ASSESSMENT OF CONFIDENCE INTERVALS
FOR RESULTS OF SEISMIC HAZARD ANALYSIS

R.B. Kulkarni (I)
R.R. Youngs (I)
K.J. Coppersmith (I)
Presenting Author: R.B. Kulkarni

SUMMARY

This paper presents a methodology to obtain confidence intervals about the expected frequencies of exceeding given levels of ground motion at a site because of seismic activity. The basic approach is to separate the effects of two major sources of uncertainty-- inherent uncertainty and statistical uncertainty. Probabilistic models are used to represent inherent uncertainty in the time of occurrence of future earthquakes and their characteristics of ground motion transmission. The statistical uncertainties in the parameters of these models are analyzed by constructing logic trees to represent combinations of possible values of each parameter. An application of the procedure to seismic hazard analysis of a nuclear power plant site is described.

INTRODUCTION

A major output of a probabilistic seismic hazard evaluation is the set of frequencies of exceeding different levels of ground motion at a given site. The calculation of these exceedance frequencies requires an estimation of the uncertainties associated with the times of occurrence of future earthquakes and the transmission of ground motion to the site from the earthquakes. It is useful to separate the effects of two major sources of uncertainties:

(i) Inherent uncertainties in the prediction of recurrences of earthquakes and their characteristics, and

(ii) statistical uncertainties in the estimation of parameters of a (stochastic) model that may be used to represent the natural phenomenon of earthquake occurence.

Inherent uncertainties cannot be reduced with the collection of additional data since the stochastic nature of the phenomenon will remain unchanged. On the other hand, statistical uncertainties can be reduced or, in concept, completely eliminated with the collection of additional data since model parameters can be estimated with increased accuracy.

The separation of the two sources of uncertainties enables the calculation of confidence intervals about the expected frequencies of exceeding specified levels of ground motion. This is particularly important in the context of critical facilities such as nuclear power plants for which acceptable

---

(I) Woodward-Clyde Consultants, Walnut Creek, California, U.S.A.

exceedance frequencies are very low (e.g., 10^{-4} per year or less). The calculation of confidence intervals provides a realistic assessment of the limitations of current scientific knowledge about seismic hazards and assists in deciding if additional data should be collected and which data would reduce the statistical uncertainties the most.

This paper presents a mathematical model for calculating the confidence intervals. A major improvement provided by the model over previous studies is in the treatment of uncertainties in the parameters of a seismic hazard model using logic trees. The logic trees attempt to relate the parameters to scientific hypotheses of earthquake phenomena, rather than to characterize the uncertainties in the parameters directly based on empirical data.

MATHEMATICAL MODEL

The mathematical model can be described in a sequence of four steps.

1. Development of Probabilistic Models to Analyze Inherent Uncertainties

A Poisson process is generally considered to be appropriate to represent recurrences of earthquakes on a known seismic source. This model assumes that earthquakes occur with some average frequency and that the probability of an earthquake during the time period t to $t + \Delta t$ depends only on the interval Δt and is independent of t. Furthermore, for a given source, an earthquake is considered equally likely to occur anywhere on the source. Although these assumptions of temporal and spatial independence may not be completely true, they are generally reasonable to evaluate seismic hazards during periods of engineering significance (e.g., 30-60 years).

Empirical data on the source-to-site attenuation of ground motion show a large amount of scatter. A lognormal probability model is generally assumed to represent the distribution of levels of ground motion at a site with a specified distance from a source and for a specified earthquake magnitude on the source.

2. Calculation of Exceedance Frequencies for Known Model Parameters

In this step the parameters of the Poisson process for earthquake recurrence and of the lognormal model for ground motion attenuation are assumed to be known with certainty. These parameters include: geometry of potential seismic sources, mean rates of occurrence of different magnitude earthquakes, maximum earthquake magnitude on each source, and median and standard deviation of ground motion levels as a function earthquake magnitude and source-to-site distance.

We want to calculate $\nu_n(z|\underline{\Theta}_n)$, the mean number of events per unit time in which ground motion level z is exceeded at the site due to seismic activity on the nth seismic source, assuming that the model parameters, $\underline{\Theta}_n$ for the nth source are known. To calculate $\nu_n(z|\underline{\Theta}_n)$, we consider the average recurrence of all possible earthquakes on the nth source. Thus,

$$\nu_n(z|\underline{\Theta}_n) = \sum_i \sum_j \lambda_n(M_i|\underline{\Theta}_n) \cdot P_n(r_j|M_i, \underline{\Theta}_n) \cdot G(z|M_i, r_j, \underline{\Theta}_n) \qquad (1)$$

in which $\lambda_n(M_i|\underline{\Theta}_n)$ = mean number of earthquakes of magnitude M_i per unit time on the nth source, given parameters $\underline{\Theta}_n$.

$P_n(r_j|M_i, \Theta_n)$ = probability that the significant source-to-site distance is r_j given an earthquake of magnitude M_i on the nth source and given parameters Θ_n, and

$G(z|M_i, r_j, \Theta_n)$ = probability that ground motion will exceed z given an earthquake of magnitude M_i at a significant distance r_j from the site and given parameters Θ_n.

Note that given the parameters Θ_n, the three quantities on the right hand side of equation 1, $\lambda_n(M_i|\Theta_n)$, $P_n(r_j|M_i, \Theta_n)$, and $G(z|M_i, r_j, \Theta_n)$ can be calculated. Consequently, the quantity $\nu_n(z|\Theta_n)$ can be calculated given the parameters Θ_n.

3. Construction of Logic Trees to Analyze Statistical Uncertainties in the Model Parameters

Because of incomplete data, the model parameters, Θ_n generally cannot be estimated with certainty. Alternative values of the model parameters, therefore, must be allowed and the likelihood that a particular level represents the true value of a parameter must be assessed. Due to limited data, the assessment of likelihoods of alternative parameter values often will have to be based on subjective judgements of scientists knowledgeable with earthquake phenomena. Our approach to the assessment of these likelihoods differs from previous studies in that we attempt to relate the parameters to scientific hypotheses of causes of earthquakes. This approach provides a more natural way for the scientists to think about the credibilities of alternative parameter values. An additional advantage of the logic tree approach is that possible dependencies among various parameters can be maintained. This simplifies the task of subjective assessment of the likelihoods since the assessor is asked to think about the likelihood of a particular value of a parameter given the true value of all other parameters which affect his opinions regarding the parameter under consideration. The next section provides more details regarding the development of logic trees.

4. Calculation of the Probability Distribution of Exceedance Frequencies

We want to calculate $\nu(z)$, the mean number of events per unit time at the site in which the level of ground motion, z is exceeded. If the model parameters were known with certainty, $\nu(z)$ can be obtained by summing the contributions of all seismic sources, i.e.,

$$\nu(z) = \sum_n \nu_n(z) \qquad (2)$$

in which $\nu_n(z)$ is the mean number of exceedance events per unit time at the site due to seismic activity on the nth source. Since the model parameters are considered to be uncertain (random variables), $\nu_n(z)$ for each n, and consequently, $\nu(z)$ are all random variables. Our approach is to obtain the probability distribution of $\nu_n(z)$ for each n and then to calculate the probability distribution of $\nu(z)$ considering it to be the sum of several random variables.

The probability mass function (PMF) of $\nu_n(z)$ is given by:

$$P[\nu_n(z) = x_k] = \sum_{m \in I(x_k)} P(\Theta_n(m)) \qquad (3)$$

in which $P(\underline{\Theta}_n(m))$ is the probability of mth terminal node of the logic tree constructed to represent alternative combinations of model parameters, $\underline{\Theta}_n$ for the nth source, $I(x_k)$ is the set of terminal nodes in this logic tree at which $\nu_n(z|\underline{\Theta}_i(m)) = x_k \pm \Delta/2$, and Δ is a suitable interval for discretizing the range of $\nu_n(z)$.

Now, to obtain the probability mass function of $\nu(z)$, we start adding contributions of different sources one by one. We will let $\nu(z|S_1)$ denote the mean number of exceedance events at the site due to seismic activity on the source S_1. Note that:

$$P[\nu(z|S_1) = x_k] = P[\nu_1(z) = x_k] \qquad (4)$$

Now we add the contribution of the next source, S_2 as follows:

$$P[\nu(z|S_1, S_2) = x_k] = \sum_j P[\nu_1(z) = x_j] \cdot P[\nu_2(z) = x_k - x_j] \qquad (5)$$

This process of adding contributions from the next source is continued until all sources are included.

Once the complete probability distribution of $\nu(z)$ is obtained, the expected value of this distribution and specified confidence intervals can be readily obtained. The expected value and variance of the probability distribution of $\nu(z)$ can be obtained more directly as follows:

$$E[\nu(z)] = \sum_n E[\nu_n(z)] \qquad (6)$$

$$\text{and} \quad Var[\nu(z)] = \sum_n Var[\nu_n(z)] \qquad (7)$$

Following the central limit theorem, the sum of several random variables, $\nu(z)$ will tend towards normal distribution as the number of individual terms in the sum becomes large. Thus, if several sources are contributing to the seismic sources at a site, the overall site exposure, $\nu(z)$ may be assumed to follow normal distribution with the mean and variance given by equations 6 and 7, respectively.

DEVELOPMENT OF LOGIC TREES TO REPRESENT PARAMETER UNCERTAINTIES

Logic trees can be used to address the statistical uncertainties in the major elements of seismic source characterization. The complexity of the logic trees is a function of several factors including, the scientific knowledge of the causes of earthquakes, the association of seismicity with geologic structure, the understanding of the geologic behavior of individual seismic sources, and the adequacy of the historical seismicity data base. For example, within intraplate tectonic environments it may be necessary to consider alternative tectonic models of the earthquake generation process, possible associations of seismicity with known geological structures, and various mechanisms and orientations of stress release. Within interplate regions, the tectonic model is usually well-established but the physical attributes of particular seismic sources must be estimated to constrain maximum magnitudes and recurrence such as fault geometry, segment length, displacement per event, and slip rate.

The nodes of the logic trees are sequenced to provide for the conditional aspect or dependencies of the source parameters and a logical progression of

assumptions regarding source definition. For example, if a particular parameter such as fault dip is conditional on a particular tectonic model, then that parameter follows the tectonic model node. Likewise, a parameter such as downdip fault width may be conditional on the assumed dip, and will therefore follow the fault dip node. By sequencing in this way, assessments of the alternative values (branches) at each node is made given that a particular condition is true. This makes the weighting or credibility assessments easier because the assessor does not need to consider more than one parameter at a time. For example, Figure 1 shows a simple logic tree to characterize a hypothetical seismic source in the northeastern U.S. The first node of the logic tree is the tectonic model, which is an explanation of the causal mechanism for earthquakes within the source zone: reactivation of preexisting faults, stress amplification associated with a pluton, or regional post-glacial rebound. The relative probability that any one of these models is correct is given beneath each branch. Each of these models has implications to the geometry of the seismic source, represented in the example as the length of rupture that may be assumed to occur, and ultimately to other parameters that relate to maximum magnitudes and earthquake recurrence. Note that the choice of rupture lengths is conditional on the assumed tectonic model and, therefore, the probabilities associated with the various choices are conditional probabilities.

Tectonic Model | Maximum Rupture Length (km)

Fault Reactivation (0.6)
- 50 (0.6)
- 30 (0.4)

Stress Amplification (0.3)
- 30 (0.2)
- 15 (0.8)
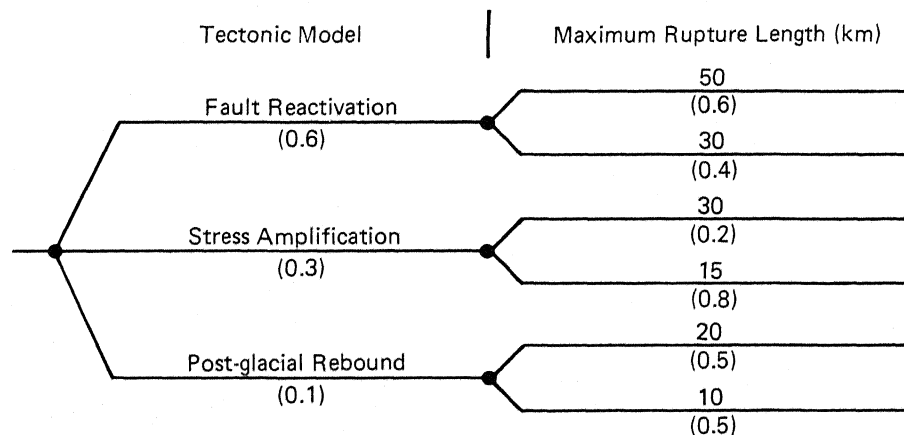
Post-glacial Rebound (0.1)
- 20 (0.5)
- 10 (0.5)

Figure 1 — Example of Source Logic Tree

The assignment of credibilities or relative weights to the seismic source parameters may be accomplished informally through interactions among experts to arrive at consensus estimates. Alternatively, formal procedures for eliciting subjective judgment may be utilized.

APPLICATION TO SEISMIC HAZARD ANALYSIS OF A NUCLEAR POWER PLANT SITE

The above procedure for quantifying the uncertainty in seismic hazard levels was applied to the evaluation of seismic hazard at the WNP-2 nuclear power plant located on the Hanford Reservation in eastern Washington State. The study was done to evaluate the probability of exceeding the SSE design peak acceleration of 0.25g.

The site lies within the Columbia Plateau which can be generally characterized as an intraplate region with low seismic activity. Large uncertainties exist as the earthquake potential of the region due to lack of correlation of instrumentally recorded seismicity with geologic structures and the catastrophic floods that flowed through the region during the late Pleistocene, removing much of the Quarternary geologic record essential for assessing fault capability and deformation rates.

Logic trees were developed for each of the potential seismic sources in the site region. The logic trees were developed by general consensus during a workshop involving twenty geologists, seismologists and geophysicists having considerable experience with the data for the Columbia Plateau. The details are described in Power et al. (Ref. 1). The development of the logic tree for the seismic source nearest to the site is outlined below.

The nearest potential seismic source to the site is the Umtanum Ridge-Gable Mtn. structural trend, which is a series of aligned folds. The major source of uncertainty concerning the earthquake potential of this structure is the primary mode of deformation. Two tectonic models have been proposed for the development of the folds in the Columbia Plateau: either primarily folding with or without secondary reverse faulting or primary reverse faulting with secondary folding. A primary fault model would imply significant down-dip fault dimensions and thus the potential for large earthquakes. A primary folding with secondary faulting model would imply limited fault dimensions and the potential for only small-to-moderate earthquakes. The choice between the two alternative tectonic models was selected as the first node in the logic tree for this source, shown in Figure 2. The available data favored folding with secondary faulting as shown by the assigned probabilities in parenthesis.

The next level of uncertainty considered was the question of segmentation of the entire structural trend in terms of coherent behavior. Three models of segmentation were considered: single source, segmented source and separate source. In the single source model the entire trend is assumed to be a coherent unit and rupture can occur anywhere along its length. In the segmented source model the entire trend is assumed to consist of five segments defined by structural and geomorphic data. Any single earthquake rupture is confined to one of the segments. However, the overall behavior of the individual segments is assumed to be strongly correlated. In the separate source model the five individual segments are assumed to act as independent, uncorrelated sources.

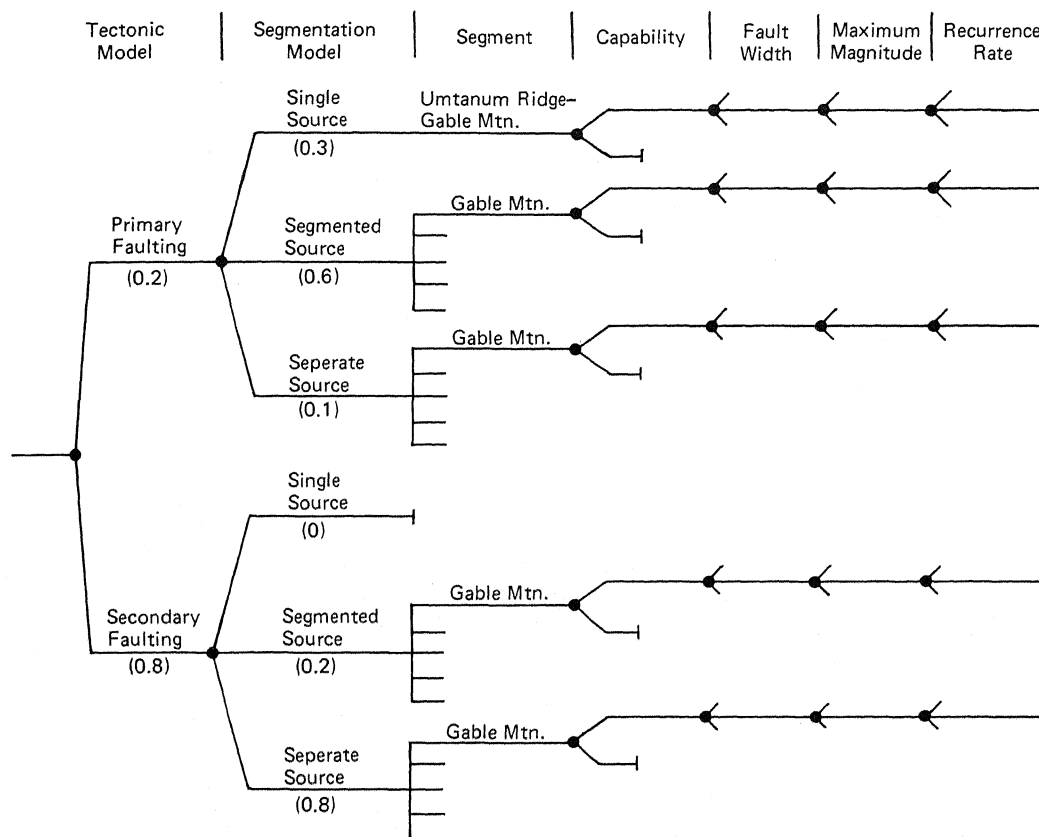| Tectonic Model | Segmentation Model | Segment | Capability | Fault Width | Maximum Magnitude | Recurrence Rate |

Figure 2 — Logic Tree for Umtanum Ridge — Gable Mtn. Structural Trend

The assignment of relative weights among the three segmentation models was made conditional on the choice of tectonic model. If primary reverse faulting is the correct tectonic model, then it was judged most likely that the structure consists of distinct segments that display common (i.e., correlated) characteristics and least likely that the segments act as completely independent, uncorrelated sources. On the other hand, if primary folding-secondary faulting is the correct tectonic model then it was judged most likely that the segments are independent and uncorrelated sources and unlikely that the entire trend acts as a single source.

Below this node in the logic tree, the branches for segmented source and separate source were expanded to treat each of the 5 source segments, as shown in Figure 2. The parameters considered for each segment were source capability, fault geometry, maximum earthquake magnitude, and earthquake recurrence rate.

In a similar fashion, logic trees were constructed for each of the significant geologic structures in the site region. For each terminal node in the trees, the rate of exceeding 0.25g was computed using equation (1). The probability density function for the exceedance rate was then computed using equation (5) assuming independence between structures. Figure 3 shows the computed density function for the annual rate of exceeding the SSE acceleration. The 90 percent confidence level on the rate of exceeding 0.25g was found to be approximately 2.6 times the expected exceedance rate.
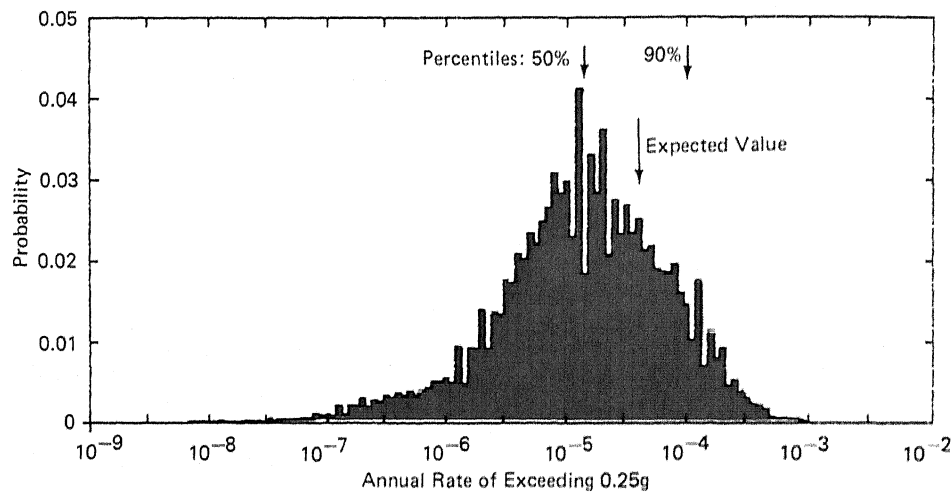

Figure 3 — Computed Probability Density Function for Rate of Exceedance

CONCLUSIONS

Separation of inherent and statistical uncertainties in seismic hazard analysis permits the calculation of confidence intervals about the expected frequency of exceeding a given level of ground motion at a site. The calculation of confidence intervals provides a realistic assessment of the limitations of current scientific knowledge about seismic hazards and assists in deciding whether additional data should be collected and which types of data would reduce the statistical uncertainties the most. A major improvement offered by the model described in this paper is in the treatment of uncertainties in the parameters of a seismic hazard model using logic trees to relate the parameters to scientific hypotheses of earthquake phenomenon. This approach provides a more defensible and intuitively appealing method for the subjective assessment of credibilities of alternative parameter levels.

REFERENCES

1. Power, M.S., Coppersmith, K.J., Youngs, R.R., Schwartz, D.P., Swan, F.H., III, 1981, Seismic Exposure Analysis for the WNP-2 and WNP-1/4 Site: Appendix 2.5K to Amendment No. 18 Final Safety Analysis Report for WNP-2, Washington Public Power Supply System, Richland Washington, 63p.