Bayesian analysis of the local intensity attenuation

R. Rotondi[1] and G. Zonno[2]

[1] *C.N.R. - Istituto di Matematica Applicata e Tecnologie Informatiche,*
Via Bassini 15, 20133 Milano, Italy
[2] *I.N.G.V. - Milano Department "Applied Seismology", Via Bassini 15, 20133 Milano, Italy*

Corresponding Author: R. Rotondi, `reni@mi.imati.cnr.it`

Abstract: *We present a method that allows us to incorporate additional information from the historical earthquake felt reports in the probability estimation of local intensity attenuation. The approach is based on two ideas: a) standard intensity versus epicentral distance relationships constitute an unnecessary filter between observations and estimates; and b) the intensity decay process is affected by many, scarcely known elements; hence intensity decay should be treated as a random variable as is the macroseismic intensity. The observations related to earthquakes with their epicenter outside the area concerned, but belonging to homogeneous zones, are used as prior knowledge of the phenomenon, while the data points of events inside the area are used to update the estimates through the posterior means of the quantities involved.*

1   Introduction

Where long historical catalogues are available, as in Italy, it is quite natural to take the macroseismic intensity as a measure of the size of earthquakes. This ordinal quantity, often measured on the 12 degrees of the MCS scale, is in practice treated, as best one can, as an integer variable on $\{1, 12\}$. In order to assess the seismic hazard in terms of intensity we have to address the problem of its attenuation. Many studies on this topic have appeared in the literature; in the large majority of these the key role is played by the deterministic function which expresses the link between the ΔI intensity decay and factors such as epicentral intensity, site-epicenter distance, depth, site types, and styles of faulting. In some cases a normally distributed random error is added to take into account the scatter of the observations around the I_s site intensity value predicted through the attenuation relationship. More emphasis is given to the uncertainty when the decay is considered an *aleatory* variable: for the intensity decay normalized on I_0, a Beta distribution with mean proportional to an attenuation law and varying deviation was first proposed by Zonno *et al.* (1995), while a logistic model was used by Magri *et al.* (*Pageoph.*, 43, 1994) to estimate the probability that the attenuation exceeds a threshold value.

We present a complete probabilistic analysis of the attenuation issue, avoiding the use of any deterministic attenuation relationship (Rotondi and Zonno, 2004). On the contrary, the emphasis here is on exploiting information from seismogenically homogeneous zones, that is, on assigning and updating prior parameters in the Bayesian framework. We give the predictive distribution of the intensity at site, conditioned on I_0 and on the d distance from the epicenter, with both discrete and continuous d. We present some validation criteria for these two classes of distributions, and then apply these criteria first to an historical event, the Camerino 1799 earthquake, to explore the goodness of fit of the model, and then to a recent event, the Colfiorito 1997 earthquake, to validate the procedure. We compare our approach

with that of Magri *et al.* (1994), estimating the decay through the mode of the respective probability distributions.

2   Description of the model

Let us assume that the variable ΔI is discrete and belongs to the domain $\{0, I_0\}$; it is reasonable to choose for $I_s = I_0 - \Delta I$, at a fixed distance, the binomial distribution $Bin(I_0, p)$ conditioned on I_0 and p:

$$Pr\{I_s = i \mid I_0 = i_0, p\} = Pr\{\Delta I = I_0 - i \mid I_0 = i_0, p\} = \binom{i_0}{i} p^i (1-p)^{i_0 - i}$$

where $p \in [0, 1]$. To account for the variability of the ground shaking even among sites located at the same distance, the parameter p has been considered a Beta distributed random variable in the Bayesian paradigm

$$Be(p; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^p x^{\alpha-1}(1-x)^{\beta-1} dx \qquad .$$

2.1   Assigning prior parameters and their updating

The procedure described here draw initial knowledge of the phenomenon from the intensity data points of earthquakes occurring in zones that are homogeneous to the area under study, from the viewpoint of seismotectonics and seismic-wave propagation.

Let us draw L distance bins $\{R_1, R_2, \ldots, R_L\}$ of width Δr around the epicenter of any earthquake of I_0 intensity in those zones. Taking a sufficiently small step we may assume that the decay process behaves in the same way within each R_j band; moreover, we denote by $\mathcal{D}_{j,0} = \left\{ i_s^{(n)} \right\}_{n=1}^{N_{j,0}}$ the set of $N_{j,0}$ felt intensities in the j-th band. As the probability of null decay $(I_s = I_0)$ is $p_j^{i_0}$, we assign the initial mean value of p_j using simply the frequency of null decay, $N_{j,0}(\Delta I = 0)/N_{j,0}$, and deduce from this value the hyperparameters $\alpha_{j,0}$ and $\beta_{j,0}$. Where there is no report of null decay, we have smoothed the valuable p_j's through the function $f(d) = (c_1/d)^{c_2}$ and estimated the coefficients c_1, c_2 by the method of least squares.

Now let us consider all the earthquakes of I_0 intensity with epicenter within the area under study; $\mathcal{D} = \bigcup_{j=1}^{L} \mathcal{D}_j = \bigcup_{j=1}^{L} \left\{ i_s^{(n)} \right\}_{n=1}^{N_j}$ denotes the set of their intensity data points, subdivided into L subsets. On the basis of this new information we update our knowledge on the attenuation process. We use Bayes' theorem to compute the posterior distribution $Be(p_j \mid \mathcal{D}_j)$, and estimate p_j through its posterior mean $\hat{p}_j = \frac{\alpha_j}{\alpha_j + \beta_j}$ where

$$\alpha_j = \alpha_{j,0} + \sum_{n=1}^{N_j} i_s^{(n)} \qquad \beta_j = \beta_{j,0} + \sum_{n=1}^{N_j} \left(i_0 - i_s^{(n)} \right) \qquad .$$

In order to let the p parameter of the binomial distribution for the intensity I_s at site vary with continuity, we smooth the estimates $\hat{p}_j, j = 1, \ldots, L$ with the method of least squares, again using an inverse power function $g(d) = (\gamma_1/d)^{\gamma_2}$. In this way it is possible to assign the probability of the intensity decay $Pr\{\Delta I \mid I_0, g(d)\}$ at any distance from the epicenter.

3 Validation

To predict the intensity points an earthquake will generate, on the basis of the knowledge accumulated before its occurrence, we can apply either a predictive probability function for all the points within every R_j band

$$Pr\left\{I_s = i \mid I_0 = i_0, \mathcal{D}_j\right\} = \binom{i_0}{i} \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j)\,\Gamma(\beta_j)} \frac{\Gamma(\alpha_j + i)\,\Gamma(\beta_j + i_0 - i)}{\Gamma(\alpha_j + \beta_j + i_0)}$$

or use a different binomial $Bin(I_0, g(d))$ probability function for the points at distance d from the epicenter, where $g(\cdot)$ denotes the smoothing inverse power function.

We propose here some measures of the degree to which a model predicts the data; for the lack of space we refer just to the predictive distribution.

• **Logarithmic scoring rule** based on the logarithm of a posterior probability (Lindley (*Stat. Science*, 2, 1987); Winkler (*Test*, 5, 1996)). We use the marginal likelihood to evaluate this measure, obtaining the following expression:

$$-\frac{1}{N_s'} \log \prod_{j=1}^{L} \prod_{\substack{n' = 1, \ldots, N_s' \\ i_s^{(n')} \in \mathcal{D}_j'}} \binom{i_0}{i_s^{(n')}} \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j)\,\Gamma(\beta_j)} \frac{\Gamma(\alpha_j + i_s^{(n')})\,\Gamma(\beta_j + i_0 - i_s^{(n')})}{\Gamma(\alpha_j + \beta_j + i_0)}$$

where N_s' is the number of sites, within the L bands R_j, at which the future event will be felt.

• $p(A)/p(B)$ **ratio** between the probability that the fitted model assesses to the realization A and the probability of the predicted value B.

The idea behind this measure is borrowed from the concept of deviance (Read and Cressie, 1988) and is based on a consideration of how much is gained from having predicted B when A occurs. If we indicate the mode of a posterior probability as the predicted value

$$i_{pred}^{(n')} = \arg \max_{i_s = 0, \ldots, i_0} Pr_{pred}(i_s \mid \cdot) \quad ,$$

the error, expressed in probabilistic terms, is given by the geometric mean of the corresponding *odds* in logarithmic scale:

$$odds_{pred} = -\frac{1}{N_s'} \log \prod_{n'=1}^{N_s'} \frac{Pr_{pred}(i_s^{(n')} \mid \mathcal{D}_j)}{Pr_{pred}(i_{pred}^{(n')} \mid \mathcal{D}_j)} \quad .$$

• **Deterministic absolute discrepancy** between observed and estimated intensities at site

$$diff_{pred} = 1/N_s' \sum_{n'=1}^{N_s'} \left| i_s^{(n')} - i_{pred}^{(n')} \right| \quad .$$

4 Some results

Let us consider the ZS4 zonation of central Italy represented in Figure 1. We have studied two earthquakes of IX intensity occurred in zone 47: the Camerino earthquake of 28 July 1799 and the Colfiorito earthquake of 26 September 1997. We have assumed that the shaded
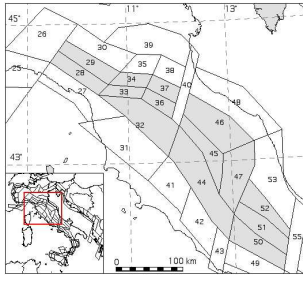
Figure 1: The ZS4 zonation of central Italy (courtesy of the Italian Group for Defence against Earthquakes).

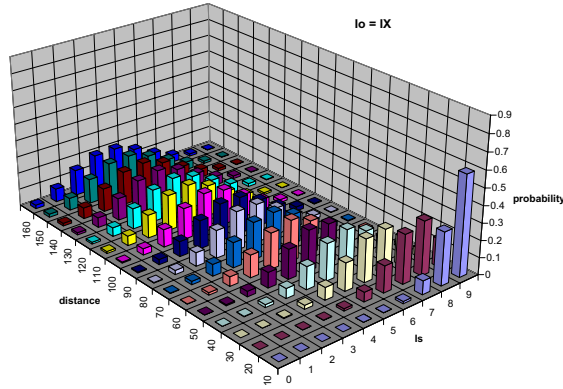| year | | $Pr\{I_s\,|\,\cdot\}$ | | |
validation	criterion	pred.	binom.	logist.
1799	scoring	*1.205	1.405	1.395
B	odds	*0.218	0.648	0.541
	discrepancy	*0.543	0.696	0.696
1997	scoring	1.575	1.789	*1.337
F	odds	*0.301	0.650	0.308
	discrepancy	0.941	*0.838	0.847
1997	scoring	1.402	1.610	*1.171
B	odds	*0.118	0.432	0.259
	discrepancy	*0.441	0.670	0.667

Figure 2: Estimate of the predictive probability function for the intensity I_s.
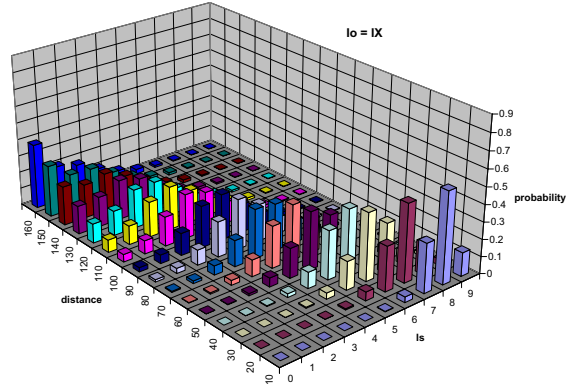

Figure 3: Estimate of the probability function for the intensity I_s drawn from the logistic model.

zones, homogeneous from the viewpoint of kinematic context and expected rupture mechanism according to the zonation, are also homogeneous from the seismic-wave propagation point of view. Figure 2 shows the estimated predictive probability function for I_S, while Figure 3 depicts the probability function produced by the logistic model. The estimation does not include the 1997 earthquake. The Table summarizes the results of the validation performed backward for the 1799 earthquake, and first forward and then backward for the Colfiorito quake respectively. The predictive distribution can be indicated as the all-around best of the probabilistic models examined.

References

Read, T.R.C. and Cressie, N.A.C. (1988) *Goodness-of-fit statistics for discrete multivariate data*, Springer-Verlag, New York.

Rotondi, R. and Zonno, G. (2004) Bayesian analysis of a probability distribution for local intensity attenuation, *Annals of Geophysics*, **47**, 5, 1521-1540.

Zonno, G., Meroni F., Rotondi R., Petrini V. (1995) Bayesian estimation of the local intensity probability for seismic hazard assessment, *Proceeding of the Fifth International Conference on Seismic Zonation*, Nice, October 17-19, 1995, 1723-1729.