11 Probabilistic Seismic Hazard Analysis

In Chapter 3, a simplified numerical example of a probabilistic hazard calculation was presented without using equations. In this chapter, the mathematical framework for PSHA is presented. Many of the concepts discussed in Chapter 3 will be repeated in this chapter but they will be explained in terms of equations.

In Chapter 3, discrete approximations were used for the variability in the earthquake magintude, source location, and ground motion. In this chapter, the hazard analysis will be developed using continuous functions. This means that instead of probabilities, we will use probability density functions as described in Chapter 4.

11.1 Mathematical Framework

The basic methodology of PSHA involves computing the how often a specified level of ground motion will be exceeded at the site. Specifically, in a PSHA the annual rate of events, ν, that produce a ground motion parameter, Sa, that exceeds a specified level, z, at the site is computed. This annual rate is also called the "annual rate of exceedance". The inverse of ν is called the "return period". (Note: we use Sa here for the ground motion parameter since most hazard is computed for spectral acceleration, but this can be any ground motion parameter)

11.1.1 Point Sources

For simplicity, we begin with the equation for the seismic hazard using point sources. Traditionally, the equation for a seismic hazard analysis due to a single source, i, has been given by

$$\nu_i(Sa > z) = N_i(M_{\min}) \int_{r=0}^{\infty} \int_{m=M_{\min}}^{M_{\max_i}} f_{m_i}(M)\, f_{r_i}(r)\; P(Sa > z \mid M,r)\; dr\, dM \qquad (11.1)$$

where r is the distance from the source to the site, M is earthquake magnitude, $N_i(M_{min})$ is the annual rate of earthquake with magnitude greater than or equal to M_{min}, M_{max_i} is the maximum magnitude, $f_m(m)$, and $f_r(r)$ are probability density functions for the magnitude, and distance which describe the relative rates of different earthquake scenarios. $P(Sa>z|M,r)$ is the conditional probability of observing a ground motion parameter Sa greater than z for a given earthquake magnitude and distance.

The form of eq. (11.1) leads to one of the common misunderstandings of PSHA. Looking at eq. (11.1), we can see the magnitude and distance for the scenario, but the number of standard deviation of the ground motion is not explicitly shown. As discussed in section 2.5, a common misunderstanding of PSHA is that it is thought to only consider the variability in the earthquake magnitude and source-to-site distance.

The ground motion variability is contained in the $P(Sa>z|M,r)$ term:

$$P(Sa > z \mid M,r) = \int_{z}^{\infty} f_{Sa}(Sa,M,r)\, dSa \qquad (11.2)$$

where $f_{Sa}(Sa,M,r)$ is the probability density function for the ground motion as defined by the ground motion model. Typically, the ground motion variability is modeled by a lognormal distribution. The ground motion model gives the median ground motion and the standard deviation in log units (see Chapter 7).

We can rewrite eq (11.2) in terms of the number of standard deviations above or below the median:

$$P(A > z \mid M,r) = \int_{\varepsilon^*(M,r,z)}^{\infty} f_{\varepsilon}(\varepsilon)\, d\varepsilon \qquad (11.3)$$

where ε is the number of standard deviations of the ground motion (above the median ground motion), $f_\varepsilon(\varepsilon)$ is the probability density function for the number of standard deviations (a standard normal distribution with mean 0 and variance 1), and ε^* is the number of standard deviations of the ground motion that leads to ground motion level z given M and r. The ε^* is given by

$$\varepsilon^*(M,r,z) = \frac{\ln(z) - \ln\left(\hat{A}(M,r,Site)\right)}{\sigma\left(M,\hat{A}(M,r,Site),Site\right)} \tag{11.4}$$

where $\hat{A}(M,r,Site)$ is the median ground motion for a given M and r based on the ground motion model (e.g. the median from the attenuation relation), and $\sigma\left(M,\hat{A}(M,r,Site),Site\right)$ is the standard deviation of the ground motion model in natural log units. In eq. (11.4), we have assumed that the ground motion model only depends on M, r, and site condition for simplicity. Many modern ground motion models will include additional terms such as style-of-faulting and depth to top of rupture. The standard deviation in eq (11.4) is shown as being dependent on the magnitude, the amplitude of the median ground motion, and the site condition. As discussed in Chapter 7, many modern ground motion models include a standard deviation that depends on the one or more of these parameters.

The integral in eq. (11.3) is just the complementary cumulative standard normal distribution given in statistical tables:

$$P(A > z \mid M,r) = \int_{\varepsilon^*}^{\infty} f_\varepsilon(\varepsilon)\, d\varepsilon = 1 - \Phi(\varepsilon^*) \tag{11.5}$$

Substituting eq. (11.5) into eq. (11.1), gives the hazard for point sources in the form typically used in PSHA calculations:

$$\nu_i(Sa > z) = N_i(M_{\min}) \int_{r=0}^{\infty} \int_{m=M_{\min}}^{M_{\max i}} f_{m_i}(M)\, f_{r_i}(r) \left[1 - \Phi\left(\varepsilon^*(M,r,z)\right)\right] dM\, dr \tag{11.6}$$

While this form of the hazard integral is computationaly efficient, it still obscures the aleatory ground motion variability. As an alternative, eq. (11.1) can be rewritten to show the ground motion variability explicitly:

$$\nu_i(Sa > z) = N_i(M_{\min}) \int_{r=0}^{\infty} \int_{m=M_{\min}}^{M_{\max i}} \int_{\varepsilon=\varepsilon_{\min}}^{\varepsilon_{\max}} f_{m_i}(M)\, f_{r_i}(r)\, f_\varepsilon(\varepsilon)\, P(Sa > z \mid M,r,\varepsilon)\, dr\, dM\, d\varepsilon \tag{11.7}$$

where $P(Sa>z|m,r,\varepsilon)$ is the probability that the ground motion exceeds the test level z for magnitude M, distance r, and number of standard deviations ε. For a given magnitude, distance, and ε, the ground motion from the attenuation relation is defined. Therefore, the term $P(Sa>z|m,r,\varepsilon)$ is either 0 or 1. That is, in this form, the probability term just sorts the scenarios into those that produce ground motions greater than z and those that produce ground motions smaller than z.

An advantage of this from of the hazard integral is that it can be directly related to the deterministic approach. In the deterministic approach, the magnitude, distance, and number of standard deviation of the ground motion need to be specified. The hazard integral in eq. (11.7) is simply constructing a suite of all possible deterministic scenarios in terms of (M, r, ε) triplets. The rate of each scenario is

$$rate(M,r,\varepsilon) = N(M_{\min})\, f_m(M)dM\ f_r(r)dr\ f_\varepsilon(\varepsilon)d\varepsilon \qquad (11.8)$$

The integrals in eq (11.8) are simply summing up the rates of the scenarios that produce ground motions greater than z. Using the $P(Sa>z|M,r,\varepsilon)$ term is just short hand for ranking the scenarios by ground motion level as discussed in chapter 3.

Another advantage of this form is that is shows that the aleatory variability in the magnitude, distance, and number of standard deviations of the ground motion are all treated in the same way. In eq. (11.8), each aleatory variable is specified by a probability density function. A rule for PSHA is that all aleatory variability is modeled by pdfs and there should be an integral over each pdf. While the integral over the ground motion pdf is implied in eq. (11.1), the explict form in eq (11.7) is easier to follow.

Mathematically, eq. (11.7) is identical to eq. (11.6). In computing the hazard, the form in eq. (11.6) is more efficient, but we will use the expanded form in this discussion. In application, the more efficient form is used.

11.1.2 Extended Sources (Faults)

For planar sources (e.g. faults), we need to consider the finite dimension and location of the rupture in to compute the closest distance which is used by the ground motion models. The scenario earthquake is specified in terms of the rupture dimension (rupture length and rupture width), and the rupture location (location along strike and location down dip). Given the rupture width, W, rupture length, L, location along strike, Loc_x, and location down dip, Loc_y, the closest distance from the source to the site can be computed. These four aleatory variables replace the single aleatory variable distance shown in eq (11.7). Typically, the rupture area and rupture width are used rather than the rupture length and rupture width because for many cases, the ruture width will be limited by the fault width, implying a correlation of length and width. This correlation can be considered implictly by computing the rupture length given the rupture area and rupture width. (That is, first compute the rupture area and width and then back calculcate the length.) For extended sources, the hazard integral is given by

$$\begin{aligned} \nu_i(Sa > z) = N_i(M_{\min}) \int_{RW=0}^{\infty} \int_{RA=0}^{\infty} \int_{x=0}^{1} \int_{y=0}^{1} \int_{m=M_{\min}}^{M_{\max i}} \int_{\varepsilon=\varepsilon_{\min}}^{\varepsilon_{\max}} & f_{m_i}(M) f_{W_i}(M,W) f_{RA_i}(M,RA) \\ & f_{Locx_i}(x) f_{Locy_i}(M,x) f_\varepsilon(\varepsilon) P(Sa > z \mid M, r_i(x,y,RA,W),\varepsilon) \\ & dW\, dRA\, dx\, dy\, dM\, d\varepsilon \end{aligned} \qquad (11.9)$$

where $f_W(M,W)$, $f_{RA}(M,RA)$, $f_{Locx}(x)$, $f_{Lcy(y)}$ are probability density functions for the rupture width, rupture area, location of the rupture along strike and location of the rupture down dip, respectively. In eq. (11.9), x and y give the location of the rupture in terms of the fraction of the fault length and fault width, respectively (e.g. x=0 is one end of the fault and x=1.0 is the other end of the fault).

The hazard integral in eq. (11.8) appears complicated, but keep in mind that all that integrals in eq. (11.8) are doing is defining a complete set of possible earthquake scenarios (magnitude, rupture dimension, and rupture location) with the full range of possible ground motions. The probability term, P(Sa>z|...) sorts out the scenarios with ground motions that exceed the test value z. A PSHA is similar to a large bookkeeping exercise. Instead of developing a small number of deterministic scenarios, a PSHA will develop thousands or millions of scenarios. We just need to keep track of each of the scenarios and its rate.

11.1.3 Hazard From Multiple Sources

For multiple seismic sources, the total annual rate of events with ground motions that exceed z at the site is the sum of the annual rate of events from the individual sources (assuming that the sources are independent).

$$\nu(Sa > z) = \sum_{i=1}^{Nsource} \nu_i (Sa > z) \tag{11.10}$$

where N_{source} is the total number of fault and areal sources. The rates are summed up over all sources because we are interested in how often severe shaking occurs at the site, regardless of what source caused the ground motion. Distinctions between ground motions from different magnitudes and distances are considered through the deaggregation discussed later in section 11.4.

11.2 Probability Models

To convert the annual rate of events to a probability, we need to consider the probability that the ground motion exceeds test level z at least once during a specified time interval. Two alternative models of earthquake recurrence probabilities were discussed in Chapter 6: the Poisson model and the renewal model. By far, the most common assumption in practice is the Poisson model.

For a Poisson process, the probability of at least one occurrence of ground motion level z in T years is given by

$$P(Sa>z|T) = 1 - \exp(-\nu(Sa>z)T) \tag{11.11}$$

(see chapter 4). For T=1 year, this probability is the annual probability.

The hazard level is often given in terms of a probability of being exceeded in T years: for example, 10% chance of being exceeded in 50 years. Using the Poisson assumption, we can convert this probability to an equivalent annual rate.

$$\nu(Sa > z) = \frac{-\ln\left(1 - P(Sa > z \mid T)\right)}{T} \tag{11.12}$$

For T=50 years, and P=0.1 (10%), then the rate, ν, is 0.0021/yr. The inverse of this rate is 475 years which is called the return period. The return period is slightly smaller than 500 yrs because there is a chance of more than one event exceeding the target level in 50 years.

If the renewal model is used, the probability of an earthquake occuring is computed for a specfied time period (e.g. the next 50 years). In this case, it is common to convert the earthquake probability from the renewal model to an equivalent Poisson rate:

$$\nu_i^{eqi}(M > m_1) = \frac{-\ln\left(1 - P_i(M > m_1 \mid T_1 > T > T_2)\right)}{T_2 - T_1} \tag{11.13}$$

Using this equivalent Possion rate, the hazard can be computed using standard software with the Poisson assumption.

11.3 Logic Trees

Scientific (epistemic) uncertainty is considered by using alternative models and/or parameter values for the probability density functions in eq. (11.9), the attenuation relation, and the activity rate. For each alternative model, the hazard is recomputed resulting in a suite of alternative hazard curves. Scientific uncertainty is typically handled using a logic tree approach for specifying the alternative models for the density functions, attenuation relations, and activity rates.

Branches on logic trees represent either-or branches. The branches represent alternative credible models. The weights on the branch represent the judgement about the credibility of the alternative models. These weights are often called probabilities, but they are better treated as evaluations of the relative merits of the alternative models (Abrahamson and Bommer, 2005). Branches in logic trees do not represent "sometimes" branches (e.g. randomness). For example, if a fault sometimes ruptures in individual segments and sometimes ruptures as multiple segments, then this variability is randomness that should be part of the probability density function in the hazard integral. The branches on the logic tree should reflect alternative estimates of the parameters and models included in the hazard integral.

To help keep track of the difference between what goes on the logic tree and what goes into the probability density functions, the terms "aleatory" variability and "epistemic" uncertainty are used (see chapter 5). Aleatory variability is the randomness part and epistemic uncertainty is the uncertainty part. The reason for using aleatory and epistemic rather than "randomness" and "uncertainty" is that randomness and uncertainty are too common of terms that are often used interchangeably. Using the terms aleatory and epistemic leads to a more consistent use of terminology.

11.4 Deaggregation of Hazard

The hazard curve gives the combined effect of all magnitudes and distances on the probability of exceeding a given ground motion level. Since all of the sources, magnitudes, and distances are mixed together, it is difficult to get an intuitive understanding of what is controlling the hazard from the hazard curve by itself. To provide insight into what events are the most important for the hazard at a given ground motion level, the hazard curve is broken down into its contributions from different earthquake scenarios. This process is called deaggregation (e.g. Bazzurro and Cornell, 1999).

In a hazard calculation, there is a large number of scenarios considered (e.g. thousands or millions of scenarios). To reduce this large number of scenarios to a managable number, similar scenarios are grouped together. A key issue is what consistutues "similar" scenarios. Typically, little thought has been given to the grouping of the scenarios. Most haard studies use equal spacing in magnitude space and distance space. This may not be appropriate for a specific project. The selection of the grouping of scenarios should be defined by the engineers conducting the analysis of the structure (Abrahamson, 2006).

In a deaggregation, the fractional contribution of different scenario groups to the total hazard is computed. The most common form of deaggregation is a two-dimensional deaggregation in magnitude and distance bins. Mathmatically, this is given by:

$$Deagg(Sa > z, M_1 < M < M_2, R_1 < R < R_2) = \tag{11.14}$$

$$\frac{\sum_{i=1}^{nSource} N_i(M_{\min}) \int_{r=R1}^{R2} \int_{m=M_1}^{M_2} \int_{\varepsilon=\varepsilon_{\min}}^{\varepsilon_{\max}} f_{m_i}(m)\, f_{r_i}(r)\, f_\varepsilon(\varepsilon)\, P(Sa > z \mid m, r, \varepsilon)\, dr\, dm\, d\varepsilon}{\nu\ (Sa > z)}$$

The deaggreation is normalized such that it sums to unity for all scenario groups. Formally, it is the conditional probability of the ground motion being generated by a earthquake with magnitude in the range M_1-M_2 and distance in the range R_1-R_2.)

The deaggregation by magnitude and distance bins allows the dominant scenario earthquakes (magnitude and distance pair) to be identified. The results of the deaggregation will be different for different probability levels (e.g. 100 yr vs 1000 yr return periods) and for different spectral periods as shown in the following example.

Using the sources shown in Figure 11-1, the results of an example hazard calculation for PGA and T=1 sec spectral acceleration are shown in Figure 11-2. The deaggreation at return periods of 500 and 10,000 years for T=1 sec are shown in Figures 11-3a and 11-3b. For the 500 year return period, the hazard is dominated by large distant earthquakes, but for the 10,000 year return period, the hazard is domianted by nearly moderate magnitude earthquakes.

11.5.1 Mode vs Median

The dominant scenario can be characterized by an average of the deaggregation. Two types of averages are considered: the mean and the mode. The mean magntiude and mean distance are the weighted averages with the weights given by the deaggregation. The equation for the mean magnitude, for example, is given by multiplying by the magnitude inside the hazard integral:

$$\overline{M} = \frac{\sum_{i=1}^{nSource} N_i(M_{\min}) \int_{r=0}^{\infty} \int_{m=M_{\min}}^{M_{\max}} \int_{\varepsilon=\varepsilon_{\min}}^{\varepsilon_{\max}} m\, f_{m_i}(m)\, f_{r_i}(r)\, f_{\varepsilon}(\varepsilon)\, P(Sa > z \mid m,r,\varepsilon)\, dr\, dm\, d\varepsilon}{\nu\ (Sa > z)} \qquad (11.15)$$

The mean distance is computed in a similar manner:

$$\overline{R} = \frac{\sum_{i=1}^{nSource} N_i(M_{\min}) \int_{r=0}^{\infty} \int_{m=M_{\min}}^{M_{\max}} \int_{\varepsilon=\varepsilon_{\min}}^{\varepsilon_{\max}} r\, f_{m_i}(m)\, f_{r_i}(r)\, f_{\varepsilon}(\varepsilon)\, P(Sa > z \mid m,r,\varepsilon)\, dr\, dm\, d\varepsilon}{\nu\ (Sa > z)} \qquad (11.16)$$

The mean has advanatages in that it is defined unambigously and is simple to compute. The disadvantage is that it may give a value that does not correspond to a realistic scenario. For example, using the deaggregation shown in Figure 11-3a , the mean distance would be about 20-30 km which does not correspond to either of the fault sources.

The mode is the most likely value. It is given by the scenario group that has the largest deaggegation value. The mode has the advantage it will always correspond to a realistic source. The disadvantage is that the mode depends on the grouping of the scenarios, so it is not robust.

11.5 Rates of Scenarios

In addition to defining the controlling scenarios, the deaggregation can also be used to convert the hazard results back to rates of occurrence of specific scenarios and ground motion levels. The hazard curve gives the rate of exceeding a ground motion level. Therefore, subtracting the hazard at two ground motion levels gives the rate of occurrence of ground motions between the two levels:

$$\nu(z_1 < Sa < z_2) = \nu(Sa > z_1) - \nu(Sa > z_2) \tag{11.17}$$

Since the deaggregation gives the fractional contribution to the hazard, then multiplying the hazard by the deaggreagation gives the rate of exceedance from the specified magnitude and distance range:

$$\begin{aligned} &\nu(Sa > z, M_1 < m < M_2, R_1 < R < R_2) = \\ &\qquad \nu(Sa > z)\, Deagg(Sa > z, M_1 < M < M_2, R_1 < R < R_2) \end{aligned} \tag{11.18}$$

The rate of occurrence is of a ground motion level from a specific scenario group is given by subtracting this rate of exceedance:

$$\begin{aligned} &\nu(z_1 < Sa < z_2, M_1 < m < M_2, R_1 < R < R_2) = \\ &\qquad \nu(Sa > z_1)\, Deagg(Sa > z_1, M_1 < M < M_2, R_1 < R < R_2) \\ &\qquad - \nu(Sa > z_2)\, Deagg(Sa > z_2, M_1 < M < M_2, R_1 < R < R_2) \end{aligned} \tag{11.19}$$

11.6 Uniform Hazard Spectra

A common method for developing design spectra based on the probabilistic approach is uniform hazard spectra (also called equal hazard spectra). A uniform hazard spectrum (UHS) is developed by first computing the hazard at a suite of spectral periods using response spectral attenuation relations. That is, the hazard is computed independently for each spectral period. For a selected return period, the ground motion for each spectral period is measured from the hazard curves. These ground motions are then plotted at their respective spectral periods to form the uniform hazard spectrum. This process is shown graphically in Figure 11-4.

The term "uniform hazard spectrum" is used because there is an equal probability of exceeding the ground motion at any period. Since the hazard is computed independently for each spectral period, in general, a uniform hazard spectrum does not represent the spectrum of any single earthquake. It is common to find that the high frequency (f>5 Hz) ground motions are controlled by nearby moderate magnitude earthquakes, whereas, the long period (T>1 sec) ground motions are controlled by distant large magnitude earthquakes.

The "mixing" of earthquakes in the UHS is often cited as a disadvantage of PSHA. There is nothing in the PSHA method that requires using a UHS. Based on the deaggregation, multiple spectra (for each important source) can be developed. The reason for using a UHS rather than using multiple spectra for the individual scenarios is to reduce the number of engineering analyses required. A deterministic analysis has the same issue. If one deterministic scenario leads to the largest spectral values for long spectral periods and a different deterministic scenario leads to the largest spectral values for short spectral periods, it is common practice to develop a single design spectrum that envelopes the two deterministic spectra. In this case, the design spectrum also does not represent a single earthquake.

The choice of using a UHS rather than multiple spectra for the different scenarios is the decision of the engineering analyst, not the hazard analyst. If it is worth the additional analysis costs to avoid exciting a broad period range in a single evaluation, then the engineer should request multiple scenario spectra from the hazard analyst.

In practice, the hazard analyst often only provides the UHS to the engineer in the hazard report. A hazard report should also include a comparison of the UHS with the spectra from the individual representative events indentified by the deaggregation. This gives the engineer the information needed to make a decision whether to evaluate multiple scenarios one at a time or to envelope the spectra from the multiple scenarios to reduce the number of analyses required. The development of scenario spectra is dicussed further in Chapter 12.

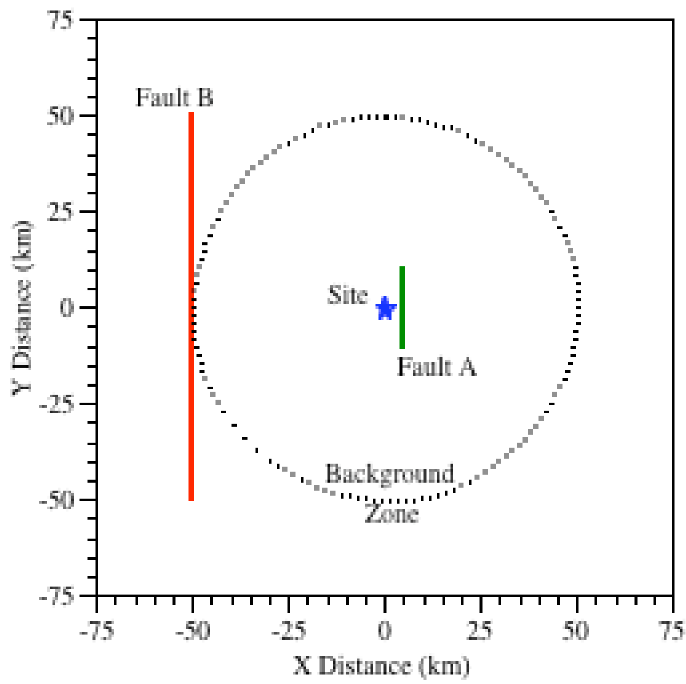

Figure 19. Source geometry used for the example.

Figure 11-1. Sources used in the example hazard

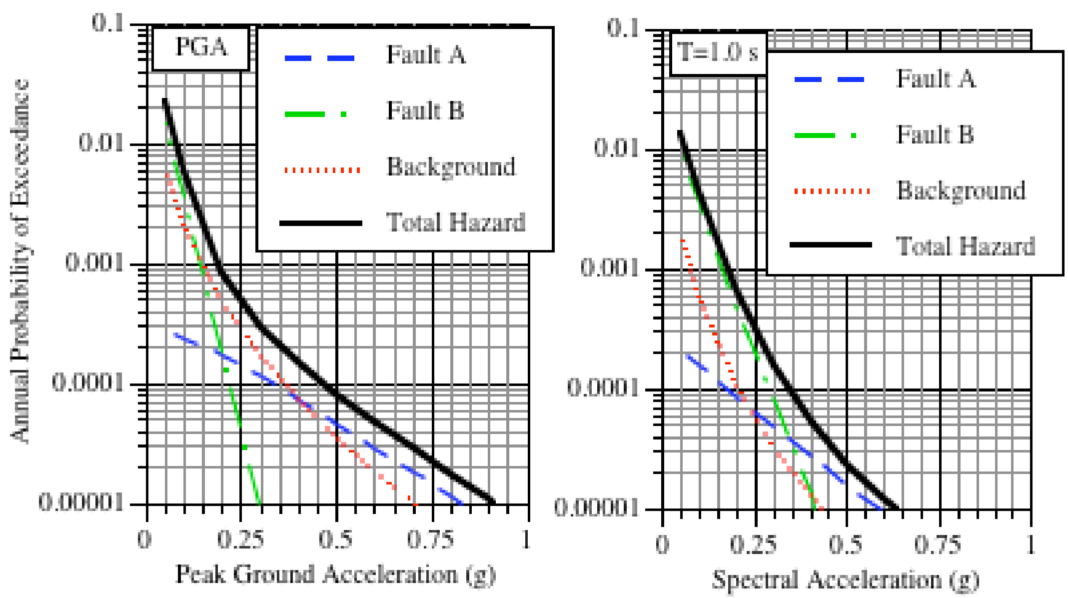

Figure11-2. Hazard curves for the example.

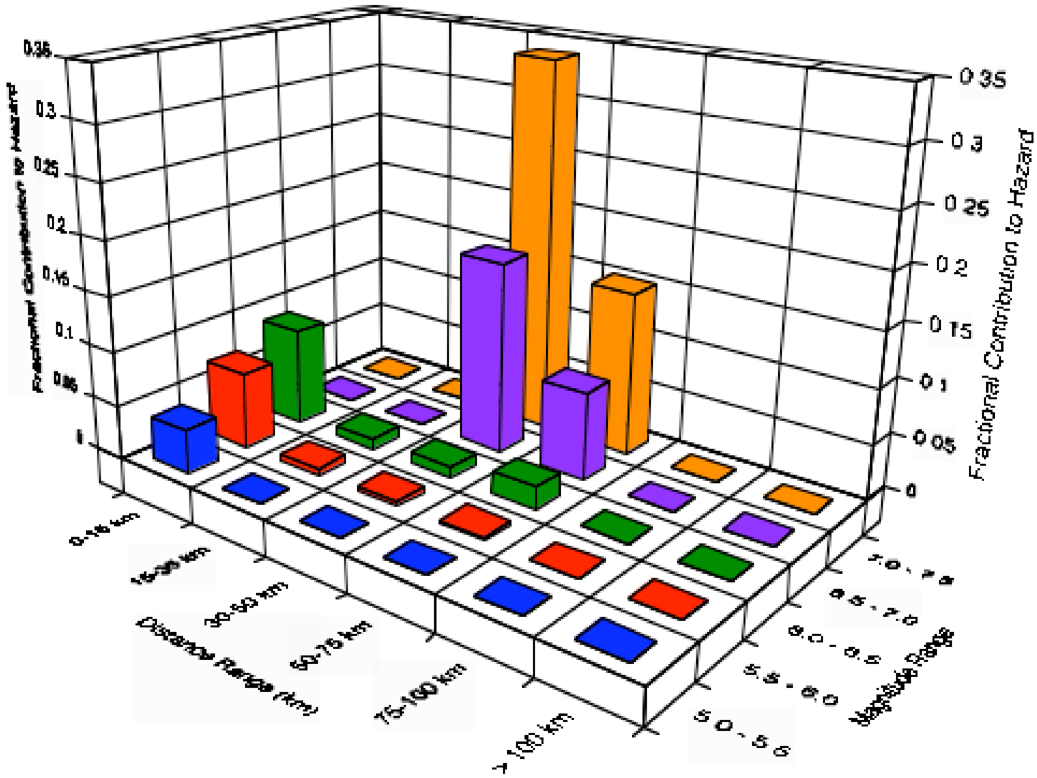

Figure 11.3a. Deaggregation for T=1 sec for a return period of 475 years.

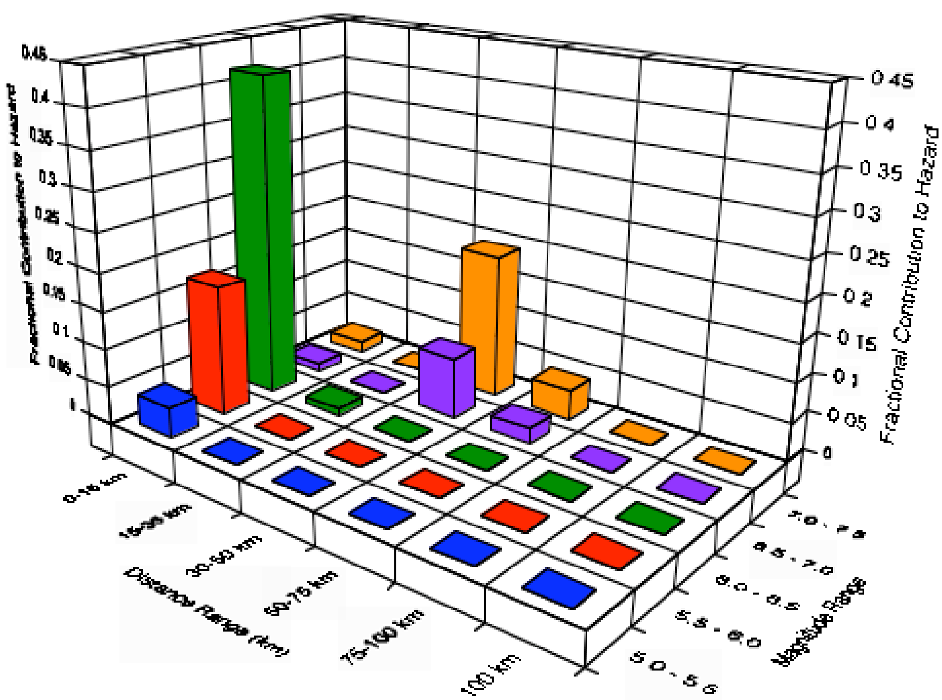

Figure 11.3b. Deaggregation for T=1 sec for a return period of 10,000 years.

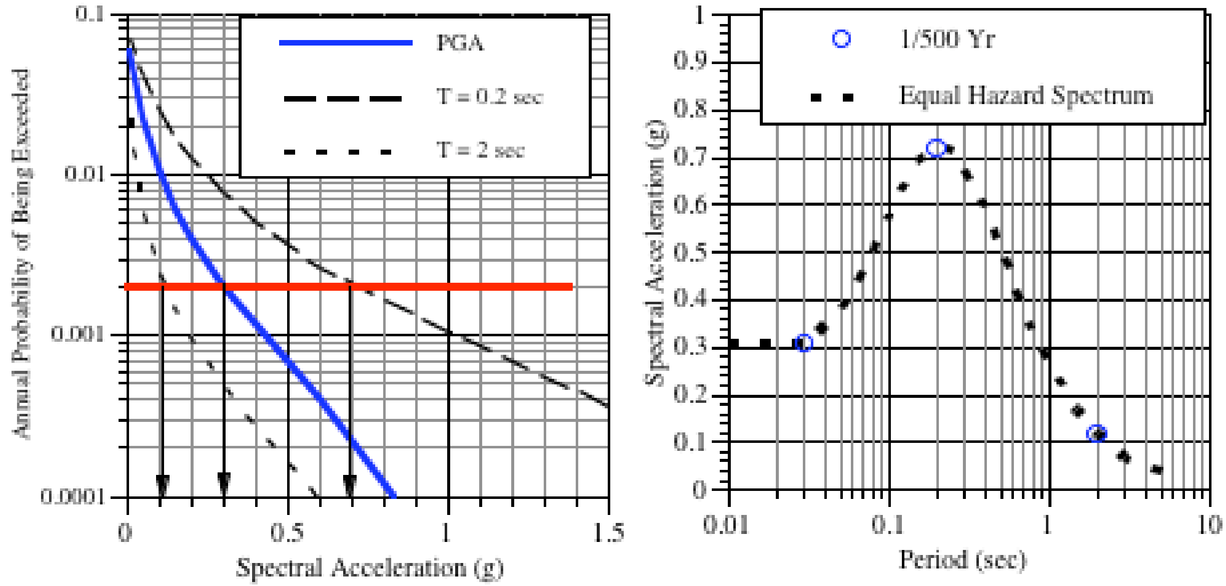

Figure 11-4. Procedure for developing equal hazard spectra. In this example, a return period of 500 years used.