Direct Calculation of the Probability Distribution for Earthquake Losses to a Portfolio

Robert L. Wesson,[a)] M.EERI, David M. Perkins,[a)] M.EERI, Nicolas Luco,[a)] M.EERI, and Erdem Karaca,[a),b)] M.EERI

We demonstrate a direct method for the calculation of the annual frequency of exceedance for earthquake losses (or the probability distribution for annual losses) to a portfolio. This method parallels the classic method of probabilistic seismic hazard analysis for the calculation of the annual frequency of exceedance for earthquake ground motions. The method assumes conditional independence of the random component of ground motions and losses at different sites for each earthquake, given magnitude, distance to the sites, and so-called interevent epsilon. Examples show that the method is realizable, and can take into account different loss functions and site conditions in the portfolio. The main advantage of this method is that it does not require a separate set of scenario earthquakes, as do Monte Carlo-based approaches, but can be calculated directly from the inputs used for hazard maps. [DOI: 10.1193/1.3159475]

INTRODUCTION

Calculation of the annual frequency of exceedance of earthquake loss for a single property is relatively straightforward. From the hazard curve, that is the annual frequency of exceedance for earthquake ground motion, it is easy to obtain an annual probability density function (pdf) for ground motion (Algermissen and Perkins 1976). Hazard curves for simplified site conditions are available for sites throughout the United States from the National Seismic Hazard Mapping website (http://earthquake.usgs.gov/research/hazmaps/), and also result from more detailed site-specific probabilistic seismic hazard analyses (PSHA). This probability density can be integrated with a probabilistic loss function for the property to obtain the annual probability density for loss, and thus the annual frequency of exceedance for loss (e.g., Cao et al. 1999). Alternatively, this calculation can be based on annual frequencies alone, converted to annual probability density.

If we are interested in the loss to a portfolio of properties, we can sum the mean annual loss for each of the properties to calculate the mean annual loss, that is, the expected loss, for the entire portfolio (e.g., *FEMA 366*, FEMA 2001). We cannot, however, calculate the full distribution of annual losses to the portfolio directly from this infor-

[a)] U.S. Geological Survey, Box 25046, Denver, CO, 80225
[b)] Swiss Re, 175 King St, Armonk, NY, 10504

Earthquake Spectra, Volume 25, No. 3, pages 687–706, August 2009; © 2009, Earthquake Engineering Research Institute

mation alone because in most realistic situations the losses are correlated.[1] In this paper we make progress toward solving this problem by considering the most well-understood, and perhaps most significant, sources of this correlation. An essential element of understanding this correlation is to understand the nature of the variability of ground motion given magnitude and distance, and to separate this variability into two parts, one part related to variability within one earthquake (intra-event variability) and variability between earthquakes (interevent variability). This separation of the variability is not required in the calculation of the hazard curve, as it only depends on the total variability, nor is it required, as discussed above, for the calculation of the mean loss. In the following, we show how a separation of the variability into these two components can be used to calculate the full distribution of losses using a direct method that parallels the classic method of PSHA and does not require a Monte Carlo simulation.

Generally, simulation-based portfolio loss modeling has consisted of simulating earthquakes one by one, obtaining random components for the site ground-motion variability, interacting with a loss function, also using random components for the loss uncertainty, and accumulating losses until some degree of stability is observed in the loss distribution. Examples include Bazzurro and Luco (2007), Crowley and Bommer (2006), and Park et al. (2007). (Sometimes sufficient stability is assumed, when a finite "representative" set of scenario earthquakes has been used.) Instead, in the development to follow, we accommodate the calculation to a process analogous to the seismic hazard calculation, interacting with loss functions in order to obtain annual loss exceedance curves.

ACCOUNTING FOR THE CORRELATION OF LOSSES

We can model earthquake losses as a random process made up of two constituent random processes: the ground motion given an earthquake and the loss given a ground motion. Obviously, the loss to a portfolio involves the losses to many different structures located at different sites, commonly with different structural types and site conditions. The loss to the portfolio can be viewed as the sum of the individual random variables describing the loss to each property or structure in the portfolio. Inherently, earthquake losses to nearby properties are correlated because of the similarity in ground motion experienced for a given earthquake. The essence of the direct method we propose is to condition on the "interevent" portion of the ground-motion variability, such that the conditional loss for an individual earthquake can be treated as the sum of independent random variables.

[1] The reader familiar with *FEMA 366* (FEMA 2001) should note that the plot therein (Figure C-1) of the annual frequency of exceedance versus the mean (or expected) annual loss for a set of uniform-hazard ground motion maps (from the USGS) is not a plot of the exceedance probability for a specified portfolio loss. The *FEMA 366* plot is simply a convenient depiction of the expected portfolio loss as a sum of the expected individual property losses, whereas the exceedance probability for the portfolio loss must account for the correlations focused on in this paper.

LOSS GIVEN GROUND MOTION

Consider first the loss given a ground motion. Let us use the term "fractional loss" to mean the cost to repair a structure, e.g., a single-family house or a bridge, divided by its total value or replacement cost. Empirical studies of loss suggest that the fractional loss, given a particular earthquake ground motion ("load"), can be characterized as a random function of that load. Recent studies suggesting relations between fractional loss and load include ATC-13 (ATC 1985), the HAZUS methodology (Kircher et al. 1997a and b), Smyth et al. (2004), Rossetto and Elnashai (2003), CUREE (Porter et al. 2002), and Wesson et al. (2004).

An important question for thought is "Given a ground motion, to what extent are the losses to similar structures independent?" In other words, suppose we consider two adjacent houses that can be reasonably assumed to experience the same ground motion; will the loss in one be related to the loss in the other? One can imagine that the houses might have any number of features in common: site condition, construction type, materials, builder, etc. We can consider each of these factors and any other that we can think of and/or have data to estimate, as explanatory variables and include them in the mean of the predictive function. *The essential point is that once we have included all the explanatory variables for which we have any data, then the remaining variability in loss given a ground motion can be reasonably assumed to be independent*. While in reality there may be important explanatory variables that are not included in the predictive function, it is difficult to estimate the resulting correlation, and here we assume that it is negligible.

GROUND MOTION

Consider next the ground motion. Ground motions from earthquakes generally decrease with distance away from the earthquake and show some dependence on magnitude. The literature describes statistical relationships, commonly called *attenuation relationships*, or ground motion prediction equations, that characterize the observations. Generally for a given distance, magnitude and other explanatory variables, the observations of ground motion are observed to follow a lognormal distribution. That is, for a given magnitude and distance, say, the attenuation relationships predict the mean of the natural logarithm of the ground motion plus a random component that is distributed as a normal distribution with zero mean and a standard deviation, σ_{total}, of the logarithms of the ground motion.

It is commonly recognized that the ground motions from two earthquakes with the same magnitude may vary systematically. From the point of view of estimating the distribution of earthquake losses to a portfolio, it is important to remember that the standard deviation, σ_{total}, is composed of two components. The first component, σ_{intra}, characterizes the randomness within any individual earthquake, that is to say, the observed variation in ground motion when the predicted value, given magnitude and distance, is taken into account. The second component, σ_{inter}, describes the between-earthquake variation in overall level of ground motion. Assuming independence between these two components, the total standard deviation is the square root of the sum of the squares of these two components.

If the total standard deviation were composed only of σ_{intra} (i.e., $\sigma_{inter}=0$), then holding the source location, site locations and magnitude fixed, the mean ground motion for the sites in a portfolio would be the same for all earthquakes. The importance of the interevent variability, σ_{inter}, is that the mean ground motion across sites for one earthquake can be significantly higher or lower than the mean for all earthquakes leading to correlated ground motions and losses for these earthquakes that will be significantly higher or lower than the mean loss. This characteristic has a strong effect on the shape (i.e., the tails) of the distribution of losses to a portfolio. It is convenient to capture this interevent variability by defining a random variable for each earthquake, ε_{inter}, that has the standard normal distribution. Then the interevent-variability can be described as $\varepsilon_{inter}\sigma_{inter}$. Thus for an earthquake with $\varepsilon_{inter}=0$, the mean of the ground motions for this earthquake would be the same as the mean for all earthquakes. For an earthquake with $\varepsilon_{inter}=1$, the mean of the ground motions this earthquake would be one standard deviation, or σ_{inter}, above the mean for all earthquakes; for $\varepsilon_{inter}=-1$, the mean for the earthquake would be one standard deviation, or σ_{inter}, below the mean for all earthquakes, and so on.

So for an individual earthquake, the ground motion at one site is given by a function of M, D, ε_{inter}, and a variation, σ_{intra}. Random effects affecting all sites and systematic differences among earthquakes are captured in the interevent variability. As with structures, an important question for thought is "Given an earthquake of magnitude M with interevent epsilon ε_{inter} and a ground motion for a site X at distance D from the earthquake, to what extent is the ground motion at site Y at the same distance similar?" In the model, the ground motions differ by the random intra-event component at each site, denoted ε_{intra}, which can reasonably be assumed to be independent. One caveat applies, however; for sites within a few kilometers of one another, local effects, such as focusing or defocusing owing to basin topography, basin reverberation, etc., may reduce the difference in the random effect or, alternatively, create a common bias. This issue will be discussed below.

The major development in the text to follow will deal with loss to a portfolio on an earthquake-by-earthquake basis, starting with one earthquake, using σ_{intra} as the random component. Because probabilistic loss to a portfolio involves taking into account all the possible earthquakes and σ_{inter} as well, the development will then be generalized to the issue of all the earthquakes.

METHOD

SEPARATING THE VARIABILITY IN GROUND MOTION

Consider the case of one earthquake of magnitude M at a fixed location, and a set of assets at sites i, with values a_i at distances D_i from the source. Let the ground motions, u_i, at each of the sites be given by

$$\log(u_i) = f(D_i, M) + \varepsilon_{total}\sigma_{total} \tag{1}$$

where $f(D_i, M)$, the attenuation relationship, is the mean natural logarithm of the ground motion as a function of distance and magnitude from the source, ε_{total} is a random term

that has the standard normal distribution with zero mean and standard deviation of one, and σ_{total} is the standard deviation of the total variability in (the natural logarithm of) ground motion (Boore et al. 1997, Abrahamson and Silva 1997). Other explanatory variables in addition to distance and magnitude, for example site condition and directivity, can be included in the mean if appropriate attenuation relationships for describing the influence of these variables on the mean are available. Below, we show an example that includes site condition and style of faulting.

We can rewrite Equation 1 as

$$\log(u_i) = f(D_i, M) + \varepsilon_{intra}\sigma_{intra} + \varepsilon_{inter}\sigma_{inter} \tag{2}$$

where the random term, ε_{total}, can be thought of as composed of two components, ε_{intra} and ε_{inter}, which are independent random variables with standard normal distributions, and σ_{intra} and σ_{inter} are the intra- and interevent standard deviations of ground motion, as described above. In hazard calculations the difference between the intra-event and interevent variability is commonly not an issue, and Equation 1 is appropriate with $\sigma_{total} = (\sigma^2_{intra} + \sigma^2_{inter})^{1/2}$. For our purposes, the difference between the intra-event and interevent variability cannot be ignored. For any one earthquake, however, we will have a particular value of ε_{inter}, and only the full range of the intra-event variability need be considered. In essence, for one earthquake we can condition on ε_{inter}, in addition to M and D_i. Thus we can write the probability density function for ground motion for one earthquake as

$$f_u(u|D_i, M, \varepsilon_{inter}) = \frac{1}{u\sigma_{intra}\sqrt{2\pi}} e^{-1/2(\log(u) - \log(u_0)/\sigma_{intra})^2} \tag{3}$$

where $\log(u_0) = f(D_i, M, \varepsilon_{inter})$ (Evans et al. 2000).

LOSS GIVEN GROUND MOTION

Next we consider the loss associated with this ground motion, using a loss (or vulnerability) function. The details of the vulnerability relation are not crucial to the argument developed below; in fact, any probabilistic form for loss given a ground motion could be used. For convenience, here we assume a scalar continuous relation, but below we address vector and discrete relations. We write the conditional probability density function for fractional loss given a ground motion as $f_l(l|u)$, where l is the loss fraction. The probability density functions for the fractional losses to the individual assets (e.g., homes) within the portfolio are given as

$$f(l_i|D_i, M, \varepsilon_{inter}) = \int_0^\infty f_l(l|u) f_u(u|D_i, M, \varepsilon_{inter}) du \tag{4}$$

Under the conditions described above, namely that we are conditioning the probability of ground motion on a given magnitude, a set of distances and ε_{inter}, the resulting ground motions and losses to the individual assets can be considered to be conditionally independent random variables, for the reasons described in the preceding section (see Pratt et al. 1995 or Schervish 1995 for a discussion of conditional independence).

PROBABILITY OF LOSS TO THE PORTFOLIO

Let the dollar losses to the elements of the portfolio be $l_i^{\$}=l_i a_i$, with corresponding probability density functions, $f_{l_i}^{\$}(l_i^{\$}|D_i, M, \varepsilon_{inter})$ obtained from Equation 4 by rescaling the fractional loss according to the dollar value of each element, a_i. Then the loss to the portfolio, L, from this one earthquake is the sum of the dollar losses to the elements of the portfolio:

$$L = \sum_i l_i^{\$}. \tag{5}$$

The probability density function for the sum of *independent* random variables is equal to the multiple convolution of the individual probability density functions (Freund 1962). Since for a given set of distances D_i, a particular magnitude M, and ε_{inter}, the $l_i^{\$}$ are assumed to be conditionally independent random variables, the conditional probability density function for the loss to the portfolio for the one earthquake can be calculated as the multiple convolution of the conditional probability density functions for the individual losses

$$f_L(L|M, \varepsilon_{inter}) = f_{l_1}^{\$} * \ \ldots \ * f_{l_n}^{\$} \tag{6}$$

where the $*$ symbol represents convolution. As will be discussed in more detail below, for the purpose of computation, we can take advantage of the fact that the Fourier transform of a multiple convolution is the product of the Fourier transforms of the functions being convolved. We can use fast Fourier transforms (FFT) to calculate the Fourier transforms of the probability densities of the loss for the individual assets and the inverse transform of their product to obtain the conditional probability density function for the loss to the portfolio.

We can then obtain the conditional probability for a loss to the portfolio greater than a particular value, L_0, for one earthquake by integrating Equation 6:

$$P(L > L_0|M, \varepsilon_{inter}) = \int_{L_0}^{\infty} f_L(L|M, \varepsilon_{inter})dL = 1 - \int_0^{L_0} f_L(L|M, \varepsilon_{inter})dL \tag{7}$$

MULTIPLE SOURCES, MAGNITUDES, AND ε_{inter}

With Equation 7 in hand, following the classical approach of PSHA, it is easy to write an expression for the frequency of losses greater than L_0, $\nu(L > L_0)$, which integrates or sums over multiple sources, events, magnitudes and ε_{inter}, weighting by the respective frequency of occurrence as

$$\nu(L > L_0) = \sum_{sources} \int_{M_{\min}}^{M_{\max}} g_s(M) \int_{\varepsilon_{inter}=-\infty}^{\infty} f_{\varepsilon_{inter}}(\varepsilon_{inter}) P(L > L_0|M, \varepsilon_{inter}) dM d\varepsilon_{inter} \tag{8}$$

where $g_s(M)$ is the magnitude-frequency relation for the sth source, and $f_{\varepsilon_{inter}}(\varepsilon_{inter})$ is the probability density function for ε_{inter}. Several important comments need to be made about Equation 8. First, even though we have written Equation 8 with a combination of

summation and integral signs, in practice the expression is calculated as a sequence of summations over bins. Second, whether we write Equation 8 with integrals or sums, because the expression is linear, we are free to change the order of integration or summation, and as will be shown in the examples it is commonly convenient to do so. Finally, $f_{\varepsilon_{inter}}(\varepsilon_{inter})$, the probability density function for ε_{inter}, is just the standard normal distribution.

This development follows the corresponding expressions for the frequency of ground motions used in the calculation of hazard curves (Cornell 1968). Continuing this analogy with the estimation of ground motion hazard (Algermissen and Perkins 1976), we can obtain an expression for the probability density function of annual loss to the portfolio as

$$f(L) = F'(L) = -\frac{d\nu(L > L_0)}{dL} e^{-\nu(L>L_0)} \tag{9}$$

METHOD OF CALCULATION

Although the explanation above has been given in terms of continuous variables, the calculations are considerably easier if we replace the continuous probability density functions by what are sometimes called "probability vectors," (c.f., Wang 1998, Wesson and Perkins 2001). These vectors simply contain as elements the discrete probabilities that the random variable will assume a value within a finite number of discrete increments. From a computational point of view this approach is particularly desirable for many reasons. First, the cumulative sum of the probability vector representing the probability density function is a vector representing the cumulative distribution function. (Vectors representing probability density functions must sum to one.) Second, by representing the conditional probability density function for a particular value of the conditioning variable as a vector, we can represent the conditional probability density function for a set of discrete values of the conditioning value as a matrix. This leads to the third reason, that by using this approach, the integral in Equation 4 can be reduced to a matrix multiplication. Fourth, with a judicious selection of intervals, numerical difficulties induced by the "spikey" nature of some probability density functions can be avoided. Finally, the convolution of the probability density functions can be very easily represented by the calculation of the discrete Fourier transforms though the use of an FFT algorithm and the multiplication of the transforms.

MATRIX FORMULATION

Briefly, for a series of ground motion levels, $u_i = u_1, u_2, \ldots, u_n$, the probability vector corresponding to the probability density for ground motion can be written

$$\mathbf{f_u} = [F(u_1), F(u_2) - F(u_1), \ldots, F(u_n) - F(u_{n-1})] \tag{10}$$

where $F(u)$ is the cumulative distribution function for ground motion u corresponding to Equation 2. Similarly, for a set of levels of fractional loss, $l_i = l_1, l_2, \ldots, l_m$, using Equation 3, we can write the matrix corresponding to the conditional probability density function for loss as

$$\mathbf{f}_{\mathbf{l}|\mathbf{u}} = \begin{pmatrix} P(l \leqslant l_1|u_1) & P(l_1 < l \leqslant l_2|u_1) & \cdots & P(l_{m-1} < l \leqslant l_m|u_1) \\ P(l \leqslant l_1|u_2) & P(l_1 < l \leqslant l_2|u_2) & \cdots & P(l_{m-1} < l \leqslant l_m|u_2) \\ \vdots & \vdots & \ddots & \vdots \\ P(l \leqslant l_1|u_n) & P(l_1 < l \leqslant l_2|u_n) & \cdots & P(l_{m-1} < l \leqslant l_m|u_1) \end{pmatrix} \tag{11}$$

and the integral in Equation 4 can be written

$$\mathbf{f}_{\mathbf{l}} = \mathbf{f}_{\mathbf{u}}\mathbf{f}_{\mathbf{l}|\mathbf{u}} \tag{12}$$

Rescaling the arguments of the $\mathbf{f}_{\mathbf{l}}$ by the value of each asset as describe above, and writing the FFT of the resulting $\mathbf{f}_{\mathbf{l}}^{\$}$ as $\Phi(\mathbf{f}_{\mathbf{l}}^{\$})$, then the conditional probability density function for the loss to the portfolio of q properties for one earthquake (i.e., for a given magnitude, set of distances, and value of ε_{inter}) can be represented as

$$\mathbf{f}_{\mathbf{l_P}}^{\$} = \Phi^{-1}(\Phi(\mathbf{f}_{\mathbf{l_1}}^{\$})\Phi(\mathbf{f}_{\mathbf{l_2}}^{\$}) \ldots \Phi(\mathbf{f}_{\mathbf{l_q}}^{\$})) \tag{13}$$

where the numerical subscripts denote the properties in the portfolio. The quantity corresponding to Equation 7, the exceedence probability for loss to the portfolio, is simply given by the cumulative sum of the vector in Equation 13. From this point the calculation proceeds in parallel with the standard approach for probabilistic seismic hazard analysis to calculate exceedance curves as in Equation 8. Note that the matrix representing the vulnerability relations in Equation 11 need only be calculated once (for each type of structure).

For a given desired level of accuracy, the calculations can be accomplished efficiently by calculating $\mathbf{f}_{\mathbf{l}}$ and $\mathbf{f}_{\mathbf{l}|\mathbf{u}}$ at logarithmic increments. Unfortunately the FFT algorithm requires a constant increment. Prior to calculating the FFT, at the same time that the fractional losses are converted to dollar loss, they can be resampled using interpolation at a constant increment.

In the case of problems involving a large number of sites, the vector for the loss to the portfolio, L, has to involve a large number of (constant) loss increments since the conditional probability distributions of losses for the whole portfolio as well as the individual structures have to be captured by the same vector. One way to achieve a desired level of accuracy with reduced computational/memory requirements is to use Equation 13 to obtain the conditional probability density functions of losses for subsets of the portfolio of structures rather than the whole portfolio. The conditional loss distributions are then resampled and convolved with each other again using Equation 13 to obtain the same function for the combined subsets of structures or ultimately for the total portfolio. This procedure involves multiple inverse FFT operations while the length of loss vectors used is smaller. The optimum choice for the length of the loss vectors or number of subsets of portfolio structures depends on the size of the problem, the desired level of accuracy, and computational requirements, i.e., memory limitations.

Table 1. Parameters of single-family homes in example

Value ($)	Small house 250,000	Large house 1,000,000
Epicentral distance, D (km)	10	0
Site condition, NEHRP Site class and V_s	B (1070 m/sec)	D (250 m/sec)
Loss function	CUREE small house (Porter et al., 2002)	CUREE large house (Porter et al., 2002; K.W. Porter, personal communication, 2007)
Ground motion parameter required for loss function	PGA	0.2 sec spectral acceleration
σ_{intra}	0.476	0.535
σ_{inter}	0.301	0.323

INTEGRATION OVER ε_{inter}

The integration in Equation 8 over ε_{inter} simply uses the pdf for ε_{inter}, which is that of the standard normal distribution. From a computational point of view we can replace the integration in Equation 8 with a summation. The sum includes values of $P(L > L_0 | M, \varepsilon_{inter})$ corresponding to intervals in a desired range in ε_{inter}, say from $-n$ to $+n$, where n is the desired maximum number of standard deviations from the mean. Each of these conditional probabilities is weighted by the probability that ε_{inter} will fall within that interval.

EXAMPLES

We illustrate the application of the methods described above by consideration of two examples. In the first, we consider an extremely simple portfolio composed of two houses, using different site conditions and two different continuous loss functions subjected to one repeated "characteristic" earthquake with ground motions represented by one attenuation relation. In the second example we consider a large portfolio with multiple discrete loss functions subjected to multiple sources and considering multiple attenuation relations.

EXAMPLE 1—TWO HOUSES, ONE EARTHQUAKE

As a first example consider a portfolio composed of two houses, one small on a stiff site and one large on a soft site. With the parameters shown in Table 1 we subject the portfolio to a magnitude 6.9 earthquake. The loss function for the small house is conditioned on peak ground acceleration (PGA), while that for the large house is conditioned on 0.2 sec spectral acceleration (SA). We use the ground motion relations of Boore et al. (1997) for these ground motion measures for the site condition at each site, but we use the estimates of σ_{intra} and σ_{inter} given by Lee et al. (2000). (We apply their estimates for 0.3 sec SA to 0.2 sec SA.)

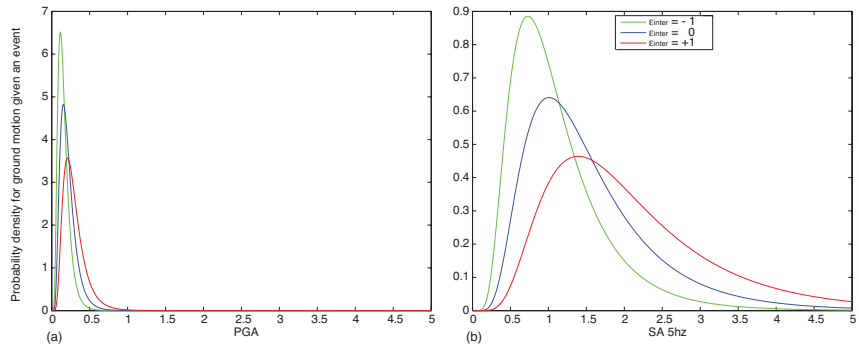

Figure 1. Probability density functions for ground motions given an earthquake a) at the site occupied by a "small house" with a V_{S30} of 1070 m/sec located 10 km from the epicenter, and b) at the site occupied by a "large house" with a V_{S30} of 250 m/sec and located at the epicenter.

Figure 1 shows the probability density functions for the ground motions at each site. Curves are shown assuming the appropriate values of σ_{intra}, but fixing $\varepsilon_{inter}=-1,0,+1$ to also show the impact of the interevent variability.[2] Obviously, the large house, located closer to the epicenter and on a much softer site experiences considerably higher ground motions. These probability densities correspond to Equations 3 and 10.

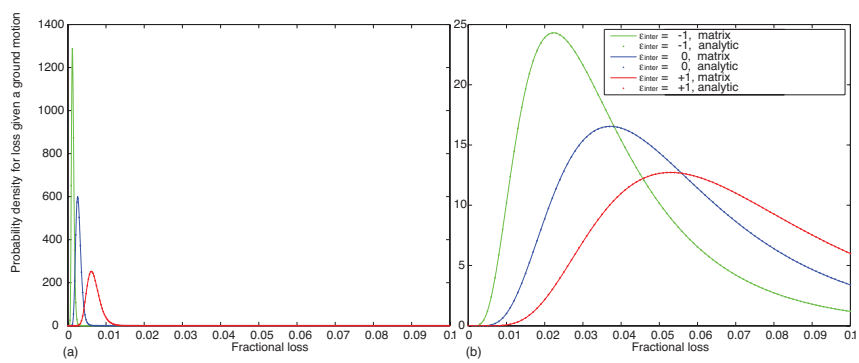

Figure 2. Conditional probability density functions (pdf's) for loss given a ground motion for a) the "small house" and b) the "large house" using loss functions derived from the CUREE Wood Frame Project (Porter et al. 2002, Porter 2007).

[2] In this example we have chosen to use loss functions for each of the two houses that depend on different measures of the ground motion. While some information about how the interevent variability of these different ground motion parameters is correlated exists (Baker and Cornell 2006), it is approximate, so here we will assume that the interevent variability is perfectly correlated between the two variables. Partial correlation can be dealt with in a manner similar to the treatment of vector descriptions of hazard, discussed in the Appendix. Obviously if the same measure of ground motion is used for the loss function for each element of the portfolio, as is the case in Example 2 to follow, this assumption is not required.

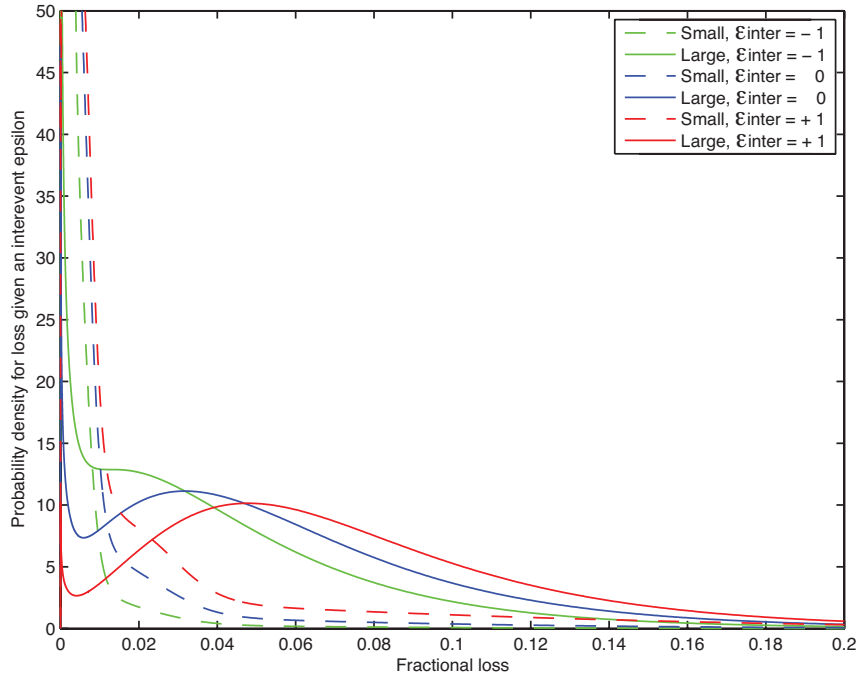

Figure 3. Probability density functions for loss given an event for the "small house" and the "large house." Curves are shown for $\varepsilon_{inter}=0$ and for interevent epsilons of plus and minus one ($\varepsilon_{inter}=\pm 1$). Notice that both houses, but particularly the small house, have large probability densities at zero loss. (The maximum for the "small house" is not shown.)

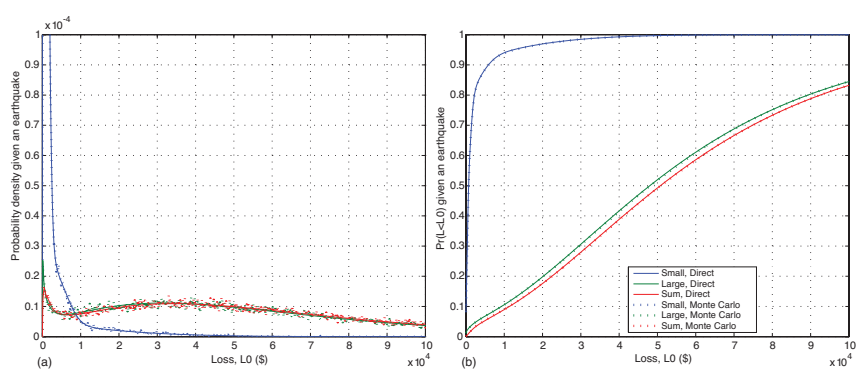

Figure 4. a) Probability density functions for loss and b) cumulative probability for loss, given an earthquake (with $\varepsilon_{inter}=0$) for each of the two houses and for the portfolio composed of both. Results are shown both for the direct method of calculation described in this paper as well as for a Monte Carlo simulation.

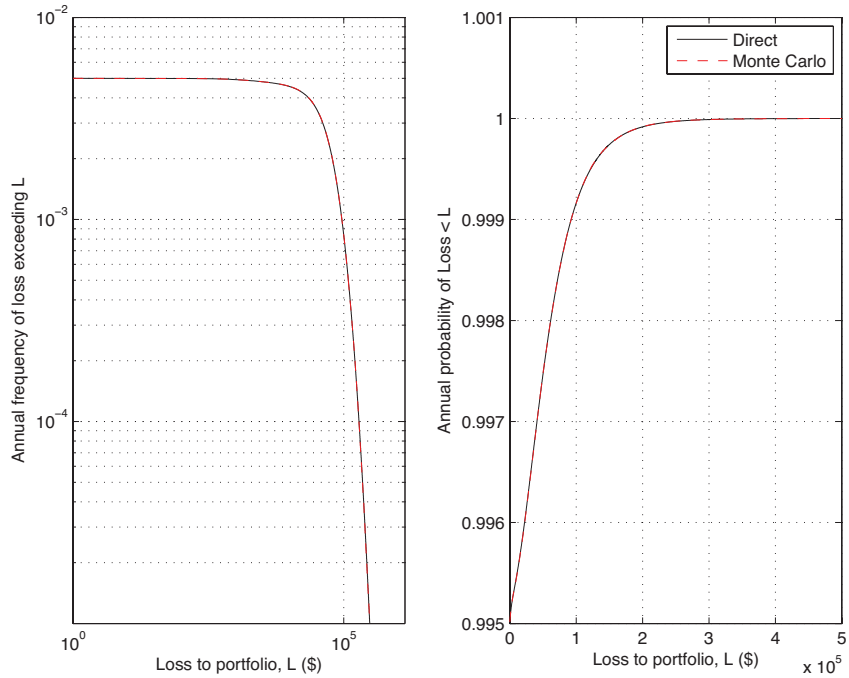

Figure 5. a) The annual frequency of exceedance for a loss greater than L for the portfolio of two houses and $\varepsilon_{inter}=0$. Results for both the direct method and a Monte Carlo simulation are shown. b) The same results as for (a), but transformed into a cumulative probability distribution assuming that the events causing the losses are Poisson distributed in time.

The method described above can be used with any loss function that can be cast in the form of conditional probability density function for loss given a ground motion (or vector of ground motions). In this example we use the relations given by Porter et al. (2002) for wood frame housing. These relations express the fractional loss as a lognormal distribution in which both the mean and the standard deviation are functions of the ground motion parameter and can be expressed as

$$f_l(l|u) = LN(\mu(u), \sigma(u)) \tag{14}$$

where l is the loss fraction, LN is the lognormal distribution, and μ and σ are the mean and standard deviation, respectively, both functions of the ground motion measure, u. Figure 2 shows the probability densities for loss given a ground motion, for the means of the ground motions shown in Figure 1, as well as for ground motions one interevent standard deviation above and below those means. Curves are plotted for both analytical estimates of the pdf's as well as from the numerical approximations using the matrix technique described in the text. The results are indistinguishable. The levels of loss expected and their variability (in linear space) are very much larger for the "large house,"

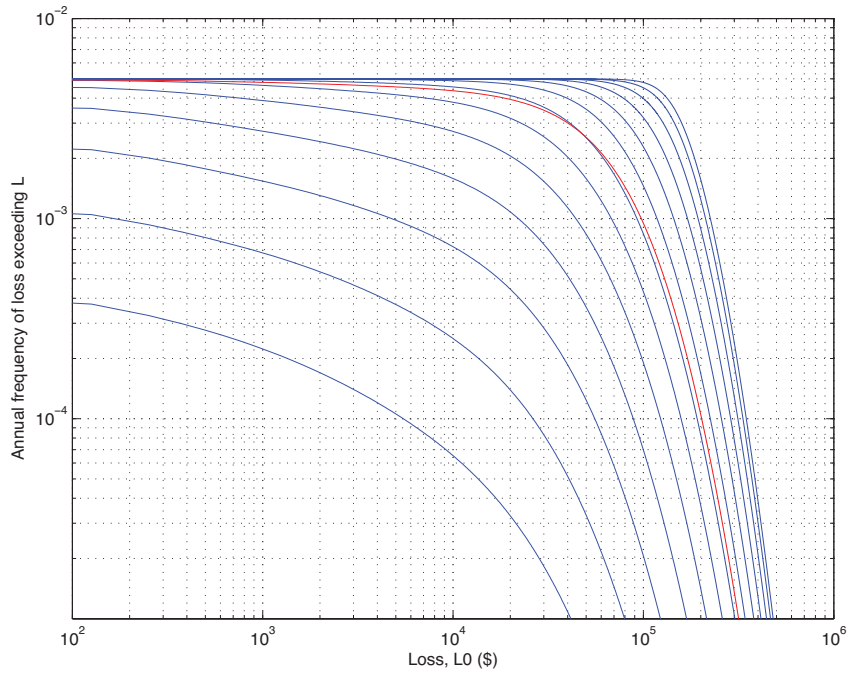

Figure 6. Exceedance curves for loss to the portfolio of Example 1 for values of ε_{inter} (blue curves) ranging from −6 to +6, that is, for six interevent standard deviations on both sides of the mean ground motion. The sum with each component weighted by its corresponding probability is shown in red. This curve is directly analogous to the "hazard curve" for ground motion that results from a classical probabilistic seismic hazard analysis.

owing to the larger ground motion (with larger variability) at its site. These probability densities correspond to $f_l(l|u)$ and Equation 11.

Figure 3 shows the probability densities for loss given three values of ε_{inter}, that is, the probability densities including all the intra-event variability, but holding ε_{inter} fixed. These probability densities correspond to Equations 4 and 12. As one would expect given the results in Figures 1 and 2, the probabilities of larger losses are higher for the "large house."

Finally, Figure 4 shows the probability density and perhaps more familiar cumulative probability for losses to each of the houses and for the portfolio of both. The probability density for the portfolio corresponds to Equations 6 and 13 and cumulative probability distribution for the portfolio correspond to Equation 7. To compare with the results of our direct calculation, we also show the results of a Monte Carlo simulation which, as can be seen, gives the same result.

To determine the annual distributions for loss we must multiply the probabilities of loss given an earthquake, by the annual frequencies of earthquakes, and also to integrate over the interevent variability. To address the first issue, consider that the magnitude 6.9

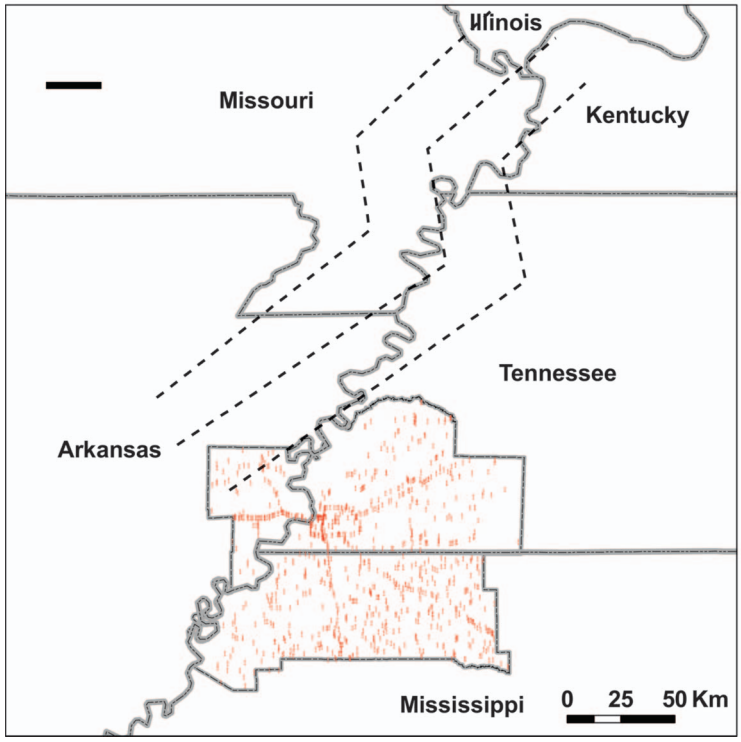

Figure 7. Map of the bridges (dots) within the Memphis Metropolitan Area. The locations of the three pseudo-faults (dashed lines) used to model New Madrid seismic zone earthquakes are also shown for reference.

earthquakes occur on average every 200 years. The resulting curves for the annual frequency of exceedance for loss, and the cumulative probability (for $\varepsilon_{inter}=0$) are shown in Figure 5.

Finally, we perform the integration over ε_{inter}. Figure 6 shows loss exceedance curves for values of ε_{inter} ranging from −6 to +6, together with the sum weighted by the probability of each. This sum represents the completion of the calculation shown in Equation 8.

EXAMPLE 2—MANY STRUCTURES, SEVERAL EARTHQUAKES, DISCRETE DAMAGE STATES

To show that the calculation of the annual loss distribution using the approach described above is feasible for a large number of properties, we calculate the results for a similar example involving 1,131 properties. We concentrate on bridges located within the Memphis, TN metropolitan area, which consists of the nine counties around Memphis, as shown in Figure 7. This region is under the threat of large magnitude damaging events from the New Madrid seismic zone (NMSZ).

The locations, values, and relevant structural properties of the bridges used are extracted from the bridge inventory of HAZUS (FEMA 2003), while the site conditions are obtained from Bauer et al. (2001). The bridge inventory has a total replacement value of $1.553B. We use the seismic fragility functions developed by Nielson (2005) for eight different bridge types commonly found in moderate seismic zones. The fragility functions use PGA as the ground motion intensity parameter and provide the probability of exceedance of four discrete damage states: slight, moderate, extensive, and complete. We used fractional losses of 0.03, 0.08, 0.25, and 1.00, respectively, for the damage states, based on HAZUS. As a result, the corresponding loss functions are conditional probability *mass* rather than *density* functions for the loss given a ground motion, where the loss can be equal to one of the five discrete values, the four fractional losses above, plus zero for no damage. This difference does not change our conceptual methodology, and since the method of calculation utilizes discretized probability vectors, nor does it change the implementation.

To represent the potential earthquakes in the NMSZ, we used a total number of 12 events on the three faults used by Frankel et al. (2002) in developing the 2002 USGS's National Seismic Hazard Maps (see Figure 7). The potential earthquakes on these faults were assigned a combined return period of 500 years. The central fault was given twice the weight (or rate, when one interprets this uncertainty as aleatory) of each of the side faults and the characteristic magnitudes used were M7.3, M7.5, M7.7, and M8.0 with weights (or probabilities) of 0.15, 0.20, 0.50, and 0.15, respectively. Also, we use the five alternative ground motion attenuation relations used by Frankel et al., with the weights that they assign to them.

We compute the probabilistic annual losses for the bridge inventory using a log standard deviation of 0.75 for ground-motion variability. We assume that interevent variability is half the intra-event variability, based on the findings of Lee et al. (2000), yielding values of $\sigma_{inter}=0.335$ and $\sigma_{intra}=0.671$.

Figure 8a shows the loss exceedance curves calculated using the direct method for each of the 12 considered events (dotted lines) and for one of the five ground motion attenuation relations considered, specifically for Campbell (2003). Note that each of these curves has already been integrated over ε_{inter}. The final loss exceedance curve for this specific attenuation relation (solid line) is simply obtained by summing the rates of exceedance of different losses (i.e., the loss exceedance curves) for each event. Note that the rate of exceedance of zero loss is equal to the total rate of occurrence for the New Madrid events, i.e., 1/500 years.

The loss exceedance curves for each of the five attenuation relations considered (thin solid lines) are plotted in Figure 8b together with their weighted mean (thick solid lines). This demonstrates that epistemic uncertainties can be quantified in our direct method in the same way that they are in probabilistic seismic hazard analysis. Also shown in the figure are the results obtained from Monte Carlo simulations (dashed lines). The agreement between the two sets of curves is excellent.

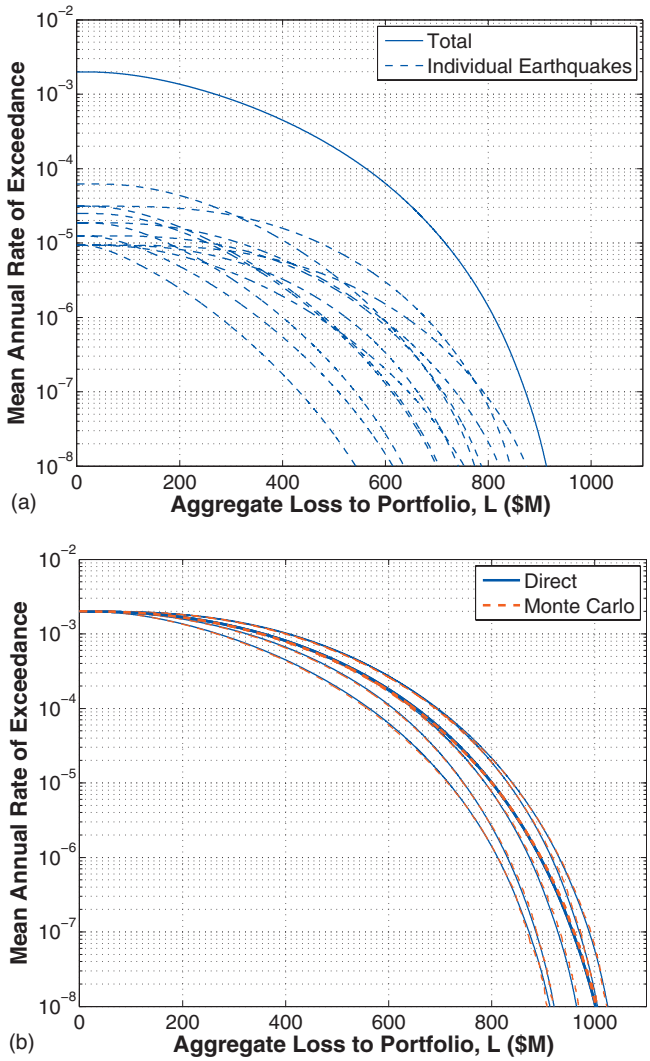

Figure 8. a) The loss exceedance curves for the portfolio of bridges calculated using the direct method with ground motions obtained from the Campbell (2003) ground motion attenuation relation. The dotted lines show the curves for individual earthquakes, while the solid line shows the summed total. b) The loss exceedance curves obtained for the five ground motion attenuation relations considered (thin lines) and their weighted mean (bold lines).

DISCUSSION

The direct method described above has several desirable characteristics. First, the direct method is, in effect, an efficient approach to sampling the high-loss tail of the distribution for aggregate loss. Second, the direct method follows directly from commonly available descriptions of hazard and therefore does not require a separate event set. Cur-

rently if it is desired to estimate the loss distribution of a portfolio using a Monte Carlo method that is consistent with a hazard estimate made using the classic approach for probabilistic seismic hazard estimation, it is necessary to produce a set of events that encapsulate the hazard model. The direct approach eliminates the need for such event sets.

Using this direct approach it is possible to calculate the loss distribution for a portfolio using a wide variety of hazard models. We show in the Appendix how it can be extended to multi-parameter, vector-based models. The approach can take into account variations in site conditions and other propagation and source effects for which attenuation relations and site data (Wills et al. 2000, for California) are available. Finally the approach can take advantage of a wide variety of loss functions—indeed, any that can be cast in terms of a conditional probability density function for loss given ground motion.

The only shortcomings apparent at present are the difficulties of incorporating the short wavelength spatial correlation in ground motion that has been documented in some cases (e.g., Evans and Baker 2006) and any correlation in the losses for a given ground motion. Conceivably these correlations could be incorporated in a manner that is analogous to the way in which our direct method already accounts for the correlation in ground motion for sites experiencing the same earthquakes, but this is a topic of future work. In the meantime one can employ Monte Carlo approaches to incorporate the additional correlations (e.g., Bazzurro and Luco 2007).

We remind the reader also that the development depending on M and D only is just a convenience. Many factors affect strong ground motion; these include site condition and style of faulting, which in fact were accounted for in the examples of this paper, as well as directivity, basin effects, and other factors. Work is currently underway to improve strong motion attenuation relationships to take into account these factors, for example, by the Next Generation Attenuation (NGA) project. For simplicity, in this paper we use the form of the attenuation relationship given in Equations 1 and 2, but this simplicity is not central to the approach described here. What is central is that for the foreseeable future strong ground motion will be characterized by a combination of physically based deterministic components, and inter- and intra-event stochastic components, where the interevent component results in the correlation considered in this paper. Physical understanding of ground motion will improve, allowing the inclusion of factors in attenuation relationships that would otherwise induce variability and correlation, but some random variability and correlation will likely remain. As these more sophisticated attenuation relationships become available, they can be substituted for the simple relationship in Equations 1 and 2.

CONCLUSIONS

A direct approach to the estimation of the complete probability distribution for earthquake losses to a portfolio is feasible and offers significant advantages. The principal advantages are that it closely parallels the classic approach for the probabilistic seismic hazard and does not require a separate set of events constructed to be an equivalent description of the hazard.

ACKNOWLEDGMENTS

The authors are extremely grateful to Tianqing Cao, Art Frankel, J.J. Love, Lixin Zeng, and two anonymous reviewers for reviews of this paper.

APPENDIX: VECTOR DESCRIPTION OF HAZARD

The method described above has assumed a loss dependent on a scalar ground motion. Hazard is beginning to be described by a vector of ground motion parameters (or intensity measures) such as 0.2 second and 1.0 second spectral acceleration, rather than one single parameter (Bazzurro and Cornell 2002, Baker and Cornell 2005). For a two-parameter case, the probability density function for the logarithms of the ground motion parameters can be described as a bivariate normal distribution with a correlation coefficient ρ, representing the correlation among the random components of the logs of the two ground motion parameters (Baker and Cornell 2006). An expression of this type would replace Equation 2 above. Similarly, the integral for the probability density function for loss in Equation 4 would be rewritten as

$$f(l_i|D_i,M,\varepsilon_{inter_a},\varepsilon_{inter_b}) = \int_0^\infty \int_0^\infty f_l(l|u_a,u_b) f_{u_a,u_b}(u_a,u_b|D_i,M,\varepsilon_{inter_a},\varepsilon_{inter_b}) du_a du_b \tag{A1}$$

where u_a and u_b are the two ground motion parameters. The condition and probability density function for loss and the probability density function for ground motion are now multivariate functions.

Starting with a vector description of hazard it is still possible to obtain a relation of the form of Equation 12. Although the description of the hazard is now a two-dimensional surface rather than a one-dimensional curve, a matrix analogous to the probability density function can be obtained using the approach described by Wesson and Perkins (2000). This involves calculating the cumulative distribution function for the bivariate normal distribution (Hull 1997), and then using an operator for mixed partial differentiation (Abramowitz and Stegun 1964). Then the rows of the matrix can be strung end to end to obtain a single vector of the form of Equation 10. Similarly the matrix analogous to the conditional probability density function of loss given the ground motion must be reformatted.

REFERENCES

Abrahamson, N. A., and Silva, W. J., 1997. Empirical response spectral attenuation relations for shallow crustal earthquakes, *Seismol. Res. Lett.* **68**, 94–127.

Abramowitz, M., and Stegun, I. A. (editors), 1965. *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, Dover Publications, New York, 1046 pp.

Algermissen, S. T., and Perkins, D. M., 1976. *A probabilistic estimate of maximum acceleration in rock in the contiguous United States*, *U.S. Geol. Surv. Open-File Rept. 76–416*.

Baker, J. W., and Cornell, C. A., 2005. A vector-valued ground motion intensity measure consisting of spectral acceleration and epsilon, *Earthquake Eng. Struct. Dyn.* **34**, 1193–1217.

——, 2006. Correlation of response spectral values for multicomponent ground motions, *Bull. Seismol. Soc. Am.* **96**, 215–227.

Bauer, R. A., Kiefer, J., and Hester, N., 2001. Soil amplification maps for estimating earthquake ground motions in the central U.S., *Engineering Geology* **62**, 7–17.

Bazzurro, P., and Cornell, C. A., 2002. Vector-valued probabilistic seismic hazard analysis (VP-SHA), in *Proceedings of the 7th U.S. National Conference on Earthquake Engineering*, Boston, MA.

Bazzurro, P., and Luco, N., 2007. Effects of different sources of uncertainty and correlation on earthquake-generated losses, *Australian Journal of Civil Engineering* **4**, 1–14.

Boore, D. M., Joyner, W. B., and Fumal, T. E., 1997. Equations for estimating horizontal response spectra and peak accelerations from western North American earthquakes: a summary of recent work, *Seismol. Res. Lett.* **68**, 128–153.

Campbell, K. W., 2003. Prediction of strong ground motion using the hybrid empirical method and its use in the development of ground-motion (attenuation) relations in eastern North America, *Bull. Seismol. Soc. Am.* **93**, 1012–1033.

Cao, T., Petersen, M. D., Cramer, C. H., Toppozada, T. R., Reichle, M. S., and Davis, J., 1999. The calculation of expected loss using probabilistic seismic hazard, *Bull. Seismol. Soc. Am.* **89**, 867–876.

Cornell, A., 1967. Engineering seismic risk analysis, *Bull. Seismol. Soc. Am.* **58**, 1583–1606.

Crowley, H., and Bommer, J., 2006. Modeling seismic hazard in earthquake loss models with spatially averaged exposure, *Bull. Earthquake Eng.* **4**, 249–273.

Evans, J. R., Baker, and J. W., 2006. Spatial covariance of ground motion in NGA data, American Geophysical Union (AGU) Fall Meeting, San Francisco, California, S12B–01.

Evans, M., Hastings, N., and Peacock, B., 2000. *Statistical Distributions, 3rd Edition*, John Wiley & Sons, Inc., New York.

Federal Emergency Management Agency (FEMA), 2001. *HAZUS99 Estimated annualized earthquake loss for the United States (FEMA 366)*, Washington, D. C.

——, 2003. *HAZUS-MH Technical Manual*, Washington, D. C.

Frankel, A., Mueller, C., Barnhard, T., Perkins, D., Leyendecker, E. V., Dickman, N., Hansen, S., and Hopper, M., 1996. *National seismic-hazard maps: documentation*, *U.S. Geol. Surv. Open-File Rept. 96–352*.

Freund, J. E., 1962. *Mathematical Statistics*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Hull, J. C., 1997. *Options, Futures, and Other Derivatives*, Prentice-Hall, Upper Saddle River, New Jersey, 720 pp.

Lee, Y., Anderson, J. G., and Zeng, Y., 2000. Evaluation of empirical ground-motion relations in southern California, *Bull. Seismol. Soc. Am.* **90**, S136–S148.

Nielson, B. G., 2005. Analytical fragility curves for bridges in moderate seismic zones, Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA.

Park, J., Bazzurro, P., and Baker, J. W., 2007. Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation, in *Proceedings of the 10th International Conference on Application of Statistics and Probability in Civil Engineering*, Tokyo, Japan.

Porter, K. A., Beck, J. L., Seligson, H. A., Scawthorn, C. R., Tobin, L. T., Young, R., and Boyd,

T., 2002. *Improving Loss Estimation for Woodframe Buildings*, CUREE Publication No. W-18, Richmond, CA.

Porter, K. W., 2007. Personal communication.

Pratt, J. W., Raiffa, H., and Schlaifer, R., 1995. *Introduction to Statistical Decision Theory*, MIT Press, Cambridge, MA.

Rossetto, T., and Elnashai, A., 2003. Derivation of vulnerability functions for European-type RC structures based on observational data, *Eng. Struct.* **25**, 1241–1263.

Schervish, M. J., 1995. *Theory of Statistics*, Springer-Verlag, New York.

Smyth, A. W., Altay, G., Deodatis, G., Erdik, M., Framco, G., Gülkan, P., Kunreuther, H., Lus, H. Mete, E., Seeber, N., and Yüzügüllü, O., 2004. Probabilistic benefit-cost analysis for earthquake damage mitigation: Evaluating measures fro apartment houses in Turkey, *Earthquake Spectra* **20**, 171–203.

Wang, S. S., 1998. *Aggregation of correlated risk portfolios: models and algorithms*, Research paper for the Casualty Actuarial Society.

Wesson, R. L., and Perkins, D. M., 2001. Spatial correlation of probabilistic ground motion and loss, *Bull. Seismol. Soc. Am.* **91**, 1498–1515.

Wesson, R. L., Perkins, D. M., Leyendecker, E. V., Roth, R. J., Jr., and Petersen, M. D., 2004. Losses to single-family housing from ground motions in the 1994 Northridge, California, earthquake, *Earthquake Spectra* **20**, 1021–1045.

Wills, C. J., Petersen, M. D., Bryant, W. A., Reichle, M. S., Saucedo, G. J., Tan, S. S., Taylor, G. C., and Treiman, J. A., 2000. A site-conditions map for California based on geology and shear wave velocity, *Bull. Seismol. Soc. Am.* **90**, S187–S208.

(Received 9 November 2004; accepted 22 January 2009)