$  zone, the sampling of the Poisson law provides the times of the different events,  $t_i^j$ , and the number of events,  $I^j$ , during the given observational time  $T_{\text{obs}}$  and according to the seismic activity rate,  $\lambda^j$ .

#### 4.3.2. Spatial repartition of the events

In PSHA, the uniform spatial model is usually used into a source zone. This source zones are then considered as homogeneous, implying that an event may occur anywhere within the zone. This assumption implies that a part of a zone, with a low seismic activity in reality, may be modeled as allowing the presence of a large number of earthquakes with large magnitudes (Musson, 2004). Moreover, with few data, some events may appear to line-up and be considered as tectonic structures. In such a case, it is difficult to differentiate a zone with a sparse seismicity (background seismicity) from a zone with a seismicity associated to tectonic structures. However, to begin with a simple case, the uniform spatial model was simulated.

#### 4.3.3. Magnitude-frequency relationship

The chosen frequency-magnitude model is the truncated exponential model. It expresses the fact that the proportion of larger earthquakes compared to smaller ones is linear over the whole size range encountered in a region, until a roll-over for magnitudes close to the maximum magnitude. This roll-over is used to prevent the maximum magnitude from becoming infinitely large as the time window considered increases. It provides, for a given magnitude  $M$ , the annual number of event,  $\lambda$ , of magnitude larger or equal to  $M$ . This relationship may be express as follow:

$$\lambda = \lambda_{\min} \cdot \frac{e^{-\beta(M - M_{\min})} - e^{-\beta(M_{\max} - M_{\min})}}{1 - e^{-\beta(M_{\max} - M_{\min})}} \quad (4.2)$$

$\lambda$ : annual number of events with magnitude larger than  $M$ ;  $\lambda_{\min}$ : annual number of events with magnitude larger than  $M_{\min}$ ;  $M_{\min}$ : minimal magnitude considered on the study;  $M_{\max}$ : maximal magnitude considered possible in the zone;  $\beta$ : coefficient of exponential decrease

The different magnitudes are generated from the seismic activity rate,  $\lambda$ , and attributed to each events  $i$  of the  $J$  zones. According to the sampling, the samples may present a more or less good fit with the theoretical magnitude-frequency law, and the extrapolation of the  $b$ -values and the maximum magnitude from a catalog realization may lead to a large variability (Page, 2011).

#### 4.3.4. Inference

The Bayesian inference is a probabilistic method which consists in calculating the plausibility of a hypothesis. Its computation is derived from the Bayes' theorem. In the Bayesian sense, a probability may be interpreted as a numerical translation of a degree of belief. In this case study, the inference allows, from the synthetical catalog, to recover the pdf for parameters of interest. A first step is the definition of priors, expressing the degree of belief about a random variable before taking into account the data. Then the method consists in evaluating the posterior probability of the model. According to Bayes' theorem, the posterior probability is proportional to the joint probability and one can determine the optimal value of the parameters and also the covariance matrix. The joint probability expresses the relationship between all the elements of the model. If the number of zones,  $J$ , is fixed, the joint probability may be expressed as follow:

$$P_j = P(\lambda_j^S, x_{\text{lim}}^j, \lambda^j, t_i^j, \vec{x}_i^j, \text{Spat.Mod}^j, \text{Temp.Mod}^j) \quad (4.3)$$

$$P_j = \prod_{j=1}^J P(\lambda_S^j) \cdot P(x_{lim}^j) \cdot P(\lambda^j / \lambda_S^j, x_{lim}^j) \cdot P(t_i^j / t_{i-1}^j, \lambda^j, Temp.Mod^j) \cdot P(x_i^j / x_{lim}^j, Spat.Mod^j) \cdot P(y_i^j / Spat.Mod^j) \quad (4.4)$$

In our case J=2, so:

$$Priors = P(\lambda_S^1) \cdot P(\lambda_S^2) \cdot P(x_{lim.}) \quad (4.7)$$

Because there are more nodes in the graphical model than the parameters and observations, we have to integrate the complete joint pdf with respect to all the intermediate nodes. This step is called the marginalization step.

Marginalization with respect to  $\lambda$ :

$$P(\lambda^j / \lambda_S^j, x_{lim}^j) = \delta(\lambda^j - (\lambda_S^j \cdot x_{lim}^j \cdot Ly_{max})) \quad (4.5)$$

$$P(\lambda_i^j, x_{lim}^j, t_i^j, \vec{x}_i^j, Spat.Mod^j, Temp.Mod^j) = \int P_j \cdot d\lambda^j \quad (4.6)$$

and we obtain:

$$P(\lambda_S^1, \lambda_S^2, x_{lim.}, \vec{x}_1^1, \vec{x}_2^2, t_1^1, t_2^2) = Priors \cdot (\lambda_S^1)^{I_1} \cdot e^{(-\lambda_S^1 \cdot x_{lim.} \cdot Ly_{max} \cdot \sum_{i=0}^{I_1} (t_{i+1}^1 - t_i^1))} \cdot (\lambda_S^2)^{I_2} \cdot e^{(-\lambda_S^2 (Lx_{max} - x_{lim.}) \cdot Ly_{max} \cdot \sum_{i=0}^{I_2} (t_{i+1}^2 - t_i^2))} \quad (4.7)$$

Marginalization with respect of  $\lambda_S^1$  and  $\lambda_S^2$  :

$$P_m = Priors \cdot (\lambda_S^1)^{I_1} \cdot (\lambda_S^2)^{I_2} \cdot e^{-\lambda_S^2 \cdot Ly_{max} \cdot Lx_{max} \cdot \sum_{i=1}^{I_2} (t_{i+1}^2 - t_i^2)} \cdot \frac{1}{a} \cdot e^{-a \cdot Lx_{max}} + \frac{1}{a} \quad (4.8)$$

with:  $a = \lambda_S^1 \cdot Ly_{max} \cdot \sum_{i=1}^{I_1} (t_{i+1}^1 - t_i^1) - \lambda_S^2 \cdot Ly_{max} \cdot \sum_{i=1}^{I_2} (t_{i+1}^2 - t_i^2)$

#### 4.3.5. First results

In order to obtain a clearer representation, the results are expressed as the energy, corresponding to  $-\log(P_j)$ . The optimal value of the inference corresponds to the minimum of the energy function and represents the optimal value of the considered parameter. The different equations are then expressed as follow:

$$U_{P_j} = -\log(Priors) - I_1 \cdot \log(\lambda_S^1) + \lambda_S^1 \cdot x_{lim.} \cdot Ly_{max} \cdot \sum_{i=0}^{I_1} (t_{i+1}^1 - t_i^1) - I_2 \cdot \log(\lambda_S^2) + \lambda_S^2 \cdot (Lx_{max} - x_{lim.}) \cdot Ly_{max} \cdot \sum_{i=0}^{I_2} (t_{i+1}^2 - t_i^2) \quad (4.9)$$

$$U_{P_m / \lambda_S^1 \lambda_S^2} = -\log(Priors) + I_{1,\log}(K_1 \cdot \sum_{i=0}^{I_1} (t_{i+1}^1 - t_i^1)) - I_1 \cdot \log(I_1) + I_1 + I_2 \cdot \log(K_2 \cdot \sum_{i=0}^{I_2} (t_{i+1}^2 - t_i^2)) - I_2 \cdot \log(I_2) + I_2 \quad (4.10)$$

$$\text{with } K_1 = x_{lim.} \cdot Ly_{max} \text{ and } K_2 = (Lx_{max} - x_{lim.}) \cdot Ly_{max}$$

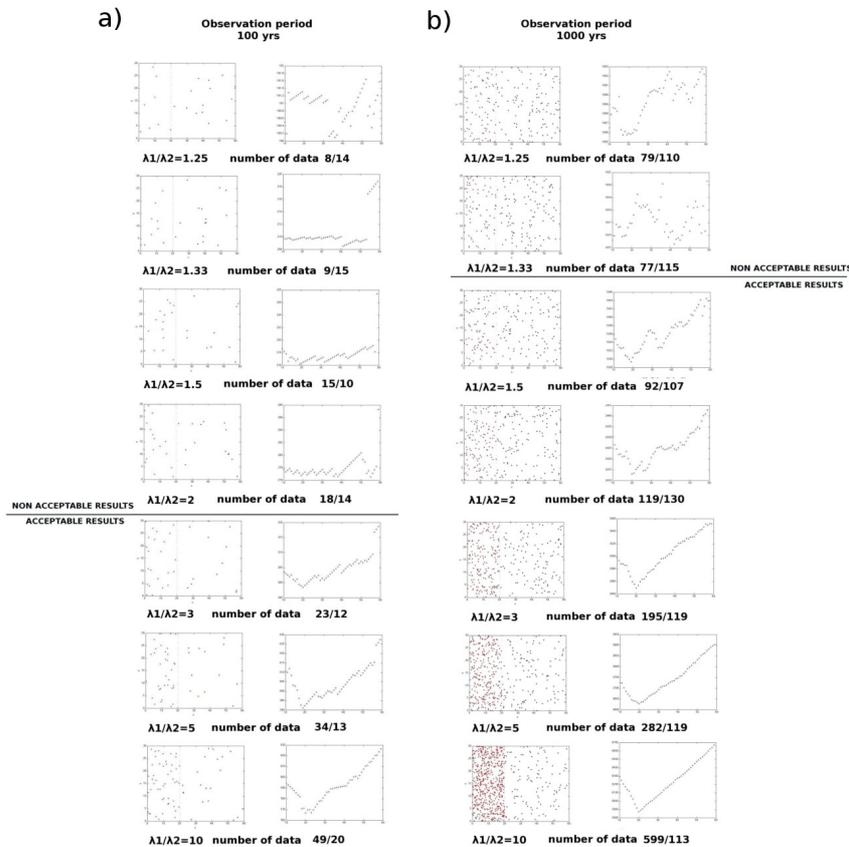
The first step consists in computing the posterior probability. Then, the Bayesian inference is used to obtain the marginal probability and to recover the limit between the two zones. These two zones are separated by a contrast in their spatial seismicity density. Several cases with different ratios between

the surface seismicity rates and different observational periods were tested to observe the behavior of the model with the number of data.

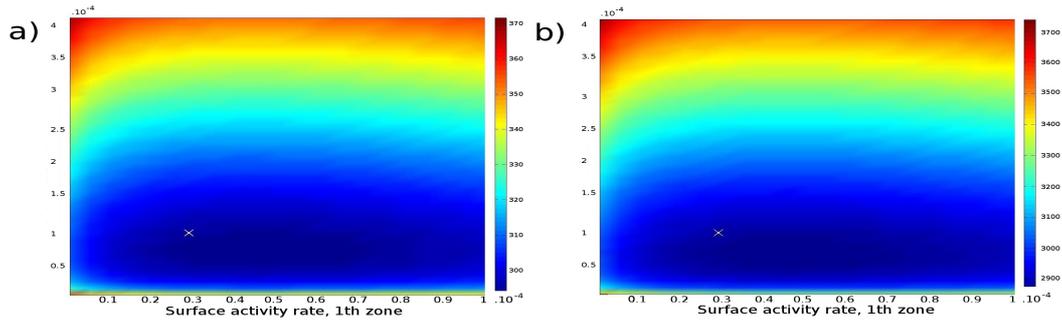
The first example (Figure 3a) is a representation of the 2 zones, for an observational period of 100 years. The limit between the two zones was placed at 20 kilometers, before generating the catalog. For these two zones (left part of the Figure 3a and 3b), we calculated the joint probability and plotted the energy function (right part of the Figure 3a and 3b). Depending on the number of data, we can obtain a minimum value for the energy function, expressing the optimal value for the limit between the two zones. We can see, on this example, that the ratio between the surface seismicity rates of the 2 zones has to be larger than 3, if we want to reach an optimal value of the limit. Some local minima may appear with a few number of data, and it is then necessary to achieve several realizations (Section variability). On the second example (Figure 3b), the limit was placed on the same location (20 kilometers), but we considered an observational period of 1000 years. The number of data points is then higher. On this example and because of the higher number of data, we can reach an acceptable solution, defined by an identifiable absolute minimum, with a ratio lower than the first example. Here a minimum value of the energy function may be reached for a ratio of 1.5.

Since we obtain an optimal value for the limit, we can evaluate both surface seismicity rates (Figure 4). The catalog was generated for an observational period of 1000 years and with a limit between the two zones at 20 kilometers. On these examples, the surface seismicity rates are  $3 \cdot 10^{-4}$  for the first zone and  $1 \cdot 10^{-4}$  for the second one, for the first case (Figure 4a), and  $4 \cdot 10^{-4}$  for the first zone and  $1,5 \cdot 10^{-4}$  for the second one, in the second example (Figure 4b).

We observe that it is possible to recover the surface seismicity rates used to generate the catalog, with a quantitative measure of the uncertainty (including the covariance).



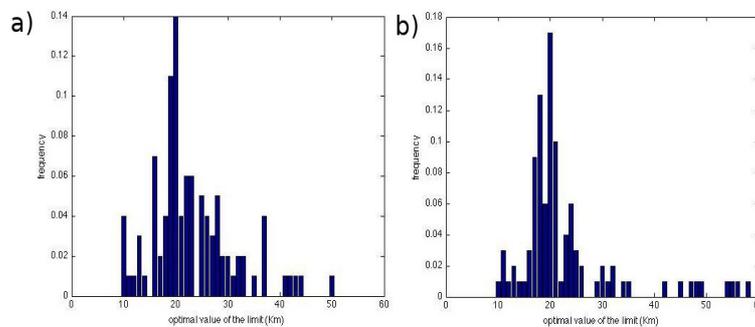
**Figure 3.** Results of the Bayesian inference to recover the geographical limit between the two zones, with different ratio between the surface seismicity rates and for an observational period of  
a) 100 years and b) 1000 years.



**Figure 4.** Results of the Bayesian inference to evaluate the surface seismicity rate of the two zones. The colour variation from the red to the blue represents the decrease of the energy function. The white dots express the true values. a) the synthetical catalog was drawn for an observational period of a) 100 years, with the surface seismicity rates of  $3.10^{-4}$  for the first zone and  $1.10^{-4}$  for the second zone (optimal values around  $3,5.10^{-4}$  for the first zone and  $1,25.10^{-4}$  for the second one. b) for an observational period of 1000 years, with the surface seismicity rates of  $4.10^{-4}$  for the first zone and  $1,5.10^{-4}$  for the second zone.(optimal values around  $4.10^{-4}$  for the first zone and  $1,6.10^{-4}$  for the second one)

#### 4.3.6. Variability

According to the realization of the catalog, the optimal value may not represent the real solution. It is then important to compare the distribution of the optimal values we can obtain. The following histograms (figure 5) express the optimal values of the limit between the two zones, for 100 realizations of catalogs. In the first example (Figure 5a), the 100 synthetical catalogs were drawn for an observational period of 100 years and with the surface seismicity rates of  $3.10^{-4}$  for the first zone and  $1.10^{-4}$  for the second one. In the second example (Figure 5b), the 100 catalogs were drawn for an observational period of 1000 years with the surface seismicity rates of  $1,5.10^{-4}$  for the first zone and  $1.10^{-4}$  for the second one.



**Figure 7.** Distribution of optimal values for the limit from 100 synthetic catalogs, drawn a) for an observational period of 100 years and with the surface seismicity rates of  $3.10^{-4}$  for the first zone and  $1.10^{-4}$  for the second one and b) for an observational period of 1000 years and with the surface seismicity rates of  $1,5.10^{-4}$  for the first zone and  $1.10^{-4}$  for the second one

These figures show the importance of using different realization of synthetic catalogs. Indeed, with only one synthetic catalog, the optimal value for the limit may be on the tail of the distribution. With the distribution shown on the figure 8b, the use of Monte-Carlo draws, with an initial value on the right of the distribution, may lead to a local minimum, around 23 Km. In terms of seismic hazard, this variability is important. Indeed, a city may be affected to a zone or another, changing its associated seismic hazard. Another reason for the importance of this variability is the association of events into the seismotectonic zones. Because of the low to moderate seismicity (implying the use of seismotectonic zoning), a major event associated to a zone or another may have a large impact on the calculation of the surface seismicity rates.

## 5. DISCUSSION AND FUTURE IMPROVEMENTS

During about 20 years, the evolution of PSHA in current practice has been slowed down by using paying software within which code source is not known and shared. But recently, initiatives have been undertaken, first in California with openSHA and after in international-wide with GEM, to propose an open source code, making it easier to incorporate new methodologies. The methodology, presented in this paper, could therefore be inserted in the hazard calculation process.

Previous assessments of epistemic uncertainties rely on weights given the different models by a panel of experts. In contrast, we propose another role for expert panels: the choice of the datasets and the priors to be used and the candidate models to be tested. Once these steps are done, the quantitative computation of the weights of the models can be performed, and is reproducible. With our contribution, we show how Bayesian Inference would be useful in Probabilistic Seismic Hazard Assessment. Other authors have pointed that out in the recent past, for renewal models (Biasi and Weldon, 2008; Fitzenz et al., 2012, Fitzenz et al, 2010, Fitzenz and al, 2007), or for a general view (Esmer, 2006).

The 2-zone-inference model allows us to determine the limit between 2 zones, characterized by a contrast of surface seismicity rate. It is then possible to evaluate the minimal number of data,  $n_{\min}$ , required to obtain an observable minimal of the energy function. Since we want to investigate seismic hazard at a local scale, where earthquakes of magnitudes larger than 5 can happen, we need to go beyond the point-source approach. Furthermore, we want to integrate geological and tectonic information into our zoning to increase its resolving power (in particular allow a good inference of the parameters even with few earthquakes). The maximal fault length, that depends on fault preferential orientation and the limit between the 2 zones, may be use to define the maximal magnitude in the draws of the magnitude from the Gutenberg-Richter law. However, the large variability in magnitude, drawn from any given Gutenberg-Richter, may prevent any conclusive result from being reached. Then, it is fundamental to go beyond this two zone case and to extend this model to  $n$  zones with a unspecified shape. To complete this model, all available data should be incorporated.

The smoothing method of Bender (1986) may be adapted in order to calculate the seismicity rates. In our method, the uncertainty, linked to the limit between the zones instead of arbitrary defined for all events, will be used to define the contribution of this source to several zones. The fact that a source, close to a seismotectonic zoning boundary, may participate in the calculation of several seismicity rates allows to smooth these rates in source zone boundaries. Thus, in proximity to a source zone boundary, resulting acceleration levels for two close sites may not differ considerably.

## 6. CONCLUSION

The objective of this approach is to model a seismotectonic zoning which 1) is reproducible and 2) preserves the information on the source and extent to the uncertainties, so as to allow to propagate them and issue recommendations for optimized future data acquisitions. An inference with two zones, differentiated by two different surface seismicity rates, was performed to obtain the geographic limits between them. To obtain an acceptable accuracy on the location of the limit between the 2 zones, the ratio of the surface activity rates has to be larger than 3, for an observational period of 100 years. Considering an observational period of 1000 years, this ratio fall down to 1.5, because of the higher number of data. This 2 zone model will be a reference in the comparison with other models, which will incorporate other available data. Future improvements will integrate the preferential orientations of faults and the geology as well. Also, we will consider a number  $n$  of zones with a unspecified shape. We emphasize that such an approach is reproducible once priors and data sets are chosen. Indeed, we will strive to incorporate expert opinions as priors, and avoid using expert decisions. Instead, the products will be directly the result of the inference, when only one model is considered, or the the result of a combination of models in the Bayesian sense.

## ACKNOWLEDGEMENT

This research and BLG are funded by the Fundação para a Ciência e a Tecnologia (PTDC/CTEGIX/101852/2008 and FCOMP-01-0124- FEDER-009326). DF benefits from the FCT Ciência 2007 program. We thank André Jalobeanu for his guidance in data partitioning and Bayesian methodology.

## REFERENCES

- Beauval, C. (2003). Analyse des incertitudes dans une estimation probabiliste de l'aléa sismique, exemple de la France. *PhD Thesis*
- Beauval, C. and Scotti, O. (2004). Quantifying Sensitivities of PSHA for France to earthquake catalogue uncertainties, truncation of ground-motion variability, and magnitude limits. *Bull. seism. Soc. Am.* **94** 1579-1594
- Bender, B. (1986). Modelling source zone boundary uncertainty in seismic hazard analysis. *Bull. seism. Soc. Am.* **76** 329-341
- Biasi, G. P. and Weldon, R. (2008). San Andreas fault rupture scenarios from multiple paleoseismic records: Stringing pearls: in Working Group on California Earthquake Probabilities, Version 2 (UCERF2) U.S.G.S., *Appendix E: Open-File Report 2007-1437 and California Geological Survey Special Report 203*
- Bommer J.J. and Scherbaum F. (2008). The Use and Misuse of Logic Trees in Probabilistic Seismic Hazard Analysis *EARTHQ SPECTRA*, **24** 997-1009, ISSN:8755-2930
- Cornell, C.A. (1968). Engineering seismic risk analysis, *Bull. seism. Soc. Am.*, **58** 1583-1606
- Kuehn N.M., Hainzl S. and Scherbaum F. (2008), Non-Poissonian earthquake occurrence in coupled stress release models and its effect on seismic hazard, *Geophys. J. Int.*, **174** 649-658 doi: 10.1111/j.1365-246X.2008.03835.x.
- Fitzenz, D.D., Jalobeanu, A. and Ferry M.A. (2012), A Bayesian Framework to Rank and Combine Candidate Recurrence Models for Specific Faults, *Bulletin of the Seismological Society of America*, To Appear in Vol **102-3** BSSA-D-11-00087
- Fitzenz, D.D., Ferry, M.A. and Jalobeanu, A. (2010) Long-term slip history discriminates among occurrence models for seismic hazard assessment, *Geophys. Res. Lett*
- Fitzenz, D. D., A. Jalobeanu, and S. H. Hickman, (2007). Integrating laboratory creep compaction data with numerical fault models: A Bayesian framework, *J. Geophys. Res.*, **112** B08410, doi:10.1029/2006JB004792
- Le Goff B., Bertil D., Lemoine A. and Terrier M. (2009). Systemes de failles de Serenne et de la Haute-Durance (Hautes-Alpes) : évaluation de l'aléa sismique. *Rapport BRGM RP-57659-FR*, 242 p, 85 ill., 15 Tab, 10Ann
- Musson R.M.W (2004). Objective validation of seismic hazard source models 13th world conference on earthquake engineering, *Vancouver, B.C, Canada*, Paper No.2492.
- Omori, F. (1894). *J. Coll. Sci. Imper. Univ. Tokyo* 7111
- Weatherill, G. and Burton, P.W. (2009). Delineation of shallow seismic source zones using K-means cluster analysis, with application to the Aegean region. *Geophysical Journal International* **176** 565-588.v
- Woo, G. (1996). Kernel estimation methods for seismic Hazard area source modelling. *Bulletin of the Seismological Society of America*, **86** 353-362
- Zöller G. , Ben-Zion Y., Holschneider M., Hainzl S. (2007). Estimating recurrence times and seismic hazard of large earthquakes on an individual fault, *Geophys. J. Int.*, **170** 1300-1310 doi: 10.1111/j.1365-246X.2007.03480.x