2003 Pacific Conference on Earthquake Engineering

Modeling magnitude distribution for local hazard evaluation: a new approach

G. Grandori & E. Guagenti

Politecnico di Milano, Italy.

L. Petrini

Università di Pavia, Italy.

ABSTRACT: For the design of ordinary buildings, many seismic codes prescribe to assume the peak ground acceleration corresponding to a return period of the order of 500 years at the site, a(500). This quantity is dominated by the size distribution of strong earthquakes that are expected in the zone. On the other hand, the number of strong earthquakes in available catalogs is generally not sufficient for the statistical validation of a magnitude distribution model, nor for a meaningful comparison between competing models. Besides this epistemic uncertainty, the problem is obviously affected by the inevitable statistical uncertainty due to the fact that the catalog is a random sample drawn from the real magnitude distribution (in the favourable hypothesis that such a distribution exists). In this paper we give a definition of the credibility of a magnitude distribution model as regards the evaluation of a(500) at a specific site. We find that the comparison between the credibilities of competing models opens new statistical prospects. In particular the hazard analysis for an Italian site shows that the systematic application of the concept of model credibility may lead to a reduction of both statistical and epistemic uncertainty.

1 INTRODUCTION

For the design of ordinary buildings at earthquake prone sites, it is generally suggested to use the peak ground acceleration (PGA) having 10% probability of being exceeded in 50 years (UBC) or, equivalently, the PGA corresponding to a return period of the order of 500 years (Eurocode). Let us call **a** this value of PGA, which is a synthetic measure of local hazard, remarkably important from the engineering point of view. This paper referes to the uncertainty in the evaluation of **a** that is connected with modeling magnitude distribution. This means that all other elements that contribute to the evaluation of **a** at the considered site (like the distribution of events in space and time, and the attenuation law) are supposed to be known and independent of the magnitude model. As a consequence, if a magnitude distribution function $F_M(m;\overline{J})$ is given, both as form and parameters ($\overline{J}$ can be a vector), then a known procedure *Z*, applied to F_M, gives the value of **a** at the site. From now on, a distribution F_M with known parameters will be briefly indicated with an upper index; hence, for instance:

$$\mathbf{a}^0 = Z(F^0). \tag{1}$$

On the other hand, a distribution that contains estimable free parameters will be indicated with a lower index, for instance F_0.

The quantity **a** at a specific site is dominated by the distribution of strong earthquakes, while the statistical data on these events are scarce: in most seismic zones the seismic record is too short to validate a model of the size distribution of strong earthquakes, so that different models can explain with the same success the available data, even if they lead to significantly different values of **a** (Kagan 1993; Wu et al. 1995).

Beyond discussion the fact that in most cases a model of the size distribution of strong earthquakes cannot be statistically validate. What we make is to seek the way to compare, on a statistical basis, the credibility of different models as regards the result of their application to the evaluation of **a** at a specific site. Such kind of credibility is measured by the "credibility index" Δ, defined as follows.

We assume, firstly, that a real magnitude distribution F^0 exists, and that its features are independent of the space and time coordinates of the events. Then the real value of **a** is

$$\mathbf{a}^0 = Z(F^0)$$

Suppose now that a certain model F_r is proposed for magnitude distribution. If a random sample with ν events is drawn from F^0, the parameters of the model can be estimated by fitting the size-ν sample. Then the procedure Z leads to an estimated value $\hat{\mathbf{a}}$. All the size-ν samples that could be drawn from F^0 give rise to the random variable $\hat{A}_r$ which is the sampling distribution of $\hat{\mathbf{a}}$ and depends on F^0, ν and F_r.

As suggested in a previous paper (Grandori et al. 1998), we call "credibility index" of the model F_r with respect to F^0 the probability

$$\Delta_r^0 = \Delta(F^0, \nu, F_r) = P\{(1-h)\mathbf{a}^0 \le \hat{A}_r \le (1+h)\mathbf{a}^0\}, \tag{2}$$

where h settles a conventional interval around $\mathbf{a}^0$. In what follows we assume h = 0.2.

Shortly: Δ_r^0 is the probability that the model F_r, starting from a random size-ν sample, leads to estimate **a** with less than 20% error in respect to $\mathbf{a}^0$.

In the case of two models F_r and F_s, note that $\Delta_r^0 > \Delta_s^0$ does not mean that F_r leads to a distribution in some way "closer to F^0" compared with F_s. It means that F_r leads to a value of **a** "closer to $\mathbf{a}^0$" in probabilistic terms, compared with F_s.

Given the real (or a conjectural) distribution F^0 and the sample-size ν, with a Monte Carlo simulation it is possible to calculate the credibility Δ_r^0 of a given model F_r.

Now the question is: can the credibility index be helpful given that the real distribution F^0 is not known? The experiments that we describe hereinafter show that it can. The experiments refer to the site Alfa, which is located in Irpinia, one of the most active zones of Italy. The seismological features of the zone, that we assume in order to define the procedure Z valid for the site Alfa, are described in the Appendix.

2 FIRST EXPERIMENT

Suppose that the real distribution F^0 has the double exponential form

$$1 - F_0 = exp\{exp[b(m_0 - u)] - exp[b(m-u)]\} \tag{I}$$

and that the real parameters are

$$\beta = 0.3,\ u = 0. \tag{I,a}$$

The lower magnitude level is $m_o = 4$ and the number of events during the last 300 years is ν = 40 (see Appendix).

The distribution F^0 is completely defined, so with Equation (1) we can calculate $\mathbf{a}^0$ at the site Alfa. We can also draw from F^0 as many size-ν random samples S^0 as we want.

Now we ask a first scientist to describe how he would estimate the quantity **a** on the basis of the information contained in one of the random samples S^0. This scientist, thanks to a special supernatural talent, “divines” the real form (I) of the magnitude distribution, that can be classified as the “right” model. Then the obvious answer of this first scientist is: from the sample S^0 I estimate the parameters of the mathematical model F_0 and so I obtain (through the procedure Z) the estimate of **a**. In order to check the credibility of this procedure, we repeated the estimate of **a** with 1000 random samples S^0; we obtained as a consequence the credibility of the “right” model: $\Delta_0^0 = \Delta(F^0, í, F_0) = 0.54$.

A second scientist is requested to solve the same problem, under the same conditions (namely just by looking at one of the samples S^0). This second scientist does not have special talents and he does not like mathematical models. He says: from the sample S^0 I derive the cumulative frequency polygon (CFP), which provides an empirical distribution F^*. From F^*, Equation (1) gives directly **a***, which is my estimate of **a**. By repeating this procedure with 1000 random samples S^0 we obtained the credibility of the “polygon” model: $\Delta_*^0 = 0.60$.

Unexpectedly, we find that for the site Alfa, if the real F^0 is defined by (I), (I,a), the credibility Δ_*^0 of the empirical “polygon” model is of the same order as the credibility Δ_0^0 of the mathematical “right” model; i.e. the two procedures are affected by the same statistical uncertainty. However, the procedure based on the mathematical model (in the absence of special supernatural talents) is also affected by an epistemic uncertainty which is difficult to control.

The same experiment has been carried out by assuming for the real parameters the values

$$\beta = 0.35,\ u = 0.4. \qquad \text{(I,b)}$$

Moreover, as an alternative hypothesis, a Weibull distribution has been assumed as real form F_0:

$$1 - F_0 = exp\left[-(rm)^a + (rm_0)^a\right], \qquad \text{(II)}$$

with three different sets of real parameters

$$a = 4.0,\ \rho = 0.21, \qquad \text{(II,a)}$$

$$a = 4.0,\ \rho = 0.24, \qquad \text{(II,b)}$$

$$a = 3.0,\ \rho = 0.24. \qquad \text{(II,c)}$$

The results obtained are summarized in Table 1, both in the case $\nu = 40$ and in the case $\nu = 20$.

Table 1. Summary of the first experiment results.

		n = 40				**n = 20**			
Real form F_0	**Real param.**	**a^0**	Δ_0^0	Δ_*^0	**Gain** $\frac{\Delta_*^0 - \Delta_0^0}{\Delta_0^0}$	**a^0**	Δ_0^0	Δ_*^0	**Gain** $\frac{\Delta_*^0 - \Delta_0^0}{\Delta_0^0}$
(I)	(I,a)	.29	.54	.60	.11	.23	.47	.62	.32
	(I,b)	.21	.61	.70	.15	.16	.58	.42	-.28
(II)	(II,a)	.38	.59	.52	-.12	.30	.47	.40	-.15
	(II,b)	.21	.69	.78	.13	.17	.62	.68	.10
	(II,c)	.42	.45	.36	-.20	.32	.39	.40	.02
				average	.01			average	.00

For all the five conjectural real distributions (which correspond to a rather wide fan of values $\mathbf{a}^0$), Δ_*^0 is of the same order as Δ_0^0, so that the average gain is near to zero in both cases ν = 40 and ν = 20. In this second case, as expected, larger fluctuations are observed; the trend, however, does not change.

The parameters of the "right" models are estimated by maximum likelihood. We carried out some check with moments method: in general it leads to lower values of Δ_0^0.

3 SECOND EXPERIMENT

Now we look for an improved version of the polygon model, with the aim of increasing its credibility. The idea is to associate the information contained in the CFP with an Hybrid mathematical model able to emphasize the role of strong earthquakes.

The mathematical model has the simple form of the characteristic magnitude model used by Wu et al. (1995):

$$1-F_M(m,J)=\begin{cases}(1-p)\dfrac{e^{-bm}-e^{-bm_1}}{e^{-bm_0}-e^{-bm_1}}+p & [m_0 \le m \le m_1]\\ p\dfrac{m_2-m}{m_2-m_1} & [m_1 \le m \le m_2]\end{cases} \tag{3}$$

The parameters of this Hybrid model H (Fig. 1) are m_1, m_2, the b value of the exponential part and the relative frequency p of characteristic earthquakes. With four parameters, we would generally expect a high statistical uncertainty, and hence a low credibility. Yet we will show that a high credibility of the model can be reached thanks to two key-provisions. One is to consider the parameters m_1 and p constant for the hazard analysis of the site, whatever the sample and the conjectural real F^0 may be. The second is to estimate the parameter m_2 from the available sample S^0 in such a way as to maximize, instead of the likelihood function, the credibility index Δ^*_H of the Hybrid model with respect to the CFP of S^0. Given m_1, p and m_2, the parameter b is simply estimated from the mean μ^0 of S^0.

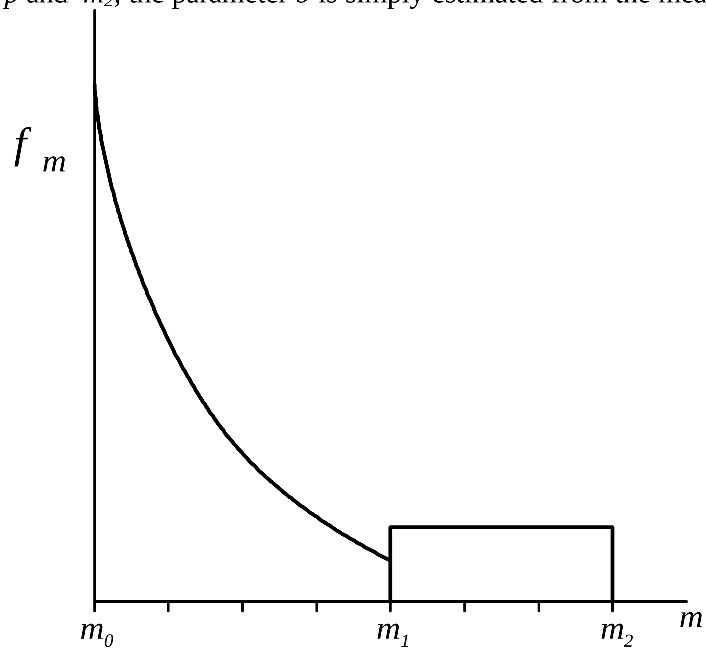

Figure 1. Probability density function of the Hybrid model

Let us now see in detail the procedure with reference to our site Alfa.

First we fixed in a rather arbitrary way the parameters m_1 and p:

$$m_1 = 5.9,\ p = 0.08. \qquad (4)$$

These values are justified <u>as a first attempt</u> by a look at many samples drawn from each considered F^0, and by what reported in the literature about characteristic earthquakes.

Then, the steps to estimate m_2 from a sample S^0 (with mean μ^0) drawn from a given F^0 are:

1. from S^0 obtain the empirical distribution F^* and calculate $\mathbf{a}^*$;
2. fix a first tentative value $m_{2(1)}$, so obtaining a one-parameter model $H_{(1)}$;
3. from F^* draw many samples S^* (with mean μ^*) and calculate the credibility $\Delta^*_{H(1)}$ of the model $H_{(1)}$ when applied to the estimate of $\mathbf{a}^*$;
4. repeat the steps 2) and 3) with different values $m_{2(i)}$, and find the "optimum" m_2, which corresponds to the maximum $\Delta^*_{H(i)}$; this is the estimate of m_2 derived from S^0

(In order to control the rare anomalous extreme values of m_2 it is convenient to use a modified value m_2' in the following cases: $m_2' = 6$ if $m_2 < 6$; $m_2' = 0.8m_2 + 0.2m^0$ if $m_2 - m^0 > 2.5$).

Being given m_1, p, m_2, the last parameter b is derived from μ^0, so that the Hybrid distribution is completely defined together with its estimate $\hat{\mathbf{a}}$.

Repeat this estimate with many random samples S^0 and get the credibility Δ^0_H of the procedure.

The Hybrid model associated with the polygon model for the estimate of the parameters will be called hereinafter the "Hybrid-polygon" model. For all the five distributions F^0 of Table 1, the credibility Δ^0_H has been calculated; it is shown in Table 2, compared with the results of Table 1.

Table 2. Summary of the second experiment results.

		n = 40						n = 20					
Real F_0	**Real param.**	$\mathbf{a^0}$	Δ^0_0	Δ^0_*	Δ^0_H	**Gain** of Δ^0_*	**Gain** of Δ^0_H	$\mathbf{a^0}$	Δ^0_0	Δ^0_*	Δ^0_H	**Gain** of Δ^0_*	**Gain** of Δ^0_H
(I)	(I,a)	.29	.54	.60	.68	.11	.26	.23	.47	.62	.70	.32	.49
	(I,b)	.21	.61	.70	.79	.15	.30	.16	.58	.42	.58	-.28	.00
(II)	(II,a)	.38	.59	.52	.71	-.12	.20	.30	.47	.40	.60	-.15	.27
	(II,b)	.21	.69	.78	.81	.13	.17	.17	.62	.68	.70	.10	.13
	(II,c)	.42	.45	.36	.57	-.20	.26	.32	.39	.40	.49	.02	.26
					average	.01	.24				average	.00	.23

The "Hybrid-polygon" model leads to a meaningful improvement of the results. The average gain, compared with the "right" model, is remarkable; and, what is also important, the gain is never negative.

The arbitrary choice of the parameters m_1 and p does not invalidate the importance of the results. On the contrary, the fact that different choices could lead both to worse and to better results means simply that the problem allows some further margin of improvement.

4 THIRD EXPERIMENT

In the preceding experiments, the conjectural real forms F_0 had two parameters and it seemed obvious that they were the free parameters of the "right" model. As a matter of fact, it is not always obvious to decide which are the free parameters of the "right" model, (Kagan et al., 2001).

Consider as a first instance the truncated exponential distribution

$$1-F_0=\frac{e^{-bm}-e^{-bm_1}}{e^{-bm_0}-e^{-bm_1}}\,, \qquad \text{(III)}$$

which is frequently favored. The value of m_1 can be chosen on the basis of physical concept, so that from the statistical point of view the model has only one free parameter. Otherwise, especially when the physics of the sources is not well known, both *b* and m_1 can be considered as free parameters.

Look (Table 3) what happens, for instance, in the case of our Irpinia site Alfa.

Table 3. Third experiment first instance results (n = 40).

Real F_0	**Real param.**	**a^0**	Δ^0_*	Δ^0_H	**"Right" model parameters**	Δ^0_0
(III)	m_1= 7 b = 1.1	.28	.70	.78	m_1 free b free	.60
					m_1 = 7.0 b free	.87
					m_1 = 7.5 b free	.48

In this case, too, the value of Δ^0_H is very good. Note that all the three values Δ^0_0 of the last column are obtained on the basis of the "divination" of the real form F_0. Therefore, they are affected by a certain level of epistemic uncertainty. In the first line both the parameters are free, so that no further epistemic uncertainty affects Δ^0_0, which is anyhow much lower than Δ^0_H. In the second line the "divination" regards <u>also the real value of the parameter m_1</u>, so that the epistemic uncertainty becomes very large. This is confirmed by the third line: if m_1 is wrong, a great reduction of Δ^0_0 is observed. In conclusion, in this case too the "Hybrid-polygon" model should be favored.

As a second instance we considered the case in which the real distribution has the Hybrid form (Eq. 3), with the real parameters shown in Table 4.

Table 4. Third experiment second instance results (n = 40).

Real F_0	**Real param.**	**a^0**	Δ^0_*	Δ^0_H	**"Right" model parameters**	Δ^0_0
(3)	m_1 = 6.1 m_2 = 6.8 b = 1.9 p = .06	.26	.66	.79	m_1 = 6.1 $\left.\begin{matrix} m_2 \\ b \\ p \end{matrix}\right\}$ *free*	.57

Again, the value of Δ^0_H is very good, in spite of the fact that m_1 and *p* in the model are different from the real ones. As regards the "right" model, with the free parameters estimated by maximum likelihood, the credibility Δ^0_0 is much lower than Δ^0_H even in the case in which only three parameters are free, and for m_1 the real value 6.1 is adopted (last column in Table 4). Obviously, by adopting the real values for two parameters (e.g. m_1 and m_2, or m_1 and *p*) it is possible to obtain $\Delta^0_0 > \Delta^0_H$. The same comments as before about the epistemic uncertainty are here appropriate.

5 DISCUSSION AND CONCLUSIONS

To look at the credibility of magnitude distribution models as defined by Equation (2) is certainly a less general approach compared with the traditional approach based on geophysical knowledge and on fitting the catalogue data. However, thanks to the accepted limitations (use of the models for the evaluation of a specific quantity at a given site) the new approach allows for more meaningful statistical comparison between models. As we have seen, at least for a specific Italian site, this may lead to a reduction of both statistical and epistemic uncertainty.

The results obtained till now do not allow for general methodological conclusions, yet they clearly support the ides that the comparison between the credibility of competing models opens new statistical prospects. The following is a paradigmatic example.

Suppose that for a given site a traditional hazard analysis has been completed and that the final decision is to adopt for magnitude distribution a specific form $F_0 = F_M (m;\ \overline{J})$. Choose a set of different vectors $\overline{J}$ such that they lead to an exhaustive picture of plausible distributions with the form F_0. Then fix tentative values of m_1 and p and calculate $\Delta^0{}_0$, $\Delta^0{}_*$, $\Delta^0{}_H$ for each vector $\overline{J}$ (eventually repeat the calculation with different values of m_1 and p in order to get better results for $\Delta^0{}_H$).

The whole of the results offers interesting information. First, looking at all values $\Delta^0{}_0$ one gets an idea about the range in which the credibility of the "right" model is likely included. Second, if the "right" model is always defeated by one of the other two models, this one should obviously be preferred.

Finally, the results may suggest a more ambitious attempt, based on the repetition of the above mentioned analysis with different forms F_0 such as to make up a reasonably exhaustive set of plausible forms F_0. If the "Hybrid-polygon" model is never clearly defeated by the "right" model, then one could say that, for the considered site, the "Hybrid-polygon" model is statistically validated.

REFERENCES:

Grandori, G., Guagenti, E. & Tagliani, A. 1998. A proposal for comparing the reliabilities of alternative seismic hazard models. *J. of Seismology*. 2: 27-35.

Kagan Y.Y. 1993. Statistics of characteristic earthquakes. *Bull. Seims. Soc. Am.*.83: 7-24.

Kagan Y.Y. & Schoenberg F. 2001. Estimation of the upper cutoff parameter for the tapered Pareto Distribution. *J. Appl. Probab.*.38(A): 911-918.

Wu S.C., Cornell C.A. & Winterstein S.R. 1995. A Hybrid Recurrence Model and its Implications on Seismic Hazard Results. *Bull. Seims. Soc. Am.*.85: 1-16.

APPENDIX

The events are supposed to be uniformly distributed over the zone of Figure 2, and the magnitude distribution is supposed to be independent of the coordinates of the events. The number of events during the last 300 years is $\nu = 40$. and the mean number of events per year is assumed $\lambda = 40/300 = 0.13$.

The local PGA depends on the epicentral distance R:

$$PGA / g = \frac{ae^{0.8M}}{(R + 25)^b}$$

with R (in Km) ≥ 10, a = 1.51, b = 1.82 (see Fig. 3).

Given a magnitude distribution, the value of **a** at the site Alfa is numerically estimated on the basis of a 40,000 years synthetic catalog. This is what we called procedure Z for site Alfa. As an alternative hypothesis, in some elaboration also the case $\nu = 20$ has been considered.

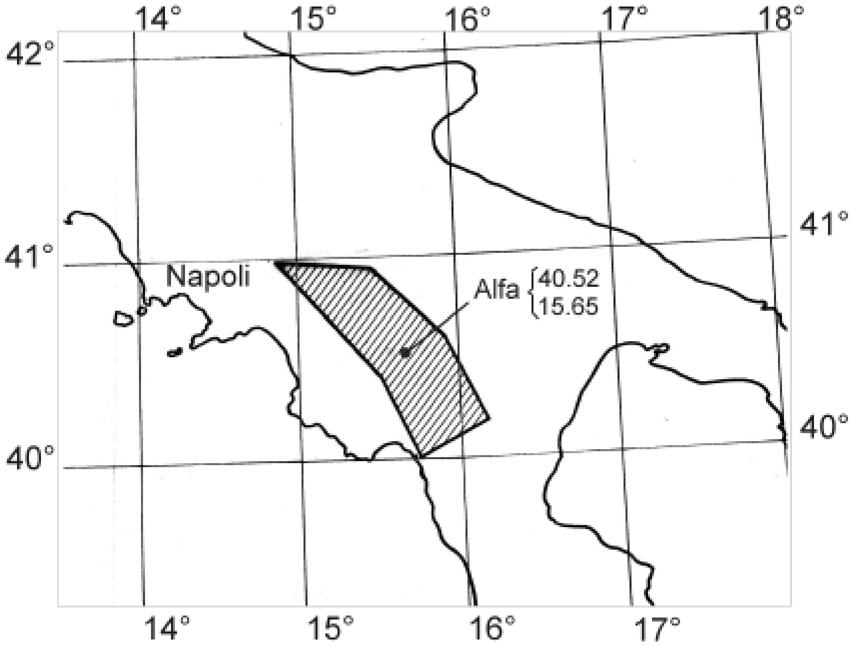

Figure 2. Position of the site Alfa in a seismic zone of Southern Italy

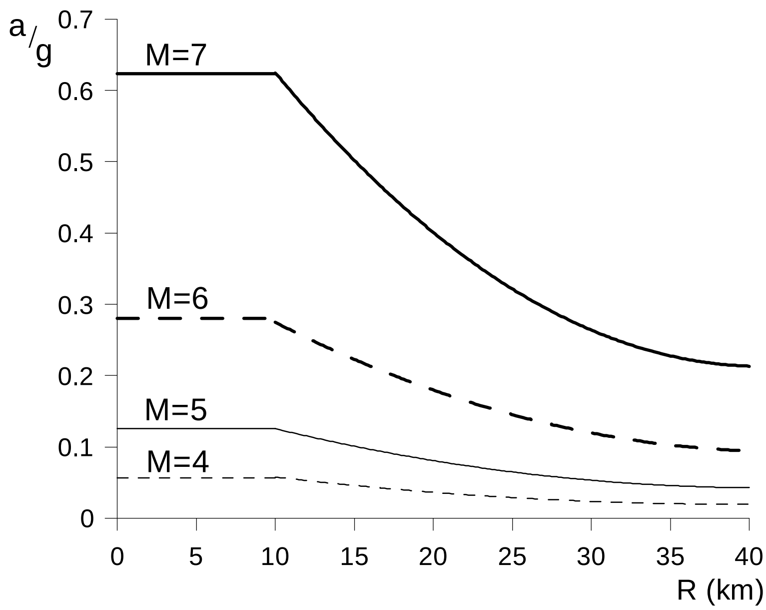

Figure 3. Attenuation law