Bulletin of the Seismological Society of America, 91, 6, pp. 1498–1515, December 2001

Spatial Correlation of Probabilistic Earthquake Ground Motion and Loss

by Robert L. Wesson and David M. Perkins

Abstract Spatial correlation of annual earthquake ground motions and losses can be used to estimate the variance of annual losses to a portfolio of properties exposed to earthquakes. A direct method is described for the calculation of the spatial correlation of earthquake ground motions and losses. Calculations for the direct method can be carried out using either numerical quadrature or a discrete, matrix-based approach. Numerical results for this method are compared with those calculated from a simple Monte Carlo simulation. Spatial correlation of ground motion and loss is induced by the systematic attenuation of ground motion with distance from the source, by common site conditions, and by the finite length of fault ruptures. Spatial correlation is also strongly dependent on the partitioning of the variability, given an event, into interevent and intraevent components. Intraevent variability reduces the spatial correlation of losses. Interevent variability increases spatial correlation of losses. The higher the spatial correlation, the larger the variance in losses to a portfolio, and the more likely extreme values become. This result underscores the importance of accurately determining the relative magnitudes of intraevent and interevent variability in ground-motion studies, because of the strong impact in estimating earthquake losses to a portfolio. The direct method offers an alternative to simulation for calculating the variance of losses to a portfolio, which may reduce the amount of calculation required.

Introduction

Why is spatial correlation important? Spatial correlation is not required to understand the mean loss to a portfolio of assets, but it is required to understand the distribution of losses about the mean, specifically to determine the variance (or standard deviation) of the losses. In finance, the standard deviation of the expected return for a portfolio of assets is commonly used as a measure of risk (cf. Bernstein, 1992; Ross *et al.,* 1996). The smaller the standard deviation, the smaller the variation about the mean, and thus, the lower the risk. This is the basis of the perennial advice about the need to diversify a portfolio, or in the vernacular, not to put all your eggs in one basket. A party exposed to earthquake risk from a portfolio of assets (e.g., an insurance company) wants to know not only the mean loss that it might suffer in a particular period, but also something about the distribution of losses. This is important so that it can manage its access to resources to cover those losses (e.g., reserves, reinsurance, catastrophe bonds, etc.). Further, a well-diversified portfolio would not be entirely composed of assets exposed to earthquake hazard in one region prone to earthquakes, say southern California, because the risk to those assets is highly correlated, that is, a loss to one asset implies likely loss to a nearby asset. Rather the portfolio should include assets from a number of regions where the risks are uncorrelated (Dong, 2000; Zadeh, 2000). The study of spatial correlation of probabilistic earthquake loss provides a tool for addressing these issues. We will explicitly address the use of spatial correlation in determining the variance of losses to a portfolio at the conclusion of this article.

A curve indicating the frequency of exceedance as a function of ground motion, termed a *hazard curve,* has become the standard measure of earthquake hazard at a site. Using site hazard curves and appropriate fragility curves, it is relatively easy to calculate a *mean* loss for a region or a portfolio of assets. For example, Cao *et al.* (1999) showed how to estimate the mean loss for a grid of sites covering a region. It should be emphasized that this approach allows only the calculation of the *mean* loss. The probability distribution function for the loss to a portfolio (e.g., fractiles and standard deviation) depends upon the correlation of losses between sites. Commonly the probability distribution of losses is calculated by simulation of possible future seismic histories (cf., Zadeh, 2000). This simulation is often conducted by Monte Carlo earthquake and magnitude draws from the same kind of source and rate models as from those that would be used in a hazard analysis. In the following, we present a direct (nonsimulation) method for the estimation of the correlation between annual probabilistic earthquake ground motion and loss at pairs of sites within a region. By the direct method, we mean that just as we use hazard analysis inputs to systematically consider each possible earthquake once at its expected annual rate of occurrence to directly build site hazard curves, in an analogous manner we build a joint hazard *surface.* Just as the usual hazard *curve* consists of exceedance frequency as a function of ground motion, our joint hazard surface consists of a matrix of joint exceedance frequencies of ground motions at two sites.

The word *hazard* is used technically for the annual frequency of exceedance, or the annual probability of exceedance of a specified ground motion. Often, especially with regard to hazard maps, *hazard* is used to describe the ground-motion level that has a specified frequency, or probability[1], of exceedance. We maintain the former usage throughout this article. To avoid confusion we use the term *probabilistic ground motion* or simply *ground motion* to refer to ground motion with a specified frequency of exceedance, or probability. When we discuss ground motions from individual events we try to be explicit. Similarly, by *annual loss* or *loss,* we mean fractional loss with a corresponding annual frequency of exceedance or probability, where the fraction is the proportion of the lost value relative to the assumed total value of the structure.

Intuitively, we would expect the spatial correlation of earthquake ground motion to depend on many factors. Among these factors are the joint proximity of sites to large faults, the frequency-magnitude distribution of earthquakes on those faults, the spatial extent of ruptures during earthquakes, the attenuation of strong ground motion with distance from the ruptures, the variance of strong motions about the mean attenuation curve (and how that variance is distributed among interevent and intraevent variability), and site conditions. Obviously, the spatial correlation of loss also depends on the building type and quality and vulnerabilities at the respective sites. The direct method for calculating spatial correlation presented here provides a useful tool for investigating the respective roles of these factors.

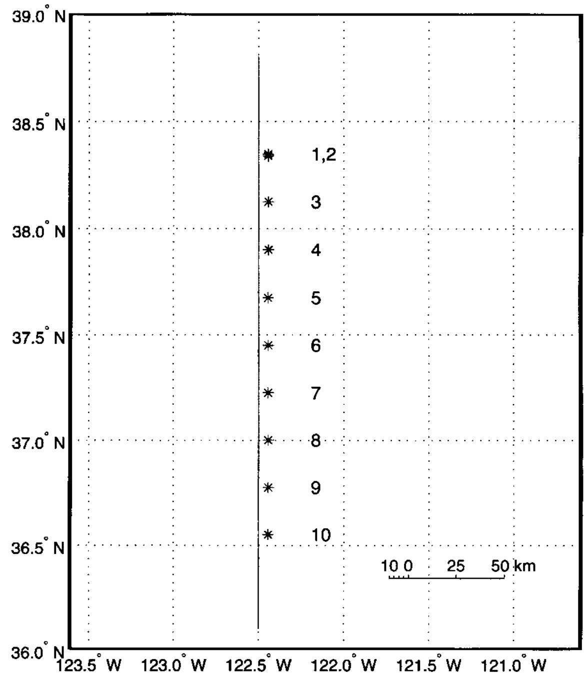

Figure 1. Map of fault and sites for calculation of ground motion and loss. The sites are all located 5 km off the fault. Sites 1 and 2 are located 1 km apart; sites 3–10 are located at 25-km intervals from site 1.

What Do We Mean by the Spatial Correlation of Earthquake Ground Motion and Loss?

To frame our discussion of the spatial correlation of earthquake motion and loss we consider a simple simulation. Again, we emphasize that we are not discussing the spatial correlation of ground motion and loss given one particular event, but the spatial correlation over all possible events. The details of the simulation will be given in a later section when we compare the results of our direct method with those of the simulation. Here we use the simulation only to develop an intuitive understanding of the central ideas of the article. Imagine 10 sites distributed along an active fault, as shown in Figure 1. The fault is 300 km long and is subject to magnitude 7.5 earthquakes with rupture lengths of 100 km that might occur anywhere along its length. Sites 1 and 2 are located 1 km apart, and the remaining eight sites are located at intervals of 25 km. When an earthquake occurs, the sites nearest the rupture will be strongly shaken, and those at greater distances will be less strongly shaken. Just the nature of the gradual decay of ground motion with distance induces a correlation of ground motion between sites that are close. Studies of strong ground motion (e.g., Boore *et al.,* 1997; Abrahamson and Silva, 1997) indicate that, in addition to the general decrease of the amplitude with distance, there is a strong random component to the observed amplitudes. (At least some of this randomness may have physical explanation, either currently understood or yet to be understood. Examples might include site effects or site effects incorrectly accounted for, focusing by geologic or crustal structure, basin edge effects, and directivity.) For any one earthquake, the observed amplitudes do not follow a smooth attenuation curve with distance but display random departures from that curve. This variability is termed *intraevent* variability. This randomness decreases the attenuation-induced correlation between sites. In addition, some earthquakes on average have higher levels of shaking than other earthquakes at comparable distances. This type of variability is termed interevent variability. Intuition may require education here—it seems possible that this interevent variability may restore some of the induced correlation lost to random intraevent variability. We will investigate this later in the article. The

[1]Assuming that the occurrence of earthquakes can be characterized as a Poisson process, the probability can be determined easily from the mean annual frequency (cf. Algermissen and Perkins, 1976).

total variability in earthquake ground motion combines these two components.

Imagine that we simulate a future history of earthquakes on this model fault over a time interval hundreds of times greater than the recurrence interval for the characteristic earthquake. Of course, we don't necessarily believe that the processes giving rise to seismicity will be stationary over hundreds of thousands or millions of years, but in order to fully sample the distributions used to characterize the seismicity, the ground motions, and the losses, many, many realizations are required. Imagine, too, that we observe ground motions at each of our sites during this simulated history, using an attenuation function and the two kinds of variability. In this simple model, in most years we observe no ground motion. In some years we record modest ground motions from earthquakes at some distance along the fault. In some few years we record very strong ground motions from earthquakes nearby.

We store all these simulated ground motions, giving us 10 sequences (one for each of the sites) with the maximum ground motion for each year at each of the sites over many thousands of years. From these 10 sequences we can calculate many standard statistics, for example, the mean and standard deviation of the maximum annual ground motion at each site. A frequency curve can be constructed to show the annual frequency of ground motions greater than u as a function of u at each site. This kind of plot is a hazard curve, which will be discussed in more detail below.

In addition, for pairs of sequences of maximum annual ground motions, we calculate the sample correlation coefficient. (By definition, the sample correlation coefficient is equal to the sample covariance of the sequences divided by the product of the two sample standard deviations.) Figure 2a shows the correlation coefficients of the maximum annual ground motion at pairs of sites. This is what we mean by the correlation of earthquake ground motion. (Again, details will be given later). Of course the correlation coefficient of site 1 with itself is 1.0, but notice that, in line with intuition, the correlation coefficients decrease as the separation distance between the sites increases.

Similarly, we believe that damage and loss are also random processes. Imagine that a similar structure is located at each of the 10 sites. Suppose further that we have a function (a fragility curve or fragility matrix) that indicates the probability of various levels of fractional loss to that structure given a particular level of ground motion. Then during our simulation, for each earthquake at each site, we take the simulated ground motion, select the fragility curve appropriate for that ground motion, and make a random draw from the appropriate probability distribution to estimate the loss. We save these losses in a sequence just as we do for ground motion, making the same assumption that each of these losses is the maximum loss in the year in which it occurs. We can calculate the same statistics on the loss sequences as for the ground motions. In particular we can calculate the correlation coefficients among pairs of the loss sequences as shown in Figure 2b. This is what we mean by the correlation of earthquake loss. Again the correlation coefficient of the losses at site 1 with itself is 1.0, but the other sites follow the sample general shape as the correlation coefficients for ground motion. The difference is that the correlation coefficients for loss are systematically lower than those for ground motion. This reflects the increased degree of randomness induced by the probabilistic fragility matrix, thereby systematically reducing the correlation. (Note that we have

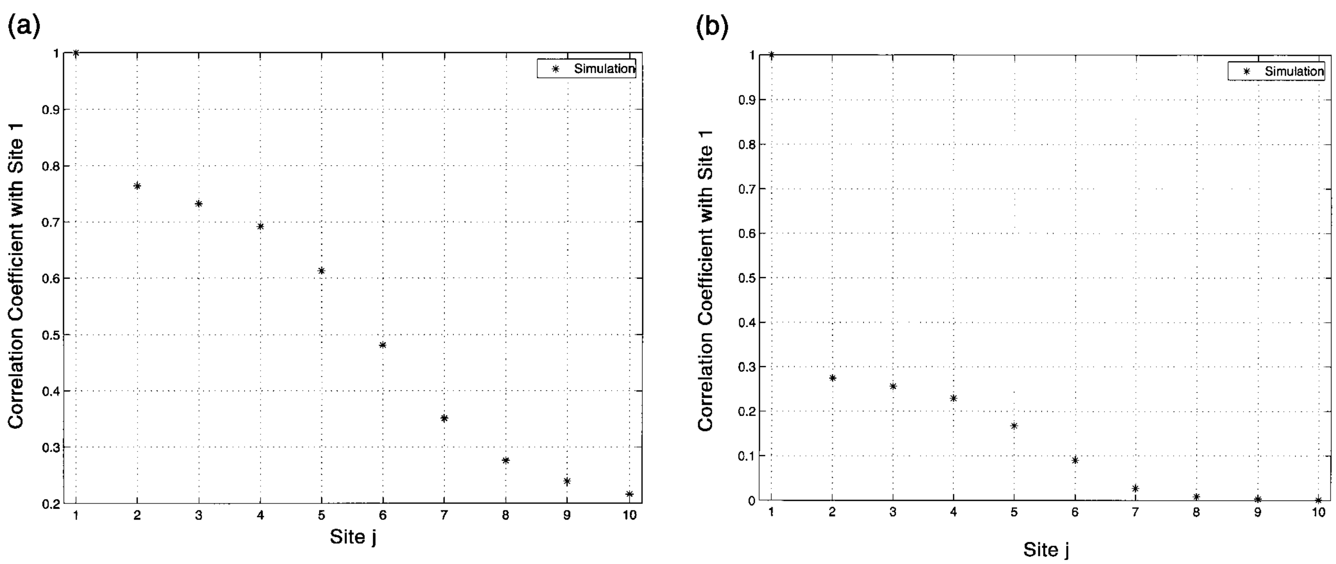

Figure 2. Correlation coefficients between pairs of sites (site 1 and site j) in simple simulation for (a) ground motion and (b) loss. The correlation coefficient for ground motion between sites 1 and 2 differs because of intraevent variability.

ignored other possible sources of correlation, such as site conditions and construction quality.)

In essence, then, we can represent earthquake ground motion and loss at our sites by these long sequences and statistics derived from them. It is more common, however, to view earthquake ground motion and loss from the point of view of an annual frequency of exceedance of ground motion and loss, or as the probability distribution for maximum annual ground motion and maximum annual loss. The frequencies (and probability distributions) can be estimated by binning the ground motions and losses in the long sequences. Alternatively, as in probabilistic hazard analyses, they can be calculated directly from the statistical properties of earthquakes and their ground motions. We now shift to this point of view and show that correlation coefficients for ground motion and loss, as well as other statistics, can be calculated from this direct perspective as well.

Correlation of Ground Motion

Following the method pioneered by Cornell (1968), earthquake hazard is commonly defined as the frequency of exceedance, or probability under a suitable model, of a particular level of intensity or ground motion (see also Reiter, 1990; Frankel *et al.,* 1996). To set the stage for deriving expressions for ground-motion correlation, we briefly restate the basic expressions of probabilistic seismic hazard analysis, considering two cases: point and fault sources of earthquakes.

Assume that the annual cumulative frequency–magnitude relation for earthquakes at the source is $N_{ann} = N_{ann}(M)$. Then the annual number of earthquakes in the magnitude interval (bin) $2\Delta M$ centered at M_l is $N_l = N_{\text{ann}}(M - \Delta M) - N_{\text{ann}}(M + \Delta M)$. The frequency—magnitude relation could be of the Gutenberg–Richter type, $\log_{10}(N_{\text{ann}}) = a_{\text{cum}} - bM$, of a characteristic earthquake type, or some other.

Frequency of Exceedance

Modern hazard maps (i.e., Frankel *et al.,* 1996) are calculated considering both point and faults sources for earthquakes; thus we consider these two cases. For point sources of earthquakes, the annual rate $\lambda_i(u_i > u_0)$ of the ground motion u_i exceeding ground motion u_0 at the ith site is determined from a sum over distance and magnitude:

$$\lambda_i(u_i > u_0) = \sum_k \sum_l N_{kl} P(u_i > u_0 \mid D_{ik}, M_l), \tag{1}$$

where the summations are over all point sources of earthquakes, k, and over all magnitude bins of interest, l, and N_{kl} now refers to magnitude bin l, for source, k. $P(u_i > u_0 \mid D_{ik}, M_l)$ is the probability that u_i at the site will exceed u_0, for an earthquake at distance D_{ik} with magnitude M_l. The expression $\lambda_i(u_i > u_0)$ is commonly referred to as the hazard curve.

Attenuation of strong motion with distance from an earthquake, given the occurrence of an event, is typically assumed to be of a form

$$\log(u) = f(D,M) + \varepsilon,$$

where u is the ground motion, $f(D,M)$ is the mean logarithm of the ground motion as a function of distance and magnitude, and ε is a random term which is approximately normally distributed with zero mean and standard deviation σ. Then the probability $P(u_i > u_0 \mid D_{ik}, M_l)$ is the area under the normal distribution between $\log(u_0)$ and infinity. Details of this calculation are discussed in Appendix A.

For a fault of length L, the hazard curve is usually calculated by considering a floating rupture zone on the fault (cf., Frankel *et al.,* 1996). That is, for each magnitude M_l, determine a corresponding rupture length d_l (for example, using the relations of Wells and Coppersmith [1994]), then sum the contributions of ruptures, sequentially offset δL along the fault. For each magnitude there will be $n_l = 1 + (L - d_l)/\delta L$ floating ruptures along the fault. Designating the closest distance from the rth rupture for magnitude M_l to the site i as D_{ilr}, then the exceedance frequency at the site is

$$\lambda_i(u_i > u_0) = \sum_l \sum_r \frac{1}{n_l} N_l P(u_i > u_0 \mid D_{ilr}, M_l). \tag{2}$$

For multiple fault sources, an additional summation can be carried out over the faults.

Having restated the exceedance frequency for a single site, we now extend the definition for two sites. Define the joint exceedance frequency at two sites i and j as the annual frequency that the ground motion u_i at site i exceeds u_{0_i}, and the ground motion u_j at site j exceeds u_{0_j}. (u_{0_i} and u_{0_j} do not have to be the same.) Then by reasoning analogous to that for the hazard curves (1) and (2) above, the *joint* exceedance frequency for a point source is

$$\begin{aligned} &\nu_{ij}(u_i > u_{0_i}, u_j > u_{0_j}) = \\ &\sum_k \sum_l N_{kl} P(u_i > u_{0_i} \mid D_{ik}, M_l, u_j > u_{0_j} \mid D_{jk}, M_l) \end{aligned} \tag{3}$$

and for a single fault is

$$\begin{aligned} &\nu_{ij}(u_i > u_{0_i}, u_j > u_{0_j}) = \\ &\sum_l \sum_r \frac{N_l}{n_l} P(u_i > u_{0_i} \mid D_{ilr}, M_l, u_j > u_{0_j} \mid D_{jlr}, M_l) \end{aligned} \tag{4}$$

where $P(u_i > u_{0_i} \mid D_i, M, u_j > u_{0_j} \mid D_j, M)$ is the joint probability. As mentioned above, the exceedance frequency at a single site is expressed as a hazard *curve* with the independent variable being the threshold ground motion. In contrast, the joint exceedance frequency at two sites is expressed as a hazard *surface* with the independent variables being the threshold ground motions at the two sites.

Following Abrahamson and Silva (1997), write the ground motion at the ith site from the qth earthquake as

$$\log_e(u_{iq}) = f(D_{iq}, M_q) + \varepsilon_{iq} + \eta_q$$

where u_{iq} is the ground motion at the ith site from the qth earthquake, D_{iq} is the distance from the ith site to the qth earthquake, and M_q is the magnitude of the qth earthquake. In addition, there are two random terms in this model: ε_{iq}, the intraevent term, which is assumed to be normally distributed with zero mean and standard deviation σ_ε; and η_q, the interevent term, which is assumed to be normally distributed with zero mean and standard deviation σ_η. Assuming independence of the two random terms[2], the total standard deviation for the model, σ, is the square root of the sum of the squares of the standard deviations of the components,

$$\sigma = \sqrt{\sigma_\varepsilon^2 + \sigma_\eta^2}.$$

Let the random parts of the two ground motions be represented by the variables:

$$X_{iq} = \log_e(u_{iq}) - f(D_{iq}, M_q) = \varepsilon_{iq} + \eta_q,$$
$$X_{jq} = \log_e(u_{jq}) - f(D_{jq}, M_q) = \varepsilon_{jq} + \eta_q.$$

Then it can be shown (see Appendix B) that the joint probability density function for X_{iq} and X_{jq} follows a bivariate normal distribution with zero means, standard deviations σ, and correlation coefficient,

$$\rho = \frac{\sigma_\eta^2}{\sigma^2} = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2 + \sigma_\eta^2}.$$

Thus the joint probability in (3) and (4) can be calculated from the cumulative bivariate normal distribution. This can be accomplished using an algorithm based on that by Hull (1997) and Odegaard (1999). (Note that for the case $\rho = 1$, the cumulative distribution is well behaved, even though the probability density function is not.)

Also note that for no interevent variability (i.e., $\sigma_\eta = 0$), $\rho = 0$, and the bivariate normal distribution reduces to the product of two univariate normal distributions. Thus the joint probability in (3) and (4) would be just the product of the probabilities of the ground motions exceeding the threshold ground motions at the individual sites. In contrast, consider the case of no intraevent variability ($\sigma_\varepsilon = 0$), and $\rho = 1$. In this case, $X_{iq} = X_{jq}$, that is, the actual ground motion for a given magnitude and distance for a given earthquake would be identical at all sites.

[2]The model of Abrahamson and Silva (1997) does not address any intraevent site correlation that might result from directivity, or other source or site effects. (Obviously, it does address the intraevent site correlation that results from the distances from the source to the sites.) Models including additional explanatory variables to include the effects of directivity (Somerville *et al.,* 1997; Abrahamson, 2000) are beginning to emerge. For such models the effect of including the additional variables is to change the mean ground motion from $f(D_{iq}, M_q)$ to a new function of all the considered variables. Similarly, if corrections for site conditions are available, they can be included in the mean ground motion. The assumption that is important for the following is that the residuals remaining after the subtraction of the mean ground motion (taking into account all the considered variables) can be considered as the sum of two independent random terms, an intraevent term and an interevent term.

Probability Distributions for Ground Motion

The cumulative probability of exceedance and the probability density functions for ground motion are easily derived from the frequency of exceedance or hazard curves assuming a Poisson occurrence distribution (Algermissen and Perkins, 1976). Since the hazard curve, $\lambda(u > u_0)$ represents the mean annual frequency of exceedance, and assuming that the distribution of events giving rise to the exceedance approximates a Poisson distribution, then the probability of zero exceedances can be obtained from the Poisson distribution. The cumulative probability distribution for the ground motion, $F(u)$, that is, the probability that the ground-motion values will **not** exceed the specified values in T years, may be written as

$$F(u) = P(u \le u_0) = e^{-T\lambda(u > u_0)}.$$

We will be concerned with the annual probabilities in the following so we will take $T = 1$. The corresponding annual probability density function for the ground-motion parameter u is

$$f(u) = F'(u) = -\frac{d\lambda(u > u_0)}{du} e^{-\lambda(u > u_0)}.$$

Although the probability density function itself is not required for the discrete integration approach described subsequently, Appendix C gives an approach to its efficient calculation if it is required for another purpose.

Derivations of the joint distribution and the joint probability density function for maximum annual ground motion follow the same general argument but require the bivariate Poisson distribution function (Johnson *et al.,* 1997). Consider the joint distribution of the arbitrary variables $X_1 = Y_1 + Y_{12}$ and $X_2 = Y_2 + Y_{12}$ where Y_1, Y_2, and Y_{12} are mutually independent Poisson random variables with means θ_1, θ_2, and θ_{12} respectively. Then the joint probability for zero events is (Johnson *et al.,* 1997)

$$P(X_1 = 0, X_2 = 0) = e^{-(\theta_1 + \theta_2 + \theta_{12})}$$

Having stated this general result, let us associate each variable with site exceedances: Let the annual number of exceedances of u_{0_i} at site i be X_1. Similarly, let the annual number of exceedances of u_{0j} at site j be X_2. Let each of these be divided into joint and independent parts so that the annual number of exceedances of u_{0_i} at site i only is Y_1, the annual number of exceedances at site j only is Y_2, and the annual number of exceedances at both sites i and site j is Y_{12}. Then the exceedances at the two sites alone are

$$Y_1 = X_1 - Y_{12}$$
$$Y_2 = X_2 - Y_{12}.$$

Note Y_1, Y_2, and Y_{12} are constructed so that they are mutually independent random variables. The mean exceedance frequencies of X_1 and X_2 are the corresponding hazard curves, denoted by λ_i and λ_j respectively. The mean exceedance frequency of Y_{12} is the joint exceedance frequency, denoted by v_{ij}. Then the mean exceedance frequencies of the Y_1, Y_2, and Y_{12} are

$$\theta_1 = \lambda_i - v_{ij}$$
$$\theta_2 = \lambda_j - v_{ij}$$
$$\theta_{12} = v_{ij}.$$

Thus

$$F(u_i,u_j) = P(u_i < u_{0_i}, u_j < u_{0_j}) = e^{-(\lambda_i+\lambda_j-v_{ij})}.$$

Note that as the positions of the two sites become closer and closer, then $\theta_1,\theta_2 \to 0$, and $\lambda_i \to \lambda_j \to v_{ij}$

The annual joint probability density function is

$$f(u_i,u_j) = \frac{\partial^2 F(u_i,u_j)}{\partial u_i \partial u_j} =$$
$$\left(\frac{\partial\lambda_i}{\partial u_i}\frac{\partial\lambda_j}{\partial u_j} - \frac{\partial\lambda_i}{\partial u_i}\frac{\partial v_{ij}}{\partial u_j} - \frac{\partial\lambda_j}{\partial u_j}\frac{\partial v_{ij}}{\partial u_i} + \frac{\partial v_{ij}}{\partial u_i}\frac{\partial v_{ij}}{\partial u_j} + \frac{\partial^2 v_{ij}}{\partial u_i \partial u_j}\right) e^{-(\lambda_i+\lambda_j-v_{ij})}.$$

The covariance of the ground-motion parameters can be calculated as

$$\mathrm{cov}(u_i,u_j) = \int\int (u_i - \bar{u}_i)(u_j - \bar{u}_j) f(u_i,u_j) du_i du_j$$

where $\bar{u}_i$ and $\bar{u}_j$ are the mean annual ground motions at sites i and j.

The correlation coefficient ρ^u_{ij} can be obtained from the covariance through

$$\rho^u_{ij} = \frac{\mathrm{cov}(u_i,u_j)}{\sigma^u_i \sigma^u_j},$$

where σ^u_i and σ^u_j are the standard deviations of the annual ground motions at sites i and j.

Correlation of Losses

Estimating the fractional loss of an asset as a function of the ground motion to which the asset is exposed is an important and active area of investigation. Estimates of fractional losses for different building types exposed to different ground-motion values have been estimated for example by ATC-13 (Applied Technology Council, 1985). Wesson *et al.* (1999) recently showed that the probability of fractional loss follows a gamma distribution, at least for single-family homes in the vicinity of the 1994 Northridge, California, earthquake. The parameters in the gamma distribution can be estimated from the ground motion. We can consider these estimates of fragility to be the conditional probabilities of loss given a ground-motion value, u, that is $f_l(l|u)$. Then the resulting probability density function for the annual loss to the asset will be

$$f_l(l) = \int f_l(l \mid u) f(u) du.$$

Similarly we can write the joint probability density for the annual losses as

$$f_l(l_i,l_j) = \int\int f_l(l_i,l_j \mid u_i,u_j) f(u_i,u_j) du_i du_j.$$

Except for special cases such as extended structures, the loss at the ith site will not depend on the ground motion at the jth site, and vice versa, except through the correlation of ground motion, so we can consider the two conditional probabilities independent and write

$$f_l(l_i,l_j \mid (u_i,u_j)) = f_l(l_i \mid u_i) f_l(l_j \mid u_j).$$

Put another way, the loss at one site depends *only* on the ground motion at that site, not on the ground motion at the other site, except in so far as the ground motions are not independent. Then

$$f_l(l_i,l_j) = \int\int f_l(l_i \mid u_i) f_l(l_j \mid u_j) f(u_i,u_j) du_i du_j,$$

which can be used to obtain the covariance of the losses,

$$\mathrm{cov}(l_i,l_j) = \int\int (l_i - \bar{l}_i)(l_j - \bar{l}_j) f_l(l_i,l_j) dl_i dl_j.$$

where $\bar{l}_i$ and $\bar{l}_j$ are the mean annual losses at sites i and j.

The correlation coefficients are

$$\rho_{ij} = \frac{\mathrm{cov}(l_i,l_j)}{\sigma^l_i \sigma^l_j},$$

where σ^l_i and σ^l_j are the annual standard deviations of the losses at sites i and j.

As will be described below, these correlation coefficients can be used to calculate the variance of the loss to a portfolio of properties. Given the mean and variance of the loss to a portfolio and an assumption about the appropriate probability distribution (one that can be sufficiently described by mean and standard deviation), then the probability of a specific loss to the portfolio can be estimated.

A Discrete Approach

Although all the integrals in the section above may be performed numerically using quadrature, it is considerably simpler to follow a discrete approach to the calculations. Let

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix}$$

be a vector of the ground-motion values of interest, that is, these are the ground-motion values for which exceedance frequencies are desired for a hazard curve. The values do not need to be evenly spaced, and will commonly be at logarithmic intervals. Then the vector

$$\mathbf{F} = e^{-\lambda(\mathbf{u})},$$

will contain as elements the cumulative probability, that is, the probability that the ground motion will be less than the value of the corresponding element of **u**. Define

$$\mathbf{f} = \begin{bmatrix} F_1 \\ F_2 - F_1 \\ \vdots \\ F_m - F_{m-1} \end{bmatrix}.$$

Assuming that **F** is calculated over a sufficient range of u such that $F_m \cong 1$, then the elements of **f** will sum to one, and **f** will behave as a discrete analog to the probability density function. The benefit of this approximation is that the mean and variance of **f** can be calculated simply through matrix multiplication rather than the more tedious numerical integration through quadrature:

$$\mu = \mathbf{u}^T \mathbf{f} \text{ and}$$
$$\sigma^2 = (\mathbf{u}^T - \mu \mathbf{1}^T)^2 \mathbf{f},$$

where the exponent notation indicates term-by-term exponentiation and $\mathbf{1} = (1,1, \ldots ,1)^{\mathrm{T}}$.

Similarly let $\mathbf{F}_{ij}$ be the matrix defined by

$$\mathbf{F}_{ij} = e^{-(\lambda_i(\mathbf{u}) + \lambda_j(\mathbf{u}) - \nu_{ij}(\mathbf{u}))}.$$

Further let $\mathbf{f}_{ij}$ be the matrix defined by the application of the difference operator to $\mathbf{F}_{ij}$ that corresponds to mixed partial differentiation (Abramowitz and Stegun, 1964), the components of which will be

$$f_{i,j} = F_{i,j} - F_{i,j-1} - F_{i-1,j} + F_{i-1,j-1}.$$

Again for an appropriately large range in **u,** the elements of $\mathbf{f}_{ij}$ will sum to one, and $\mathbf{f}_{ij}$ is the discrete analog of the joint probability density function. In particular,

$$\mathrm{cov}(\mathbf{f}_{ij}) = (\mathbf{u} - \mu_i \mathbf{1})^T \mathbf{f}_{ij} (\mathbf{u} - \mu_j \mathbf{1}).$$

Similarly let

$$l = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_n \end{bmatrix}$$

be a vector of loss levels, and let

$$\mathbf{f}^l = \begin{bmatrix} \Pr(0 \le l \le l_1) \\ \Pr(l_1 < l \le l_2) \\ \Pr(l_2 < l \le l_3) \\ \vdots \\ \Pr(l_{n-1} < l \le l_n) \end{bmatrix},$$

then we can define a fragility matrix **G** relating the probability of loss in each loss increment to the probability of ground motion in each ground-motion increment

$$\mathbf{G} = \begin{bmatrix} \Pr(0 \le l \le l_1 \mid 0 \le u \le u_1) & \Pr(0 \le l \le l_1 \mid u_1 < u \le u_2) & \ldots & \Pr(0 \le l \le l_1 \mid u_{m-1} < u \le u_m) \\ \Pr(l_1 < l \le l_2 \mid 0 \le u \le u_1) & \Pr(l_1 < l \le l_2 \mid u_1 < u \le u_2) & \ldots & \Pr(l_1 < l \le l_2 \mid u_{m-1} < u \le u_m) \\ \vdots & \vdots & \ddots & \vdots \\ \Pr(l_{n-1} < l \le l_n \mid 0 \le u \le u_1) & \Pr(l_{n-1} < l \le l_n \mid u_1 < u \le u_n) & \ldots & \Pr(l_{n-1} < l \le l_n \mid u_{m-1} < u \le u_m) \end{bmatrix}$$

Then we can write

$$\mathbf{f}^l = \mathbf{G}\mathbf{f}$$

and

$$\mathbf{f}_{ij}^l = \mathbf{G}\mathbf{f}_{ij}\mathbf{G}^T.$$

The mean, standard deviation, and covariance are exactly analogous to those for ground motion

$$\mu^l = \mathbf{l}^T \mathbf{f}^l \text{ and}$$
$$\sigma^{l^2} = (\mathbf{l}^T - \mu^l)^2 \mathbf{f}^l,$$
$$\mathrm{cov}(\mathbf{f}_{ij}^l) = (\mathbf{l} - \mu_i^l \mathbf{1})^T \, \mathbf{f}_{ij} (\mathbf{l} - \mu_j^l \mathbf{1}).$$

One of the difficulties encountered in the straightforward numerical integration of the expressions in the preceding section is that both the probability density functions for ground motion and the conditional probability for loss given low levels of ground motion are very sharply peaked near zero. The discrete formulation presented here avoids any numerical problems by focusing on the differences in the cumulative distributions. In addition, the discrete formulation can be calculated much, much more rapidly.

A Simple Example

To illustrate this approach, let us return to our simple example of 10 sites arrayed along a fault as shown in Figure 1. The sites are all located 5 km off the fault. Sites 1 and 2 are located 1 km apart; sites 3–10 are located at 25-km intervals from site 1. Structures at each site are assumed to be characterized by the same fragility matrix. (By this construction, the sites will have similar, but not identical hazard and loss curves. The probabilistic ground motions and losses at the sites near the center of the fault are greater than those at the sites near the ends of the fault. The higher ground motions and losses near the center result from more frequent exposure to earthquakes at close distances.) For this example the frequency-magnitude relation is assumed to be an earthquake with a magnitude 7.5, a recurrence time of 300 yr, and a rupture length (from Wells and Coppersmith, 1994) of 100 km. The attenuation relation assumed is that of Boore *et al.* (1993, 1994) for peak ground acceleration (referred to in the following as BJF). The variabilities for BJF are given in Table 1.

Losses at each site were assumed to be characterized by the loss model developed by Wesson *et al.* (1999) to characterize fractional losses to single family homes in southern California (Table 2). This model was developed from insured losses from the 1994 Northridge earthquake. The model is specified as a gamma distribution in which the two parameters of the gamma distribution depend on the peak ground acceleration. In essence the model gives the conditional probability distribution for fractional loss given the ground motion. Small ground motions lead to gamma distributions sharply peaked near zero loss. Larger ground motions lead to broader distributions leading to larger probabilities of loss at higher loss levels (Fig. 3).

Hazard and loss were calculated at each site by two methods. First, parameters describing the hazard and loss were calculated by the direct method described in this article. Second, corresponding parameters were calculated from the simple Monte Carlo simulation described at the beginning of this article. The mean and standard deviations of the ground motion were also calculated by numerical integration. In the calculation using the direct method, hazard was calculated at logarithmic intervals of 0.01 and 0.005 from −2.5 to 0.7 in log ground motion. Fractional losses were calculated at logarithmic intervals of 0.01 and 0.0025 from −2.5 to 0.5. Calculations in MATLAB on a Sun Sparc Ultra 10 with a 300 MHz processor for all the parameters characterizing the ground motions and losses at the 10 points and their interrelationships requires about 1.5 hr for an interval of 0.01. Satisfactory agreement between the simulation and direct methods for correlation coefficients was obtained for the coarser spacing. However, satisfactory agreement for the mean and standard deviation of loss required the finer intervals.

In the simulation, 65,000 earthquakes were simulated corresponding to about 10^8 yr at the assumed recurrence interval. We make the assumption that each of these earthquakes corresponds to the maximum ground motion and loss experienced in one year. Two random draws were made to characterize each simulated earthquake. The first draw was from a uniform distribution to position the rupture along the fault. All ruptures were assumed to be 100 km in length. The second draw was made to a normal distribution to determine the contribution of interevent variability. At each site, two additional draws were made to determine the ground motion (that is, taking into account intra-event variability) and loss. First, a draw was made from a normal distribution to determine the contribution to the ground motion from intraevent variability. The draws for the interevent contribution and the intraevent contribution, together with the distance to the site from the randomly located rupture, determined the ground motion through BJF. A second draw for the site was made from a gamma distribution, determined by the ground motion, as in Figure 3, to determine the fractional loss at the

Table 1
Standard Deviations for the Boore-Joyner-Fumal (1993, 1994) Attenuation Model

Interevent	Intraevent	Total
0.08	0.211	0.226

Quantities are $\log_{10}$ (peak ground acceleration, g); To convert to $\log_e$, divide quantities above by $\log_e 10$.

Table 2
Parameters of Conditional Probability of Loss

The conditional probability of loss is that given by Wesson *et al.* (1999)

$$f(l \mid u) = \gamma(l, a, b),$$

where
l is loss fraction, u is ground motion in peak ground acceleration, and $\gamma(x,a,b)$ is the gamma distribution with shape parameter a and scale parameter b:

$$\gamma(x,a,b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-\frac{x}{b}}.$$

$\Gamma(x)$ is the gamma function,
$\log_{10}(a) = -8.5610 + 7.8202q - 1.7065q^2$, and
$\log_{10}(b) = -1.5365 + 1.0836q - 0.3277q^2$, where
$q = \log_{10}(u)$,
and u is given in percent g (e.g., from 0 to 100).

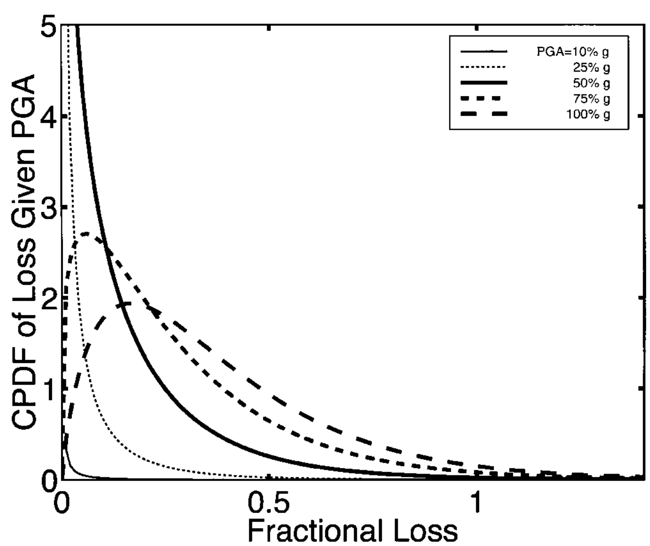

Figure 3. Conditional probability of loss (predicted fragility curves) given ground motions as used in example (Wesson *et al.,* 1999).

site. Ground motions and losses were stored for all sites for all simulated earthquakes. The assumption is made that the ground motion and loss for each of these earthquakes was the maximum for the year in which they occurred. Then standard statistics were calculated from the results, means, and standard deviations for each site and correlation coefficients among sites for both ground motion and loss.

The hazard curves for all sites are shown in Figure 4. The joint hazard frequencies for sites 1 and 2 and sites 1 and 10 are shown in Figure 5a and b. The difference between the joint hazard frequencies for these two pairs of sites is shown in Figure 5c. The marginal frequencies of the joint hazard curve are just the hazard curves. (Note, however, that Fig. 5 is plotted on a linear scale in contrast to Fig. 4). The mean and standard deviation of the ground motion and loss as calculated from the direct approach, and those calculated from the simulation and numerical integration, are all shown in Figure 6. As mentioned previously, the higher ground motions and losses near the center of the fault result from more frequent exposure to earthquakes at close distances. The standard deviation is increasing roughly in proportion to the mean largely because of the geometry and recurrence properties of the model. At the center of the fault the frequency of higher ground motions is higher, and the standard deviation reflects the broadening of the distribution, which has many zero values for years with no events. If the model were such that there were an event in every year, the relation between the mean and standard deviation would depend on the compactness of the distribution of ground motion alone.

The correlation coefficient between the ground motion at site 1 and each of the other sites is shown in Figure 7. Calculations for the direct method are shown for the BJF attenuation relationship assuming (1) interevent and intraevent variability as given by BJF, (2) total variability as given by BJF, but all assumed to be interevent variability, and (3) total variability as given by BJF, but all assumed to be intraevent variability. Agreement between the results from the direct calculation and those of the simulation are quite good.

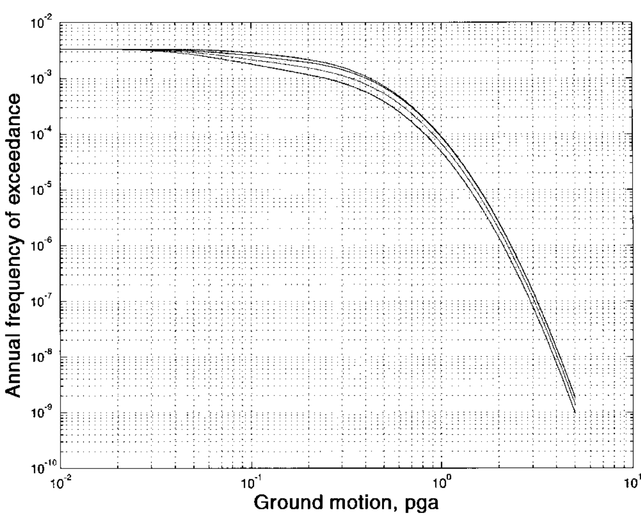

Figure 4. Hazard curves for all the sites determined by direct calculation. The curves showing higher hazard are those for sites near the center of the fault; those showing lower hazard are for sites near the ends of the fault.

Two interesting observations can be made from Figure 7. First, in line with intuition, the correlation of ground motion falls off with the distance separating the sites. The falloff is most rapid at about the separation distance corresponding to one earthquake rupture length (100 km, the distance between site 1 and site 6). Second, holding the total variability fixed, increasing the intraevent variability decreases the correlation coefficients. These observations underscore the importance of understanding how variability in ground motion is apportioned between interevent and intraevent variability. For example, the risk of a large loss to a portfolio is significantly greater if all the variability is interevent than if all the variability is intraevent.

Correlation coefficients between the loss at site 1 and each of the other sites are shown in Figure 8. Obviously because of the additional degree of randomness introduced by the fragility matrix based on the gamma distribution, the correlation coefficients (except obviously for site 1 with itself) are less than those for ground motion. Otherwise, the observations made previously for ground motion pertain to loss as well.

Factors Contributing to Spatial Correlation

The factors contributing to spatial correlation of ground motion and loss are explored in Figure 9. Partitioning of variability, the rate of ground-motion attenuation with distance, and the spatial extent of the source (e.g. the rupture dimension) all play key roles. Not examined here are construction type and quality and site condition, also key factors in the correlation of loss. Figure 9a shows the spatial correlation of ground motion for all variability interevent and no variability for three source configurations: the 100-km rupture length described previously, a point source rupture, and an infinite rupture. For each of these source configurations, no variability and all variability interevent give the same results for ground motion. This may appear surprising at first until one realizes that since the correlations are among sites within each event, the interevent variability plays no role. For no variability, the ground motion is simply a deterministic function of the magnitude and distance. For all interevent variability, the ground motions for each source position differ only by a multiplicative factor from event to event. Thus the spatial correlation of ground motion is the same for no variability and all interevent variability. For an infinite source with all our sites at a constant distance of 5 km and either no variability or all interevent variability, ground motions at all sites are identical, giving rise to a

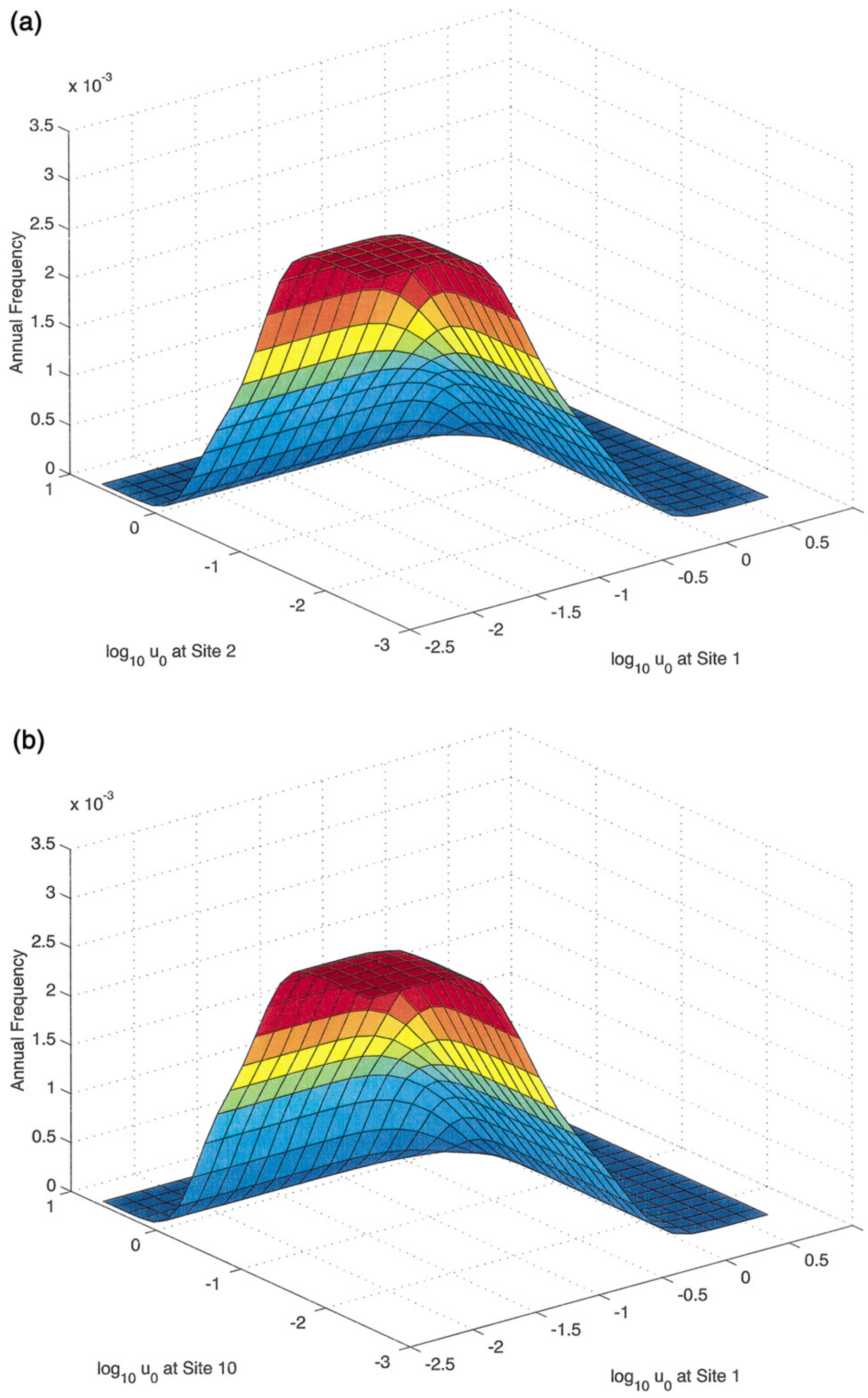

Figure 5. Joint frequency of exceedance for (a) sites 1 and 2, (b) sites 1 and 10, and (c), (*continued on next page*).

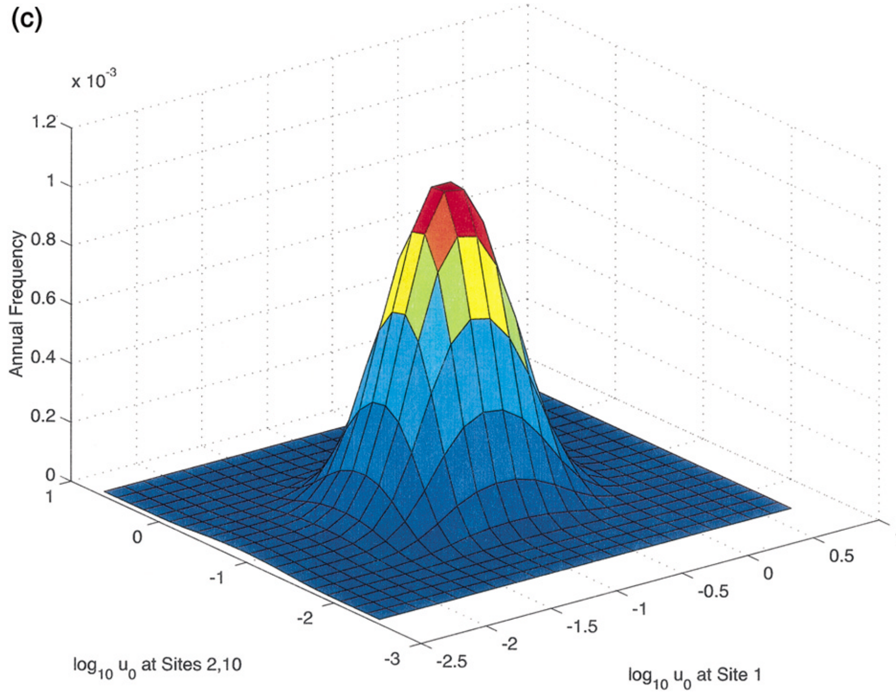

Figure 5. (*Continued*) (c) the difference between the joint exceedance frequencies for sites 1 and 2, and sites 1 and 10.

correlation coefficient of 1.0 at all sites. The spatial correlation for the 100-km rupture length is as described previously, falling off with distance with a maximum slope at an intersite distance of about a rupture length. However, for a point source of rupture the spatial correlation decays more rapidly with distance, reflecting only the attenuation of ground motion with distance from the site.

Practically speaking, in addition to the effects accounted for in the simple BJF model, ground-motion site correlations will decrease for sites that do not share the same geologic site conditions, and loss correlations will decrease for structures that do not share the same fragility function. Similarly, just as site ground-motion correlation is induced at nearby sites, merely because of the similarity in predicted ground motion from the attenuation function, so, too, attenuation functions expressing directivity effects (e.g., Somerville, 1997; Abrahamson, 2000) will affect the induced correlation. Ground-motion site correlation (and loss correlation) will increase when both sites experience increased ground motion owing to that directivity (or decrease when one site experiences the increase in ground motion and the other site does not).

For the spatial correlation of loss shown in Figures 9b and c, the source configurations play similar roles. However the correlation coefficients for all interevent variability (Fig. 9b) are significantly higher than for those for no variability (Fig. 9c). This apparently results from the nonlinear effects of the fragility curves. For events with high ground motions, the losses will be significantly greater, leading apparently to higher spatial correlation.

Conditional Probability

Another way to view the relationship between losses at pairs of sites is through conditional probability. For example, given the annual loss at one site, it might be useful to calculate the probability distribution for annual loss at another site. All the information required to make this calculation is contained within the discrete analogs to the probability density functions. The conditional probability density is defined as (e.g., Freund, 1962):

$$f_l(l_j \mid l_i) = \frac{f_l(l_i, l_j)}{f_l(l_i)}.$$

Using the discrete method outlined previously it is straightforward to calculate the discrete analog to this conditional density. Figure 10 shows the results of such a calculation for the conditional probabilities for loss at sites 2–9, given four levels of loss at site 1. Each figure part shows the probability that the losses at the ith site will be greater than a specified loss level, as a function of that loss level. (This probability is one minus the cumulative probability for the conditional loss.) Interestingly, the conditional probabilities for loss at sites 3 and 4 are larger than those for site 2. This is because sites 1 and 2 are toward the end of the fault and sites 3 and 4 are toward the middle. Sites toward the middle generally experience higher ground motion and loss than sites near the end of the fault as explained above. More to the point it is more likely that the event causing a loss at sites 1 and 2 is toward the middle of the fault thereby causing

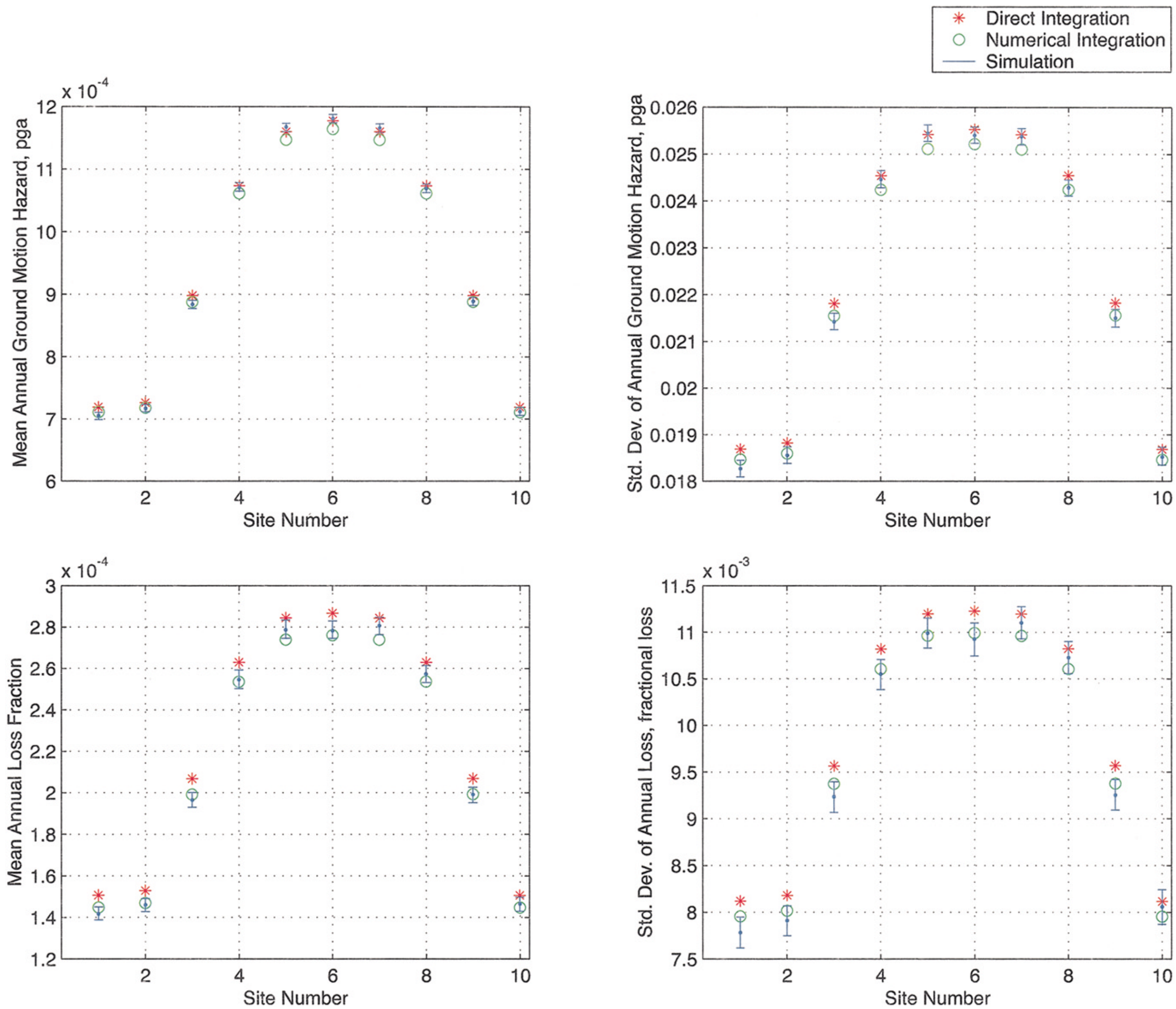

Figure 6. Mean and standard deviation of ground motion and loss at all sites determined from direct method, simulation, and numerical integration. The 95% confidence intervals on the results of the simulation are estimated by boot strapping.

larger ground motions and losses at sites 3 and 4. This is true given all four loss levels at site 1. For a larger universe of sources, including sources to the other side of site 1 from sites 3 and 4, this effect would diminish.

As the losses at site 1 increase, the conditional probability for losses at nearby sites also increases as might be expected intuitively. What is perhaps more interesting is that as the loss levels at site 1 increase, the conditional probability for losses at distant sites actually decreases. This occurs because when an event is near sites 1 and 2 it is farther from the distant sites.

From all the pairwise conditional probabilities it would be possible to obtain an improved approximation of the shape of the probability density for the sum of the losses in the portfolio. Whether this is a practical approach is a subject for future investigation.

Losses to a Portfolio

Annual losses to a portfolio of assets can be estimated by combining the means, variances, and correlations among the losses at the individual sites using the theorem for the mean and variance of the sum of random variables. Let the value of the asset at each site in the portfolio be a_i with the fractional loss represented by the random variable l_i, with mean u_j and variance σ_i^2. Then the total loss to the portfolio is

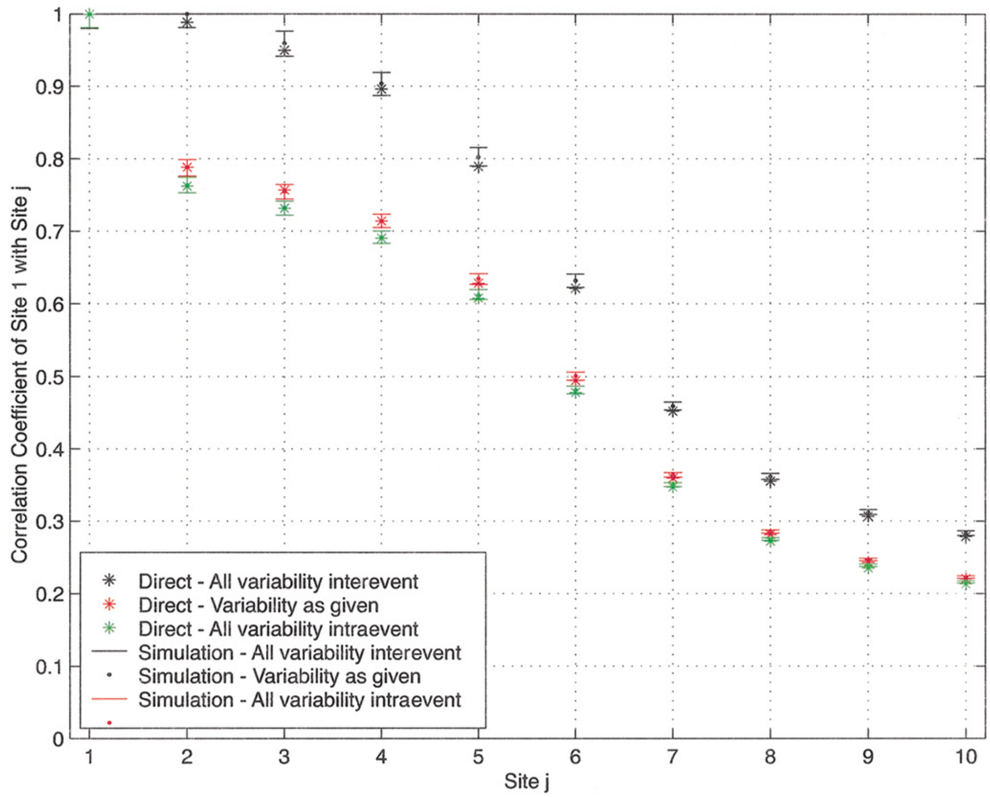

Figure 7. Correlation of ground motion between site 1 and all other sites determined from both direct calculation and simulation. "Variability as given" refers to calculations done with the components of variability as given by BJF. "All variability interevent" refers to calculations done with the same total variability as BJF, but all contributed by interevent variability. "All variability intraevent" refers to calculations done with the same total variability as BJF, but all contributed by intraevent variability.

$$L = \sum_{i=1}^{n} a_i l_i$$

and the mean annual loss and the mean variance of the loss to the portfolio can be estimated (cf., Freund, 1962) as:

$$\mu_p = \sum_i a_i \mu_i,$$

$$\sigma_p^2 = \sum_i \sum_j a_i a_j \sigma_i \sigma_j \rho_{ij},$$

$$\sigma_p^2 = \sum_i a_i^2 \sigma_i^2 + 2\sum_{i>j} \sum a_i a_j \sigma_i \sigma_j \rho_{ij},$$

where ρ_{ij} are the correlation coefficients between the losses at the ith and jth sites. The second equation for σ_p^2 is important because it demonstrates the contribution of the unsystematic risk (the first term) and the systematic risk (the second term). The systematic risk can be reduced through the selection of sites whose losses are uncorrelated (i.e., that is through appropriate diversification) (cf., Marshall and Bansal, 1993). Thus the smaller the spatial correlation, the lower the variance of portfolio and the lower the risk.

For the simple case described previously with all the weights, a_i, assumed to be 0.1, the mean and standard deviation of a portfolio combining similar properties at each site are shown in Table 3. The mean and standard deviation for the losses in the portfolio are fit within about 1.5%.

The direct method thus provides a method to calculate the mean and variance of losses to a portfolio, but what if anything else can be said about the distribution of losses? In contrast, simulation can provide an estimate of the entire distribution. Obviously if the distribution of losses is well described by the mean and variance, then the problem is essentially solved. For example, if the distribution of losses to the portfolio can be well fit with a two-parameter, analytical distribution, then knowledge of that distribution plus the mean and variance would suffice. This subject remains an area for additional research.

Conclusions

We present a feasible method for the direct calculation of joint exceedance frequencies at pairs of sites. These frequencies can in turn be used to calculate the joint probability

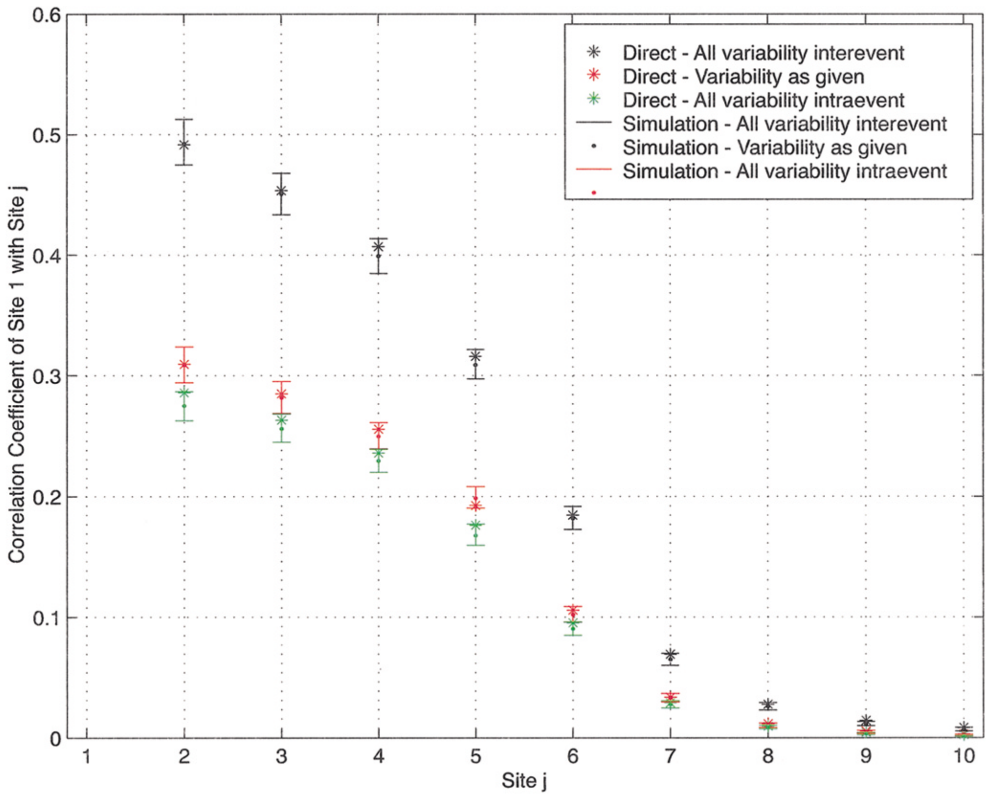

Figure 8. Correlation of loss between site 1 and other sites determined from both direct calculation and simulation. (Correlation coefficients of Site 1 with itself are one for all cases. These are omitted to permit expansion of the vertical scale.) "Variability as given" refers to calculations done with the components of variability as given by BJF. "All variability interevent" refers to calculations done with the same total variability as BJF, but all contributed by interevent variability. "All variability interevent" refers to calculations done with the same total variability as BJF, but all contributed by interevent variability.

distributions for ground motion and loss and the correlation coefficients of ground motion and loss at pairs of sites. This method is verified by comparison of results with those of a simple simulation. The method is potentially useful for gaining additional insights into the contributions of various factors giving rise to correlation and to estimating the effect of correlation on the distribution of losses to a portfolio. The direct method offers an alternative to simulation for calculating the variance of losses to a portfolio, which may reduce the amount of calculation required. It may be possible, for example, to calculate the pairwise, joint exceedance surfaces for points on a grid, which corresponds to a hazard map, for use in calculating the variances in losses to a wide variety of portfolios.

Acknowledgments

This work was initially stimulated by the wonderful book by Peter Bernstein, Capital Ideas. Conversations with Lixin Zeng further stimulated our interest and thinking. Conversations with Erdal Safak, Mark Linker, Charlie Scawthorn, Weimin Dong, and Craig Taylor helped clarify our thinking. Mark Petersen has been especially encouraging and helpful. The work could not have been begun or completed without the patient support and encouragement of Art Frankel. The article benefited from careful reviews by Tianqing Cao and Paolo Bazzuro, for which we are grateful. We are particularly grateful to Dr. Cao for pointing out difficulties that had crept into equations in Appendices A and C.

References

Abrahamson, N. A. (2000). Effects of rupture directivity on probabilistic seismic hazard analysis, in *Proc. 6th International Conference on Seismic Zonation*, Palm Springs, California, November, 2000.

Abrahamson, N. A., and W. J. Silva (1997). Empirical response spectral attenuation relations for shallow crustal earthquakes, *Seism. Res. Lett.* **68,** no. 1, 94–127.

Abramowitz, M., and I. A. Stegun, (Editors) (1964). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables,* National Bureau of Standards, Applied Mathematics Series 55, 1046 pp.

Algermissen, S. T., and D. M. Perkins (1976). A probabilistic estimate of maximum acceleration in rock in the contiguous United States, *U.S. Geol. Surv. Open-File Rep. 76–416,* 45 pp.

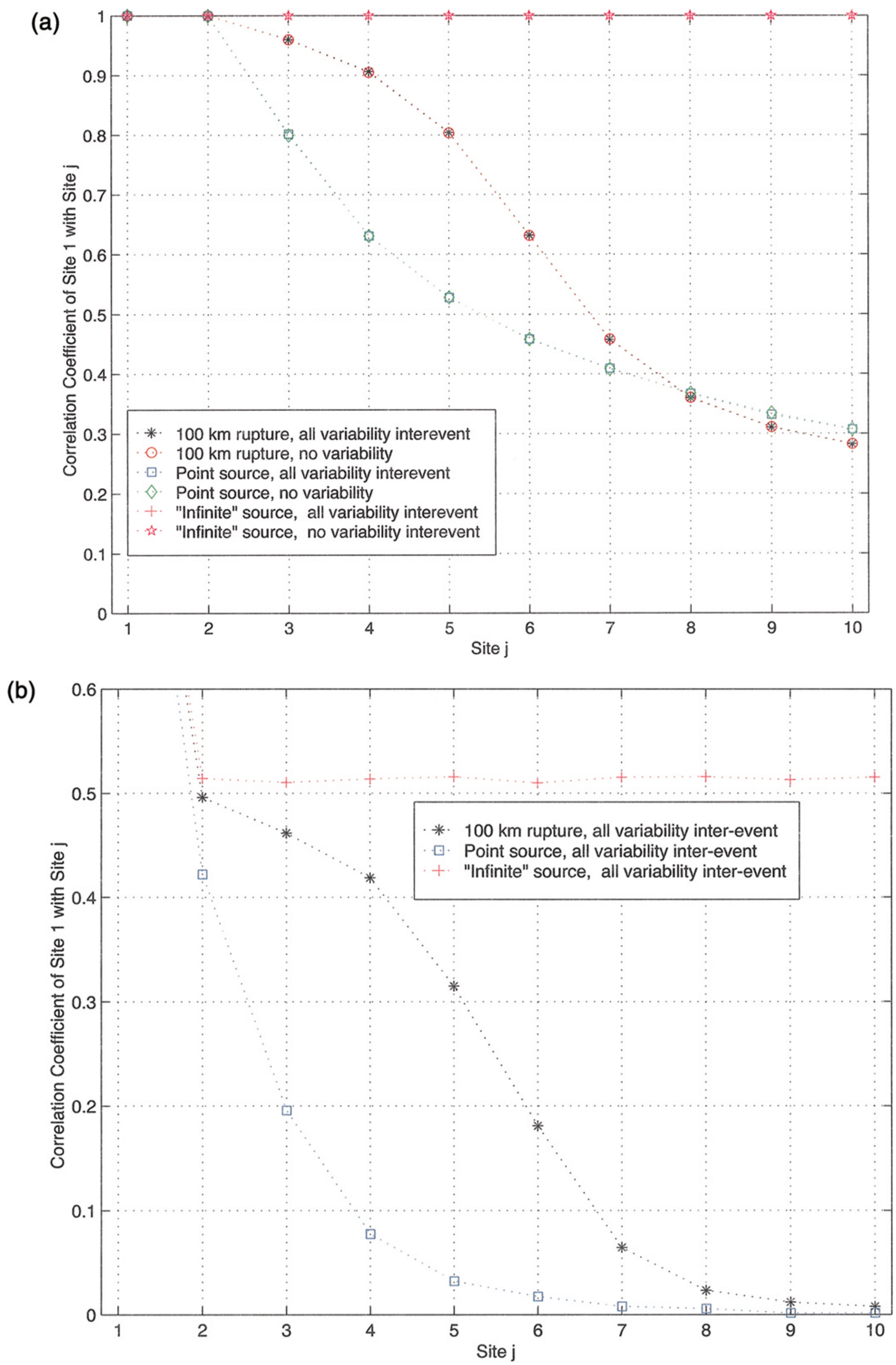

Figure 9. Factors contributing to spatial correlation: (a) ground-motion correlation for point, 100-km, and infinite ruptures for all ground-motion interevent variability and no ground-motion variability. (b) Loss correlation for all ground-motion interevent variability (c) Loss correlation for no ground-motion variability.

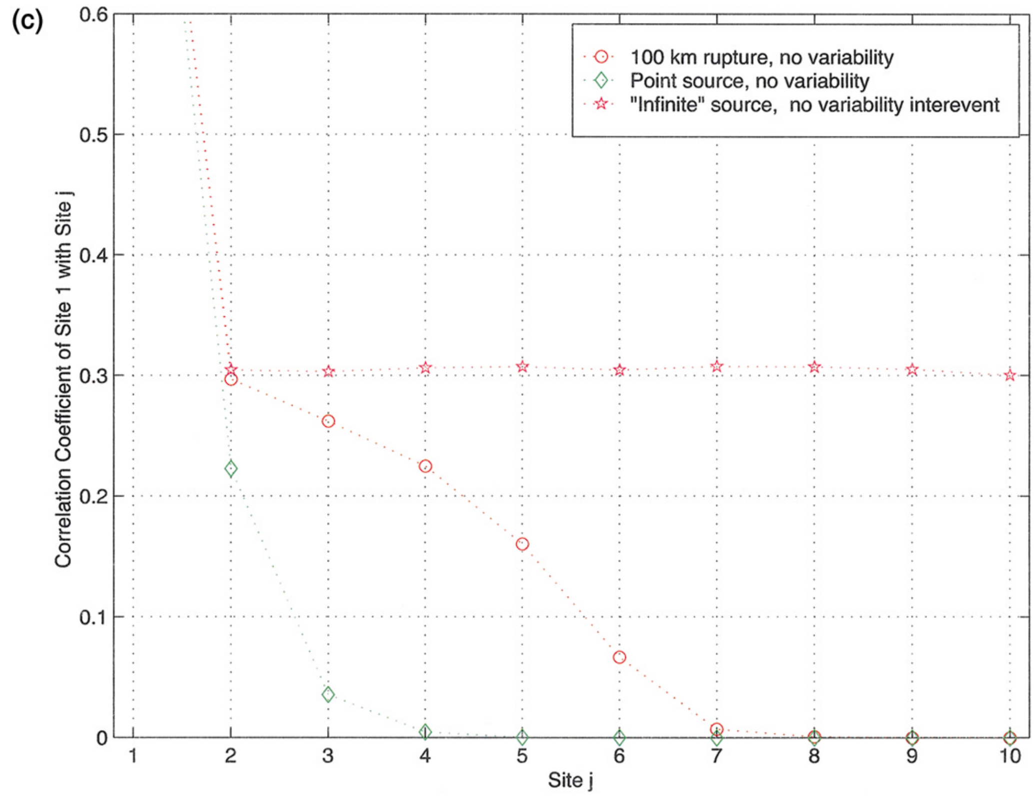

Figure 9. (*Continued*)

Applied Technology Council (1985). ATC-13, Earthquake Damage Evaluation Data for California, Redwood City, California, 492 pp.

Bernstein, Peter (1993). *Capital Ideas: The Improbable Origins of Modern Wall Street,* The Free Press, New York, 340 pp.

Boore, D., W. Joyner, and T. Fumal (1993). Estimation of response spectra and peak accelerations from western North American earthquakes: an interim report, *U.S. Geol. Surv. Open-File Rep. 93–509,* 72 pp.

Boore, D., W. Joyner, and T. Fumal (1994). Estimation of response spectra and peak accelerations from western North American earthquakes, Part 2, *U.S. Geol. Surv. Open-File Rep. 94–127,* 40 pp.

Boore, D., W. Joyner, and T. Fumal (1997). Equations for estimating horizontal response spectra and peak acceleration from Western North American earthquakes: a summary of recent work, *Seism. Res. Lett.* **68,** no. 1, 128–153.

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

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

Dong, W. (2000). The insurance sector, in *Financial Management of Earthquake Risk,* EERI Endowment Fund White Paper, Earthquake Engineering Research Institute, Oakland, Chapter 1, Chapter 8, 77–84.

Frankel, A., C. Mueller, T. Barnhard, D. Perkins, E. Leyendecker, N. Dickman, S. Hanson, and M. Hopper (1996). National seismic-hazard maps: documentation June 1996, *U.S. Geol. Survey Open-File Report 96–532,* 110 pp.

Freund, J. E. (1962). *Mathematical Statistics,* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 390 pp.

Hildebrand, F. B. (1962). *Advanced Calculus for Applications,* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 646 pp.

Hull, J. C. (1997). *Options, Futures and Other Derivatives,* Third Ed., Prentice-Hall, Upper Saddle River, New Jersey, 572 pp.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1997). *Discrete Multivariate Distributions,* John Wiley & Sons, Inc., New York, 299 pp.

Marshall, J. F., and V. K. Bansal (1993). *Financial Engineering,* Second Ed., Kolb Publishing Company, Miami, Florida, 728 pp.

Odegaard, B. A. (1999). Financial numerical recipes, *http://finance.bi.no/~bernt/gcc_prog/algoritms/algoritms/node59.html*

Reiter, L. (1990). *Earthquake Hazard Analysis: Issues and Insights,* Columbia University Press, New York, 254 pp.

Ross, S. A., R. W. Westerfield, and J. Jaffe (1996). *Corporate Finance,* Irwin, Chicago, 729 pp.

Somerville, P. G., N. F. Smith, R. W. Graves, and N. A. Abrahamson (1997). Modification of empirical strong ground motion attenuation relations to include amplitude and duration effects of rupture directivity, *Seism. Res. Lett.* **68,** 1999–122.

Wells, D. L., and K. J. Coppersmith (1994). New empirical relations among magnitude, rupture length, rupture width, rupture area, and surface displacement, *Bull. Seism. Soc. Am.* **84,** 974–1002.

Wesson, R. L., E. V. Leyendecker, R. V. Roth, Jr., and D. M. Perkins (1999). Losses and ground motions from the 1994, Northridge, California, earthquake, *EOS Trans. Am. Geophys. Union* **80,** F685–F686.

Zadeh, M. M. (2000). Understanding risk management, *Financial Management of Earthquake Risk,* EERI Endowment Fund White Paper Earthquake Engineering Research Institute, Oakland, California, Chapter 1, 1–14.

Appendix A

Probability of Exceedance from Ground-Motion Variability

In analyzing the attenuation of strong ground motion, some authors (e.g., Boore *et al.,* 1993, 1994) use logarithms to the base 10, while others (e.g., Boore *et al.,* 1997; Abra-

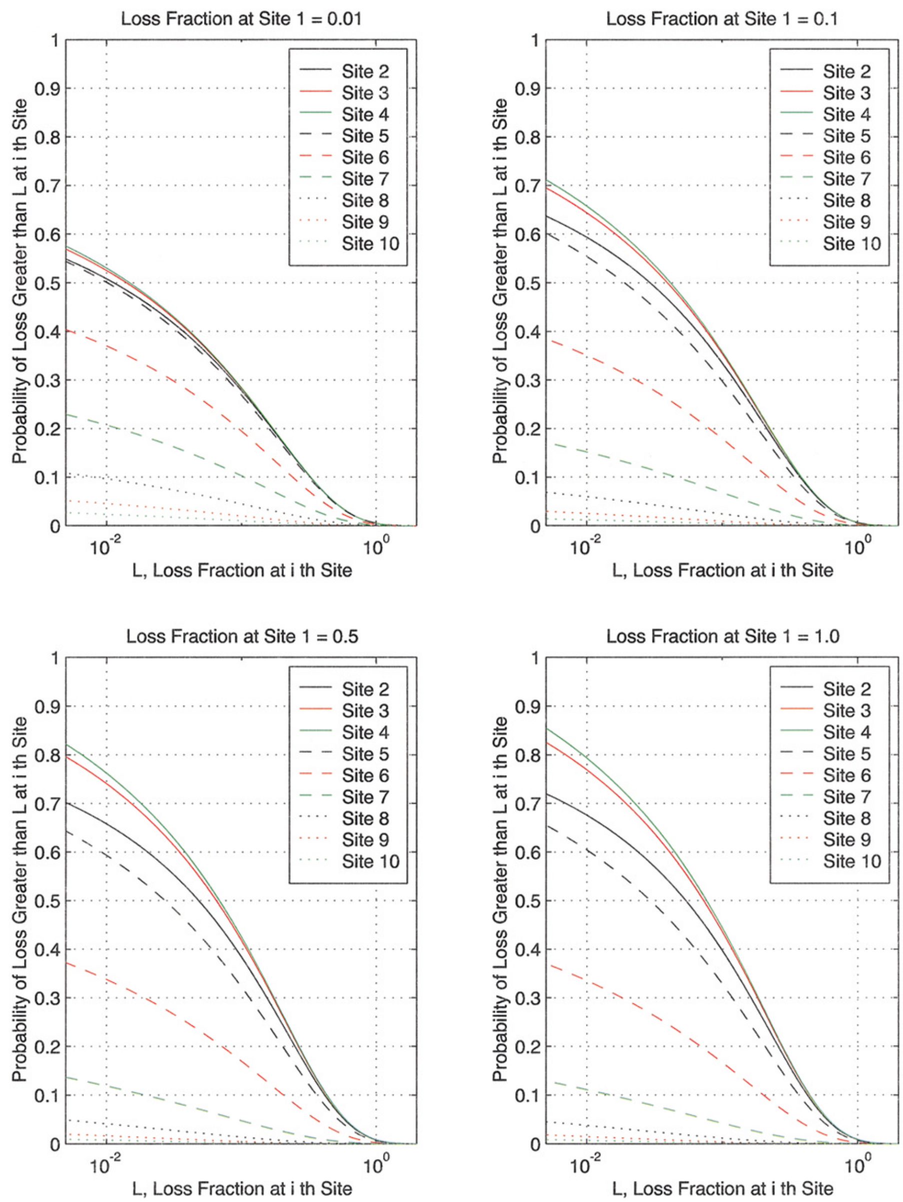

Figure 10. Conditional probability of loss at sites 2–10 given loss at site 1.

Table 3
Comparison of Portfolio Loss Parameters using BJF Variability as Given

	Mean	Standard Deviation
Simulation	2.18×10^{-4}	4.74×10^{-3}
Direct	2.21×10^{-4}	4.79×10^{-3}

hamson and Silva, 1997) use logarithms to the base e. For a typical attenuation curve, such as that given by Boore *et al.* (1993, 1994), the ground-motion parameter is estimated by $\log_{10}(u) = f(D, M)$, where $\log_{10}(u)$ is assumed to be normally distributed with standard deviation σ. The probability of the ground motion exceeding u_0 is then equal to the area under the normal curve (cf., Freund, 1962; Abramowitz and Stegun, 1964),

$$P\big(u > u_0 \mid D_{ik}, M_l\big) = \tfrac{1}{\sqrt{\pi}} \int_{z_0}^{\infty} e^{-z^2}\, dz = \tfrac{1}{2}\big(1 - \mathrm{erf}(z_0)\big) \tag{A1}$$

where

$$z_0 = \frac{\log_e u_0 - f(D,M)\log_e 10}{\sqrt{2}(\log_e 10)\sigma} \tag{A2}$$

and erf(z) is the error function.

Appendix B

Joint Distribution for Ground Motions

Let

$$x = \eta + \varepsilon_x,$$
$$y = \eta + \varepsilon_y,$$

where η, ε_x, and ε_y are drawn from independent normal distributions with zero mean:

$$f(\eta) = \frac{1}{\sqrt{2\pi}\sigma_\eta} \exp\left(-\frac{\eta^2}{2\sigma_n^2}\right)$$
$$f(\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \exp\left(-\frac{\varepsilon^2}{2\sigma_\varepsilon^2}\right).$$

We want the joint probability density for x and y. The probability density functions for x and y are the densities for the sum of independent normal variables, being, respectively,

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$
$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{y^2}{2\sigma^2}\right) \tag{B1}$$

where $\sigma^2 = \sigma_\eta^2 + \sigma_\varepsilon^2$, a well-known result (e.g., Feller, 1971).

Consider

$$x - y = \varepsilon_x - \varepsilon_y.$$

Noting that ε_{ix} and ε_{iy} are drawn from independent, but identical, normal distributions, by the same previous argument we can write

$$f(x-y) = \frac{1}{2\sqrt{\pi}\sigma_\varepsilon} \exp\left(-\frac{(x-y)^2}{4\sigma_\varepsilon^2}\right).$$

Note that the variance of this distribution is $2\sigma_\varepsilon^2$. Then we can obtain the covariance of x and y indirectly, by writing (Freund, 1962)

$$\mathrm{var}(x-y) = \mathrm{var}(x) + \mathrm{var}(y) - 2\,\mathrm{cov}(x,y).$$

We have already obtained the variances in this equation and can solve for the unknown covariance. Noting that the variance of x and of y is σ^2, and that $\sigma^2 = \sigma_\eta^2 + \sigma_\varepsilon^2$, this expression can be rewritten as

$$2\sigma_\varepsilon^2 = (\sigma_\eta^2 + \sigma_\varepsilon^2) + (\sigma_\eta^2 + \sigma_\varepsilon^2) - 2\,\mathrm{cov}(x,y).$$

Solving for the last term, we obtain

$$\mathrm{cov}(x,y) = \sigma_\eta^2,$$

or, using the definition of correlation coefficient

$$\rho_{x,y} = \frac{\mathrm{cov}(x,y)}{(\mathrm{var}\,x \cdot \mathrm{var}\,y)^{1/2}} = \frac{\sigma_\eta^2}{\sigma^2} = \frac{\sigma_\eta^2}{(\sigma_\eta^2 + \sigma_\varepsilon^2)}.$$

Thus the joint distribution we seek has normal distributions (B1) as marginal distributions and correlation coefficient ρ. It can be easily verified that the bivariate normal distribution with zero means, standard deviations σ, and correlation coefficient ρ has these properties, and therefore it is the desired joint distribution.

Appendix C

Calculating Derivatives of Hazard Curves

From a computational point of view it is more accurate to calculate $d\lambda(u > u_0)/du$ directly rather than to differentiate the hazard curve numerically. This can be done straightforwardly by differentiating (1) and (2) to obtain for point sources

$$d\lambda(u > u_0)/du_0 = \sum_k \sum_l N_l dP(u > u_0 \mid D_{k,} M_l)/du_0$$

and for a fault

$$d\lambda(u > u_0)/du_0 = \sum_l \sum_r \frac{1}{n_l} N_l dP(u > u_0 \mid D_{lr,} M_l)/du_0,$$

where the derivative of the probability with respect to the ground-motion threshold is determined from differentiating equation (A1) in accord with Leibnitz's rule (Hildebrand, 1962):

$$dP(u > u_0 \mid D_{lr,} M_l)/du_0 = \frac{-e^{-z_0^2}}{\sqrt{2\pi}(\log_e 10)\sigma u_0}.$$

U.S. Geological Survey
Central Region Geologic Hazards Team
P.O. Box 25046
Denver, Colorado 80225

Manuscript received 5 December 2000.