A RE-EVALUATION OF THE CURRENT SEISMIC HAZARD ASSESSMENT METHODOLOGIES

H. C. Shah (I)
W. M. Dong (II)
Presenting Author: H. C. Shah

SUMMARY

This paper first reviews the widely used seismic hazard assessment methodologies and then, specifically, evaluates the problems and pitfalls of a typical currently used empirical probabilistic approach. For each step of this approach, from source modeling and recurrence relationship to attenuation law, suggestion to improve our current procedures is made. This includes the use of pattern recognition theory in source modeling, maximum entropy principle in information processing, a refined Bayesian model for incorporating geophysical and geological information for recurrence relationship, and a geophysical-based theoretical approach for evaluating the attenuation law. Finally, an approximate uncertainty analysis for the whole seismic hazard methodology and its influence on the engineering decisions are presented.

INTRODUCTION

During the last 15 years, methods for evaluating seismic hazard have reached a certain level of maturity. Some of the empirical methods have been used by researchers and practicing engineers in a routine manner. Most of the major engineering projects use these methods in developing lateral load-resisting criteria. Generally, the hazard evaluation follows four steps: source modeling, recurrence relationship, attenuation law, and site severity assessment (Fig. 1). There are two different approaches in pursuing this hazard evaluation: deterministic and probabilistic. Deterministic approach does not take into account the uncertainties in the size, location and attenuation of the seismic events. Instead, it adopts most conservative values for all these factors such as maximum magnitude, shortest distance, etc., resulting in extremely large design requirements. For most structures, these highly conservative design values cannot be justified economically for use. In recent years, even for critical facilities, probabilistic methods have been used. Probabilistic approach incorporates uncertainties in seismicity and attenuation, and provides the probability distribution of seismic severity parameters at the site. Even for this probabilistic approach, there are many problems that need re-evaluation and improvements. In this paper we will review this approach step by step, and present some suggestions for improvement.

SOURCE MODELING

In the regions where seismic sources are relatively well defined along the plate boundaries or faults, the active fault approach is used for seismic

(I) Professor of Civil Engineering, Stanford University, Stanford, CA.

(II) Visiting Scholar, The John A. Blume Earthquake Engineering Center, Stanford University, Stanford, CA.

hazard evaluation. Seismic sources could be modeled as point, line, or area source (Fig. 1a). In this case, the seismicitiy is usually relatively high and the source-to-site distance is reasonably well defined. However, for many introplate regions, the seismicity is diffused over a large area without any identifiable faults. In these regions, tectonic province approach is used which delineates different provinces as area sources. Each province is assumed to have a homogeneous seismicity, resulting in large uncertainty. Due to lack of sufficient historical and geological evidence, there is no unique and consistent way to delineate the provinces. Expert opinions often play a key role in the source configuration, resulting in personal biases and some arbitrariness. Considering the successful development of pattern recognition in many fields, the authors presume that the clustering analysis combining with fuzzy information from the experts can be used in order to reduce the arbitrariness and uncertainty (Ref. 1). For example, in delineating basically "round" provinces, the clustering analysis can be conducted by minimizing the objective functional

$$J_w(U,v) = \sum_{i=1}^{c} \Big(\sum_{\underline{x}_k \varepsilon \underline{u}_i} w_k \, \|\underline{x}_k - \underline{v}_i\|^2 \Big) \qquad (1)$$

where U is the ensemble of the provinces,

- c is the number of provinces,
- $\underline{u}_i$ is the i^{th} province,
- $\underline{v}_i$ is the center vector of privince i,
- $\underline{x}_k$ is the vector of historical epicenter,
- $\underline{w}_k$ is the weight of each epicenter, considering the magnitude and the confidence of the data.

For other shapes of provinces, the appropriate clustering criterion has to be used to get a rational delineation.

Usually the expert opinion is very vague, mostly using language descriptions. For example, when determining the activity of a fault, one has to investigate the ground surface features which is described by visual inspection, imagery technique, and drilling and trenching, etc. The information obtained from these methods is usually very imprecise and not well defined. Hence fuzzy logic is needed to infer the activity from that information. Also such a logic can help to combine opinions of different experts in a consistent manner.

SOURCE SEISMICITY

The assessment of source seismicity has mostly depended on the recorded events, but the data base on a given source is often incomplete and nonhomogeneous in time. In the past, due to low population density and lack of interest in earthquake activity, only large events were recorded. With increasing instrumental coverage, intermediate and small earthquakes have been recorded with more frequency, producing an apparent increase in seismicity with time which biases the statistics from the catalog of data. A data adjustment is usually necessary to get a reliable hazard analysis.

For the occurrence of earthquakes, there are two kinds of models: time-independent and time-dependent models. Traditionally, the homogeneous Poisson model has been used to describe the occurrence of earthquakes. The common Gutenberg & Richter (GR) relationship $\ln N(m) = \alpha + \beta m$ implies the exponential distribution of magnitude with unlimited maximum magnitude. The parameters are determined by the least-squares (LS) method. The fact that the magnitude in a specific region has an upper bound leads to a modified GR relationship by truncating the magnitudes at the maximum possible magnitude (Fig. 1b) (Ref. 2). This truncated model has good agreement with world-wide data as well as with data for smaller regions. It should be noted that the parameters α and β estimated by the LS would always deviate from the population distribution parameters if the population would follow the modified GR relationship (Ref. 3). On the average, the estimates α and β would be greater than the population parameters. Figure 2 shows the results for a specific case. If the observed mean rate α is used in the modified GR model, the larger estimated β from LS would underestimate the probability of major events. On the other hand, using the estimated mean rate α would increase the energy level. Figure 2c shows the ratio between energy release of the population and the GR model. To avoid this bias, the authors have suggested in a recent paper (Ref. 4) that the Maximum Entropy Principle (MEP) or maximum likelihood method be used to get an unbiased estimate of parameter β, i.e.,

$$f_M(m) = \frac{\beta\, e^{-\beta m}}{e^{-\beta M_{min}} - e^{-\beta M_{max}}} \qquad M_{min} \le m \le M_{max} \tag{2}$$

where β is determined by the constraint condition

$$\int_{M_{min}}^{M_{max}} m\, f_M(m)\, dm = \frac{1}{\beta} + \frac{M_{min}\, e^{-\beta M_{min}} - M_{max}\, e^{-\beta M_{max}}}{e^{-\beta M_{min}} - e^{-\beta M_{max}}} = \overline{M} \tag{3}$$

in which $\overline{M}$ is the observed mean magnitude = $\sum n_i m_i / \sum n_i$.

If another constraint of mean seismic moment $\overline{M_o}$ is added, the MEP would lead to a generalized GR relationship.

Since the historical data are usually scarce, one would not expect a reliable prediction based only on the data, especially for large events. As an example, estimation of seismic hazard in the San Francisco Bay Area from data obtained between 1907 and 1983 will underestimate the hazard. On the other hand, data from 1900 to 1983 may overestimate the hazard if the mean return period for large events is more than 200 years. In such cases, it would be prudent to use geological and/or geophysical information to anchor the rate of occurrence of large events.

There are three methods by which geological and geophysical information about the occurrence of large earthquakes can be used. One way is to use the average seismic moment to "anchor" the values of M_{max}, α, and β (Ref. 5). The geophysical data are assumed to have no uncertainty. This method does not consistently combine the short-range historical data with long-range geological and geophysical data. The second method uses the concept of modified maximum entropy. Based on the information on M_{max} and its average return period, the probability distribution function for magnitudes is obtained. This distribution has the property that it is minimally biased and is consistent with the type and level of information (Ref. 4). The third method considers the uncertainty in the geological/geophysical information. Combination of such data with historical data is done by incorporating relative "weights" through Bayes' model (Ref. 6). Any one of these methods achieves one objective and that is to use all the available information. This gets us away from relying too much on short historical data.

All the models discussed above assume that earthquakes (small, medium, or large) are independent events. The elastic rebound theory and the available evidence suggest that the assumption of independence may be practical but not realistic. To get around this problem, researchers have suggested time-dependent models. Savy et al. (Ref. 7) have suggested a nonhomogeneous Poisson process to model the time-dependence of seismic events. Nishioka and Shah (Ref. 8) have used a simple Markov chain model to estimate probabilities of occurrence of events over a specified time t. Patwardhan et al. (Ref. 9) have suggested a semi-Markov process in which the period of quiescence is dependent upon the magnitude of the last and the next future event. Recently, "time-predictable" and "slip-predictable" models have been suggested (Ref. 10). The time-predictable model assumes that an earthquake would occur when the accumulated stress reaches a certain threshold. Since the rate of stress increase is assumed to be constant, the time to the next event is predictable but the magnitude is random (Fig. 3a). The slip-predictable model assumes that after each event, the stress along the fault drops to zero. The magnitude of the next event depends on the time from the last event. The longer the holding time, the bigger the event due to a constant rate of increase of stress (Fig. 3b).

All these time-dependent models under investigation seem to be more realistic than the time-independent models currently used. However, we have to be careful in our unconditional adoption of these models. They require estimation of many constants and parameters that are currently either not known or the data base is so small that they cannot be estimated with any reasonable reliability or confidence. Thus estimation of hazards from such "realistic" models may still be full of assumptions and uncertainty. A false sense of security may be one of the dangers that we should beware of.

ATTENUATION OF GROUND MOTION

The type and amount of attenuation of seismic ground motion depends on many factors such as the size of the event, the type of fault mechanism, transmission path, distance and local soil condition of the site. The commonly used empirical attenuation relationship incorporates some of these parameters but generally leaves out important variables such as the azimuth between the source and the site, and the parameters that identify the fault rupture mechanism. Equation 4 shows the commonly used empirical attenuation function

$$PGA = f(M, R, b_1, b_2, b_3, c) \tag{4}$$

where PGA = Peak Ground Acceleration,
M = magnitude,
R = distance from the source to the site,
b_1, b_2, b_3 = regression constants, which depend on the type of data, site condition, transmission path, etc.,
C = saturation effect, depending on magnitude.

The uncertainty in explaining the "load effect" at a site due to an earthquake using such crude empirical equations is considerable. Large error terms are common, indicating that attempts to quantify site severity parameter are at best crude. The problems in using Eq. 4 are two-fold. First, the ground motion severity cannot be truly represented by a single parameter such as the

PGA. Second, the equation leaves out important contributing factors such as the azimuth, stress drop, velocity of rupture, etc. In recent years, attempts have been made to rectify this but without much success due to lack of reliable data. Generating ground motion severity parameters by using geophysical models such as the normal mode analysis (Ref. 11) may provide one bit of additional input to our data base. Such analytically generated values, which would be functions of distance, magnitude, azimuth and geophysical properties of fault rupture may be helpful in refining and improving our attenuation relationships. Some recent work on the use of pattern recognition may also provide a better tool in processing the data from various diverse source.

UNCERTAINTY ANALYSIS

Due to limited knowledge of source and propagation mechanism and scarcity of data, there are large uncertainties in each step of the seismic hazard assessment. In source modeling, identification of a tectonic province may depend on expert opinion or a consensus of many opinions. This averaging results in large errors. Even for tectonic plate boundary regions, source configuration is not always properly and accurately modeled. Spatial and temporal distributions of occurrence and their homogeneity or lack of it may introduce further uncertainties. Methods of combining expert opinion, either through Bayesian analysis or through fuzzy set theory, also result in more uncertainty. How does one get a handle on this uncertainty level? This is one of the most important unresolved aspects of probabilistic hazard analysis.

As an approximate estimation, assume that the site severity parameter (e.g., PGA) has Type-II extreme value distribution:

$$F_A(a) = \exp\left[-\left(\frac{a}{u}\right)^{-k}\right] \tag{5}$$

with $k = 2.3$–3.3 (Ref. 12), then the coefficient of variation for A would be

$$V_A = \sqrt{\frac{P(1 - 2/k)}{P^2(1 - 1/k)} - 1} = 0.57\text{–}1.38 \tag{6}$$

which, compared to the coefficients of variation for other structural loads (0.07–0.37), is much larger.

If one further considers the uncertainties in duration and frequency context, the overall coefficient of variation could be even larger (perhaps, 1.0–1.5). However, it should be noted that the engineers should not be discouraged by this large coefficient of variation. The coefficient of variation is not an absolute measure of uncertainty. For instance, as is well known, the uniform distribution for PGA from 0 to an upper limit (say, 1.0g) has the greatest uncertainty, but its coefficient of variation, V_A, is about 0.577. Suppose that by the hazard analysis we have a truncated Type-II extreme value distribution for PGA, as follows:

$$F_A(a) = \exp\left[-\left(\frac{a}{0.177}\right)^{-2.3}\right] \qquad 0 \leq a \leq 1.0\text{g} \tag{7}$$

The coefficient of variation, V_A, for this case is about 0.68. Note that from a very fuzzy knowledge of uniform distribution where the COV was 0.58, the extreme value Type-II distribution on PGA gives COV of 0.68. Thus the extra effort in obtaining a better assessment on seismic hazard results in larger COV. To the uninitiated observer, this may seem contradictionary.

It does not mean that the analysis increases the uncertainty. For example, let us look at the decision problem of selecting the design level a^*. Suppose that we have the following total cost function, C:

$$C = \begin{cases} = (1 + 0.5a^*)\ C_o & a \leq a^* \\ = [(1 + 0.5a^* + 1.2(a - a^*)]\ C_o & a > a^* \end{cases} \tag{8}$$

where C_o is the cost without seismic consideration.

Increasing the design level increases the initial incremental cost of seismic design. If the future earthquake PGA is greater than the design PGA a^*, the loss associated with this overload could be of the form $1.2(a - a^*)C_o$. If $a^* = 0$ (i.e., no seismic design consideration) and if $a = 1.0g$, it can be assumed that the structure will be a total loss, amounting to C_o plus an additional loss due to the cost of demolition and cleaning up. In such a case, using optimization theory, the expected total cost and design level would be as follows:

For uniform PGA distribution,

$$a^* = 0.58g \quad \text{and} \quad E[C] = 1.36\ C_o$$

For truncated extreme value Type-II distribution,

$$a^* = 0.23g \quad \text{and} \quad E[C] = 1.18\ C_o$$

The above example demonstrates that by using the available information, we can improve our knowledge and hence reduce uncertainty in seismic design decisions. Even though the COV indicates otherwise, the design level and expected cost for extreme value Type-II distribution are lower than the corresponding values for uniform distribution. In passing, it should also be mentioned that a consistent method for expressing uncertainty is through entropy. The entropy of the above uniform distribution is 6.2, whereas the entropy of the truncated extreme value Type-II distribution is 5.0. Entropy is a good measure of uncertainty.

CONCLUSION

It is sometimes quite amusing and often of great concern when we see users of the current probabilistic risk analysis models put unreasonale confidence in the numbers they get. Often, we see engineers, planners, regulators, and public officials argue about the level of peak ground acceleration one would get for, say, a 100 year return period. A 10-20% variation from the estimated value is sometimes argued between the various parties as if the analyst had the ability to "fine-tune" the numbers. This "over-reliance" or "faith" in the results seem to be inversely proportional to the level of understanding one has about the uncertainty in each step of the hazard analysis.

There is also a trend among users and developers of probabilistic hazard analysis procedures to "get more out" of the data than the data can provide. This is especially true in source modeling and in attenuation studies. There seem to be as many attenuation relationships as there are researchers in seismic hazard analysis. They all use the same data and they all come out with conceptually similar empirical models with minute differences in constants or the numerical procedures. We seem to be all trying to squeeze water out of stones. In our opinion, the time has come when we should look at the available information from an entirely fresh perspective instead of redoing everything with an "epsilon" type of variation, just to get a method or an equation which is given the name of the researcher. Use of pattern recognition in sorting past intensity data and in combining recurrence information from historical and geologic data bases is one such fresh approach. Use of fuzzy set theory may not provide the ultimate in combining expert opinion, but it will provide a new and fresh look at the way we do things. Only through such innovative and imaginative tools will we be able to improve our ability to reduce uncertainty in seismic hazard estimation.

REFERENCES

1. Bezdek, J. C. (1982). Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York.
2. Cornell, C. A. and E. H. Vanmarcke, (1969). The major influences on seismic risk, Proc. 4th World Conf. Earthquake Engineering, Santiago, Chile, Vol. 1.
3. Bender, B. (1983). Maximum likelihood estimation of b values for magnitude grouped data, BSSA, Vol. 73, No. 3.
4. Dong, W. M., H. C. Shah, and A. B. Bao, (1983). Use of maximum entropy principle in earthquake recurrence relationships, Proc. of U.S.-Japan Cooperative Research on Generalized Seismic Risk Analysis and Development of a Model Seismic Format.
5. Anderson, J. G. and J. E. Luco (1983). Consequences of slip rate constraints on earthquake occurrence relations, Bull. Seism. Soc. Am., Vol. 73.
6. Dong, W. et al. (1983). Utilization of energy flux, seismic moment and geological information in Bayesian seismic hazard models, Proc. of U.S.-Japan Cooperative Research on Generalized Seismic Risk Analysis and Development of a Model Seismic Format.
7. Savy, J. B., H. C. Shah, and D. Boore (1980). Nonstationary risk model with geophysical input, J. Struc. Div., ASCE, Vol. 106.
8. Nishioka, T. and H. C. Shah (1980). Application of the Markov chain on probability of earthquake occurrence, Proc. of J.S.C.E., No. 298.
9. Patwardhan, A. S., R. B. Kulkarni, and D. Tocher (1980). A semi-Markov model for characterizing recurrence of great earthquakes, Bull. Seism. Soc. Am., Vol. 70.
10. Kiremidjian, A. S. and T. Anagnos (1983). Stochastic models for earthquake occurrence. Technical Report, The John A. Blume Earthquake Engineering Center, Stanford University.
11. McCann, M. W. (1981). A Bayesian geophysical model for seismic hazard. Technical Report, The John A. Blume Earthquake Engineering Center, Stanford University.
12. Ellingwood, B. et al. (1980). Development of a probability based load criterion for American National Standard A58, NBS Special Publication 577.

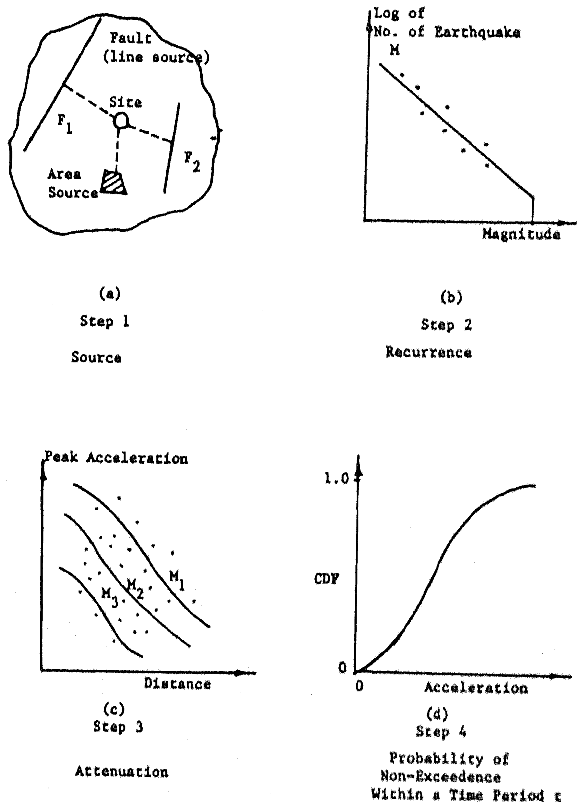

Figure 1. General steps of seismic hazard analysis.

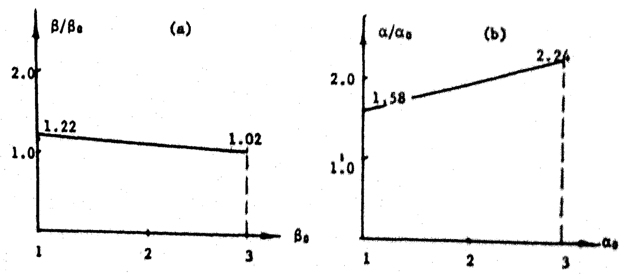

Figure 2. Estimates of parameters by LS.

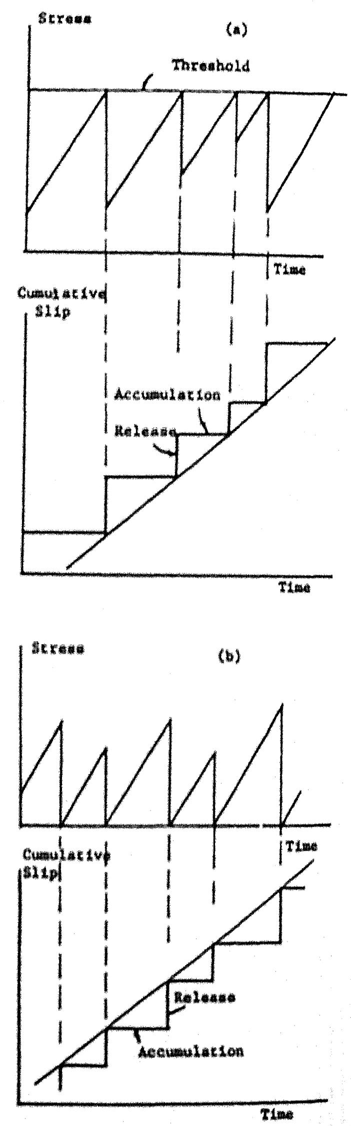

Figure 3. Time-dependent occurrence models.