

A METHOD TO ESTIMATE INTENSITY OCCURRENCE PROBABILITIES IN LOW SEISMIC ACTIVITY REGIONS

J. J. EGOZCUE, A. BARBAT, J. A. CANAS AND J. MIQUEL

Universidad Politécnic de Cataluña, ETSICCP, Jordi Girona Salgado 31, 08034 Barcelona, Spain

AND

E. BANDA

Instituto Jaime Almera, CSIC, Spain

SUMMARY

The estimation of site intensity occurrence probabilities in low seismic activity regions has been studied from different points of view. However, no method has been definitively established because several problems arise when macroseismic historical data are incomplete and the active zones are not well determined. The purpose of this paper is to present a method that estimates site occurrence probabilities and at the same time measures the uncertainties inherent in these probabilities in low activity regions. The region to be studied is divided into very broad seismic zones. An exponential intensity probability law is adjusted for each zone and the degree of uncertainty in the assumed incompleteness of the catalogue is evaluated for each intensity. These probabilities are used to establish what may be termed 'prior site occurrence models'. A Bayesian method is used to improve 'prior models' and to obtain the 'posterior site occurrence models'. Epicentre locations are used to recover spatial information lost in the prior broad zoning. This Bayesian correction permits the use of specific attenuation for different events and may take into account, by means of conservative criteria, epicentre location errors. Following Bayesian methods, probabilities are assumed to be random variables and their distribution may be used to estimate the degree of uncertainty arising from (a) the statistical variance of estimators, (b) catalogue incompleteness and (c) mismatch of data to prior assumptions such as Poisson distribution for events and exponential distribution for intensities. The results are maps of probability and uncertainty for each intensity. These maps exhibit better spatial definition than those obtained by means of simple, broad zones. Some results for Catalonia (NE of Iberian Peninsula) are shown.

1. INTRODUCTION

Probabilistic methods of evaluating seismic hazard have been developed over the last two decades and they have been used successfully in intermediate and high seismic activity regions.¹⁻⁴ Usually, the seismic parameter to be modelled is magnitude or maximum acceleration and, at times, is intensity. Although these probabilistic methods have also been applied in low seismic activity regions, the results seem to be inconclusive.

The difficulties stem from (a) incompleteness of historical macroseismic catalogues which span several centuries, (b) seismic sources that are either too broad and, therefore, roughly determined or excessively detailed with irrelevant information, (c) insufficient knowledge of attenuation laws, especially in the case of high intensities, and (d) uncertainty in epicentre locations and intensities in both the historical and the instrumental catalogues. This is the case in many low seismic activity regions in which hazard modelling should take into account the uncertainty factor as well as all other available information. Moreover, since large earthquakes have very long return periods in low activity regions very long catalogues are needed to estimate probabilities or average return periods. Usually, long span catalogues give information only about epicentral intensities and, consequently, a method for seismic hazard assessment in low activity regions should use intensity as the main seismic parameter in order to obtain reliable estimates for large events.

The use of intensity as a hazard parameter has important shortcomings. The definition of intensity is mainly based on observed damages or movements of objects, buildings or soil observed by people.

Consequently, intensity takes into account the event itself as well as subjective effects and damages that are building or object dependent. Therefore, intensity describes hazard as well as some kind of risk. The time dependent character of intensity is also remarkable because buildings, structures, people's psychology and recording media have drastically changed over one thousand years. Uncertainties of intensity data are not directly treated here but we point out two features that tend to moderate errors in the intensities of old events. Intensity is, by definition, a discrete parameter; thus accuracy is reduced to one intensity degree. Further, intensities of old events are mainly evaluated in towns or monasteries and macroseismic epicentres are often assigned to such places. The weakness of old structures tends to increase the effects of earthquakes but we expect that physical epicentres are not frequently located at these towns because they were scarce due to the lack of population. So that these two circumstances counteract in evaluating old intensities.

The method suggested in this paper uses standard information given by catalogues and can add specific data regarding attenuation of individual earthquakes or errors in epicentre location. Furthermore, the method is capable of controlling the degree of uncertainty in the estimated probability in accordance with the assumed deficiency of the data.

Using Bayesian techniques a method is developed to evaluate site event probability. Initially, a broad zone approach is used to establish a local model of occurrence probabilities. This may be termed a 'prior model'. The problem with this approach is that its spatial definition is low and, therefore, an improvement is required. This is achieved by means of a Bayesian correction which adds information regarding the geographic epicentre distribution. The resulting probabilities are termed the 'posterior site model'.

The last part of this paper discusses the results obtained when this approach was applied to Catalonia (NE of Iberian Peninsula)

2. EVALUATING EVENT PROBABILITY

An occurrence model is developed by using a table of probabilities of annual occurrence of events for each intensity. If the events to be considered are epicentral intensities in a zone, it is a zonal occurrence model (ZOM). Alternatively, if the events are intensities observed at a site, it is a site occurrence model (SOM). The task is to estimate the site occurrence model for an arbitrary point within the studied area. In addition, a measure of uncertainty for each probability in the model will be calculated.

The method may be divided into three steps. First, the ZOM is evaluated for each previously defined seismic zone in the region under investigation, as well as the uncertainty of each probability. Secondly, the SOM is calculated on the basis of the effect that each zone produces at the site. This site model is similar to those obtained by standard zoned probabilistic methods.^{5,6} However, it differs in that the uncertainty of each probability is computed from the uncertainties of the zonal model. Since the site model is corrected by a Bayesian technique, it is termed the prior SOM. Thirdly, a Bayesian correction for the prior site model is carried out to obtain the posterior SOM. This correction is made without taking into account the initial broad zoning.

As is frequently assumed in Bayesian methods,⁷ all probabilities are considered as random variables. The distribution of probabilities, as random variables, is extremely useful when establishing parameters for the associated uncertainty. These parameters may be either the distribution itself or other characteristics such as mean and variance or probability intervals.

It is clear that, if the posterior site model is computed at the nodes of an appropriate grid covering the whole area, the results can be used to obtain probability and uncertainty maps that contain the basic information of the hazard assessment.

2.1. Zonal occurrence models

Some broad seismic zones have to be defined in the studied area. Initially, these zones are assumed to be seismically homogeneous and statistically independent. Usually these zones have to be defined by subjective criteria, and most of the known epicentres must be included within them. Although geological knowledge has been used to define the zones, the above mentioned hypotheses regarding homogeneity and statistical independence are rough approximations.

Some kind of model is needed to describe the seismic occurrence, both in time and intensity, in each zone. These models have been extensively studied; the most popular intensity or magnitude models are the Gutenberg and Richter exponential model and its modifications^{2,8-11} and those based on the extreme value distribution.¹²⁻¹⁸ Time occurrence has been modelled usually by Poisson or Markov processes.¹⁹⁻²³ However, in low hazard regions, it is not advisable to work with an excessive number of parameters, because the number of events recorded in the catalogue for each zone is insufficient in order to obtain a reliable parameter estimate. Although the zones are very broad, we assume that it is not possible to estimate more than two parameters for each zone. Accordingly, doubt arises regarding the use of more complex stochastic models. For instance, more than two parameters are needed in order to model time-clustering of events.

Nonetheless, an examination of catalogues frequently reveals that the time-clustering of events does in fact exist. This phenomenon is seldom active for longer periods than one year, especially, in low activity regions. This difficulty may be overcome by considering the probabilities of the event of maximum epicentral intensity greater than or equal to i , in each zone over an annual period. This is denoted p_i , $i = i_0, \dots, i_m$, where i_0 , i_m are the minimum and maximum intensities of interest (V to XII are used in this case); a superscript to p_i stands for the zone and is used when necessary. The set of probabilities p_i , $i = i_0, \dots, i_m$ defines the ZOM. Note that $1 - p_{i_0}$ is the annual probability of the maximum intensity being less than i_0 .

Usually, low hazard areas in Europe have macroseismic catalogues dating from the late Middle Ages for the highest intensities (IX, X, XI, XII). The information regarding low intensities (V, VI) is highly incomplete up to the nineteenth century. The quality of information regarding intermediate intensities increases throughout the seventeenth and eighteenth centuries and, thus, the different catalogue spans may be taken as being complete for each intensity.

As a result, the evaluation of the p_i 's has a different quality for each intensity i . In order to include the evaluation variability, p_i is assumed⁷ to be a random variable, whose distribution is determined by its mean $\langle p_i \rangle$ and variance $\text{Var}(p_i)$. The variances $\text{Var}(p_i)$, or other distribution parameters such as variation coefficients, are appropriate measures of uncertainty when estimating p_i 's.

As we have previously stated, the exponential model, and its modifications, are commonly used to model the magnitude or intensity distributions. There is experimental evidence regarding the performance of these laws in very extensive areas, but for small areas their validity has to be tested with the local data. In low seismic activity regions this testing is meaningless because of the lack of data.

Nevertheless, it seems preferable to use a simple, general model than to use a distribution-free model. As we have previously stated, the use of more than two parameters per ZOM is not advisable. Accordingly, a two parameter exponential model for the mean probabilities is assumed for each zone in the prior ZOM. There may be a mismatch between the exponential model and the available data, but it is to be expected that the Bayesian correction will reduce the deviation. In any case, deviations of data from the assumed model will result in an increase of the variance of estimates, as is explained below.

The exponential model for mean probabilities is given by

$$\langle p_i \rangle = \exp(a - bi), \quad i = i_0, \dots, i_m \quad (1)$$

where a and b are to be estimated from the data. It should be noted that the parameter a is not determined by b because no hypothesis of truncated exponential distribution is made for $\langle p_i \rangle$'s. The ZOM will be completely defined when values for $\text{Var}(p_i)$, $i = i_0, \dots, i_m$, are derived from completeness and uniformity of the catalogue and deviations of data from (1).

For each zone the evaluation of a , b is performed by minimizing

$$\sum_{i=i_0}^{i_m} w_i \left| \exp(a - bi) - \frac{M_i}{N_i} \right|^2 \quad (2)$$

where M_i is the number of years in which the maximum epicentral intensity was greater than or equal to i in the zone; N_i is the number of years in the catalogue for intensity i . The weight w_i is assumed to be the inverse of subjective variance of p_i . The solution of the problem in equation (2) is achieved by means of a non-linear optimization.

It is worthwhile pointing out that, almost certainly, $M_i = 0$ for the X, XI and XII intensities, and this is valuable information. Consequently, it is not advisable to use the logarithmic equivalent of equation (2) to obtain a linear regression problem to estimate a and b . In fact, in this last approach, there are two alternatives: (a) to assume that probabilities for non-observed intensities are zero, or (b) to extrapolate the fitted model (1) for all intensities. Alternative (a) is not advisable for Bayesian methods because these zero probabilities can not be corrected by using data. Further, this zero approach might be unrealistic because the maximum observed intensity might be less than the possible maximum intensity. Alternative (b) usually overestimates the probabilities of intensities that have not been observed. This was a reason to introduce three parameter quadratic exponential laws.⁹ The minimization of equation (2), when it is compared with the log-linear standard regression, has two main effects. Normally the b , estimated by using equation (2), is less than that obtained by minimizing the logarithmic equivalent of equation (2). The other effect is that the probabilities of the high intensities are not zero although no event of such intensities has been observed.

Figure 1 shows the sample probabilities M_i/N_i (full line) for the Catalan Coast (Figure 4) compared with the fitted models by using equation (2) (dotted-dashed line) and by using log-linear regression (dotted line).

The variances of the p_i 's may be estimated subjectively, as has been done in determining the weights w_i in equation (2). However, an improvement of those estimates can be designed in order to take into account time deviations of the sample catalogue from the estimated exponential model of equation (1). Once a and b have been estimated from equation (2), we define another estimator of the variance of p_i , denoted by $\text{Var}(p_i)$, by partitioning the observation time period for the i th intensity, N_i , into k_i subintervals of t_i years. If M_{ij} is the number of years in which an event of intensity greater than or equal to i has been observed in the subinterval j , the selected estimator is

$$\text{Var}(p_i) = \frac{1}{k_i} \sum_{j=1}^{k_i} \left| \frac{M_{ij}}{t_i} - \exp(a - bi) \right|^2, \quad i = i_0, \dots, i_m \quad (3)$$

Some conditions should be taken into account when selecting k_i :

- (i) Null variances are not admissible because uncertainty is always present.
- (ii) For fixed M_i , $N_i = k_i t_i$, the maximum sample variance has to be obtained when all events are observed in the same subinterval; the minimum sample variance has to be obtained when the number of events is the same in each subinterval. In order for these maximum and minimum values to be obtainable from the data t_i must be greater than or equal to M_i and k_i must be less than or equal to M_i .
- (iii) A feasible value of k_i would be $k_i = M_i$, because, for zones in low activity regions, M_i is usually a small number and t_i would satisfy condition (ii).

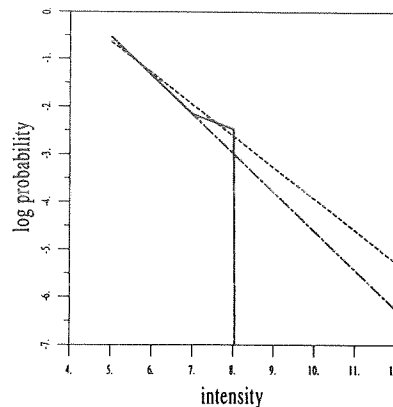


Figure 1. Sample probabilities M_i/N_i (full line) compared with exponential model fitted by minimizing equation (2) (dotted-dashed line) or log-least squares (dotted line) for Catalan Coast

The variances estimated by equation (3) reflect incompleteness and the irregularities of the catalogue comparing the data with a stationary binomial occurrence model. Deviations of data from exponential model (1) are also taken into account in equation (3). However, the estimator (3) is only approximate, especially when the number of events of intensity i is small. On the other hand equation (3) is not adequate when M_i is less than two because no information is available about irregularities of the catalogue. In these cases it is preferable to use a Bayesian estimate of the variance using a uniform prior probability distribution of p_i .

2.2. Prior site occurrence models

A SOM determines the probabilities of the annual maximum site intensity being greater than or equal to i . These probabilities are denoted by q_i , $i = i_0, \dots, i_m$, for each site. Consequently, $1 - q_{i_0}$ is the annual probability of the maximum intensity felt at the site being less than i_0 . We assume that the q_i 's are random variables with a distribution determined by its mean $\langle q_i \rangle$ and its variance $\text{Var}(q_i)$. To calculate $\langle q_i \rangle$ and $\text{Var}(q_i)$ from the zonal occurrence models, a mean attenuation law has to be adopted.

In low seismic activity zones, if foreshocks and aftershocks are excluded, we expect that the zonal annual maximum intensity is reached only once per year, especially for intermediate and high intensities. Consequently, it can be assumed that the influence of a zone at a certain site can be evaluated from the zonal models. Although the q_i values obtained will be slightly lower than the true values, this effect will be counteracted by the Bayesian correction presented below.

The locus of epicentres of intensity i whose observed intensity at the site $i - k$ is a circular ring centred at the site with the radii determined by the mean attenuation law. We denote such a circular ring by C_k (Figure 2). If $A(\cdot)$ is the area of the argument, the parameters of the site occurrence model $\langle q_i \rangle$, $\text{Var}(q_i)$, $i = i_0, \dots, i_m$, can be calculated approximately, assuming independence and homogeneity of zones, from

$$\langle q_i \rangle = \sum_{s=1}^S \sum_{k=0}^{i_m-i} \frac{A(C_k \cap Z_s)}{A(Z_s)} \langle p_{i+k}^s \rangle \quad (4)$$

$$\text{Var}(q_i) = \sum_{s=1}^S \sum_{k=0}^{i_m-i} \left| \frac{A(C_k \cap Z_s)}{A(Z_s)} \right|^2 \text{Var}(p_{i+k}^s) \quad (5)$$

where the Z_s , $s = 1, \dots, S$ are zones and \cap denotes intersection. $A(C_k \cap Z_s)$ is the area in Z_s in which an $(i + k)$ -intensity epicentre would be felt at the site with intensity i . The quotient of areas in equations (4) and (5) is the probability of an i -event felt at the site coming from an $(i + k)$ -epicentre in the zone Z_s , based on the assumed homogeneity of the zone.

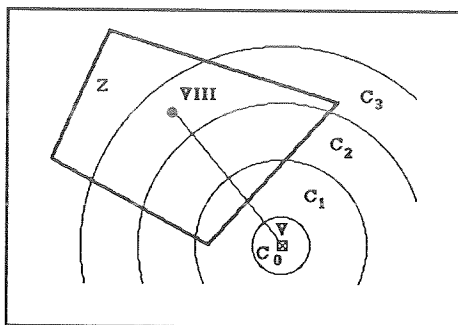


Figure 2. A VIII intensity epicentre is observed in the zone Z , and it is felt at the site with intensity V because the epicentre is in the circular ring C_3 . Other circular rings defined by the attenuation law are shown

To obtain the exact expression, occurrences of earthquakes of intensities i_0, \dots, i_m are assumed to be Poisson distributed in each zone Z_s with parameter λ_s . In this hypothesis the probabilities q_i are

$$q_i = 1 - \exp \left\{ \sum_{s=1}^S \sum_{k=0}^{i_m-i} \log \left(1 - p_{i+k}^s \frac{A(C_k \cap Z_s)}{A(Z_s)} \right) \right\} \quad (6)$$

for $i = i_0 + 1, \dots, i_m$. For $i = i_0$, it has a slightly different form because the Poisson parameter appears explicitly:

$$q_{i_0} = 1 - \exp \left\{ \sum_{s=1}^S \left[-\lambda_s \frac{A(C_0 \cap Z_s)}{A(Z_s)} + \sum_{k=1}^{i_m-i_0} \log \left(1 - p_{i_0+k}^s \frac{A(C_k \cap Z_s)}{A(Z_s)} \right) \right] \right\} \quad (7)$$

First order Taylor series expansion of exponentials and logarithms in equations (6) and (7) reduce these formulae to equation (4). This first order approximation is very accurate when the Poisson parameter is less than unity (one example is reported in Section 3). Therefore, the approximation is suitable for low seismic activity regions.

2.3. 'Posterior' site occurrence models

It is feasible to construct maps of isoprobability curves for each intensity, based on the prior SOM located at the nodes of a grid. These sorts of maps are strongly influenced by the selected zones. If the zones are broad, a smooth spatial variability of the occurrence probabilities is obtained inside each zone, producing loss of details. These results have an important shortcoming: they show less spatial information than the epicentre location maps. This information was lost in the estimation procedure of the ZOMs because spatial information was reduced to two parameters per zone. Bayes' theorem permits the recovery of the lost spatial information by means of a procedure which does not take into account the zones. This procedure is termed posterior correction.

If it is known that, in n_i years of observation on a site, the observed maximum intensity has been greater than or equal to i in m_i years, the likelihood of these binomial data can be expressed as

$$L(m_i|q) = \binom{n_i}{m_i} q^{m_i} (1-q)^{n_i-m_i} \quad (8)$$

The distribution of q_i is assumed to be beta type.⁷ This hypothesis is justified because:

- (i) the rank of the beta distribution is $[0, 1]$ as corresponds to a probability;
- (ii) the beta distribution is very flexible and its shape is fully determined by its mean and variance;
- (iii) the beta distribution is the conjugate prior for the binomial likelihood function according to the Bayes' theorem; consequently, the posterior distribution will also be a beta distribution.

The posterior correction of the prior SOM results from the application of Bayes' theorem: the prior probability density function of q_i with parameters r'_i, t'_i is

$$f'_i(q) = B'_i q^{r'_i-1} (1-q)^{t'_i-r'_i-1}, \quad B'_i = \frac{\Gamma(t'_i)}{\Gamma(r'_i)\Gamma(t'_i-r'_i)} \quad (9)$$

where $\Gamma(\cdot)$ is the Euler gamma function and the $(')$ stands for prior distribution. Bayes' theorem implies that the posterior distribution is given by

$$f''_i(q) = B''_i q^{r''_i-1} (1-q)^{t''_i-r''_i-1}, \quad B''_i = \frac{\Gamma(t''_i)}{\Gamma(r''_i)\Gamma(t''_i-r''_i)} \quad (10)$$

$$r''_i = r'_i + m_i, \quad t''_i = t'_i + n_i$$

where $('')$ stands for the posterior distribution. The new mean and variance are easily calculated from the r'' and t'' for each intensity. The relationship between the beta parameters and the mean and variance of the

probability is given by

$$\langle q \rangle = \frac{r}{t}, \quad \text{Var}(q) = \frac{r(t-r)}{t^2(t+1)} \quad (11)$$

where r , t are either the prior parameters or the posterior ones for each intensity.

The effect of posterior correction on the beta probability density is cleared up in Figures 3(a) and (b). In the first one, the full curve represents a prior distribution of probability for which the parameters are $r' = 6 \cdot$ and $t' = 30 \cdot$. Thus the prior mean probability would be 0.2. After an observation of 4 events in 70 years—sample mean of 0.057—we would obtain posterior parameters $r'' = 10 \cdot$ and $t'' = 100 \cdot$ and the density represented by the dashed curve, for which the mean probability of 0.1 and the variance have been decreased due to the increasing of the years of observation. This shows how sample information corrects the prior estimate if they are not in accordance. Note that, if the number of years of observation was greater, the variance of the posterior distribution would be less by equation (11). Figure 3(b) shows another example; the full line corresponds to a prior beta distribution with $r' = 0.5$ and $t' = 5 \cdot$ (mean probability 0.1), whereas the dashed line is the posterior beta distribution given a sample of 3 events in 15 years (sample mean 0.2). Corresponding posterior parameters are $r'' = 3.5$, $t'' = 20 \cdot$ and the posterior mean probability is 0.175.

Unfortunately, site information is seldom available, especially if the site is not a city or a town. However, if the epicentral intensity and the attenuation law are known, the effect of an earthquake at the site can be deduced. In fact, the attenuation law determines the circular rings C_k defined in Section 2.2 (Figure 2). Once the epicentral distance is known, we can identify what C_k contains the epicentre. Thus intensity felt at the site is estimated by subtracting the value of that k from the epicentral intensity. Although this deduced intensity is only approximate, it does allow the introduction of some maximality criteria to avoid possible errors. Macroseismic epicentre locations may be affected by errors of 10 or 15 km. One of the possible corrections would be to displace the epicentre closer to the site to maximize the induced site intensity. Specific attenuation laws may also be used for well known earthquakes, including directional attenuation when available.

When the posterior correction is carried out over a grid of sites and the isoprobability curves are drawn, the spatial resolution of the map increases, showing most of the epicentre spatial distribution in the zones. This procedure allows updating of the variances of the occurrence probabilities and it is also feasible to depict variation coefficient maps to assess the uncertainty of the mean occurrence probability.

However, the posterior correction has to be undertaken carefully to avoid an inconsistent posterior SOM which may result in mean probabilities that increase with their corresponding intensity. This may occur

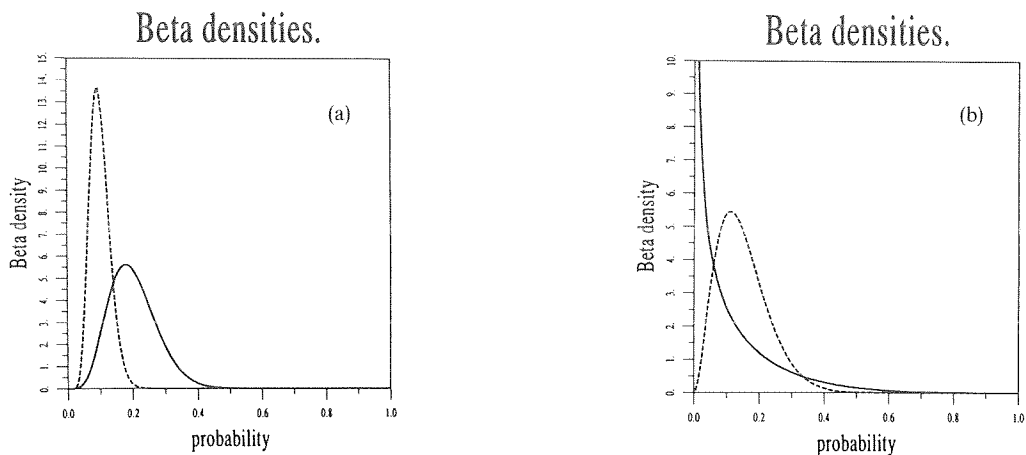


Figure 3. Two examples of posterior correction of beta densities. Full curves are prior densities for occurrence probabilities and dashed curves are posterior ones given data: (a) 4 events in 70 years; (b) 3 events in 15 years

because the observation time span of catalogues increases with intensity. This is unrealistic because it is expected that $\langle q_i \rangle$ should be greater than $\langle q_{i+1} \rangle$.

Another undesired effect of a posterior correction might be that the SOM presents a sudden decrease in occurrence probability for intensities higher than those which have been observed at the site. This type of SOM reflects the data correctly but the results obtained are not in agreement with the qualitative features of the definition of intensity. This type of undesirable SOM may be avoided by carrying out a smoothing of the number of local observations and adjusting them to a decreasing exponential model as follows:

$$m_i^* = \exp(c_0 - c_1 i) \quad (12)$$

where the constants c_0 and c_1 are found by fitting equation (12) to the data $m_{i_0}, m_{i_0} + 1, \dots, m_{i_m}$. The smoothed observations m_i^* are in general fractionary but may be accepted and used as in equation (10). This kind of smoothing usually avoids the above mentioned anomalies.

Theoretical inconsistency may be argued against the posterior correction because part of the information used to obtain the prior SOM is also used to calculate the posterior correction. This argument is valid because the numbers M_i and N_i , used to construct the prior ZOM as seen in equation (2), are correlated with $m_i, m_{i-1}, m_{i-2}, \dots$, used in the posterior correction (10).

However, this correlation between the zonal data M_i, N_i and the local data m_i, m_{i-1}, \dots is low when the zonal areas under consideration are considerably larger than the areas of rings C_k defined by the attenuation laws. It is true that, depending on the distribution of epicentres within the zone, m_i, m_{i-1}, \dots represent only small parts of M_i . This is another reason for considering broad zones when obtaining the ZOM.

In regions of medium to low seismic activity it may be advisable to construct the ZOM using the historical catalogue and to use the instrumental catalogue to perform the posterior correction. In this way the data used in the prior model and the posterior correction are independent.

3. APPLICATION OF THE METHOD TO CATALONIA (NE OF IBERIAN PENINSULA)

What follows is a brief description of the data compiled by the method presented in this paper when applied to the NE of the Iberian Peninsula.

The seismic catalogue of the NE part of the Iberian Peninsula has two well defined parts: a short instrument catalogue, dating from 1910 to the present, and an extensive macroseismic historical catalogue that covers the period from the tenth century to the early twentieth century.

For the instrumental catalogue, it may be assumed that almost all events of intensity greater than or equal to V are noted, with their date, hour, epicentre location (instrumentally determined in 40 per cent of cases), and sometimes magnitude (20 per cent), focal depth (15 per cent) and macroseismic intensity map (20 per cent). Although the epicentres have been located instrumentally, it may be observed that the instrumental epicentre sometimes differs from the macroseismic epicentre by several kilometres (10, 15 or even more); these errors are due to the low quality of the seismic network in these areas up to the 1980's. Furthermore, most of the focal depths are only orientative. Magnitudes are assumed to be more accurate but they are too low to base the seismic hazard assessment on them.

The historical catalogue contains the list of dates, intensities and epicentres collected (directly or indirectly) by some authors and institutions,²⁴⁻²⁷ and some intensity charts, most of them reconstructed at the beginning of our century and re-elaborated²⁸ in 1982. The loss of information is evident for intermediate intensities (V, VI, VII) in the earlier centuries and it is difficult to evaluate for the eighteenth and nineteenth centuries. In addition, this loss seems to be different for zones with different historical population.

Five broad seismic zones have been defined in the NE of the Iberian Peninsula and its surroundings, as shown in Figure 4 together with the epicentres considered in the catalogue. These zones are named Western Pyrenees, Eastern Pyrenees, Western Pre-Pyrenees, Catalan Coast and Iberian System. It is clear that the Balearic Islands, and the zones of Alicante and Murcia have very little influence on Catalonia and can be ignored without this producing any significant changes in the results.

The zones have been defined by taking into account the epicentre distribution, some geological criteria and previous studies in the area.²⁹⁻³¹ Table I shows the estimated values of a, b in formulae (1) and (2) for each

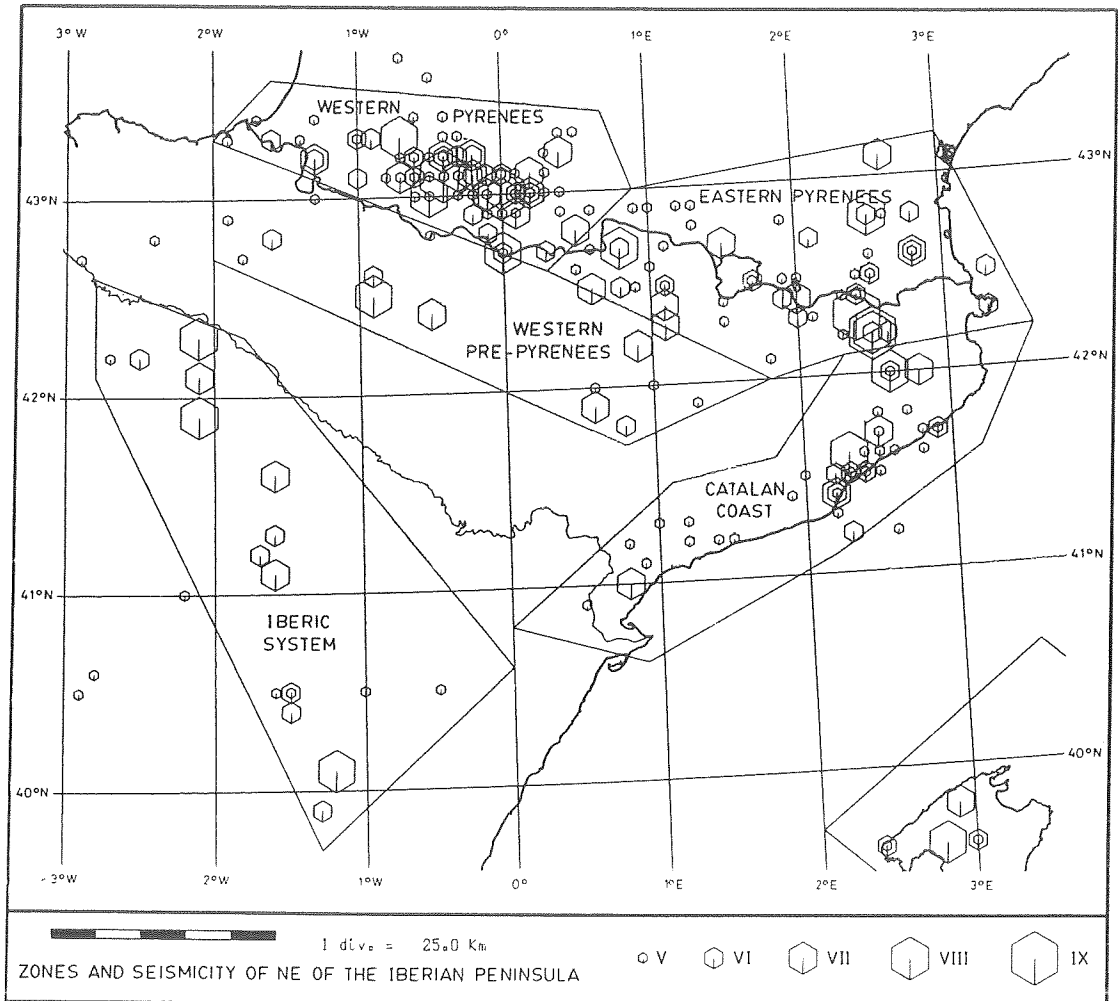


Figure 4. Macroseismic epicentres of the NE of the Iberian Peninsula. After- and foreshocks have been removed. MSK intensity is indicated by the size of hexagons. The zones used to obtain the ZOMs are shown

Table I. Zone parameters a , b from equations (1) and (2)

Zone	a	b
Western Pyrenees	4.254	1.009
Western Pre-Pyrenees	3.663	1.198
Eastern Pyrenees	4.922	1.264
Catalan Coast	8.189	1.883
Iberic System	3.190	1.018

zone. Different catalogue starting dates have been used for each zone and intensity, as shown in Table II.

The obtained ZOMs are compatible with other results for the Iberian Peninsula; for instance, recently García *et al.*³² have estimated the Poisson parameter and the Richter parameter b for some standard zones using the maximum likelihood method of Kikjo and Sellevoll.²³ Although the selected zones are slightly

different, the zonal occurrence probabilities found are close to those calculated by the present method and they are well within the calculated 90 per cent probability interval. Table III compares these results for the Eastern Pyrenees and Catalan Coast. The estimated mean return periods for these two zones show that the Catalan Coast has significant activity for intensity V but it decreases rapidly for higher intensities. The Eastern Pyrenees shows more activity in higher intensities than the Catalan Coast.

It should be observed that the mean return period increases exponentially with intensity but the limits of the 90 per cent probability interval increase linearly (Table III). This is due to the exponentially increasing variance of the return period for intermediate intensities.

The general attenuation law was obtained from the known isointensity curves collected in the NE of Spain and SE of France.²⁶ Figure 5 shows the mean radius of these isointensity curves—labelled Mean atten.—and the attenuation model for the greatest earthquake in the NE of the Iberian Peninsula with epicentre in the surroundings of Queralps in 1428.³³ The first one was used to calculate the prior SOMs. Posterior SOMs have been calculated using different attenuation laws, deduced from isoseismal maps²⁸ if they were available. If no information about attenuation was available, the mean attenuation law (Figure 5) was used. A significant example of specific attenuation law is the above mentioned one for the 1428, X intensity event of Queralps.

Table IV shows the performance of the approximation in equation (4) when compared with the exact formulae (6) and (7); it shows the results obtained in Barcelona (41·40N, 2·20E) and Girona (41·95N, 2·80E) for different intensities.

Figure 6 shows the prior SOM map for intensity VI. Figures 7 to 10 show the posterior SOM maps for intensities V, VI, VII and VIII. Figure 11 is the iso-variation coefficient map for the intensity VI. They were calculated over a triangular grid of 512 nodes. The isoprobability curves are labelled with the decimal logarithm of the probability and the iso-variation coefficient curves with the natural value. The logarithmic

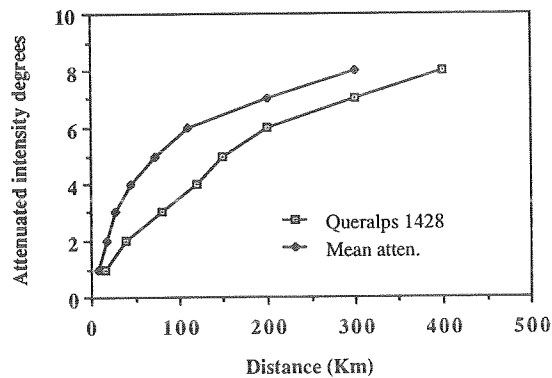


Figure 5. Mean attenuation law for the NE of the Iberian Peninsula. I is the observed intensity, I_e the epicentral intensity and $I_e - I$ is the intensity difference

Table II. Starting dates of catalogues of each intensity for the considered zones

Zone	Intensity				
	V	VI	VII	VIII	Other
Western Pyrenees	1900	1850	1700	1400	1000
Western Pre-Pyrenees	1900	1850	1700	1400	1000
Eastern Pyrenees	1900	1850	1700	1400	1000
Catalan Coast	1900	1850	1700	1400	1100
Iberic System	1900	1900	1850	1800	1300

Table III. Comparison of return periods (years) for Eastern Pyrenees and Catalan Coast. First column as calculated by García *et al.* (1987). Second column gives the prior mean return periods calculated by the present method ($1/\langle p_i \rangle$). Third column is the 90 per cent probability interval

Eastern Pyrenees

Intensity (MSK)	Return period (García <i>et al.</i>) (years)	Mean return period (years)	90% return period interval (years)
V	4.4	4.0	[2.7, 7.1]
VI	11.8	14.3	[3.3, 5×10^5]
VII	34.2	50.6	[7.2, ∞]
VIII	114.9	179.0	[8.9, ∞]
IX	742.5	633.2	[9.8, ∞]

Catalan Coast

V	6.5	3.4	[1.4, 47.6]
VI	24.6	22.3	[3.8, ∞]
VII	105.1	146.8	[7.6, ∞]
VIII	727.1	964.5	[11.3, ∞]
IX	∞	6337.	[157., ∞]

Table IV. Prior mean occurrence probabilities for Barcelona as calculated by means of the approximate formula (4) and the exact formulae (6) and (7)

Intensity	Barcelona		Gerona	
	Prior mean prob. eq. (4)	Prior mean prob. eq. (6) (7)	Prior mean prob. eq. (4)	Prior mean prob. eq. (6) (7)
V	0.8562 10^{-2}	0.8484 10^{-2}	0.9966 10^{-2}	0.9894 10^{-2}
VI	0.1287 10^{-2}	0.1298 10^{-2}	0.1714 10^{-2}	0.1730 10^{-2}
VII	0.2030 10^{-3}	0.2033 10^{-3}	0.3287 10^{-3}	0.3291 10^{-3}
VIII	0.2917 10^{-4}	0.2918 10^{-4}	0.6702 10^{-4}	0.6703 10^{-4}
IX	0.4207 10^{-5}	0.4208 10^{-5}	0.1395 10^{-4}	0.1395 10^{-4}

labelling was adopted because the numbers indicate either log-mean probability, when considered negative, or log-mean return period; the interval between curves is 0.5 logarithmic units to obtain a reasonable spatial distribution of isoprobability curves although a fractionary label does not correspond to a standard return period (for instance, a label 2.5 indicates a return period of 316 years).

Different attenuation laws were used to compute the posterior correction. When no special information was available the mean attenuation law on Figure 5 was used. The macroseismic epicentres were displaced 15 km closer to the site and the instrumental epicentres were displaced only 10 km.

The isoprobability curves in the prior SOM map for intensity VI (Figure 6) match the different zones and reveal the above mentioned low spatial definition. Posterior results (Figures 7 to 10) show an increased spatial definition. For instance, they show a small area, to the NE of Barcelona (approximately 41.6N, 2.4E), that has higher probabilities than other parts of the Catalan Coast. This is in agreement with the spatial distribution of epicentres shown in Figure 4. It is also worth pointing out that borders of zones have no significant influence in these posterior maps. For instance, Catalan Coast has been well divided into two

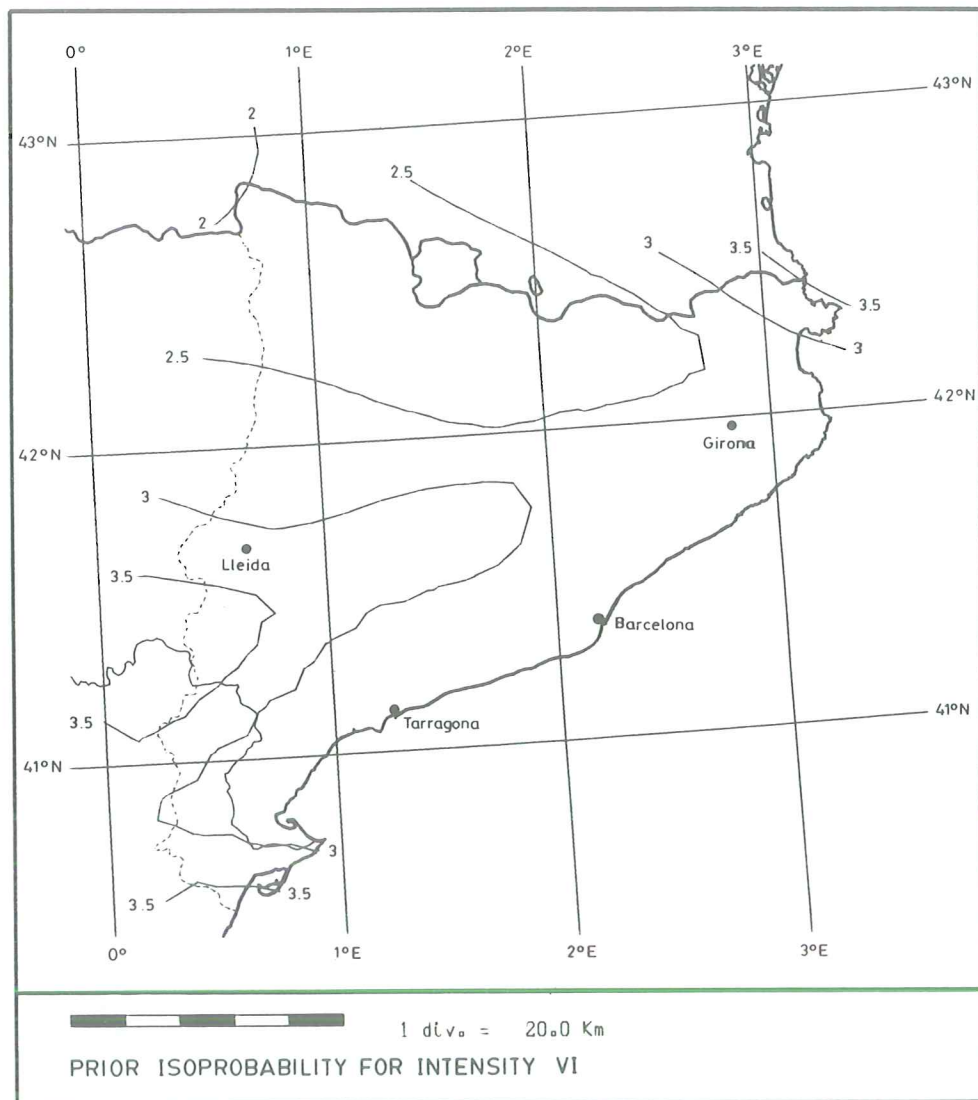


Figure 6. Prior SOM map for intensity VI. The isoprobability lines are numbered with the log return period (return period in years)

zones regarding hazard in the posterior maps; the border of Catalan Coast with Eastern Pyrenees does not match the isoprobability curves in posterior maps.

Figure 11 shows areas with variation coefficient greater than 1. This means that mean probabilities or mean return periods in these areas are almost meaningless. Consequently, the hazard in such areas must be estimated by probability intervals or probability distributions.

Figure 12 shows the prior and posterior SOM for Barcelona in terms of logarithmic mean return periods ($1/\langle q_i \rangle$) and they are compared with the sample return period for intensities felt in Barcelona, calculated from the attenuation laws. The prior SOM return periods seem clearly overestimated when compared with the posterior SOM return periods or the sample return periods. This is the effect of the non-homogeneous zonation.

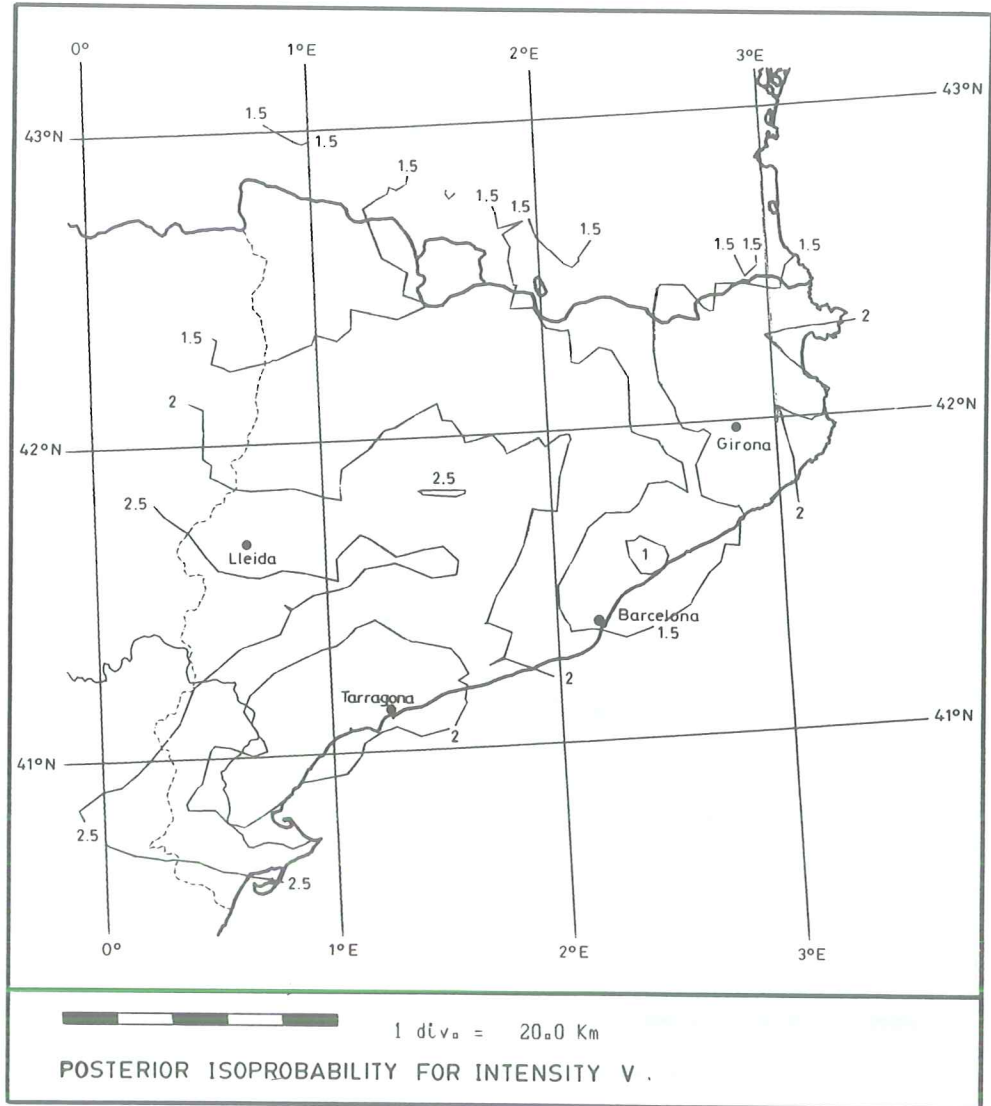


Figure 7. Posterior SOM map for intensity V. The isoprobability lines are numbered with the log return period (return period in years)

Figure 13 shows the posterior SOM with the upper and lower bounds of the 90 per cent probability interval for the return periods in Barcelona. It is worthwhile pointing out the large uncertainty that these results exhibit. The sample return periods are also shown.

4. CONCLUSIONS AND COMMENTS

In low seismic activity regions accurate zonation is difficult because of the lack of data concerning earthquakes. Furthermore, the estimation of zonal parameters is virtually impossible owing to the scarcity of data in each zone. Consequently, a broad zonation approach is preferable. However, this implies the

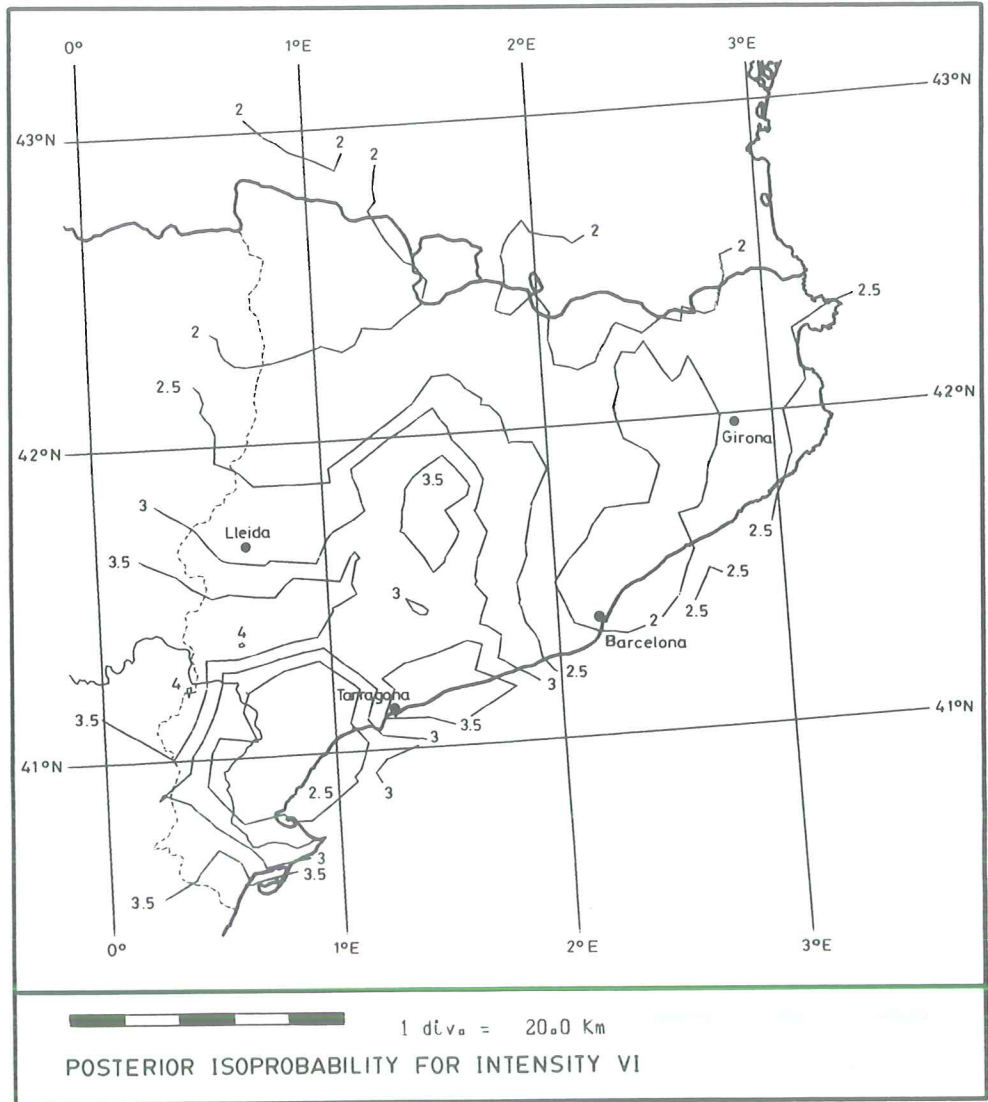


Figure 8. Posterior SOM map for intensity VI. The isoprobability lines are numbered with the log return period (return period in years)

assumption that these broad zones are homogeneous and, therefore, site probability maps will have very low spatial definition, even lower spatial definition than that an epicentre chart has.

To avoid the undesired effect of the broad zone approach, a Bayesian correction of spatial information can be performed. By assuming the site occurrence probabilities as beta distributed random variables, the spatial information of epicentres is recovered through the binomial likelihood of the intensities felt at the site. Furthermore, the correction can be used to incorporate individual attenuation information or to compensate for errors of the epicentre location. The resulting corrected maps show a good spatial definition compared to uncorrected maps.

The posterior correction estimates approximately the variance of the site occurrence probabilities. The variance of the posterior site occurrence probabilities is estimated taking into account the statistical variance

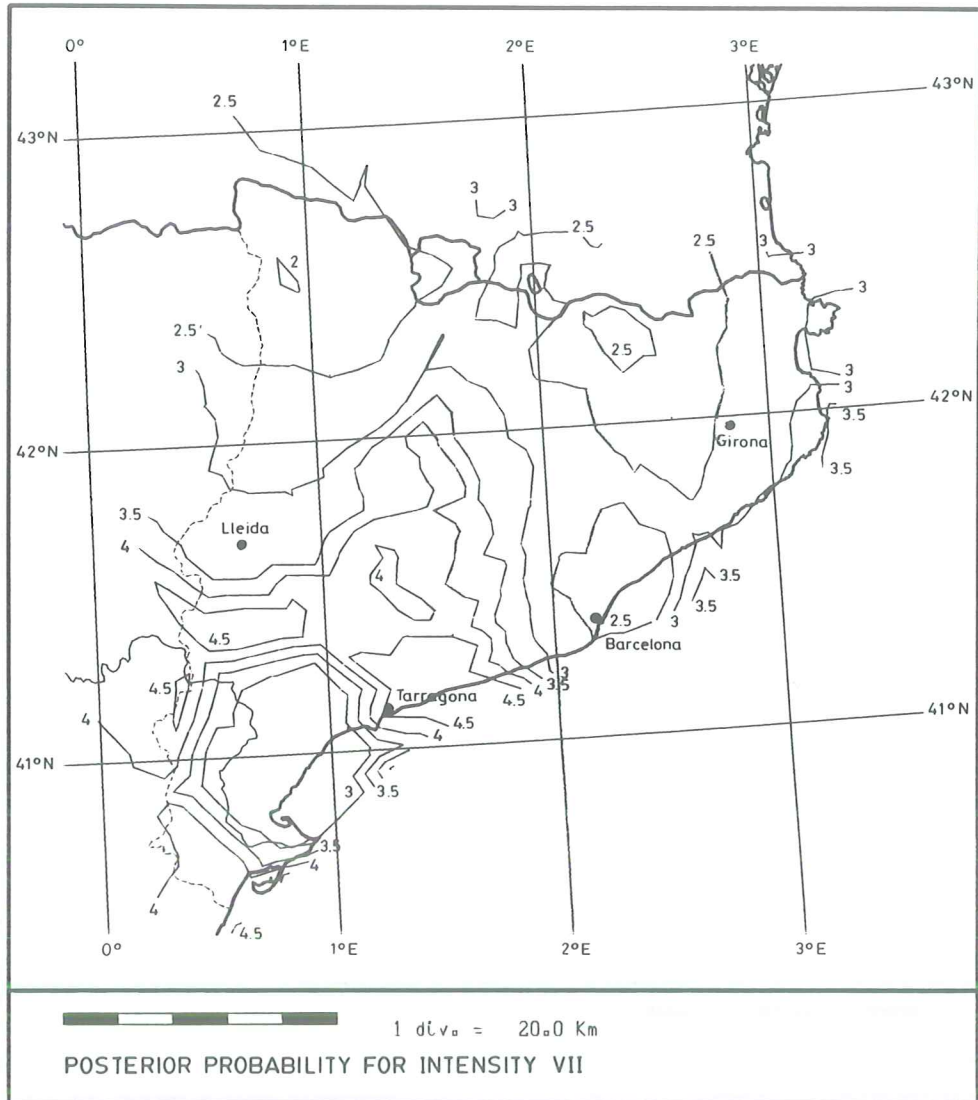


Figure 9. Posterior SOM map for intensity VII. The isoprobability lines are numbered with the log return period (return period in years)

of probability estimators, incompleteness of the catalogue, and the mismatch of data to prior assumptions. Probability intervals for the estimated site occurrence probabilities or return periods can also be estimated. These parameters measure the reliability of mean annual probability and they show how mean return periods may be used for engineering purposes.

Incompleteness or inhomogeneity of catalogues is handled in two different ways; by estimating the variance of zonal occurrence probabilities—equation (3)—and by limiting the span of catalogues for each intensity and zone. Deviations of data from the prior assumption of homogeneity of the zones produce, through posterior correction, increments in the variances of occurrence probabilities.

Formula (4) permits the prior site occurrence probabilities to be deduced from the zonal occurrence probabilities. This formula is an approximation of the true probability, but its performance is very good when the considered region has a low seismic activity.

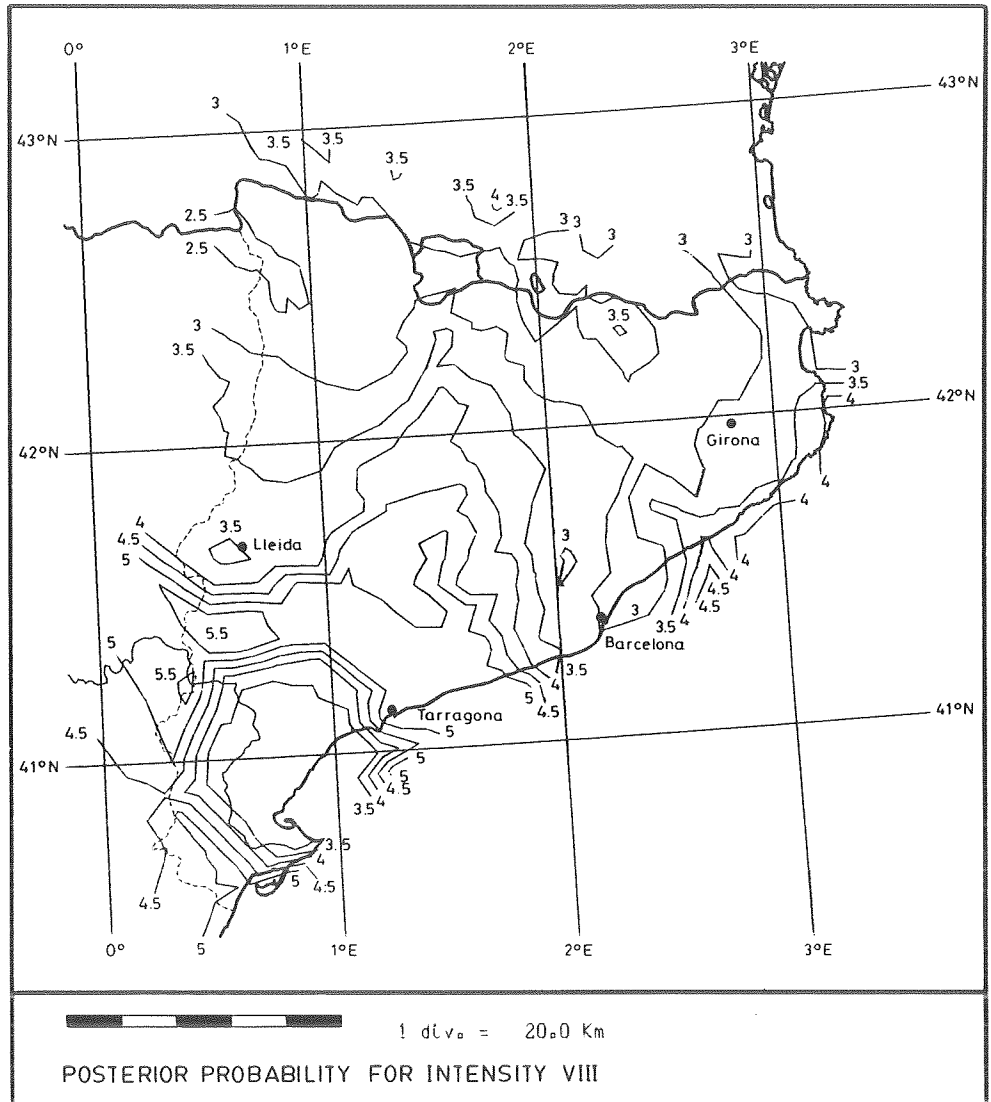


Figure 10. Posterior SOM map for intensity VIII. The isoprobability lines are numbered with the log return period (return period in years)

The uncertainty of attenuation laws and epicentre locations is not controlled by probabilistic reasoning. It is avoided by displacing the epicentres some distance closer to the site. These changes in the epicentre locations tend to increase the site occurrence probabilities; thus, they are conservative assumptions.

The large uncertainty in the occurrence models demands further research in hazard prediction, because the statistical analysis of the macroseismic data seems to be insufficient in low seismic activity regions.

ACKNOWLEDGMENTS

The authors would like to acknowledge the help of Vera Pawlovski in assessing the probabilistic aspects of this study, the drawing computer programs of Fermín Navarrina and the discussion of the results with Mariano García. The authors also thank the reviewers for the suggestions regarding presentation of this paper.

This study was partially supported by the Servei Geològic of the Generalitat de Catalunya and the Universidad Politècnica de Catalunya.

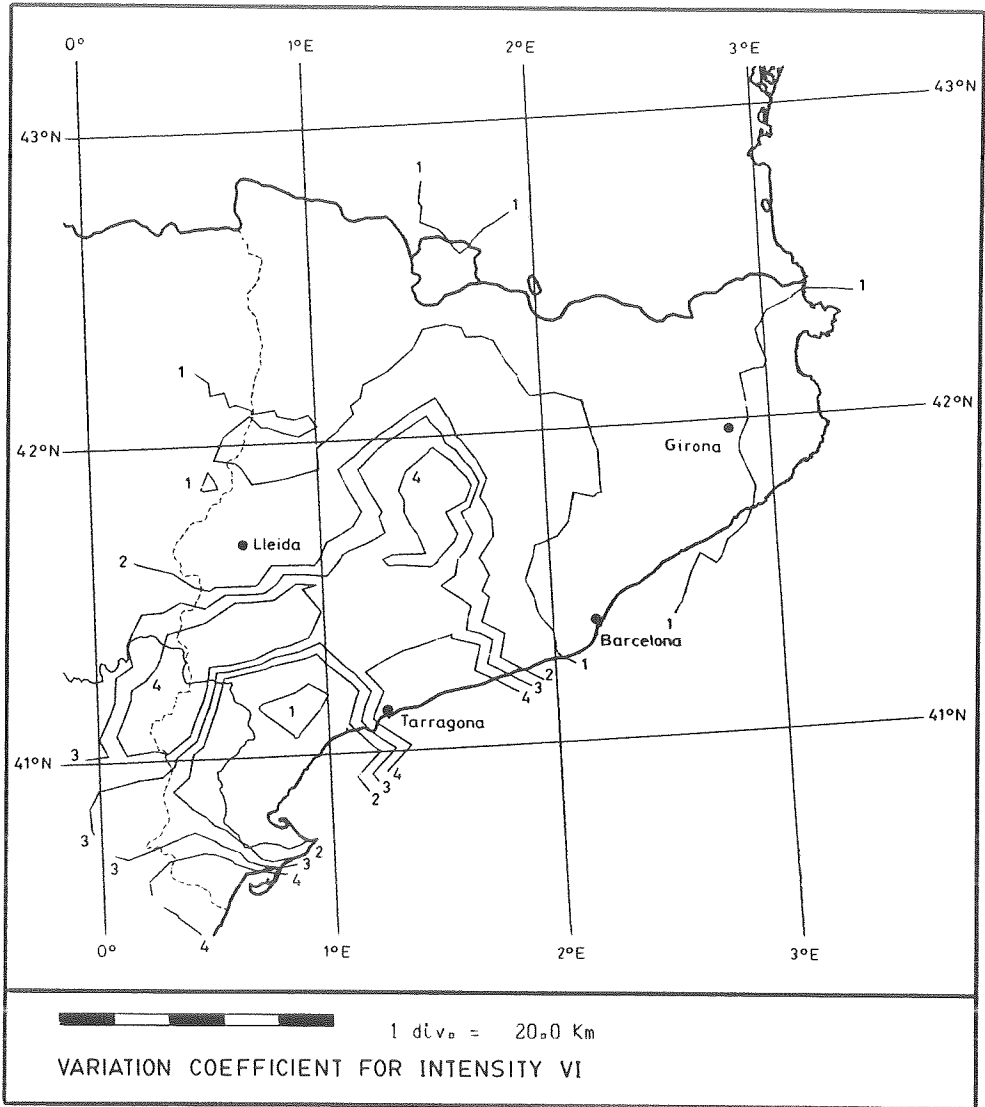


Figure 11. Iso-variation coefficient map for intensity VI. The isolines are numbered with the natural value of the variation coefficient

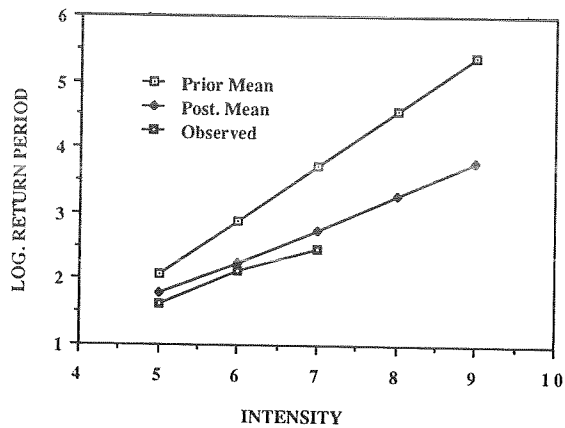


Figure 12. Comparison between the prior and posterior SOM for Barcelona. The sample return period is shown for intensities V, VI, and VII

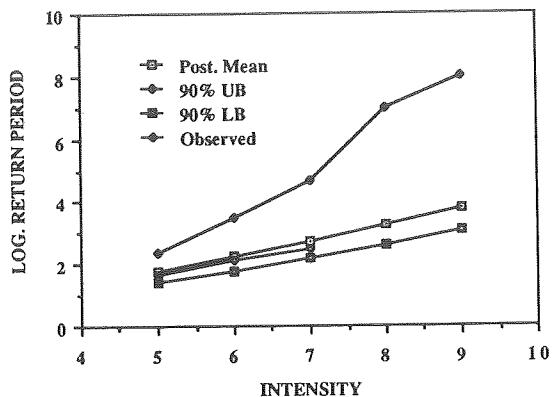


Figure 13. Posterior SOM for Barcelona with the upper and lower bounds of the 90 per cent probability interval. The sample return period is also shown

REFERENCES

1. C. Lomnitz, 'An earthquake risk map of Chile', *Proc. 4th world conf. earthquake eng.* Santiago, Chile (1969).
2. R. K. McGuire 'Effects of uncertainty in seismicity on estimates of seismic hazard for the east coast of United States', *Bull. seism. soc. Am.* **67**, 827-848 (1977).
3. C. A. Cornell and H. A. Merz, 'Seismic risk analysis of Boston', *J. struct. div. ASCE* **101**, 2027-2043 (1975).
4. C. P. Mortgat and H. C. Shah, 'A Bayesian model for seismic hazard mapping', *Bull. seism. soc. Am.* **69**, 1237-1251 (1979).
5. R. K. McGuire, 'FORTRAN computer program for seismic risk analysis', U.S.G.S. *Open File Report 67-76*, U.S.A. 1976.
6. C. A. Cornell, 'Engineering seismic risk analysis', *Bull. seism. soc. Am.* **58**, 1583-1606 (1968).
7. J. R. Benjamin and C. A. Cornell, *Probability, Statistics and Decision for Civil Engineers*, McGraw-Hill, New York, 1970.
8. C. F. Richter, *Elementary Seismology*, Freeman, San Francisco, 1958.
9. H. A. Merz and C. A. Cornell, 'Seismic risk analysis based on a quadratic magnitude-frequency law', *Bull. seism. soc. Am.* **63**, 1999-2006 (1973).
10. C. A. Cornell and H. A. Merz, 'Seismic risk analysis of Boston', *J. struct. div. ASCE* **101**, 2027-2043 (1975).
11. J. Lomnitz-Adler and C. Lomnitz, 'A modified form of Gutenberg-Richter magnitude-frequency relation', *Bull. seism. soc. Am.* **69**, 1209-1214 (1979).
12. E. J. Gumbel, *Statistics of Extremes*, Columbia University Press, New York, 1958.
13. L. Knopoff and Y. Kagan, 'Analysis of the theory of extremes as applied to earthquake problems', *J. geophys. res.* **82**, 5647-5657 (1977).
14. K. W. Campbell, 'Bayesian analysis of extreme earthquake occurrences. Part I. Probabilistic hazard model', *Bull. seism. soc. Am.* **72**, 1689-1706 (1982).
15. Z. J. Gan and C. C. Tung, 'Extreme value distribution of earthquake magnitude', *Phys. earth planet. interiors* **32**, 325-330 (1983).
16. A. Kijko, 'A modified form of the first Gumbel distribution: Model for the occurrence of large earthquakes. Part I, Derivation of the distribution', *Acta geophys. Pol.* **30**, 333-340 (1982).
17. A. Kijko, 'Is it necessary to construct empirical distributions of maximum earthquake magnitudes?', *Bull. seism. soc. Am.* **74**, 339-347 (1984).
18. A. Kijko, 'Extreme value distribution of earthquake magnitude', *Phys. earth planet. interiors* **37**, 285-287 (1985).
19. C. Lomnitz, 'Poisson processes in earthquake studies', *Bull. seism. soc. Am.* **63**, 735 (1973).
20. D. Vere-Jones, 'Stochastic models for earthquakes sequences', *Geophys. j. r. soc.*, **42**, 811-826 (1975).
21. S. Shlien and N. Toksoz, 'A branching Poisson-Markov model of earthquake occurrences', *Geophys. j. r. soc.* **42**, 49-59 (1975).
22. A. S. Patwardhan, R. B. Kulkarni and D. Tocher, 'A semi-Markov model for characterizing recurrence of great earthquakes', *Bull. seism. soc. Am.* **70**, 323-347 (1980).
23. A. Kijko and M. A. Sellevoll, 'Bayesian estimation of seismic parameters for extreme historical and complete instrumental data', *Seismo-Series, 7*, Seism. Obs. U. Bergen, Norway (1986).
24. J. Galbis, *Catálogo Sísmico de la Zona Comprendida entre los Meridianos 5E y 20W de Greenwich y los Paralelos 45N y 25N*, Tomos I y II, Inst. Geográfico, Catastral y de Estadística, Madrid, 1932 and 1940.
25. E. Fontserè and J. Iglésies *Recopilacio de Dades Sísmiques de les Terres Catalanes entre 1100 i 1906*, Fundació Salvador Vives Casajuana, Barcelona, 1971.
26. E. Suriñach and A. Roca, *Catálogo de Terremotos de Cataluña y Zonas Adyacentes, 1100-1979*, Publicación de la Cátedra de Geofísica 190, U. Complutense, Madrid, 1982.
27. J. Mezcuca and J. M. Martínez, *Sismicidad del Area Ibero-Mogrebi*, Inst. Geogr. Nacional, Publ. 203, Madrid, 1983.
28. J. Mezcuca, *Catálogo General de Isosistas de la Península Ibérica*, Inst. Geogr. Nacional, Publ. 202, Madrid, 1982.
29. E. Suriñach and A. Roca, 'Sismicidad en la región N. E. de la Península Ibérica', *Rev. Geofísica* **41**, 23-36 (1985).
30. A. J. Martín, 'Riesgo sísmico en la Península Ibérica', *Tesis Doctoral*, U. Politécnica Madrid, 1983.
31. A. Roca and A. Udias, 'Sismicidad y riesgo sísmico en Cataluña y Pirineos', *Rev. geofísica* **35**, 183-207 (1976).
32. M. García, M. J. Jiménez and A. Kijko, 'Seismic hazard parameters estimation in Spain from historical and instrumental catalogues. Paper presented to IASPEI (IUGG), Vancouver, Aug. (1987).
33. E. Banda and A. Correig, 'The Catalan earthquake of February 2, 1428', *Eng. geol.*, **20**, 87-89 (1984).