Probabilistic Seismic Hazard Analysis (PSHA)

A Primer

Written by Edward (Ned) H. Field

These notes (available at http://www.relm.org/tutorial_materials) represent a somewhat non-standard treatment of PSHA; they are aimed at giving an intuitive understanding while glossing over potentially confusing details. Comments and suggestion are highly encouraged s (to field@usgs.gov).

The goal of probabilistic seismic hazard analysis (PSHA) is to quantify the rate (or probability) of exceeding various ground-motion levels at a site (or a map of sites) given all possible earthquakes. The numerical/analytical approach to PSHA was first formalized by Cornell (1968). The most comprehensive treatment to date is the SSHAC (1997) report, which covers many important procedural issues that will not be discussed here (such as the use of "expert opinion", the meaning of "consensus", and how to document results). Except where otherwise noted, the SSHAC report represents the best source of additional information (that I know of). It's a must-read for anyone conducting PSHA.

Traditionally, peak acceleration (PGA) has been used to quantify ground motion in PSHA (it's used to define lateral forces and shear stresses in the equivalent-static-force procedures of some building codes, and in liquefaction analyses). Today the preferred parameter is Response Spectral Acceleration (SA), which gives the maximum acceleration experienced by a damped, single-degree-of-freedom oscillator (a crude representation of building response). The oscillator period is chosen in accordance with the natural period of the structure (roughly number_of_stories/10), and damping values are typically set at 5% of critical (see Figure 1).

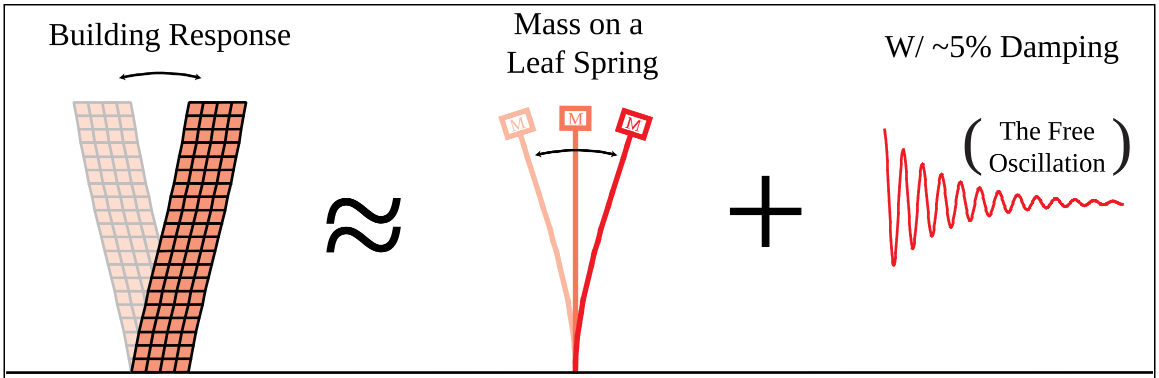


Figure 1. The response-spectrum value is the peak motion (displacement, velocity, or acceleration) of a damped single-degree of freedom harmonic oscillator (with a particular damping and resonant period) subjected to a prescribed ground motion.

To keep things simple, PGA will be used as the ground-motion parameter here (the analysis is otherwise equivalent).

PSHA involves three steps: 1) specification of the seismic-hazard source model(s); 2) specification of the ground motion model(s) (attenuation relationship(s)); and 3) the probabilistic calculation.

Seismic-Hazard Source Model:

Stated most simply, the seismic-hazard source model is a description of the magnitude, location, and timing of all earthquakes (usually limited to those that pose a significant threat). For example, a source model might be composed of N total earthquake scenarios (E_n), where each has its own magnitude (m_n), location (L_n), and rate (r_n):

$$E_n = E(m_n, L_n, r_n).$$

This total set might be composed of subsets that represent earthquakes on particular faults. Furthermore, m_n might represent the single characteristic magnitude of a specific fault, or it might represent a discrete value sampled from a fault or region that has a continuous (e.g., Gutenberg-Richter) distribution of events.

The location term L_n is usually given as a point or a rectangular surface, although any arbitrarily complex surface could be used. One of the biggest bookkeeping challenges in PSHA is specifying the location of events on a fault (e.g., the location of magnitude 5 events over a large fault that has changing strike and dip).

The rate term r_n represents the annual rate of the earthquake scenario (or one over the average repeat time). Technically speaking, the probability of the scenario over some specified time period should really be given; this would allow the implementation of time-dependent models that might imply greater or lesser likelihood than the long-term behavior. However, time-dependent models are usually implemented by converting the conditional probability into an equivalent Poissonian, time-independent rate (see box below for an example from WGCEP (1995)). Therefore, we keep the rate term r_n with the understanding that it may represent and "effective" Poissonian rate.

Example of Converting Time-Dependent Probability into an Effective Time-Independent Rate

WGCEP (1995) ascertained that the average repeat time of earthquakes on the San Bernardino segment of the San Andreas Fault is 146 years, giving a long-term rate of ~0.007 events per year. The Poissonian (time-independent) probability of having more than one earthquake over T years is:

$$P_{pois} = 1 - \exp(-r \cdot T)$$

Thus, the Poissonian probability for a San Bernardino segment event in the next 30 years is ~19%. However, considering that it had been ~184 years since the last event, they applied a time-dependent model (that assumed repeat times are log-normally distributed) and came up with a 30-year conditional probability of 28% ("conditional" because the probability would drop if the event occurred the next day, whereas the Poissonian probability never changes). Substituting the conditional probability (P_{cond}) for P_{pois} in the above equation, an effective Poissonian rate can be solved for as:

$$r_{eff} = -\ln(1 - P_{cond})/T$$

Thus, WGCEP (1995) came up with an effective 30-year rate of 0.011 events per year for the San Bernardino segment, which was then applied in the PSHA as a proxy for time dependence.

So again, the seismic-hazard source model is simply a list of scenarios, each with an associated magnitude, location, and effective rate. A more detailed overview of some particular seismic-hazard source models can be found at: http://www.relm.org/tutorial_materials.

Ground Motion Model (Attenuation Relationship)

The ground motion model used in PSHA is referred to as an Attenuation Relationship. Given the typically large number of earthquakes and sites considered in an analysis, attenuation relationships must be simple and easy to compute. Several have been developed by various practitioners (see the Jan/Feb, 1997 issue of *Seism. Res. Lett.* devoted to this topic). The most basic attenuation relationship gives the ground motion level as a function of magnitude and distance, but many have other parameters to allow for a few different site types (e.g., rock vs soil) or styles of faulting. Different relationships have also been developed for different tectonic regimes. All are developed by fitting an analytical expression to observations (or to synthetic data where observations are lacking). An example of the relationship developed by Boore, Joyner, and Fumal (1997) is shown in Figure 2 below.

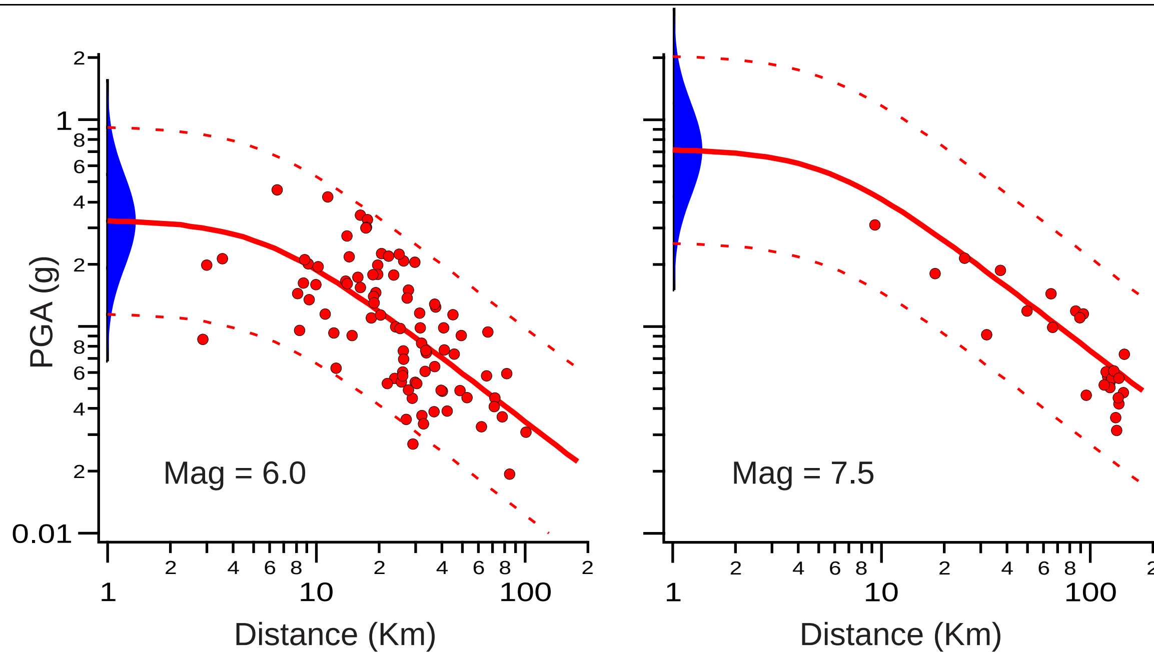


Figure 2. Median value (solid line) and 95% confidence region (dashed) predicted by the Boore, Joyner, and Fumal (1997) attenuation relationship for strike slip earthquakes and soil site conditions. The analytical expression for this particular case is:

$$\ln(PGA) = 0.53(M-6) - 0.39\ln(D^2+31) + 0.25$$

where D is the distance. The circles are southern California observations for events within 0.2 magnitude units of that computed. The blue normal curve represents the distribution of predicted values at 1 km distance.

As Figure 2 reveals, the observations scatter significantly about the predicted values. The present generation of curves represent this uncertainty with a log-normal distribution (that is, ln(PGA) has a normal or Gaussian distribution). The standard deviation of ln(PGA) is typically ~0.5, meaning that 95% of observations will fall within a factor of ~2.7 ($\exp(1.98 \cdot 0.5)$) of the median predicted PGA.

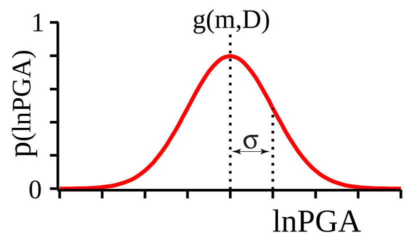
Although several different attenuation relationships are available, they often differ significantly in terms of predicted values where data are lacking (e.g., short distances to large earthquakes, as in Figure 2 where there is only one observation within 10 km for the magnitude 7.5 case). They also differ in whether sediments amplify or de-amplify PGA at high ground motion levels. Unfortunately, PSHA in seismically active regions is often most sensitive to the conditions where data lack and discrepancies exist. For more information, see SCEC's Phase III report for an exhaustive evaluation of attenuation relationships for southern California, as well as how they can be improved to account for site effects (http://www.scec.org/ phase3).

The Probabilistic Calculation

With the seismic-hazard source model and attenuation relationship(s) defined, the probabilistic-hazard calculation is conceptually simple. In practice, however, things can get messy. Besides the non-triviality of defining the spatial distribution of small earthquakes on large faults (mentioned above), there is also the problem that different attenuation relationships use different definitions of distance to the fault plane. Having noted these troublesome details, they will be glossed over here.

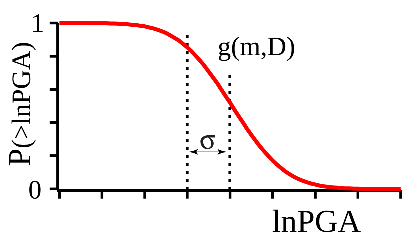
Consider the PSHA calculation for a particular site. The seismic-hazard source model has provided N earthquake scenarios E_n, each of which has an associated magnitude (m_n), location (L_n), and rate (r_n). From the scenario location we can determine the distance D_n to our site (again, ignoring the details of this calculation here). With m_n and D_n, the attenuation relationship tells us that the distribution of possible ground-motion levels for this scenario is:

$$p_n(\ln PGA) = \frac{1}{\sigma_n \sqrt{2\pi}} e^{-(\ln PGA - g(m_n, D_n))^2 / 2\sigma_n} = \qquad (1)$$



where $g(M_n,D_n)$ and σ_n are the mean and standard deviation of lnPGA given by the attenuation relationship. What engineers want to know is the probability of exceeding each lnPGA, so we integrate as follows:

$$P_n(> \ln PGA) = \frac{1}{\sigma_n \sqrt{2\pi}} \int_{\ln PGA}^{\infty} e^{-(\ln PGA - g(m_n, D_n))^2 / 2\sigma_n} d\ln PGA = \qquad (2)$$



Again, this gives the probability of exceeding each lnPGA were the event to occur. If we multiplying this by r_n, we get the annual rate at which each lnPGA is exceeded (R_n) due to the scenario:

$$R_n\,(>\ln PGA) = r_n P_n(>\ln PGA) \tag{3}$$

Then, summing over all N scenarios we get the total annual rate of exceeding each lnPGA:

$$R_{tot}(>\ln PGA) = \sum_{n=1}^{N} R_n(>\ln PGA) = \sum_{n=1}^{N} r_n P_n(>\ln PGA) \tag{4}$$

Finally, using the Poissonian distribution, we can compute the probability of exceeding each ground-motion level in the next T years from this annual rate as:

$$P_{pois}(>\ln PGA, T) \;=\; 1 - e^{-R_{tot} T} \tag{5}$$

This is referred to as the hazard curve.

As an example, consider a source model composed of two scenarios. The first event is an m 6.0 earthquake that occurs every 22 years, and the second is an m 7.8 earthquake that occurs every 300 years. Both are strike slip events, and located 10 km from the site of interest which is soil. The Boore, Joyner, and Fumal (1997) attenuation relationship gives median PGAs of 0.19 and 0.50 g for scenarios 1 & 2, respectively, and a standard deviation of 0.52 for lnPGA. The annual rates of exceedance for each of these scenarios (R_1 and R_2) as well as the total (R_{tot}) are shown in Figure 3, and the Poissonian probability of exceeding each ground-motion level over a 50-year period is shown in Figure 4.

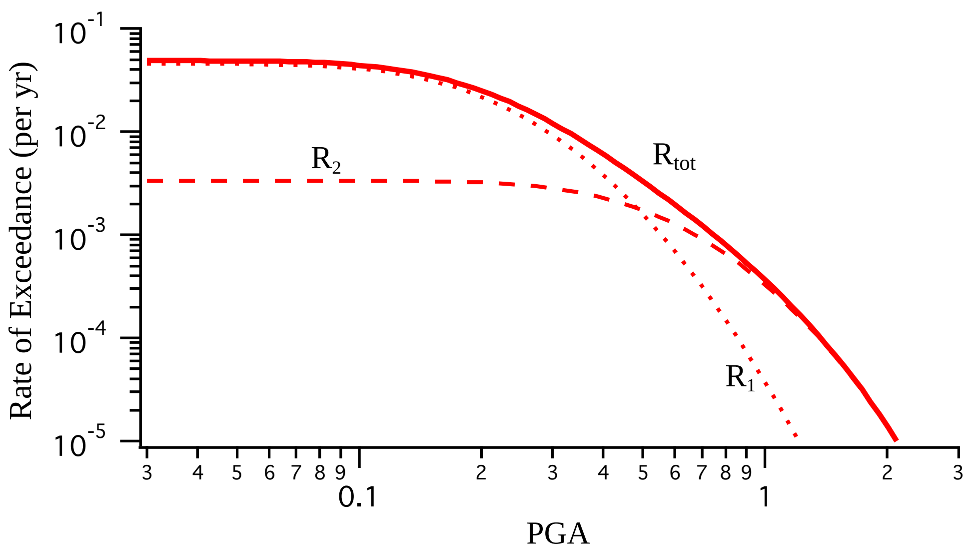


Figure 3. PGA exceedance rates for the two scenarios described in the text, as well as the total.

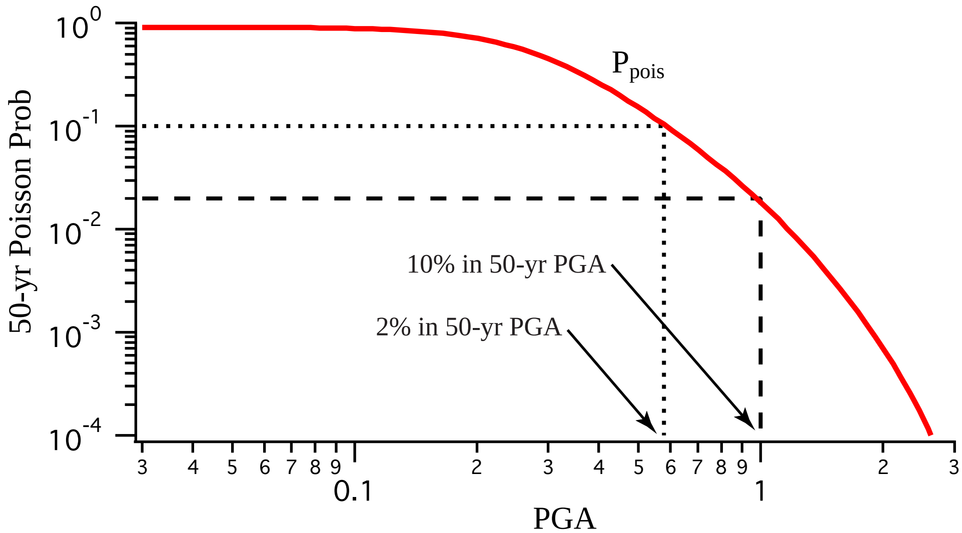


Figure 4. The hazard curve (50-year Poissonian probability of exceedance) for the two-scenario source model described in the text. The 10% and 2% chance of exceedance PGAs are 0.58 and 1.0 g, respectively.

It is often desirable to further reduce a hazard curve to a single value. This is typically done by choosing the PGA that has a 10% (or 2%) chance of exceedance in 50 years, as shown with the dotted line in Figure 4 (dashed line for the 2% level). One can then generate a map of these values for the region of interest. Alternatively, if the analysis has been done for several response spectral periods (Figure 1), rather than just PGA, then one can make a "uniform hazard spectrum" (e.g., a plot of the 10%-in-50-year spectral acceleration as a function of building period).

Handling Uncertainties

Note: The definition and treatment of uncertainties is not well established; the following is my take on the present literature. In addition to SSHAC (1997), see Anderson and Brune (1999)

As its name implies, PSHA can account for any uncertainties provided they are quantifiable. There are generally two categories of uncertainties that need to be handled differently:

"Epistemic" Uncertainties:
- "Of, related to, or involving knowledge" (Am. Heritage Dict.).
- Resulting from an inadequate understanding.
- With time (e.g., additional observations) these uncertainties can be reduced and the true value ascertained.
- e.g., the true mean and standard deviation of magnitude 7.8 strike-slip events recorded at a 10 km distance.

"Aleatory" Uncertainties
- "Dependent on chance" (Am. Heritage Dict.).
- Due to the intrinsic variability of nature.
- Over time, all values will eventually be sampled.
- e.g., the scatter, about the mean, of PGAs observed from all magnitude 7.8 strike-slip events recorded at 10 km.

Although this distinction sounds pleonastic, it's believed by most to be quite important. Aleatory uncertainties should be handled (e.g., averaged) in computing the exceedance rates (R_{tot} above), whereas epistemic uncertainties should be handled after computing the hazard curves (P_{pois} above).

Attenuation relationships exemplify this distinction. The aleatory uncertainty is reflected in the range of possible PGAs for the next earthquake of given magnitude and distance, as represented by the mean and standard deviation. This range of values is averaged over (integrated) in equation (2); that is, before computing the hazard (P_{pois}, equation (5)). However, as discussed previously, there are several different published attenuation relationships, each reporting a different mean and standard deviation at a given magnitude and distance. This represents epistemic uncertainty, as time will eventually tell what the true mean and standard deviation are. In this case we compute the hazard (P_{pois}, equation (5)) separately for each attenuation relationship, giving rise to a family of hazard curves.

Another example is in the two-scenario hazard model exemplified above. As stated, our site was subjected to magnitude 6.0 events every 22 years, and magnitude 7.8 events every 300 years (on average). Assuming both scenarios involve the same fault segment (e.g., Parkfield), this means that ~93% of events on the segment are magnitude 6.0 and ~7% are magnitude 7.8. Here, the question of what magnitude will occur next is aleatory in that both values will eventually occur. As appropriate, the contribution from each was summed in computing the total exceedance rate (R_{tot}, equation (4)).

Alternatively, suppose only one magnitude ever occurs on the segment, and all we know is that it's either a magnitude 6.0 event every 22 years, or a magnitude 7.8 event every 300 years. The uncertainty is now epistemic, as time will tell us which value it is. Here, we compute the hazard (P_{pois}, equation (5)) for each scenario separately, giving rise to the two curves shown in Figure 5.

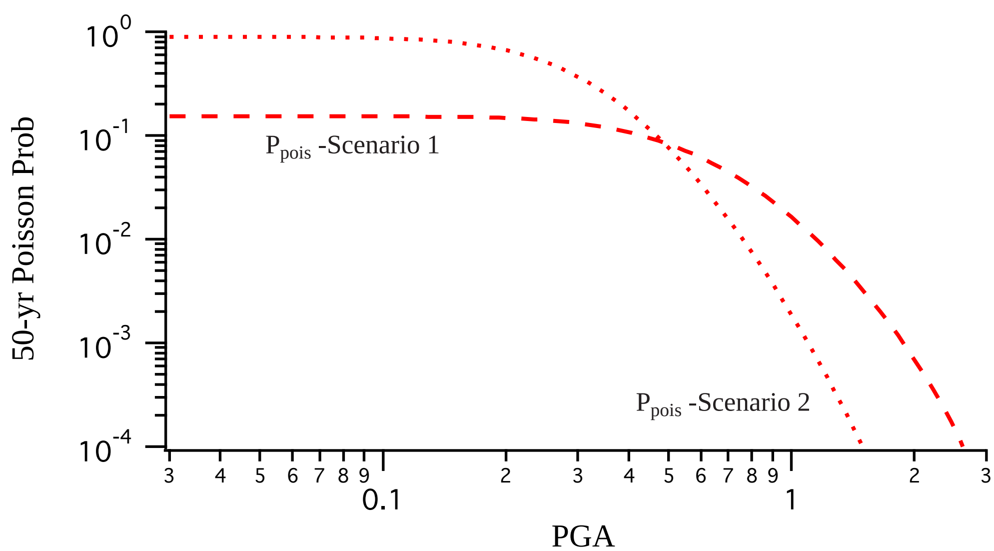


Figure 5. Hazard curves for the two scenarios treated as epistemic uncertainties.

The two hazard curves in Figure 5 versus only one if Figure 4 reflects an important difference. If there are no epistemic uncertainties, then we end up with only one hazard curve that has no uncertainty. If there is epistemic uncertainty, then we end up with a family (or distribution) of hazard curves (in Figure 5 we have two choices, or a binomial distribution); that is, the hazard curve itself is uncertain. Again, time will reveal the true hazard curve, but for now we are left with a distribution. To reduce the distribution to a single, representative hazard curve, we could simply take the average. However, reporting only this would be misleading if there is a large spread in the family of curves. Alternatively, one could report some fractile(s) of the distribution (e.g., the curve for which we are 95% confident that the true curve lies below).

Epistemic uncertainties have traditionally been handled with logic trees (discrete possible values with associated weights) while aleatory uncertainties are often handled with probability distributions (as in equation (2)). However, this distinction is one of convenience, as either type of uncertainty can, in principle, be treated with either continuous or discrete distributions.

Finally, it's important to note that the distinction between epistemic and aleatory uncertainty is model dependent. In fact, some argue that there is no such distinction in reality (aleatory uncertainty being ultimately due to a lack of knowledge). Such philosophical ramblings aside, the distinction is definitely useful in practice. Any model that purports to provide PGA as a function of only magnitude and distance will obviously have an intrinsic (aleatory) uncertainty that no amount of further research will reduce. On the other hand, in the future we could conceivably expand the attenuation relationship to include more and more parameters (e.g., stress drop), and thereby reduce the aleatory uncertainty to zero.

Other Issues

To keep things intuitively simple, many subtle issues have been glossed over in these notes. One obvious example is that Figure 4 implies a finite probability for PGA exceeding 2 g (which some might argue is physically impossible). In fact, the normal distribution used in equations (2) and (3) has a nonzero probability over all ground-motion levels. Therefore, some practitioners truncate the distribution at some upper and lower level (e.g., at +/- 3σ). Another important consideration is any correlation between epistemic uncertainties (e.g., correlation between a- and b-value estimates for a Gutenberg Richter distribution). Again, see the SSHAC (1997) report for more information on such issues.

As stated, many important computational details have been glossed over here. In fact, the myriad of subtle decision can cause significantly different results (generally defined as more than 10%) to be obtained by different practitioners attempting an identical analysis.

References

Anderson, J.G. and J.N. Brune (1999). Probabilistic hazard analysis without the ergodic assumption, *Seism. Res. Lett.* **70**, 19-23.

Cornell, C.A. (1968). Engineering seismic risk analysis, *Bull. Seism. Soc. Am.*, **58**, 1583-1606.

SSHAC (Senior Seismic Hazard Analysis Committee). Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts, *US Nuclear Regulatory Commission report* **CR-6372**, Washington DC.

WGCEP (Working Group on California Earthquake Probabilities) (1995). Seismic hazards in southern California: probable earthquakes, 1994 to 2024. *Bull. Seism. Soc. Am.*, **85**, 379-439.