Uncertainty and Spatial Correlation Models for Earthquake Losses

Nilesh Shome, Nirmal Jayaram & Mohsen Rahnama
Risk Management Solutions Inc., Newark, California, USA

SUMMARY
Insurance companies typically estimate earthquake losses for buildings using catastrophe loss models. In this study, we analyze the insurance loss data from the Northridge earthquake to improve existing loss models. Conventionally, the distribution of loss is modelled using the Beta distribution. The claims data, however, show that the Beta distribution combined with Dirac-delta functions at 0 and 1 to model the probabilities of zero and complete loss improves the fit to the observed loss distribution. In addition, the spatial correlation between the losses at two locations is developed for modelling the loss distribution for a portfolio of buildings. The components of the correlation model are estimated using claims data and earthquake intensity data. Finally, the loss assessment for a portfolio of buildings is carried out using the improved uncertainty and spatial correlation models in order to demonstrate the importance of these models in assessing and managing seismic risks.

Keywords: Portfolio loss, loss uncertainty, spatial correlation, seismic loss distribution, Northridge earthquake.

INTRODUCTION

Insurance companies typically estimate seismic losses for individual buildings and building portfolios using catastrophe loss models (e.g., RiskLink® by Risk Management Solutions, Inc.). For a given earthquake, these loss models typically provide estimates of the expected loss for the building(s) of interest as well as the variance in the loss estimate. These models are developed using a combination of insurance claims data and analytical models. In this study, the Northridge earthquake claims data from a number of insurance companies are analyzed in order to further improve the existing seismic loss models. Conventionally, the distribution of a building loss for a given event is modelled using the Beta distribution parameterized by the loss mean and variance (Benjamin and Cornell 1970). But the claims data show that the Beta distribution combined with Dirac-delta functions at 0 and 1 to model the probabilities of zero and complete loss fits the observed loss distribution data better. In this study, in addition to developing models for the mean and variance of losses, models are also developed for the probabilities of zero and complete loss for various building classes in order to completely parameterize the loss distribution.

While the mean and the variance of the loss are usually sufficient to parameterize the loss distribution at a given site, it is important to model the spatial correlation between the losses at two locations for modelling the loss distribution for a portfolio of buildings. The correlation model proposed in this study has two components: 1) a constant global component and a 2) distance-dependent local correlation component. The spatial correlation model is developed from the claims loss residuals where the residual is defined as the difference between the observed loss at a site and the mean loss conditioned on the hazard at that site. Since the detailed loss data are available only from one event, ground-motion (GM) intensity data is used to adjust the model to take into account the variation in the results from multiple events. Typically the catastrophe models neglect the distance-dependent local correlation component between the loss residuals at two close-by sites. But Northridge data clearly demonstrates the presence of distance-dependent correlation between loss residuals, particularly

visible at reasonably closely-spaced locations.

Finally, the loss assessment is performed for a portfolio of buildings using the improved uncertainty and spatial correlation models in order to demonstrate the importance of these models in assessing and managing seismic risks.

LOSS DATA

The 1994 Northridge earthquake caused an unprecedented loss in Los Angeles for a moderate-sized (M_w 6.7) earthquake. The loss to the insurance industry was about $12.5B (Insurance Information Institute). The wide spread loss also provided a wealth of loss data that have been used over the years by the insurance industry to calibrate and validate earthquake loss models. In this study, we have revisited the development process to further improve the method in which loss models consider the uncertainty distribution of loss at one location and the correlation between losses at multiple locations. There are some challenges in using the existing Northridge loss data for parameter assessment. Later on this manuscript describes how these challenges are addressed in the current study.

In this study, loss data from six insurance companies primarily for the single family dwelling (SFD) occupancy are used. The total insured value (TIV) of the buildings which were subjected to at least 0.2g spectral acceleration (S_a) at period 0.3s is $57B and the total gross (GR) loss (i.e., losses after applying deductible) is $2.8B. We only consider losses due to damage to only structural and non-structural building components. The losses exclude damage to contents (e.g., furniture) and additional living expenses (ALE) associated with building damage (e.g., expenses associated with hotel stay due to being temporarily displaced from a damaged home).

While location-level loss data provides the most accurate estimates of the loss model parameters, we only have the zip code information about the insured properties for insurer confidentiality reasons. So it is not possible to accurately capture the variability of the losses within a zip code due to the variation in the GM intensities or soil conditions within zip codes. The other issue which is common to any insurance loss data is the availability of only the gross losses, which are the loss estimates obtained after applying deductibles to the incurred losses. This type of data, commonly known as (left) "censored" data, needs to be dealt appropriately for accurate estimation of the loss distribution and correlation parameters. This issue is even more important in this study due to the high deductibles associated with US earthquake insurance policies (the average deductible was around 10% for earthquake insurance policies during the Northridge event versus 2% for typical US hurricane insurance policies for residential properties). In order to estimate the distribution and correlation parameters, the entire loss, which includes deductible, is first calculated. This is also known as ground-up (GU) loss in the insurance industry. The distribution of the GU loss ratio (ratio of GU loss to TIV) for all the six insurance companies is shown in Fig. 1. Note that since the claims data reports losses below deductible as zero loss, GU =0 when GR=0. Hence the probability of loss below the average deductible (around 10%) in the figure is not accurate.

An additional piece of information that is required for parameterizing the loss functions is the ground-motion intensity (S_a) at the building locations. Here, GM intensities are obtained from USGS' ShakeMap, which is developed based on the observed intensities at the recording stations (ShakeMap 1994). Since the focus is only on the losses to SFD buildings, the S_a at T=0.3s, which is close to the fundamental period of low-rise SFDs, is considered.

LOSS DISTRIBUTION

In this section, we will estimate the GU loss distribution as a function of spectral parameter (S_a at T=0.3s). Generally it is assumed that the loss distribution follows a Beta distribution (ATC-13 1985), in part because the Beta distribution can be limited between 0 and 1 and is flexible in regard to the shape of its probability density function. The Beta distribution is completely defined by the first two central moments of the sample loss data, namely, the mean and the variance. In loss calculations, the losses are represented using damage ratio (DR), which is the ratio of GU loss to the TIV, as functions of the S_a. The mean DR as a function of S_a is known as the building vulnerability function. Fig. 2 shows the variation of DR as a function of $S_a(T$=0.3s) for SFD buildings as observed during the Northridge earthquake. The vertical stripes in the figure are indicative of the distribution of DRs in a zip code. In order to calculate the first two moments of the sample data as a function of GM intensity for vulnerability calculations, sample loss data are first grouped into different S_a bins. It is assumed that the buildings constructed between the 1933 Long Beach earthquake and the 1975 State Building Code Act requiring 1973 UBC as the minimum construction code for all buildings in communities throughout the state have similar vulnerability characteristics. This construction period is approximately HAZUS (2003) defined moderate-seismic SFD buildings. Hence the losses for buildings constructed during this period are grouped together for developing building vulnerability for the 1933-1975 construction period. For illustration, the distributions of DR at a low and a high intensity for the moderate-seismic SFD are shown in Fig. 3. Note that the bump at the zero damage is both because of buildings not seeing any loss at that intensity and because the GU loss is assumed to be zero when the GR loss equals 0 (i.e., no additional information about the GU loss is available when it is below the deductible). Also Fig. 3(b) shows that there is a small bump in the distribution at complete damage (DR=100%) and we expect that the bump would will increase progressively at higher intensities. The complete damage is due to collapse of buildings at high intensity GM and also due to writing off by the insurers of high damaged buildings even though the buildings are not collapsed. The maximum damage at which the insurance adjusters write off varies with the loss-adjustment policy and the under-writing guideline of the insurance companies.

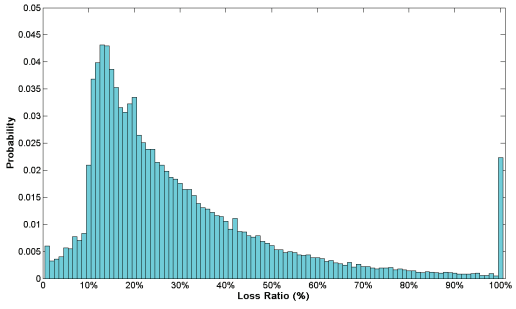


Figure 1. Distribution of percentage of non-zero GU loss ratio (GU/TIV) for all the companies.

When the standard Beta distribution is fitted by equating the first two moments of the distribution (mean and variance) to the moments of the observed loss data, Fig. 3 demonstrates that it will not be possible to accurately capture the bumps in the loss distribution at the end. In order to capture the probability of no loss (F_0) and of complete loss (F_1), it is necessary to estimate these parameters separately in addition to the first two moments of the data. In other words, the parameters F_0 and F_1 are estimated from the loss data in addition to the two standard beta parameters in order to define the loss distribution. In general, it is a common practice to fit a four-parameter distribution by matching the first four theoretical and sample moments.

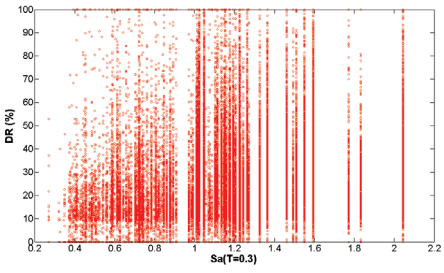
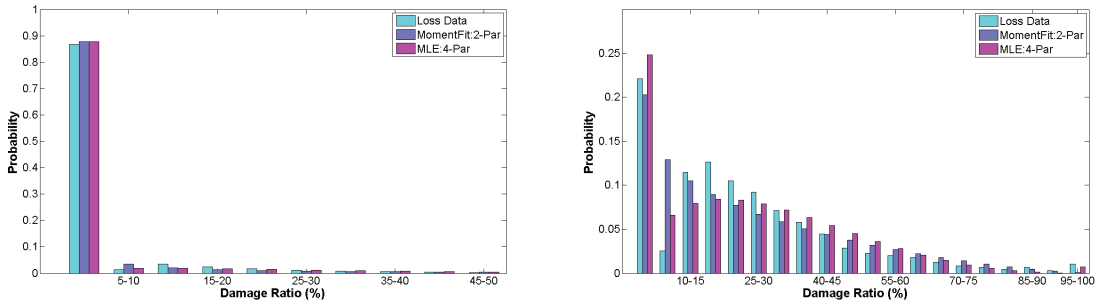


Figure 2. *DR* (GU/TIV) as a function of $S_a(g)$ at T=0.3s

The third and the fourth moments, known as skewness and kurtosis, however, are notorious for their large standard errors. Estimating these moments requires sample sizes that are rarely available at moderate to high intensities that drive the insurance loss calculations. In this study, we propose an alternative approach that employs only the first two moments. In addition, the probabilities of zero (F_0) and complete losses (F_1) are used in place of skewness and kurtosis. It is quite common of having no loss data for deriving all the parameters for buildings with all different construction classes and age groups at all the GM intensity ranges of interest. So parameters like F_0 and F_1, which are easy to understand and can be developed from engineering calculations, are more suited for catastrophe risk modeling compared to skewness and kurtosis.



(a) Low intensity (b) High intensity

Figure 3. Distribution of damage ratio at low and high ground-motion intensities

The 4-parameter Beta probability distribution used in this study is a mixture of discrete (delta functions at 0 and 1) and continuous random variables (standard Beta distribution between 0 and 1), whose cumulative density function (CDF) is described below:

$$F(x) = F_0 + (1 - F_0 - F_1).F_{beta}(x;\alpha,\beta) + \int_0^x \delta(x-1) \quad (1)$$

where δ denotes the Dirac Delta function, which is defined by:

$$\delta(x) = \begin{cases} \infty, & x = 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

such that

$$\int_{-\infty}^{+\infty} \delta(x)dx = 1 \quad (3)$$

where $F_{beta}(\cdot)$ is the cumulative Beta distribution function between 0 and 1, and α and β are the parameters of Beta distribution. The probability density function (PDF) of 4-parameter Beta distribution is given below:

$$f_{beta}(x) = \begin{cases} F_0, & x = 0 \\ (1 - F_0 - F_1) \cdot f_{beta}(x;\alpha,\beta), & 0 < x < 1, \\ F_1, & x = 1 \end{cases} \quad (4)$$

where $f_{beta}(\cdot)$ is the Beta density function.

Censored Data: Maximum Likelihood Estimation (MLE) of Distribution Parameters

As discussed before, insurance loss data cannot be used directly to estimate the parameters of the distribution due to relatively large deductibles associated with earthquake insurance policies (i.e.,

"censoring" of the loss data below the policy deductible). In particular, the value of F_0 will be considerably overestimated if the loss data are used as such because the zeros in the loss data generally correspond to GU loss being below deductible and not being zero as such. Maximum Likelihood Estimation (MLE) can be used to fit the parametric distribution (Eqn. 1) to censored data sets, by explicitly accounting for the portion of the distribution that is below deductible. The general formulation of the likelihood function for loss data which are censored below deductible is as follows (Kendall and Stuart 1977):

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} f(x_i|\boldsymbol{\theta}) \cdot \prod_{j=1}^{N_c} F(D_j|\boldsymbol{\theta}) \quad (5)$$

where $f(x|\boldsymbol{\theta})$ is the probability density function (here Beta), $F(x|\boldsymbol{\theta})$ is the cumulative distribution function, $\boldsymbol{\theta}$ is the vector of distribution parameters (here α, β, F_0 and F_1), N is the number of loss data above deductible and N_C is the number of loss data points below deductible D_j. In MLE, the distribution parameters are estimated by maximizing the likelihood function.

Fig. 3 shows the MLE fit of 4-parameter Beta distribution to the loss data at a low- and a high-intensity GM. The two-parameter Beta distribution fit to the data by the commonly used method of moments is also shown in the figure for comparison. A comparison of the loss results from these distributions with those observed will be shown later on. The process of MLE fit is repeated at different intensity bins for estimating the 4-parameters of the distribution as a function of intensity. This gives us the vulnerability functions for estimating losses in catastrophe risk models.

SPATIAL CORRELATION

The distribution model discussed in the previous sections is required for estimating the loss distribution for a single building. In order to estimate the loss distribution for a portfolio of buildings, we need information about the correlation between losses for all the pairs of buildings. The spatial correlation model developed in this study has two components: 1) a constant global correlation term (ρ_G) and 2) a distance-dependent local correlation term (ρ_L) (Fig. 4). The global component is the same for any two pair of sites. This component takes into account the correlation of losses between two locations during an event induced by the unique characteristics of that event not captured by the catastrophe model (e.g., stress drop). The local component, on the other hand, depends on the separation distance between the sites. It is highest at the zero separation distance and reduces to zero at large separation distances. This correlation at two closely-spaced sites is due to similarity in local site effects (similar terrain, similar soil conditions etc.), similarity in the path effects (e.g., earthquake wave propagation path), similarity in the design and quality of constructions, and many other factors, which are not explicitly considered in the model.

The spatial correlation model is developed from the residuals of DR. The residual is the difference between the observed loss at a site and the mean damage ratio conditioned on the hazard at that site. The *DR* at a site can be written as follows:

$$DR_{i,j} = g\left(s_{a(i,j)}, \boldsymbol{\eta}\right) + \varepsilon_{i,j} + \beta_j \quad (6)$$

where $DR_{i,j}$ is the damage ratio at location i during event j, $g(\cdot)$ is the vulnerability function defining the mean *DR* as a function of spectral intensity, $s_{a(i,j)}$, $\boldsymbol{\eta}$ denotes the set of parameters defining the characteristic of the building at location i (e.g., construction type, building age, etc.) that will impact the mean damage ratio, , $\varepsilon_{i,j}$ is the residual error term that denotes the uncertainty in the damage ratio that varies from one location to another, and β_j denotes the inter-event error term in the damage ratio that is constant for a given event and does not vary from one location to another. A positive value of β implies that the event in consideration produces above average damage ratios at all the locations. The sum of the two residuals is the total residual. The formulation shown in Eqn. 6 is called a mixed-effects regression model (Brillinger and Preisler 1984), which helps to separate the total residuals in

different components, and is commonly used in modelling earthquake hazards (e.g., Campbell and Bozorgnia, 2008).

Let σ denote the standard deviation of the residual term ε and τ denote the standard deviation of the event term β. It can be shown that the correlation between the damage ratios at sites i and i' from event j equals:

$$\begin{aligned}\rho(h) &= \frac{\mathrm{E}[(\mathrm{DR}_{i,j}-\overline{\mathrm{DR}_{i,j}})\cdot(\mathrm{DR}_{i',j}-\overline{\mathrm{DR}_{i',j}})}{\tau^2+\sigma^2} \\ &= \frac{\tau^2}{\tau^2+\sigma^2} + \frac{\sigma^2\cdot\rho_\epsilon(h)}{\tau^2+\sigma^2} \\ &= \rho_G + \rho_L(h)\end{aligned} \qquad (7)$$

where $\rho(h)$ denotes the correlation between the damage ratios at locations i and i' that are separated by distance h, and $\rho_\varepsilon(h)$ denotes the correlation between the residuals ε_i and $\varepsilon_{i'}$. $\rho_\varepsilon(h)$ arises because even the residual error term that varies across sites is not independent from one site to another. It is observed in the claims data that the residual error term has distance-dependent correlation, with the correlation decreasing with distance.

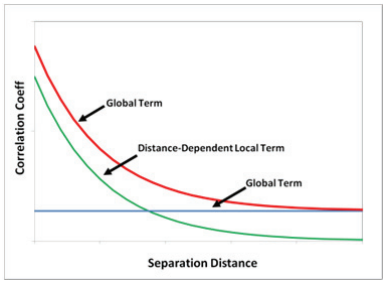


Figure 4. Correlation model

Note that the first term in Eqn. 7, which is independent of separation distance, arises because the damage ratios at both the sites contains the identical event term (β) which creates some correlation between the damage ratios. The second term arises because the residual term (ε) itself is correlated from one location to another. The first term is referred to as global correlation (ρ_G) in Fig. 4, while the second term is the distance dependent local correlation (ρ_L).

The above formulation implicitly assumes that the correlation is both stationary and isotropic, which implies that the correlation is only a function of the separation distance and does not depend on the actual location of the sites or the azimuth of the line connecting the sites. The validity of these assumptions is discussed in detail for GM intensities by Jayaram and Baker (2009). It is assumed here that these assumptions are also applicable to building losses, which are functions of GM intensities. But these assumptions cannot be verified because of lack of location-level loss data from multiple earthquakes (although verified for losses from multiple hurricanes for large number of insurance companies in Gulf states and Florida).

Local Correlation (ρ_L)

It is commonly assumed that the local correlation between loss residuals at two close-by sites is negligible. But it is seen from the Northridge data (Fig. 5) that there is a strong correlation between loss residuals at reasonably closely-spaced locations. The local correlation is estimated using a semivariogram, which is commonly used in geo-statistics to measure the spatial dependence in a random field. Let $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n]$ denote a spatially-distributed vector of the residual error term ε, and let $\tilde{\epsilon}$ denote the normalized residual term, which equals ε divided by σ. For correlation computation, it is more convenient to work with the normalized residuals than the residuals since the normalized residual has unit variance, although their correlations are the same. The semivariance is a measure of dissimilarity between the random variables and is related to the spatial correlation between the $\tilde{\epsilon}s$. Let i and j denote two locations separated by distance h. The semivariance, $\gamma(i,j)$, is defined as

$$\gamma(i,j) = \frac{1}{2}E[\tilde{\epsilon}_i - \tilde{\epsilon}_j]^2 \qquad (8)$$

The computation of the semivariance as defined in Eqn. 8 requires repeated measurements of ε at

locations i and j. This is rarely available (for instance, it is very rare to have a large number of earthquake recordings at two sites i and j). It is common, therefore, to use the stationary semivariogram instead, which is computed by assuming that the random function $\tilde{\epsilon}$ is both stationary and isotropic. Based on these assumptions, it can be stated that $\gamma(i,j)$ is independent of i and j, and depends only on h. Hence $\gamma(i,j)$ is represented by $\gamma(h)$. $\gamma(h)$ can now be computed simply using all pairs of $\tilde{\epsilon}$s in the data set that are separated by distance h, as follows:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [\tilde{\epsilon}_i - \tilde{\epsilon}_{i+h}]^2 \quad (9)$$

where $(\tilde{\epsilon}_i, \tilde{\epsilon}_{i+h})$ denotes a pair of epsilons at sites separated by distance h, and $N(h)$ denotes the number of such pairs.

Let $\grave{\epsilon}$ denote the total residual which equals the sum of the event term (β) and the error term (ε) normalized by σ. Hence, Eqn. 6 can be written as follows

$$\grave{\epsilon}_i = \frac{\epsilon_i + \beta}{\sigma} = \left(DR_i - f(s_{a(i)}, \boldsymbol{\eta})\right)/\sigma \quad (10)$$

In the above equation, the event notation j has been left out since these calculations are performed by event. Further, the semivariance of $\grave{\epsilon}$ from a single event based on Eqn. 10 equals

$$\begin{aligned} \gamma(h) &= \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[\frac{DR_i - f(s_{a(i)}, \boldsymbol{\eta})}{\sigma} - \frac{DR_{i+h} - f(s_{a(i+h)}, \boldsymbol{\eta})}{\sigma} \right]^2 \\ &= \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[\frac{DR_i - f(s_{a(i)}, \boldsymbol{\eta}) - \beta}{\sigma} - \frac{DR_{i+h} - f(s_{a(i+h)}, \boldsymbol{\eta}) - \beta}{\sigma} \right]^2 \\ &= \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [\tilde{\epsilon}_i - \tilde{\epsilon}_{i+h}]^2 \end{aligned} \quad (11)$$

In other words, ε, $\grave{\epsilon}$ and $\tilde{\epsilon}$ all have the same semivariogram. The spatial correlation is related to the semivariance as follows (e.g., Goovaerts 1997, Jayaram and Baker 2009):

$$\rho_{\varepsilon}(h) = 1 - \gamma(h) \quad (12)$$

This $\rho_{\varepsilon}(h)$ can be used in combination with Eqn. 7 to obtain the local correlation, $\rho_L(h)$. We are using a semivariogram to estimate correlation as against correlogram directly because the semivagriogram, which averages squared differences of the residual, tends to filter out the influence of a spatially varying mean.

While Eqn. 11 provides the empirical semivariance estimate, it is necessary to fit a smooth continuous function through the empirical estimates for prediction purposes. An exponential model is selected as the predictive model for fitting the semivariance to the loss data as a function of separation distance (h). In particular,

$$\gamma(h) = nI(h > 0) + (a - n) \cdot (1 - e^{-3h/b}) \quad (13)$$

where a and b are the sill and the range of semivariogram respectively, n is called the nugget effect, and $I(h>0)$ equals 1 when $h>0$ and 0 otherwise. Note that the exponential model has been selected over the Gaussian model since it has been found that the exponential model works better for modelling the spatial correlation of GM intensities (Jayaram and Baker 2009). It is expected that the same holds true for correlation of losses which is a function of GM intensity (also supported by insurance loss data for multiple hurricanes in Florida).

The estimation of the semivariogram is particularly tricky since we generally have irregularly distributed data and typically have few data pairs at shorter distances. But modelling the correlation at the shorter distances accurately is of paramount importance for estimating the joint distribution of loss

for portfolios of buildings. In this study, the exponential model is ensured to fit the empirical data well at short distances, even if this requires somewhat misfitting the data at large separation distances. The losses at large separation distances have small correlations, and in addition, the losses at widely separated sites have little impact on each other due to 'shielding' of their influence by more closely-located sites (Goovaerts 1997).

The loss data are used to fit the model in Eqn. 13 and the resulting fit is shown in Fig. 5. This fit, however, cannot be used directly since this is based on only one event. It has been shown by Jayaram and Baker (2009) that there are significant differences in the correlation of GM intensities from one earthquake to another. Fig. 5 shows the site-to-site correlation of GM intensities for Northridge earthquake and the average correlation from a number of earthquake events (Jayaram and Baker 2009). The Northridge correlations are seen to be smaller than the average correlations. In anticipation that the hazard correlation trend will also be reflected in the loss correlation, the correlation values based on the Northridge loss data are multiplied by the ratio of the intensity correlation for average earthquakes and Northridge at each separation distance in order to obtain the proposed model for this study.

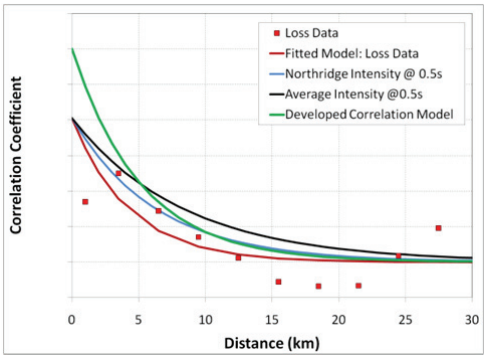


Figure 5. Correlation in Northridge loss data

In addition, the correlation model in Fig. 5 needs further refinement since only the zip codes of the sites are available to us. Since the zip-to-zip distances are used for calculating the distance between sites, the correlation at zero distance is underestimated. The correlation at zero distance as shown in Fig. 5 actually represents the correlation at the average site-to-site distance (d_{avgZip}) within zip codes. Hence, it is assumed that the zero distance shown in Fig. 5 is actually d_{avgZip}, and the semivariogram is simply extrapolated to 0 distance, thereby reducing the nugget and increasing the correlation at zero separation. Based on this logic, the correlation at zero separation is increased by around 50% corresponding to a d_{avgZip} of roughly 2km.

Global Correlation (ρ_G)

As described earlier, the global correlation term captures the characteristics (not explicitly considered in the catastrophe model) that cause different events to produce above or below average losses at all the locations (Eqn. 6). In order to estimate the global correlation, it is necessary to have site-level loss data from multiple earthquake events. Unfortunately, we have detailed earthquake loss data only from the Northridge earthquake. Hence ρ_G is estimated from GM intensity data which are available in abundance, and it is assumed that the loss global correlation will be comparable to the hazard global correlation. Based on Eqn. 7, this is done by first estimating the standard deviation of the residual term (σ) and the event term (τ), and subsequently using these estimates to calculate ρ_G.

Eqn. 6 which is called the *mixed-effects regression model* is commonly used for developing ground-motion prediction equations (GMPE). This formulation facilitates separating the total residuals into inter-event (β) and intra-event (ε) residuals. The ground-motion modelers also provide estimates of the standard deviations of these residual terms, which vary from period to period. The variation of ρ_G with period (T) based on the σ and τ from Campbell and Bozorgnia (2008) and Jayaram and Baker (2011) is shown in Fig. 6. Note that the former approach does not consider the correlation of intensities at the nearby locations, whereas the latter one does in the estimation of the variance of the residuals. The consideration of correlation reduces ρ_G by about 70%. This result is adopted in this study for ρ_G. The final correlation model is shown in Fig. 5.

LOSS RESULTS

The 4-parameter distribution model and the correlation model developed in this study are used for calculating the losses for a portfolio of buildings. The portfolio comprises of a random selection of 10,000 policies out of 250,000 total policies provided to us by the insurance companies. The portfolio loss is calculated based on the expected S_a as estimated by the USGS ShakeMap at the building sites for the Northridge earthquake for the following cases:. :

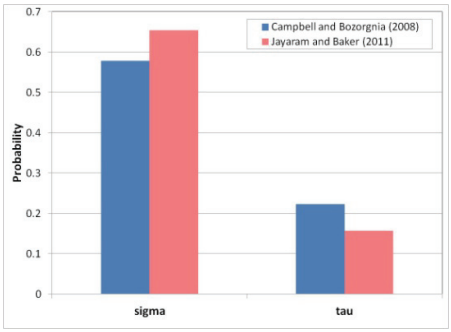


Figure 6. Global correlation based on Campbell and Bozorgnia (2008) and Jayaram and Baker (2011).

1. *Case 1-Existing Approach*: Loss distribution is defined by a 2-parameter Beta where the parameters are estimated by the method of moments. Site losses are assumed to be independent.
2. *Case 2- Developed Approach*: Loss distribution is defined by a 4-parameter Beta where the parameters are estimated by MLE by taking into account that the losses below deductible are censored. Site-to-site loss correlation is defined by a local term and a global term (Eqn. 7).

As described before, the parameters of the distribution are estimated as a function of S_a by grouping the observed loss data to different S_a bins. These results are used for calculating the parameters at different values of S_a at the building sites. Random loss samples are simulated from the two cases of distribution of losses only for buildings (i.e., no consideration of content and ALE losses) in Northridge earthquake. Note that the losses at different sites in Case 1 are simulated independently. The losses in Case 2, on the other hand, are correlated as defined in Fig. 5. The losses of structural and non-structural components of buildings constructed during 1933 to 1975 only are shown here for illustrating the difference in the results from the two different cases. In addition, the process is repeated 10 times in order to minimize the sample-to-sample variation in the results. In all the simulations, the GU and GR losses of the portfolio of buildings are calculated and the results are compared with those observed during the Northridge earthquake. Fig. 8 shows the comparison of loss distributions for zero and non-zero loss for Cases 1 and 2. The GU distribution from Case 2 is very close to that observed from the claims data, whereas the conventional approach (Case 1) results in a different distribution because of the deficiency of the 2-parameter Beta distribution in terms of modelling zero losses. The GU loss distribution for Case 2, however, should be lower than those observed because GU= 0 when loss is below deductible. The GR loss distribution, however, is quite similar for all the cases. Fig. 9 compares only the non-zero loss distribution for different cases with the observed data. The figure clearly demonstrates that the distribution from Case 2 is much closer to the observed for both GU and GR losses.

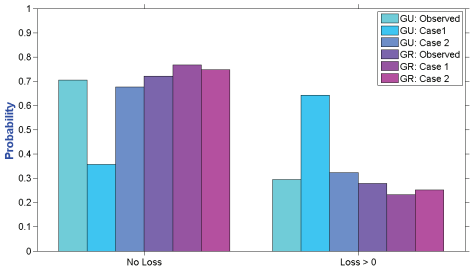


Figure 7. Comparison of non-zero and zero loss distribution for different cases with the observed.

The other advantage of using the 4-parameter distribution is that it improves the efficiency of the loss assessment procedure, particularly when a simulation approach is used. In simulation, when the loss distribution is defined by the 4-parameter Beta distribution, the loss samples are drawn by inverting the Beta distribution only when the losses are neither zero nor 100%, as defined by F_0 and F_1. In other words, the Beta inversion is only done (100-F_0-F_1)% of the time, unlike for the 2-parameter distribution where the Beta inversion has to be done 100% of the time. For portfolio loss calculations, this can reduce the computation time significantly for the simulation of loss samples. We have observed50-250% reduction in the computation time for large portfolios with around 100-300k insurance policies when analyses were carried out for large number of earthquake events for the calculation of the exceedance probabilities of losses for insurance risk management. .

CONCLUSION

This study explored the use of a new 4-parameter loss distribution model and a spatial correlation model intended at improving the loss estimation for building portfolios. In particular, the use of the 4-parameter loss model was aimed at better modeling the probabilities of observing zero and complete damage at a given ground-motion intensity. These probabilities have been estimated using claims data from the Northridge earthquake. In addition, a spatial correlation model was developed, which has a distance-independent global term which is induced by unmodeled event characteristics that affect all the locations during a given event. The correlation model also has a distant-dependent local term, which is induced by many unmodeled local parameters, e.g., local site condition, building characteristics, etc., defining the site-to-site correlation of loss during an event. The parameters of the correlation model were also estimated from the insurance loss data. Since the loss data was only available for a single event, the estimated correlation parameters were adjusted to take into account the variation among the events by using the corresponding information from GM intensities. Also the global correlation parameter is developed purely based on the GM intensity data since this parameter cannot be estimated without having loss data from multiple events. These models were then used to calculate the loss distribution for a portfolio of buildings. The results showed that the distribution of loss results from the developed models is superior to that estimated using the conventional 2-parameter Beta distribution with no site-to-site correlation term. The developed model is also efficient when loss calculations follow simulation approach.

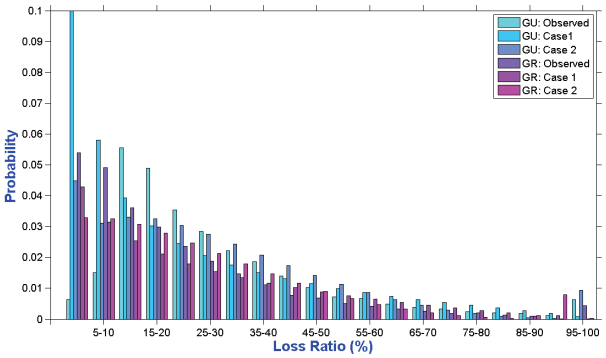


Figure 8: Comparison of non-zero loss distribution for different cases with the observed in Northridge earthquake.

REFERENCES

ATC-13 (1985). Earthquake damage evaluation data for California. Applied *Technology Council*, Redwood City.

Brillinger, D. R., and H. K. Preisler (1984). An exploratory analysis of the Joyner–Boore attenuation data, *Bulletin of Seismological Society of America*, **74**, 1441–1450.

Benjamin, J. R. and Cornell, C. A. (1970). Probability, Statistics and Decision for Civil Engineers, McGraw-Hill, Inc., New York.

Federal Emergency Management Agency (2003). Multi-hazard loss estimation methodology, Earthquake model, *HAZUS-MH MR3-Technical Manual*, Washington, DC.

Campbell, K. W. and Bozorgnia, Y (2008). NGA Ground Motion Model for the Geometric Mean Horizontal Component of PGA, PGV, PGD and 5% Damped Linear Elastic Response Spectra for Periods Ranging from 0.01 to10s. *Earthquake Spectra*. **24(139)**, 139-171.

Federal Emergency Management Agency (2003). Multi-hazard loss estimation methodology, Earthquake model, *HAZUS-MH MR3-Technical Manual*, Washington, DC.

Goovaerts, P (1997). Geostatistics for Natural Resources Evaluation. Oxford University Press: Oxford, NY.

Jayaram, N., and Baker, J.W. (2011). Considering spatial correlation in mixed-effects regression, and impact on ground-motion models. *Bulletin of the Seismological Society of America*, **100(6)**, 3295-3303.

Jayaram, N., and Baker, J.W. (2009). Correlation model for spatially distributed round-motion intensities. *Earthquake Engng Struct. Dyn*. **38**:1687–1708

Kendall and Stuart (1977). The advanced theory of statistics, Vol. 1. Charles Griffin, London.

Insurance Information Institute. The Ten Most Costly Catastrophes, United States. (http://www.iii.org/facts_statistics/catastrophes-us.html).

ShakeMap (1994). Shakemap scNorthridge. United States Geological Survey (http://earthquake.usgs.gov/earthquakes/shakemap/sc/shake/Northridge/).