Investigation of the Evidence of Inhibition of Very Strong Ground Motion

L. Al Atik
Linda Alatik Consulting, San Francisco, CA 94110

N. Abrahamson
University of California, Berkeley, CA 94720

15 WCEE
LISBOA 2012

SUMMARY:
The current state of practice of probabilistic seismic hazard analysis (PSHA) uses lognormal distribution to model ground motion variability. At low annual probabilities of exceedance, unbounded lognormal distribution leads to very high ground motion values. A common practice in PSHA has been to truncate the upper tail of ground motion variability at 2 to 3 standard deviations above the median without any technical basis for this truncation. Other researchers used empirical ground motion data to examine the inhibition of very large ground motions using total residuals and the assumption that they are independent which ignores the correlation of the total residuals of a single earthquake. In this paper, we use the NGA West-1 and preliminary NGA West-2 datasets to statistically investigate the evidence of inhibition of very strong ground motion. Our results indicate that the lognormal distribution is appropriate for estimating the occurrence of very large peak ground accelerations (PGA).

Keywords: Engineering seismology, Ground motion, NGA West, Probabilistic seismic hazard

1. INTRODUCTION

Ground motion variability is commonly modelled with a lognormal distribution. The lognormal distribution is unbounded; that is, the probability of exceeding any ground motion level is always nonzero regardless of the intensity of the ground motion. At low annual probabilities of exceedance (large return periods), PSHA results are controlled by the upper tail of the ground motion distribution. This can lead to large ground motion values at sites of critical facilities such as nuclear power plants that require the use of such low annual probabilities of exceedance

In recent years, researchers and practitioners have attempted to address the issue of unrealistically high ground motion values at low annual probabilities of exceedance. A common practice in PSHA has been to truncate the upper tail of the ground motion variability at 2 to 3 standard deviations above the median without any technical basis for this truncation. Strasser *et al.* (2008) studied empirical strong ground motion data and showed that there is no statistical reason for truncating the lognormal distribution. Rhoades *et al.* (2008) used empirical ground motion data from Japan and New Zealand to examine the inhibition of very large ground motions. They compared the actual numbers of exceedance of given accelerations to those predicted by empirical ground motion models. Ignoring the correlation of the total residuals of a single earthquake in their approach, Rhoades *et al.* (2008) concluded that there is statistical evidence of inhibition of very large ground motions. Abrahamson and Wooddell (2010) modified the Rhoades *et al.* (2008) approach to properly account for the correlation of residuals and applied it to the Abrahamson and Silva (2008) (AS08) NGA West-1 dataset. They concluded that there is no evidence of inhibition of very strong ground motions in the AS08 model.

Huyse *et al.* (2010) studied the tail of the distribution of recorded PGA from 2 Chi-Chi aftershocks and of total residuals of PGA of the AS08 model to evaluate the validity of the lognormal assumption in modelling low probability ground motions. Huyse *et al.* (2010) also ignored the correlation of residuals of a single earthquake and concluded that the generalized Pareto distribution (GPD) is a more appropriate model for the distribution of the upper tail of ground motion variability. They recommended the use of a composite distribution model in PSHA, lognormal up to a certain threshold ground motion and GPD beyond the threshold, to limit the ground motion at low annual probabilities of exceedance. The application of the peak-over-threshold analysis in Huyse *et al.* (2010) and the Rhoades *et al.* (2008) approach are based on the fundamental assumption that the total residuals are independent and identically distributed. This assumption is not appropriate for most ground motion datasets because the total residuals from multiple recordings of a single earthquake are correlated through the event term.

In this paper, we demonstrate the correlation of total residuals of ground motion models. We also investigate the appropriateness of the GPD for modelling the upper tail of the ground motion distribution. For this purpose, we use the AS08 NGA West-1 dataset and model as well as the preliminary subset of the NGA West-2 dataset selected by Abrahamson and Silva (AS) for the update of the AS08 model. We apply the peak-over-threshold analysis to the within-event residuals of the data to obtain the parameters of the GPD fit to the upper-tail of the residuals. We then compare the actual number of times given peak ground accelerations are exceeded in the datasets to predicted numbers using the composite distribution model suggested by Huyes *et al.* (2010) and using the lognormal distribution. Correlations of total ground motion residuals are properly addressed in our analysis.

2. CORRELATION OF TOTAL GROUND MOTION RESIDUALS

The AS08 ground motion prediction equation (GMPE) incorporates the effects of soil nonlinearity on the median and the standard deviations of the model and has the following statistical form:

$$y_{es} = f(\vec{X}_{es}, \vec{\theta}) + \delta W_{es} + \delta B_{es}, \tag{2.1}$$

where $f\left(\overrightarrow{X}_{es}, \vec{\theta}\right)$ is the median ground motion model, $\overrightarrow{X}_{es}$ is the vector of independent parameters for the recording at station *s* from earthquake *e*, $\vec{\theta}$ is a vector of the model coefficients determined by the regression. δW_{es} and δB_{es} are the within-event and between-event residuals with standard deviations ϕ_{es} and τ_{es}, respectively. Subscripts *e* and *s* in the between-event and within-event standard deviations notations refer to magnitude dependence and soil nonlinearity effects, respectively. Eqn. 2.1 can be written in terms of the normalized within-event and between-event residuals:

$$y_{es} = f(\vec{X}_{es}, \vec{\theta}) + \varepsilon_{es}^{W} \phi_{es} + \varepsilon_{e}^{B} \tau_{es}. \tag{2.2}$$

In this formulation, the normalized between-event residual, ε_e^B, is constant at all sites that recorded the same earthquake. We can also define the observed between-event residual for low levels of outcrop rock motion (linear site response) as:

$$\delta B_{e0} = \varepsilon_e^B \tau_0, \tag{2.3}$$

where τ_0 is the between-event standard deviation of the observed ground motion for linear site conditions and is constant. For a site located on soil underlain by rock, the site-specific between-event residual is, therefore, related to the linear site conditions between-event residual by:

$$\delta B_{es} = \delta B_{e0} \frac{\tau_{es}}{\tau_0}. \tag{2.4}$$

The total ground motion residual, δ_{es}, can be written as:

$$\delta_{es} = \delta W_{es} + \delta B_e = \delta W_{es} + \frac{\delta B_{e0}}{\tau_0}\tau_{es}. \tag{2.5}$$

Eqn. 2.5 shows that total residuals of multiple recordings of a single earthquake, *e*, are not independent because they are correlated through the common normalized between-event residual, $\frac{\delta B_{e0}}{\tau_0}$. This correlation of total ground motion residuals has been repeatedly ignored in analyses and leads to erroneous results. Within-event and between-event residuals are independent but total ground motion residuals are not.

3. ANALYSIS APPROACH

Peak ground acceleration (PGA) residuals of the AS08 NGA West-1 dataset with respect to the AS08 GMPE were used in this analysis. Residuals at a spectral period of 0.01 seconds of the AS selected subset of the preliminary NGA West-2 dataset with respect to an updated version of the AS08 model were also used herein. Both datasets consist of recorded ground motion from shallow crustal earthquakes (mainshocks and aftershocks) in active tectonic regions.

Given that the within-event residuals are independent and identically distributed, we apply the peak-over-threshold analysis suggested in Huyse *et al.* (2010) to the within-event residuals to evaluate the suitability of the lognormal and the GPD distributions in modelling the behaviour of the upper tail of ground motion distribution. In the first step of the peak-over-threshold analysis, the threshold, λ, that marks the start of the tail portion of the within-event residuals is estimated using the plot of the mean conditional excess function, $E\left(\delta W - \lambda \mid \delta W > \lambda\right)$, versus the threshold level. The mean conditional excess function is the sum of excesses over the threshold level divided by the number of cases that exceed this threshold. If the conditional mean excess function is a linear function of the threshold level, then the empirical data follow a GPD. The threshold, λ, is selected as the start of the last linear segment of the mean conditional excess plot. The shape and scale parameters of the GPD, *c* and δ, associated with the threshold level, λ, are estimated by fitting a GPD to the within-event residuals above λ. We note that the analysis of the within-event residuals distribution tail faces challenges such as the scarcity of extreme data and choosing the threshold or beginning of the tail. These aspects will be discussed in the next section.

The next step of the analysis involves comparing the actual number of exceedances of observed ground motion in the dataset to predicted exceedances assuming lognormal distribution and composite distribution model. The composite distribution model follows the lognormal distribution before the threshold and the GPD in the tail region. Based on this model, the probability that a given earthquake with magnitude *m* generates a PGA at a distance *r* exceeding a particular value a_0 is given in Huyse *et al.* (2010) as:

$$P(Y \geq \ln(a_0) \mid m, r) = \begin{cases} 1-\left(1-p_{tail}\right)\dfrac{\Phi(Z)}{\Phi(z_\lambda)} & \ln(\mathrm{a}_0) - \mu \;\leq\; \lambda \\ p_{tail}\left[1-F_{GPD}(z_0)\right] & \ln(\mathrm{a}_0) - \mu > \lambda \end{cases}, \tag{3.1}$$

where $y = ln\left(PGA\right)$ is a random variable with mean μ and total standard deviation σ_T. $Z = \dfrac{\ln\left(a_0\right) - \mu}{\sigma_T}$ is a standard normal random variable, $z_\lambda = \dfrac{\lambda}{\sigma_T}$ corresponds to the threshold residual, $z_0 = \ln(a_0) - \mu$, p_{tail} is the fraction of the recorded data in the tail region, ϕ is the cumulative distribution function (cdf) of the standard normal distribution and $F_{GPD}(z_0)$ is the cdf of the GPD written as:

$$F_{GPD}(z_0) = 1 - \left[1 + c\frac{z_0 - \lambda}{\delta}\right]^{-1/c}. \tag{3.1}$$

As proposed in Abrahamson and Wooddell (2010), the correlation of the total residuals is addressed by treating the between-event residuals as known and using the variability of the within-event residuals in computing the expected probabilities of exceedance of specific ground motions. Since the majority of strong motion recordings are not at short distance in the datasets used, the between-event residual on rock, δB_{e0}, should not be impacted by the inhibition of very strong ground motion. The inhibition of very large ground motions would primarily affect the within-event residuals. The ratio of observed to expected number of recordings above a given ground motion threshold can be written as:

$$Ratio(a_0) = \frac{\sum_{e=1}^{nEqk} \sum_{s=1}^{n\mathrm{Rec}_e} I\left(y_{es} > \ln(a_0)\right)}{\sum_{e=1}^{nEqk} \sum_{s=1}^{n\mathrm{Rec}_e} P\left(y_{es} > \ln(a_0) \mid \mu_{es}, \delta B_{es}, \phi_{es}\right)}, \tag{3.3}$$

where *nEqk* is the number of earthquakes in the selected subset of the NGA West-2 dataset, $nRec_e$ is the number of recordings for earthquake *e*, y_{es} is the natural logarithm of the ground motion at station *s* from earthquake *e*, and a_0 is the test ground motion level. $I\left(y_{es} > \ln(a_0)\right)$ is a function that selects the ground motion values that exceed a_0; it is equal to 1 if $y_{es} > \ln(a_0)$ and zero otherwise. $P\left(Y_{es} > \ln(a_0) \mid \mu_{es}, \delta B_{es}, \phi_{es}\right)$ is the conditional probability that the ground motion will exceed the test level a_0 given the median ground motion (μ_{es}), between-event residual (δB_e), and within-event standard deviation (ϕ_{es}). Assuming a lognormal distribution of the within-event residuals, the probability term in Eqn. 3.3 can be written as:

$$P\left(Y_{es} > \ln(a_0) \mid \mu_{es}, \delta B_{es}, \phi_{es}\right) = 1 - \Phi\left(\frac{\ln(a_0) - (\mu_{es} + \delta B_{es})}{\phi_{es}}\right). \tag{3.4}$$

For a truncated lognormal distribution of the within-event residuals at δW_0, the probability term is written as:

$$P\left(Y_{es} > \ln(a_0) \mid \mu_{es}, \delta B_{es}, \phi_{es}\right) = 1 - \frac{\Phi\left(\frac{\ln(a_0) - (\mu_{es} + \delta B_{es})}{\phi_{es}}\right)}{\Phi\left(\frac{\delta W_0}{\phi_{es}}\right)}. \tag{3.5}$$

Assuming a composite distribution model, the probability term of Eqn. 3.3 can be written as:

$$P(Y_{es} \geq \ln(a_0) \mid \mu_{es}, \delta B_{es}, \phi_{es}) = \begin{cases} 1 - (1 - p_{tail})\dfrac{\Phi(Z)}{\Phi(z_\lambda)}, & \ln(\mathrm{a}_0) - (\mu_{es} + \delta B_{es}) \leq \lambda \\ p_{tail}\left[1 - F_{GPD}(z_0)\right], & \ln(\mathrm{a}_0) - (\mu_{es} + \delta B_{es}) > \lambda \end{cases}, \tag{3.6}$$

where $Z = \dfrac{\ln(a_0) - (\mu_{es} + \delta B_{es})}{\phi_{es}}$ is a standard normal variable, $z_\lambda = \dfrac{\lambda}{\phi_{es}}$ corresponds to the within-event residual threshold, and $z_0 = \ln(a_0) - (\mu_{es} + \delta B_{es})$.

4. RESULTS

Within-event residuals of the AS08 NGA West-1 dataset and the preliminary AS NGA West-2 dataset were evaluated for evidence of inhibition of very large ground motions using the two-step approach described in the previous section. The mean conditional excess plot of the within-event residuals of PGA for NGA West-1 is presented in Fig. 4.1 and shows that the within-event residual threshold is 1.35 (2.7 times the within-event standard deviation). However, at this threshold level, only 13 points are available to evaluate the tail statistics of the within-event residuals. In order to increase the tail sample size, we also consider the within-event residual threshold at 1.09 and 0.92 corresponding to 2.2 and 1.86 times the within event standard deviation as shown in Fig. 4.2 and 4.3, respectively. For the NGA West-2 dataset, the mean conditional excess plot in Fig. 4.4 shows that the GPD threshold is 1.65 (3 times the within-event standard deviation) with only 10 available data points in the tail region. It is also clear in Fig. 4.4 that the threshold cannot be considered at a smaller within-event residual. The GDP parameters and the percentage of the data in the tail region are summarized in Table 4.1. The within-event residual upper bounds of the GPD fits, calculated as $\lambda - \delta/c$, are also included in Table 4.1.

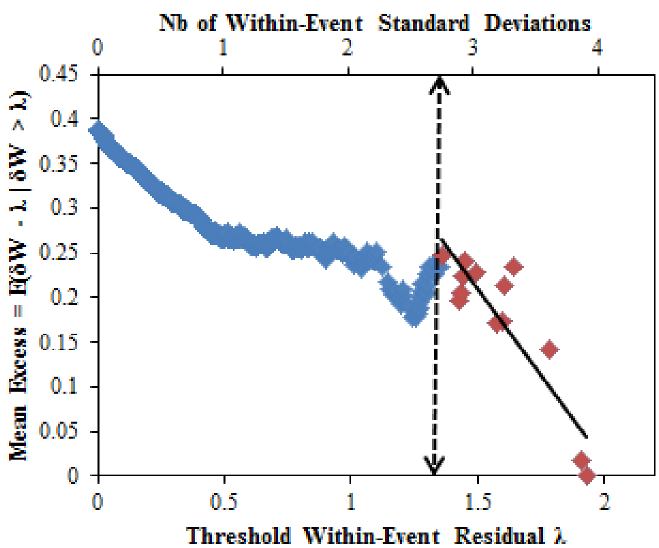

Figure 4.1. Mean conditional excess plot of the within-event residuals of PGA in the AS08 NGA West-1 dataset. A threshold λ = 1.35 (2.7 standard deviations) can be identified as the beginning of the last linear part of the plot (GPD fit 1).

As mentioned in Huyse *et al.* (2010), the threshold selection constitutes a significant challenge in tail statistics. If the threshold is selected too low, a larger number of data points that do not belong to the tail will be included in the peak-over-threshold analysis and will result in a biased estimate of the shape parameter. On the other hand, if the threshold is selected too high, fewer data points will be included in the tail portion of the distribution and will lead to statistically unstable estimates of the shape parameter. As a result, a significant degree of uncertainty is associated with the selected threshold level. Table 4.1 shows that the shape parameter of the second GPD fit of the NGA West-1 dataset is close to zero, which makes the distribution close to being an exponential distribution. Moreover, for the NGA West-2 dataset, only 0.33% of the dataset is in the tail region which makes the estimates on the GPD in the tail region unreliable. Therefore, we only consider GPD fit 3 of the NGA West-1 which has the largest tail portion in the rest of the analysis. Fig. 4.5 shows the probability density function of the composite distribution using NGA West-1 GDP fit 3 compared to the lognormal distribution. Fig. 4.5 indicates that the composite distribution is bounded but has a fatter tail than the lognormal distribution.

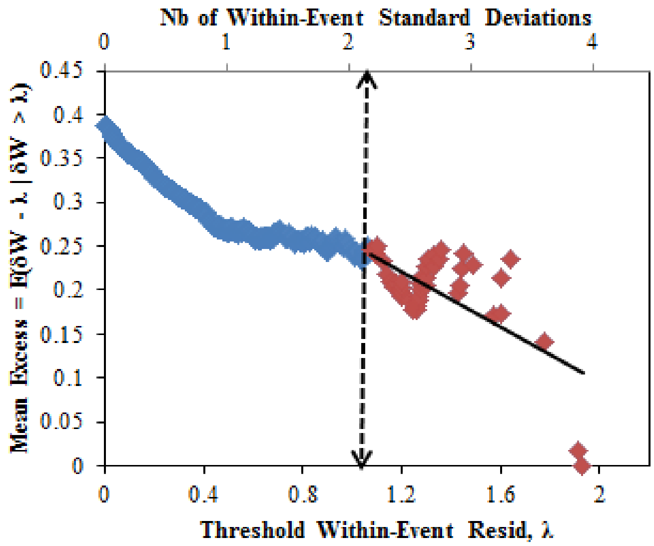

Figure 4.2. Mean conditional excess plot of the within-event residuals of PGA in the AS08 NGA West-1 dataset. A threshold λ = 1.09 (2.2 standard deviations) can be identified as the beginning of the last linear part of the plot (GPD fit 2).

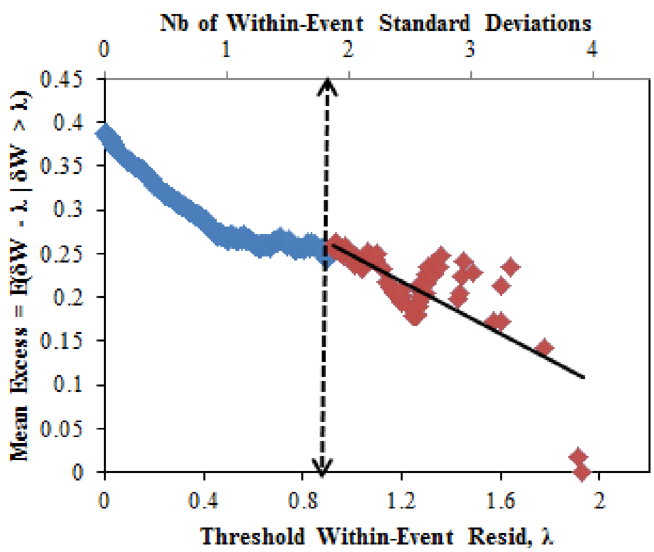

Figure 4.3. Mean conditional excess plot of the within-event residuals of PGA in the AS08 NGA West-1 dataset. A threshold λ = 0.92 (1.86 standard deviations) can be identified as the beginning of the last linear part of the plot (GPD fit 3).

Fig. 4.6 shows a comparison of the GPD fit 3 to the NGA West-1 within-event residuals. The lognormal distribution with parameters calculated from the statistical moments of all the NGA West-1 within-event residuals are also shown in Fig. 4.6. Fig. 4.6 shows that the GPD better fits the tail of the within-event residuals than the lognormal distribution. It is important to note here that large within-event residuals are not necessarily associated with large ground motions. Fig. 4.7 shows the within-event residuals greater than 0.9 versus recorded NGA West-1peak ground accelerations and indicates that a large portion of the within-event residuals greater than 0.9 are associated with PGA less than 0.2g. Therefore, while Fig. 4.6 might suggest that the distribution of the within-event residuals is bounded; this does not necessarily means that very large ground motions are inhibited.

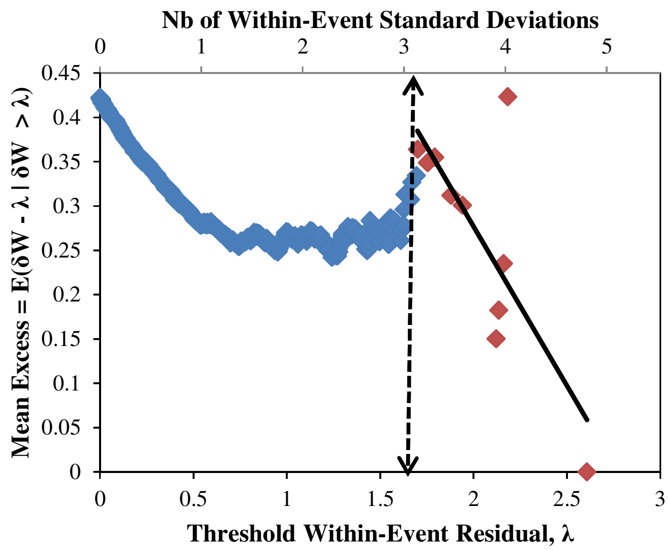

Figure 4.4. Mean conditional excess plot of the within-event residuals at period of 0.01 seconds in the preliminary AS NGA West-2 dataset. A threshold λ = 1.65 (3 standard deviations) can be identified as the beginning of the last linear part of the plot.

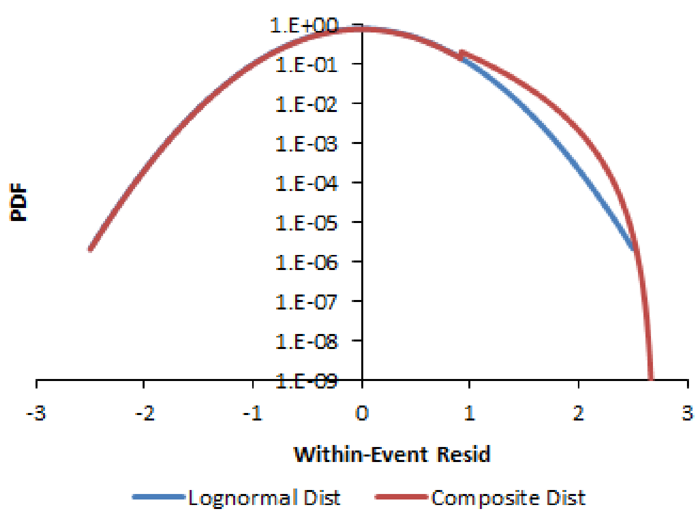

Figure 4.5. Probability density function of the lognormal distribution and the composite distribution model using GPD fit 3

Table 4.1. Tail statistics of the GPD fit to the within-event residuals of AS08 NGA West-1 and AS NGA West-2 datasets

Dataset	GPD Fit	Number of Data Points	% Tail	Threshold	Scale	Shape	Within-event Residual Upper Bound
NGA West-1	1	13	0.94%	1.35	0.28	-0.21	2.70
NGA West-1	2	46	3.34%	1.09	0.27	-0.06	5.52
NGA West-1	3	87	6.32%	0.92	0.30	-0.17	2.70
NGA West-2	1	10	0.33%	1.65	0.55	-0.45	2.87

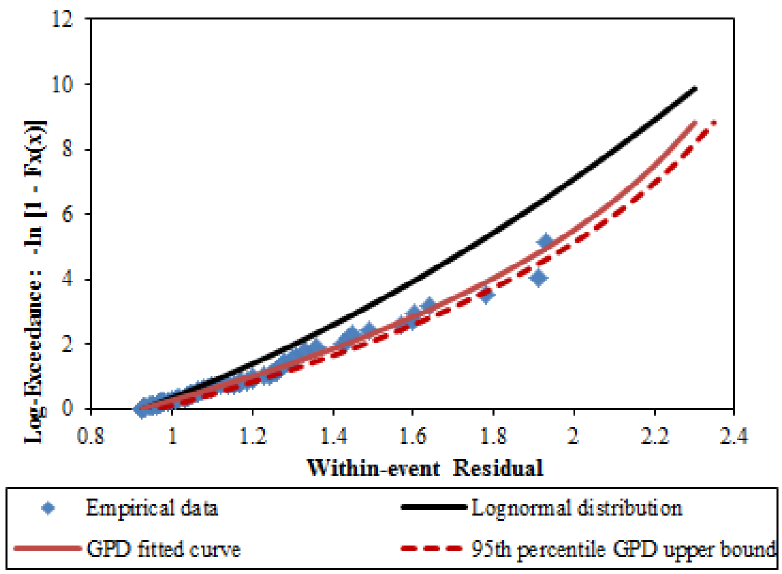

Figure 4.6. Comparison of GPD fit 3 to the NGA West-1 within-event residuals with 95% confidence upper bound to the lognormal distribution.

Next, we apply the modified Rhoades *et al.* (2008) approach as suggested in Abrahamson and Wooddell (2010) to evaluate the appropriateness of the lognormal and composite models in estimating the actual number of exceedances of specific ground motion levels. Fig. 4.8 compares the observed number of exceedances of PGA to the expected numbers of exceedances using the lognormal distribution and composite distribution with GPD fit 3 for the within-event residuals. The blue dashed curves show the 95% confidence interval. Fig. 4.9 shows the ratio of the actual to expected number of exceedances of PGA using the lognormal distribution and the composite distribution with GPD fit 3 for the AS08 model. Fig. 4.8 and 4.9 show that the lognormal distribution leads to a better estimate of the number of exceedances of PGA, whereas the composite distribution over-predicts the number of exceedances at small and large peak ground accelerations. Moreover, no evidence of inhibition of strong ground motions is observed in Fig. 4.8 and 4.9. For PGA greater than 1g, the lognormal distribution underestimates the observed number of exceedances while the composite distribution overestimates the observed number of PGA exceedances.

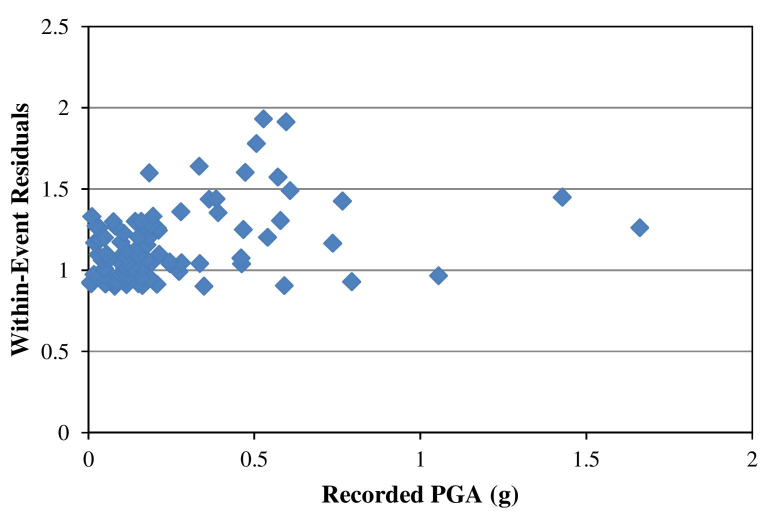

Figure 4.7. Distribution of the within-event residuals greater than 0.9 versus PGA for the AS08 NGA West-1 model

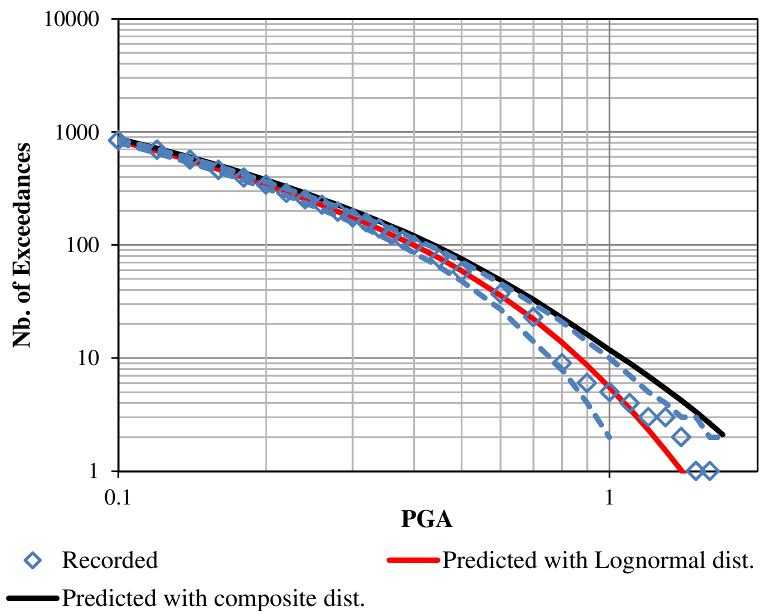

Figure 4.8. Observed and expected number of exceedances of PGA using the lognormal distribution and the composite distribution with GPD fit 3. The 95% range is shown as the blue dashed lines.

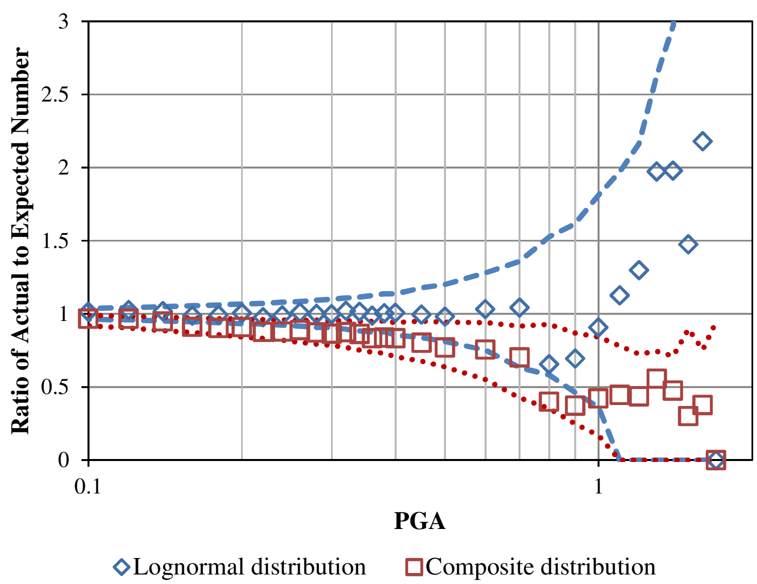

Figure 4.9. Ratio of actual to expected number of exceedances of PGA using the lognormal distribution and the composite distribution with GPD fit 3. The 95% range is shown as the blue dashed lines for the lognormal distribution and as the red dashed lines for the composite distribution.

5. CONCLUSIONS

Total residuals of empirical ground motion models are not independent due to the correlations of recordings from a single earthquake through the between-event residual. For the AS08 model, the effects of this correlation are significant. Analysis of the tail distribution of the residuals without accounting for their correlation is erroneous and often generates misleading results regarding the inhibition of strong ground motions.

Analysing the tail distribution of within-event residuals, we find that a composite ground motion distribution that follows the lognormal distribution up to a certain threshold and then the bounded GPD beyond the threshold is not appropriate for predicting the observed number of exceedances of PGA in the AS08 dataset. While a GPD distribution appears to well fit the upper tail of within-event residuals, this should not imply that large ground motions are inhibited because large within-event residuals do not necessarily correspond to large ground motion values. A lognormal distribution provides better estimates of the actual number of ground motion exceedances. Moreover, we find that there is no statistical evidence of inhibition of strong ground motions in the AS08 model. Physical limits of ground motion were not addressed in this paper. If such limits exist, the AS08 dataset was not sufficient to observe their effects.

ACKNOWLEDGEMENT
This study was sponsored by the Pacific Earthquake Engineering Research Center (PEER) and funded by the California Earthquake Authority, the California Department of Transportation, and the Pacific Gas & Electric Company. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsoring agencies.

REFERENCES

Abrahamson, N. A. and W. J. Silva (2008). Summary of the Abrahamson and Silva NGA ground-motion relations. **Earthquake Spectra** 24: 67-98.

Abrahamson, N. A. and K. Wooddell (2010). Evaluation of evidence for inhibition of very strong ground motions in the Abrahamson and Silva Next Generation Attenuation ground-motion model. Bull. Seism. Soc. Am. 100: 2174-2184.

Huyse, L., R. Chen, and J. A. Stamatakos (2010). Application of Generalized Pareto Distribtuion to constrain uncertainty in peak ground accelerations. **Bull. Seism. Soc. Am**. 100: 87-101.

Rhoades, D. R., J. X. Zhao, and G. H. McVerry (2008). A simple test for inhibition of very strong shaking in ground-motion models. **Bull. Seism. Soc. Am.** 98: 448-453.

Strasser, F. O., J. J. Bommer, and N. A. Abrahamson (2008). Truncation of the distribution of ground-motion residuals. **Journal of Seismology** 12: 79-105.