INCORPORATING PARAMETER UNCERTAINTY INTO ATTENUATION RELATIONSHIPS

Robb Eric S. Moss[1] and Armen Der Kiureghian[2]

ABSTRACT

Strong ground motion attenuation relationships estimate the mean and variance of ground shaking as it decreases with distance from an earthquake source. Current relationships use "classical" regression techniques that treat the input variables or parameters as exact, neglecting the uncertainties associated with the measurement of ground acceleration, moment magnitude, site-to-source distance, shear wave velocity, etc. This leads to a poorly constrained estimate of the uncertainty of strong ground motions. This paper discusses the work in progress on; a) estimating the statistics of parameter uncertainty, and b) incorporating the parameter uncertainty into the regression of strong motion attenuation data using a Bayesian framework. The results are an improved understanding of the uncertainties inherent in the phenomena of strong ground motion attenuation, a reduced and better defined model variance, and better constrained estimates of rarer events associated with ground accelerations towards the tail of the distribution.

Introduction

This paper describes ongoing research into measurement error related to strong ground motion parameters and estimated variance related to strong ground motion attenuation predictions. The current statistical method for developing an attenuation relationship is univariate regression on a database using a fixed-effects or random-effects model (e.g., Boore et al., 1997; Abrahamson & Silva, 1997; Campbell & Bozorgnia, 2003). This methodology assumes that the input parameters are exact.

There exists, however, measurement error in the input parameters. For instance, the moment magnitude of a particular seismic event is calculated using a non-unique inversion process resulting in an unspecified amount of uncertainty. This can be seen in the differences in reported moment magnitudes by seismology labs such as USGS and Harvard (Moss, 2003). Differences in inversion techniques used over time have also led to uncertainty in the moment magnitude (Kagan, 2002).

The geometric mean of the peak ground acceleration is an input parameter that considers both horizontal directions of ground shaking. This parameter has measurement error that is a function of the orientation of the strong motion seismometer in relation to the geometry of the fault

[1] Asst. Prof., Dept. Civil and Environmental Engr., California Polytechnic State University, San Luis Obispo, CA
[2] Prof., Dept. Civil and Environmental Engr., University of California Berkeley, CA

rupture. The horizontal motions can be numerically rotated providing statistical estimates of the median and standard deviation as a function of azimuth.

Measurement errors also exist in the input parameters that define site-to-source distance, site class as measured by shear wave velocity in the upper 30 meters, and other input parameters.

Treating these input parameters as inexact instead of exact leads to a better understanding of the sources of uncertainty that propagate through the regression analysis. This also results in reduced overall model variance. A Bayesian framework allows for the treatment of input parameters as inexact, and provides the mathematical flexibility to use any type of functional model form (Der Kiureghian, 1999; Gardoni et al, 2002; Moss et al., 2003).

For this study a Bayesian regression methodology has been formulated for estimating strong motion attenuation using existing published attenuation equations. This paper uses Boore et al. (1997) for a feasibility study, comparing regression results with and without measurement error in the input parameters. Future goals of this research include; collecting more statistical data on the input parameters and evaluating state-of-the-art attenuation equations using a single database to measure relative performance.

Quantifying Measurement Error

The first step in including measurement error (i.e., parameter uncertainty) into a predictive model, in this case a strong motion attenuation relationship, is to evaluate and quantify that uncertainty. There are two forms of uncertainty, epistemic and aleatory uncertainty. Aleatory uncertainty is the inherent randomness that is a function of the phenomena that the model strives to predict. Aleatory uncertainty is inherent in nature and cannot be influenced by the observer or the manner of observation. This type of uncertainty cannot be reduced. Epistemic uncertainty is a function of our lack of knowledge, incomplete description of the phenomena in the model, measurement errors, and/or lack of sufficient measurements to fully capture the phenomena. Epistemic uncertainty is reducible. This study aims at reducing the epistemic uncertainty in attenuation relationships by incorporating the uncertainty in the input parameters, the measurement error, into the regression analysis.

Peak Ground Acceleration

Uncertainty in the ground acceleration can be observed in the variability of the peak values in orthogonal directions. This is a property of the orientation of the rupture plane, complexity of the rupture plane, nature and geometry of the rupture, travel path complexities, surface topography, and other site effects. Ground acceleration measurements also contain uncertainty that is a function of the orientation of the strong motion seismometer in relation to the geometry of the fault rupture.

To capture the uncertainty in strong ground motions, acceleration time histories for different events were evaluated. The motions were rotated through a sweep of 90 degrees, using the method described by Penzien & Watabe (1975), and the geometric mean, median, and coefficient of variation were measured. A 90 degree rotation of orthogonal motions provides a full sweep of the recorded motion in the horizontal direction.

Figure 1 shows three plots of the processed Hayward Bart Station recording from the Loma Prieta earthquake. The first shows a 90 degree sweep of the peak acceleration in both horizontal directions, the average, the geometric mean, the median geometric mean, and the minimum covariance angle. The second plot shows the frequency histogram of the normalized geometric mean, that is the geometric mean divided by the median of the geometric mean. The third plot shows the cumulative frequency distribution of the normalized geometric mean.

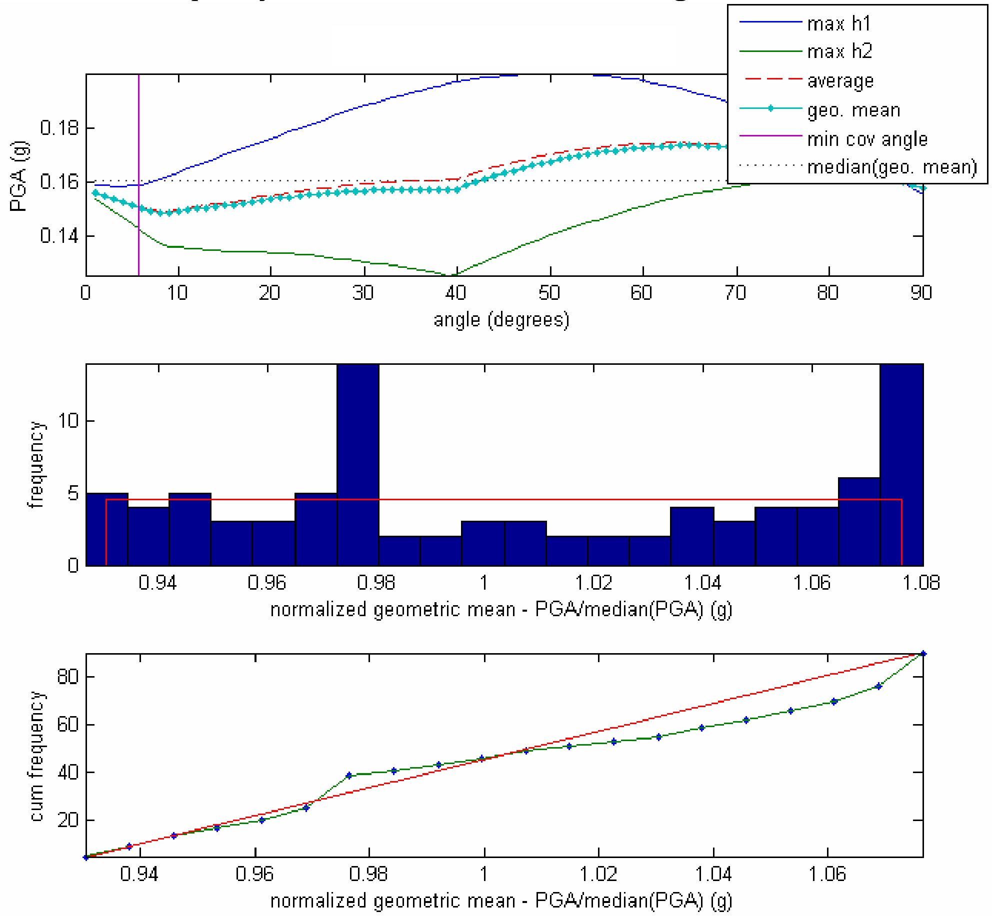

Figure 1. Statistical results of Loma Prieta – Hayward Bart Station motion, typical of the motions evaluated to date.

It can be seen in the frequency histogram that the randomness of the recorded motion throughout the rotated angles does not follow a common theoretical probability distributions (e.g., normal or gaussian distribution). This result is typical of motions evaluated in this study so far. Plotted for comparison is the uniform frequency distribution and the uniform cumulative distribution, respectively.

For this preliminary analysis an average sample median and standard deviation was calculated from the motions evaluated so far. This was used as the estimated measurement error in the subsequent regression analysis.

Moment Magnitude

The uncertainty of the moment magnitude can be attributed mainly to the inversion process used to calculate the seismic moment, and thus the moment magnitude. Moment magnitude is reported by seismology laboratories following an event, and iterated on for a week or two until the final revised value is reported. Calculating the moment magnitude involves an inverse problem to determine the seismic moment. The uncertainty in these calculations comes from the non-uniqueness of the inversion process.

Uncertainty in moment magnitude has also been shown to be a function of time. Kagan (2002) has estimated the standard deviation of the moment magnitude as a function of the inversion technique used to calculate the seismic moment. The accuracy and compatibility of different inversion techniques has improved over time, thereby providing a reduced standard deviation as we approach the present.

Uncertainty in the moment magnitude was quantified for the NGA (Next Generation Attenuation) project funded by PEER (Pacific Earthquake Engineering Research). The standard deviation of moment magnitude was estimated from multiple reported magnitudes for each event where they existed. The standard deviation reported in the NGA dataset was based on the consideration of statistical standard deviation, time, and quality of the data and method used to derive magnitude (Chiou, 2005).

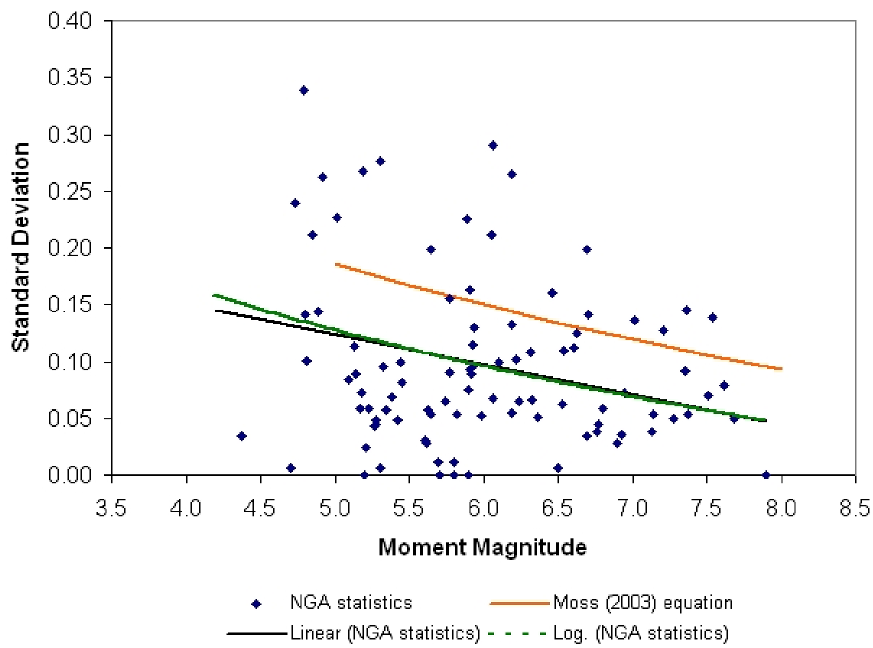

Figure 2. Standard deviation of moment magnitude. Curves are regressed on the standard deviations reported in the NGA dataset.

Figure 2 shows magnitude versus standard deviation as reported in the NGA dataset. There is a large amount of scatter in the data, but a decrease in uncertainty with an increase in magnitude can be observed. This trend was conjectured by Moss (2003) based on the logic that for the inversion of seismic moment the dimensions of the fault plane and the amount of slip associated with larger magnitude events tend to be easier to define than with smaller magnitude events.

Uncertainty also stems from different inversion techniques used: partial or complete waveforms, regional or teleseismic recordings, and different Green's functions. Bigger magnitude events also have more stations recording the event (bigger sample size), generally have a higher signal to noise ratio, and different seismology labs may be using some of the same stations resulting in correlated results.

Shown in Figure 2 are a linear regression line, logarithmic regression line, and the equation from Moss (2003). All three curves exhibit a similar slope, although the intercepts of the regression lines are lower. For this preliminary analysis the logarithmic regression line was used to estimate the uncertainty associated with moment magnitude for the subsequent regression analysis.

Other Parameters

The measurement errors associated with other input parameters have not been evaluated yet, as we are still in the preliminary stages of this research. In particular, uncertainty in the site class as measured by V_{S30} (the shear wave velocity in the upper 30 meters) appears to have some impact on the model variance. Also, measurement errors associated with the site-to-source distance, and the rake angle of the rupture plane may be quantified for future analyses. For acceleration, not just the peak acceleration but spectral acceleration values throughout the frequency range need to be evaluated. These are topics that will be covered in subsequent stages of this research.

Regression Analysis

Predicting strong ground attenuation uses a univariate-type model. It is univariate because only one quantity of interest is to be predicted from a set of measurable variables $x=(x_1,x_2,\ldots x_n)$. The quantity of interest in this case is the spectral acceleration. The general univariate model can be written as,

$$Z = Z(x,\Theta) \tag{1}$$

where Θ denotes a set of model parameters used to fit the model to the observed data. In this study various models, based on attenuation relationships proposed previously in the literature, will be used. The generalized univariate model can then be written as,

$$Z(x,\Theta) = \hat{z}(x,\Theta) + \varepsilon \tag{2}$$

where $\hat{z}(x,\Theta)$ is the selected attenuation relationship and ε is a random normal variate with zero mean and unknown standard deviation that is the model error term. Aleatory uncertainty is found in the measured variables x and partly in the error term ε. Epistemic uncertainty is found in the model parameters Θ and partly in the model error term ε.

Model Uncertainty

In this model formulation the error term ε captures the imperfect fit of the model to the measurements. The imperfect fit may be due to inexact model form or due to missing variables. The missing variables can be considered inherently random and that portion of the model error term is aleatory uncertainty. The portion of the model error term that is from the inexact model form is epistemic uncertainty.

Measurement Error

Measurement error tends to comprise a large portion of the epistemic uncertainty in geoscience problems. This uncertainty comes from imprecise measurement of the variables $x=(x_1,x_2,\ldots x_n)$. These measurement errors are treated as statistically independent normally distributed random variables with zero mean (assuming unbiased measurement errors) and measurable standard deviation. The errors are incorporated as $x_i = \hat{x}_i + e_{xi}$ where $\hat{x}_{i\,xi}$ is the measured value and e_{xi} is the measurement error.

Statistical Uncertainty

The size of the sample *n* will influence the accuracy of the model parameters Θ. The larger the sample size the less epistemic uncertainty introduced into the model parameters. In this case, there is a limited amount of strong motion recordings for model fitting.

Parameter Estimation through Bayesian Updating

A Bayesian framework is used to estimate the unknown model parameters (i.e., regression). The Bayesian approach is useful because it incorporates all forms of uncertainty related to the problem of strong ground motion attenuation into the regression analysis.

Bayes rule is derived from simple rules of conditional probability, yet the simplicity portends little of the power of the Bayesian technique. Bayes rule can be written as (Box & Tiao, 1992),

$$f(\Theta) = c \cdot L(\Theta) \cdot p(\Theta) \tag{3}$$

where; $f(\Theta)$ is the posterior distribution representing the updated state of knowledge about Θ, $L(\Theta)$ is the likelihood function containing the information gained from the observations of x, $p(\Theta)$ is the prior distribution containing our *apriori* knowledge about Θ, and $c = [\int L(\Theta) \cdot p(\Theta) \cdot d(\Theta)]^{-1}$ is the normalizing constant.

The likelihood function is proportional to the conditional probability of the observed events, given the value of Θ. The likelihood function incorporates the objective information that, in this case, are the measurements associated with strong ground motion attenuation. The prior distribution can include subjective information known about the distributions of Θ. The posterior distribution incorporates both the objective and subjective information into the distributions of the model parameters. The process of performing Bayesian updating involves formulating the likelihood function, selecting a prior, calculating the normalizing constant, and then calculating the posterior statistics.

The prior distribution tends to be the most controversial issue for detractors of Bayesian methods. Box & Tiao (1992) have shown that the use of a non-informative prior can lead to an unbiased, data-driven estimate of the model parameters. A non-informative prior allows the data, through the likelihood function, to dominate the posterior distribution, thereby minimizing the role of the subjective information. A non-informative prior, by definition, has no effect on the shape of the posterior distribution and is used when no prior information about the parameters is available. Gardoni *et al.*, (2002) have shown that for a univariate model where the unknown parameters Θ are the coefficients in a linear expression and the standard deviations σ

of ε, the noniformative prior simplifies to ,

$$p(\sigma) \propto \frac{1}{\sigma} \tag{4}$$

The mean vector M_Θ and covariance matrix $\Sigma_{\Theta\Theta}$ can be calculated from the posterior distribution of Θ. Computation of these statistics and the normalizing constant is non-trivial, requiring multifold integration over the Bayesian kernel. Importance sampling, a sampling algorithm as described in Gardoni (2002), was used to efficiently perform these calculations.

Likelihood Function

As defined above the likelihood function is proportional to the conditional probability of observing a particular event given a value of Θ. In order to formulate the likelihood function a limit-state must be defined to provide a threshold for defining the probability of observation.

For this feasibility study the attenuation relationship from Boore et al. (1997), is used as a basis for the likelihood function. Boore et al. (1997) was chosen because the database used in the regression was provided in the paper. The function form of this attenuation relationship is,

$$\log(Y) = \theta_1 + \theta_2(M_w - 6) + \theta_3(M_w - 6)^2 - \theta_4 \ln(\sqrt{R_{jb}^2 + \theta_5^2}) - \theta_6 \ln(V_s / \theta_7) \tag{5}$$

where Y represents the spectral acceleration value, M_w is the moment magnitude, R_{jb} is the Joyner-Boore distance, V_S is the shear wave velocity in the upper 30 meters, and the θ's are the model parameters. Boore et al., (1997) determined the parameters of this model by using “classical” regression with a two step procedure.

To present this attenuation relationship as a limit-state function, the equation is rearranged to describe the most likely location of a threshold given a value of Θ. This limit-state would be where the threshold lies at the zero mean of the error term at a value of Z_i for a given x_i. This thereby minimizes the error on either side of the threshold at that point. From Equation 2, $Z_i = \hat{z}(x_i, \theta) + \varepsilon_i$ or $\varepsilon_i = g_i(\theta)$ where $g_i(\theta) = Z_i - \hat{z}(x_i, \theta)$ and ε_i is the model error term at the i^{th} observation. The attenuation relationship of Campbell et al., (1997), shown in Equation 5, then becomes,

$$g(\Theta) = \log(Y) - [\theta_1 + \theta_2(M_w - 6) + \theta_3(M_w - 6)^2 - \theta_4 \ln(\sqrt{R_{jb}^2 + \theta_5^2}) - \theta_6 \ln(V_s / \theta_7)] \tag{6}$$

The likelihood function for the problem of strong ground motion attenuation is the product of the probabilities of observing n values with the limit-state co-located with the zero mean of the error term. Given exact measurements and statically independent observations, the likelihood can be written as,

$$L(\theta, \sigma_\varepsilon) \propto P\left[\bigcap_{i=1}^{n} \{g_i(\theta) = \varepsilon_i\}\right] \tag{7}$$

where σ_ε is the standard deviation of the error term ε. Given that ε is a standard normal variate, Equation 7 can be written as,

$$L(\theta, \sigma_\varepsilon) \propto \prod_{i=1}^{n} \left\{\frac{1}{\sigma_\varepsilon} \varphi\left[\frac{g_i(\theta)}{\sigma_\varepsilon}\right]\right\} \tag{8}$$

where φ is the standard normal distribution function. When measurement errors are considered the likelihood function becomes,

$$L(\theta,\sigma_{\varepsilon}) \propto \prod_{i=1}^{n}\left\{\frac{1}{\hat{\sigma}_{\varepsilon}(\theta,\sigma_{\varepsilon})}\varphi\left[\frac{\hat{g}_{i}(\theta)}{\hat{\sigma}_{\varepsilon}(\theta,\sigma_{\varepsilon})}\right]\right\} \tag{9}$$

The above formulation was used to estimate the statistics of the model parameters, Θ, and the model error, ε, for the given functional form of the attenuation relationship and the given database. These estimated terms are analogous to the coefficients solved for using "classical" regression in Boore et al., (1997). The means and standard deviations of the coefficients are used to define the predictive model. The model variance is found using a second order Taylor series expansion about the mean point.

Feasibility Study Results

The results of the feasibility study, using the functional form of the attenuation relationship and the database from Boore et al., (1997), are shown in Figure 3. This figure shows a comparison of Boore et al., versus preliminary results from this study.

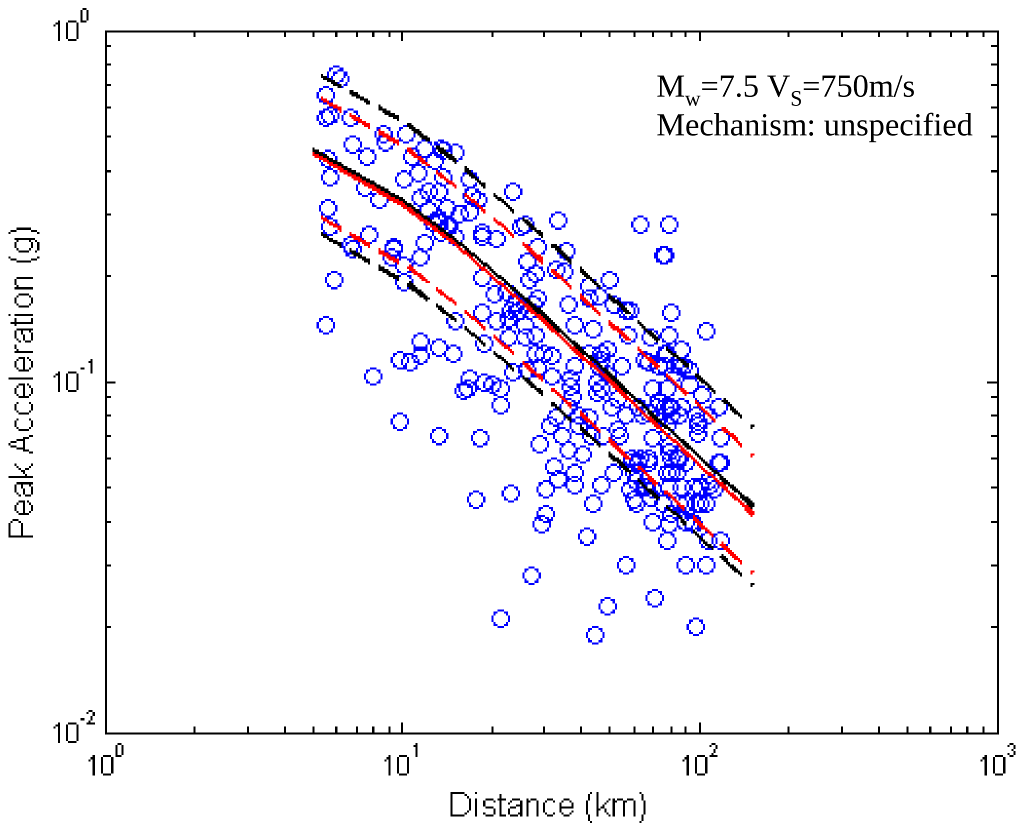

Figure 3. Comparison plot of attenuation relationship estimated using "classical" regression with exact parameters versus Bayesian regression that incorporates parameter uncertainty. The black curves are from Boore et al. (1997), the red curves from this study. Plus/minus one standard deviation curves are shown as dashed lines.

There is a slight difference in the limit-states or mean regression curves found using "classic" versus Bayesian regression. This is due to the influence of including inexact parameters in the Bayesian regression analysis. More important is the reduced model standard deviation found using Bayesian regression. The standard deviation is reduced because the parameter uncertainty, or measurement error of the input parameters, is quantified and incorporated in the analysis. By including the additional information of parameter uncertainty we achieve an improved (i.e., reduced) estimate of the model standard deviation.

For this preliminary study, Bayesian regression was initially performed without parameter uncertainty to confirm that a similar standard deviation was calculated as Boore et al. Then the Bayesian regression was performed including parameter uncertainty, with a coefficient of variation (standard deviation divided by the mean) of ~0.10 for moment magnitude and ~0.30 for peak acceleration. As shown in Figure 3 the total model standard deviation (the square root of the model variance) of the natural log of peak ground acceleration, σ_{lnY}, is 0.386, compared to 0.520 from Boore et al. The earthquake to earthquake component of the standard deviation, σ_e, is the same for the two studies.

Summary

Presented here is a method to incorporate parameter uncertainty, or the measurement error associated with the input parameters, into strong ground motion attenuation relationships. This method uses Bayesian regression for incorporating inexact parameters into the regression analysis. A feasibility study was carried out using the functional form of the attenuation relationship and the database from Boore et al., (1997). The results of this feasibility study demonstrate that a reduced or better-constrained model variance is one benefit of the presented methodology. As part of this ongoing research study; further analysis of the statistics of strong ground motion attenuation, exploration of other benefits of using a Bayesian approach such as model optimization and correlation analysis, and analysis using other attenuation models will be carried out in the future.

Acknowledgements

Components of this study are supported by the PEER (Pacific Earthquake Engineering Research center) Lifelines Program in conjunction with the NGA (next generation attenuation) study. This support is gratefully acknowledged. Thanks to Dr. Alison Faris for her assistance in trouble-shooting the Bayesian regression algorithm.

References

Ang, A.H-S. & Tang, W. H. (1975) **Probability Concepts in Engineering Planning and Design, Vol. 1.** John Wiley and Sons, Inc.

Abrahamson, N.A., & Silva, W.J. (1997). "Empirical Response Spectral Attenuation Relations for Shallow Crustal Earthquakes." Seismological Research Letters, Vol. 68, No. 1.

Boore, D.M., Joyner, W.B., and Fumal, T.E. (1997). "Equations for Estimating Horizontal Response Spectra and Peak Acceleration from Western North American Earthquakes: A Summary of Recent Work." Seismological Research Letters, Vol. 68, No. 1.

Box, G.E.P. & Tao, G.C. (1992) **Bayesian Inference in Statistical Analysis.** John Wiley and Sons, Inc.

Campbell, K.W., and Bozorgnia, Y. (2003) "Updated Near-Source Ground-Motion (Attenuation) Relations for the Horizontal and Vertical Components of Peak Ground Acceleration and Acceleration Response Spectra." Bulletin of Seismological Society of America, Vol. 93, No. 1.

Chiou, Brian (2005). Personal Communication, Aug.

Der Kiureghian, A. (2000) "A Bayesian framework for fragility assessment." **Applications of Statistic and Probability-Civil Engineering Reliability and Risk Analysis,** eds. R.E. Melchers, and M.G. Stewart, Balkema, Rotterdam, pp. 1003-11010.

Gardoni, P., Der Kiureghian, A., and Mosalam, K.M. (2002). "Probabilistic Capacity Models and Fragility Estimates for Reinforced Concrete Columns based on Experimental Observations." Journal of Engineering Mechanics, October.

Moss, R.E.S. (2003) CPT-Based Probabilistic Assessment Of Seismic Soil Liquefaction Initiation. Ph.D. Dissertation, University of California Berkeley.

Moss, R. E. S., Cetin, K. O., and Seed, R. B. (2003) "Seismic Liquefaction Triggering Correlations within a Bayesian Framework ." 9^{th} International Conf. Application of Statistic and Probability in Civil Engineering.

Penzien, J., & Watabe, M. (1975). "Characteristics of 3-Dimensional Earthquake Ground Motions." Earthquake Engineering and Structural Dynamics, Vol. 3.