Earthquake Engineering, Tenth World Conference © 1992 Balkema, Rotterdam. ISBN 90 5410 060 5

For an adequate use of intensity data in site hazard estimates: Mixing theoretical and observed intensities

M. Mucciarelli & L. Magri
ISMES, Bergamo, Italy

D. Albarello
Earth Science Department, University of Siena, Italy

ABSTRACT: A new approach to the problem of seismic hazard analysis at the site based on intensity data is proposed, in order to obtain more reliable estimates and to take into account the effect of uncertainties involved. A proper use of the historical data allows for a correct retrieval of large data set containing precious information on the most severe shocks, characterized by long return periods. The proposed formulation is based on the use of a distribution function describing (for each earthquake) the probability of being felt with a given intensity at the site, conditioned by its distance from the macroseismic epicenter and by the epicentral intensity. A methodology has been developed in order to combine such probabilities, to compute usual seismic hazard statistics. This approach has been developed mainly to overcome the difficulties in hazard estimates due to the use of ordinal and discrete data (macroseismic intensity) and to avoid misleading results due to the assumption that intensity can be treated as a real number (continuous distribution estimators, attenuation relationships, etc.).

INTRODUCTION

Our knowledge of seismogenetic processes has been greatly increased in the last years, but not enough to allow for an effective estimate of seismic hazard based only on deterministic approaches.
In this situation, a possible estimate of hazard can be obtained by the analysis of site seismic history. Following this approach, it is necessary to use historical macroseimic data, particularly in regions characterized by long return periods.
Using this kind of data two main problems arise. First of all, historical data are qualitative and they can be summarized by the use of macroseismic scales, which are ordinal and discrete.
Second, the attribution of the severity level to the earthquake effects is affected by uncertainties due to the difficoult interpretation of historical information. It is well known that many problems are connected to the translation of old sparse documents in a single intensity value, related to the historical context existing when the source text was written (for a more detailed discussion of current topics in the field of historical seismology see e.g. Vogt, 1991).
In general, for a given site, three possible situations are encountered:

a. Data are sufficient for an univocal attribution of intensity value. This is particularly the case of recent earthquakes or for the sites where effects have been the most intense (macroseismic epicenters);

b. Data allow only the determination of a an interval of acceptable intensity values, possibly attributing to each intensity value a probability value (this is particularly the case of historical data)

c. Data are not available at the site but only, for instance, at the epicenter. This is the a frequent situation, above all when for ingeneering purposes it is necessary to estimate seismic hazard in sites with few inhabitants, where historical information is generally lacking (e.g. dam sites, power plants, etc.).

Due to these problems, the characterization of each intensity felt at the site has to be performed in probabilistic terms, expressing the

probability level associated to each intensity class.
As a consequence, even the measure of the simplest parameters of seismic activity (e.g., the number of eartquakes felt with a given intensity) has to be substituted by a probabilistic estimate of the same parameter.
The aim of this work is the developement of a formalism which main characteristic are the following:

1) Combined use of observed and calculated site intensities.
2) Treatment of intensities as ordinal number without a defined metric.
3) Care in intrinsic variance of data, use of uncertainties as additional information and estimate of the significance level of the results.
4) Proper treatment of completeness periods, assuming different duration for each intensity class.
5) No model imposed a priori to the data (e.g. G-R Law, Poissonian behaviour, etc.).
6) Simplicity and easiness of implementation in an unexpensive computer program.

PROBABILITY LEVELS OF THE FELT INTENSITY.

For each earthquake felt at the site, a probability density function p(I) is defined. This function assigns to each intensity value I the probability that this value represents the true estimate of the felt intensity. In other words, p(I) represents the confidence attributed to the statement that the value I is significant of the true severity of shaking. To assure the probabilistic character of the function p(I), the following assumption must be fulfilled

$$\sum_{i=I_{min}}^{I_{max}} p(i) = 1 \qquad (1)$$

where I_{min} and I_{max} are respectively the minimum and the maximum values of the adopted macroseismic scale.
If data are sufficient for an univocal attribution of intensity equal to I, the function p(i) will be 0 for each value different from I and 1 in correspondence of I.
For each probability density function p(i) it is possible to define the exceedance probability P(i) defined as

$$P(I) = \sum_{j=I}^{I_{max}} p(j) \qquad (2)$$

Several procedures can be adopted to define the function p(I) for each earthquake felt at the site. In the case of observed intensities, the subjective judgement of the researcher which attributes the value of intensity contains an intrinsic uncertainty that must be taken into account.
As observed above, data at the site are not always available for all the earthquakes occurred. In general, information exist about those sites where the effects have been the strongest (epicentral sites). The experimental evidence that severity levels tends to monotonically decrease with distance from the epicentral site it is widely accepted. On this basis it is possible to define a probability density function $R(I/r,I_o)$ where R is the probability that the earthquake has been felt at the site with intensity greater or equal to I, conditioned to the epicentral distance (r) and epicentral intensity (I_o).
The function R may be determined on empirical basis. Its shape is presumably related to the geological features of the region under study and thus R has to be considered as a characteristic of the region where the hazard estimates have to be performed.
Assuming that also the intensity at the epicenter is characterized by a probability density function pe (Io), the probability Ps(I) that the earthquake has been felt at the site with intensity greater or equal to I will be given by the combination of $R(I/r,I_o)$ and $pe(I_o)$. These two probabilities can be considered independent and then the exceedance probability Ps(I) will be given by:

$$Ps(I) = \sum_{j=I}^{I_{max}} p_e(j) . R(I/r,j) \qquad (3)$$

It is worth noting that in the framework of the proposed formalism, the probability density function R represents the equivalent of the attenuation laws largely described in the literature (e.g., Kovesligethy, 1906; Blake, 1941; Howell e Shultz, 1975; Berardi et al., 1990).

On the contrary of the usual attenuation laws, the function R has a probabilistic character wich takes into account the uncertainties involved when this kind of laws are used. These uncertainties are directly implied in the estimate of the intensity at the site with the confidence level associated to it.
For the italian region, Mucciarelli et al. (1990) proposed a relationship of this kind:

$$R' = \frac{e^{a(Ao)+b(Ao).\ln(r)}}{1 + e^{a(Ao)+b(Ao).\ln(r)}} \quad (4)$$

where for each attenuation Ao=1,2..n

$$a(Ao) = 1.00 + 1.95.Ao \quad (5)$$

$$b(Ao) = -1.15 - 0.16.Ao \quad (6)$$

are two empirically determined relationships.

COMPUTATION OF SEISMIC HAZARD AT A SITE

The probability $R'(Ao/\ln(r))$ given by eq (4) can be easily used to estimate a catalogue of site effects where and when historical information are lacking. Then it will be possible to compute mean return periods and probabilities of exceedance vs. time for each intensity level using the probability of each earthquake to be felt wit a given intensity instead of the usual number of felt events. All the demonstrations given in the following represent an example of possible applications of the method, and can be reduced to the usual statistics on the number of events if the proability levels are set to only two possible values: 0 and 1.
Mean return periods and occurrence rates.
The simplest estimators for the mean return period $t(I)$ and the occurrence rate $g(I)$ of an earthquake of intensity greater or equal to I are:

$$t(I) = \frac{T}{N(I)} \quad (7)$$

$$g(I) = \frac{1}{t(I)} = \frac{N(I)}{T} \quad (8)$$

where T is the sampling period (i.e. the duration of the seismic catalog) and $N(I)$ is the number of earthquakes of intensity greater or equal to I expected during time T.
The computation of the expected number of events $N(I)$ may be performed as follows. Let x_i be the Bernoullian random variable describing the exceedance of intensity I for the i-th event in a seismic catalogue, containing n events recorded during the span T, being xi respectively equal to 1 or to 0 if the intensity at the site due to that earthquake whas greater or not than intensity I. The probability density function of x_i is:

$$f(x_i) = p_i^{x_i} \, (1-p_i)^{1-x_i} \qquad i=1,\dots,n \quad (9)$$

The probability p_i, if direct historical information permit to assign a precise intensity value at the site, assumes a value equal to 0 or 1. In the case that uncertainty occurs on historical data, pi may be assigned by expert subjective judgment. If no historical observation are available for a given earthquake at the site, pi coincides with the probability function R' shown in (6) evaluated for the i-th event in the seismic catalogue:

$$p_i = R'(Io_i - I/\ln(r_i)) \quad i=1,..n \quad (10)$$

In eq. (10) Io_i and r_i are respectively the epicentral intensity and the epicentral distance of the i-th earthquake in the sequence. More in general, the probability pi may me assumed equal to $P_s(I)$ as given by eq. (3).
The expected number $N(I)$ is now:

$$N(I) = E(\sum_{i=1}^{n} x_i) = \sum_{i=1}^{n} E(x_i) \quad (11)$$

Considering that the average of the Bernoullian random variable xi is just:

$$E(x_i) = p_i \quad (12)$$

at last the formulae (7) and (8) may be written as:

$$t(I) = \frac{T}{\sum_{i=1}^{n} p_i} \quad (13)$$

$$g(I) = \frac{1}{t(I)} = \frac{\sum_{i=1}^{n} p_i}{T} \qquad (14)$$

An approximate evaluation of the standard deviations s_t and s_g of the two estimators $t(I)$ and $g(I)$ is provided by the following formulas:

$$s_t^2 = \frac{n . s_t^2}{N(I)^2} + \frac{T^2 s_N^2}{N(I)^4} \qquad (15)$$

$$s_g^2 = \frac{s_N^2}{T^2} + \frac{N^2\ n\ s_t^2}{T^4} \qquad (16)$$

where s_t and s_N are the standard deviations respectively of the inter-event times sequence $(t_i)i=1,\ldots,n-1$ and of the estimator $N(I)$ of the expected number of events with intensity at the site greater or equal to I:

$$s_t^2 = \frac{\sum_{i=1}^{n} (t_i - \bar{t})^2}{n-1} \qquad (17)$$

$$s_N^2 = \sum_{i=1}^{n} p_i(1-p_i) \qquad (18)$$

It is worth noting that no assumptions has been made on the probability density function of the intensity in order to compute the mean return period and the occurence rate. Hypoteses like the Gutemberg-Richter law may eventually be verified a posteriory, for instance by observing the allignment of the return periods in a logarithmic plot. This hypothesis, in fact, even if it is a general law concerning homogeneous seismic sources, may be not fullfilled in the case of intensities felt at a site from many sources having different seismic behaviour.
At last, estimators (13) and (14) may be computed considering different complete parts of the catalog, depending on the reference intensity value. As far as the duration of the catalog is concerned, it is important to remark that different results may be obtained if in the duration itself the period between the beginning of completeness and the first felt event, and/or the period the last felt event and the end of the catalog is taken into accont.

Probability of exceedance vs time.

A methodology for the computation of the probability of exceedance vs time will be first proposed for a time series of earthquakes of well known intensities at a site (probabilities pi equal to 0 or 1), and then the methodology will be extended for all cases.
The probability $P(t)$ of recording an earthquake of intensity greater or equal to I in a time span t may be computed as

$$P(t) = 1 - Po(t) \qquad (19)$$

where $Po(t)$ is the probability of no events in time t.
The probability $Po(t)$ may be computed easily under two different hypoteses:
a) the beginning of the time interval t coincides with a seismic occurrence, and the events are indipendent;
b) the beginning of the time interval t may occur in every moment with equal probability.
It must be noted that, in the hypotesis of stationariety of the random process, the two different estimates of $Po(t)$ shown in (a) and (b) are equivalent in convergency, that is for $N, T \to +\infty$.
In general it is possible to evaluate:

$$Po(t) = \sum_{i=1}^{n} Po(t/t_i)\pi(t_i) \qquad (20)$$

where $Po(t/t_i)$ is the probability of recording no events during the time t, conditioned by the hypothesis, having probability $\pi(t_i)$, that the beginning of the interval t will occur in the interval ti of the time series $(t_i)_{i=1,\ldots,n}$.
The probability $p(t_i)$ may be evaluated as:

$$p(ti) = \begin{cases} \dfrac{1}{n} & \text{(hypothesis a)} \\ \dfrac{t_i}{T} & \text{(hypothesis b)} \end{cases} \qquad (21)$$

The probability $Po(t/t_i)$ is computed as:

$$Po(t/ti) = \begin{cases} \wp_i \\ \wp i \dfrac{t_i - t}{t_i} \end{cases} \qquad (22)$$

$$\wp i = \begin{cases} 0 & \text{for } ti \therefore t \\ 1 & \text{for } ti > t \end{cases}$$

From (20), (21) and (22) it is readily possible to obtain:

$$Po(t) = \begin{cases} \dfrac{1}{n} \sum_{i=1}^{n} \wp_i & \text{(hypothesis a)} \\ & (23) \\ \dfrac{1}{T} \sum_{i=1}^{n} \wp_i (t_i - t) & \text{(hyp. b)} \end{cases}$$

In the hypothesis of a Poisson random process the two formulas shown in (23) are equivalent estimators of Po(t), and they converge to exp(-gt) for N, T-> $+_\infty$. This may be verified, for instance, by the Montecarlo method: if a pseudo-random Poisson time series is generated, it is possible to compare the results of the two formulas (26) with the theoretical value exp(-gt).
The two estimators proposed in (23) are of general use, and they need no hypotheses concerning the funcional shape of the probability density function of the random process. Anyway care should be used for the estimator for condition (a), because of the assumption of the stochastic independence of the events, which often may not be considered suitable.
In the case that the intensities recorded at the site are affected by uncertaintes, the formulas (23) may be computed only for hypothesis (b), because situation (a) implicitly needs the knowledge of well determined time intervals t_i. In analogy to formula (26), in this case it is possible more generally to write:

$$Po(t) = \sum_i \frac{t_i}{T} \wp_i \quad (24)$$

Here the intervals ti are not observed inter-event times, but they are all the separate intervals to which different probability values $\wp_i$ correspond, being

$$\wp_i = \Pi_j (1-p_j) \qquad (25)$$

the probability that no one of the j events occurring in the time t exceed the intensity I.
Of course it is not easy to compute manually the probability Po(t) proposed in formula (24) and (25), but anyway it is possible to do it by a computer routine that scans the time series, divides the catalog in many equal-probability steps, and computes probability Po(t).
Also for the computation of probability of exceedance in time, it is important to highlight the fact that no a priori hypotheses on the probability density functions of the random process have been made.

The intensity data are often used in a not proper way for seismic hazard estimates. In fact the semi-qualitative character of the intensities do not permit the use of techniques developed for instrumentally measured data like magnitudes. Nevertheless intensity data must be used whenever available, because they supply information about seismic events whose return periods are longer than the length of instrumental catalogs.
To this purpose a new approach has been proposed, where to every intensity value a probability of recording is associated. In the developed method only the intensity values prescribed by the intensity scale are used, avoiding all the problems related to a discrete and non metric scale.
The main advantages consist of:
-mixed treatement of observed felt intensities and theoretical data obtained by probabilistic attenuation relationships;
-correct use of the uncertainties in the seismic hazard estimates;
-full consideration of the expert's judgement relevant to the attribution of the intensities and to the involved uncertainties;
-easy implementation in a personal computer code.

References

Berardi R., Mucciarelli M. and Magri, L., 1990. Do different experts and computer program agree on the interpretation of the same intensity map? Proceedings of the XXIII EGS Assembly, Barcelona, in press.

Blake, A., 1941. On the estimation of focal depth from macroseismic data. Bull. Seis. Soc. Am., 31, 225-231

Cox D.R., Lewis P.A.W., 1966. The statistical analysis of series of events. Mathwen, 158-160.

Howell B.F. e Shultz T.R., 1975. Attenuation of Modiefied Mercalli Intensity with distance from epicenter. Bull. Seis. Soc. Am., 651-665

Koveslígethy (Von), R.. 1906. Seismonomia. Boll. Soc. Sism. Ital., XIV.

Mucciarelli, M., Albarello, D. e Magri, L, 1990, Un metodo alternativo alle leggi di attenuazione nelle stime di rischio sismico, proceedings of IX Conference of CNR-GNGTS, Rome, in press.

Vogt, J. 1991, Some glimpses at historical seismology, in M. Stucchi, D. Postpischl and D. Slejko Eds. "Investigation of historical earthquakes in Europe", special issue, Tectonophysics, 193, 1-3.