OVER-PREDICTION OF EXTREME MOTIONS BY GROUND-MOTION MODELS

G.H. McVerry[1] , D.A. Rhoades[1] and J.X. Zhao[2]

[1] *Principal Scientist, Institute of Geological & Nuclear Sciences, Lower Hutt, New Zealand*
[2] *Senior Scientist, Institute of Geological & Nuclear Sciences, Lower Hutt, New Zealand*
Email: g.mcverry@gns.cri.nz

ABSTRACT :

There is often questioning of the credibility of extreme earthquake ground-motion predictions that sometimes arise in probabilistic seismic hazard analyses. Several attempts to explain the most extreme ground-motion predictions have sought departures of the upper-tail residuals from the log-normal probability distributions that govern most of the data range. Such departures are usually found to be only marginal. The approach also has the disadvantage that extreme upper-tail residuals need not be associated with the strongest motions found in the dataset or predicted by the model. This paper reports a more direct approach. For the dataset from which the ground-motion model was derived, the actual numbers of exceedances of various accelerations (e.g. 0.5g, 0.75g, 1g…) were compared to their expected numbers found by summing the probabilities of their exceedance according to the model across the data points. The method is illustrated for models from New Zealand and Japan, for both peak ground acceleration and 5% damped response spectrum accelerations. Both models show similar trends, increasingly over-predicting the numbers of exceedances as the acceleration increases. For the Japanese model with a very large dataset of 4695 records from 269 earthquakes, the ratio of actual to predicted exceedances of peak ground acceleration values falls from close to one for accelerations up to about 0.3g to about 0.15 at 1.0g. If the over-prediction of rates of the stronger motions is shown to be a general feature of ground-motion models, the results are of profound importance for seismic hazard estimates for critical facilities.

KEYWORDS: Extreme ground-motions, probabilistic hazard analyses

1. INTRODUCTION

Seismic hazard estimates for critical facilities often consider motions with annual probabilities of exceedance from about 10^{-3} to 10^{-6} or lower. In highly-seismic parts of the world, these low probabilities of exceedance may lead to peak ground acceleration estimates well in excess of 1g. There is considerable interest in the reasonableness of such high accelerations. Both statistical and physically-based approaches have been used to investigate extreme motions. Statistical studies of the most extreme ground-motion predictions have generally taken the approach of analyzing the upper tail of the distributions of residuals between recorded data and the fitted earthquake ground-motion attenuation models. Such studies have sought departures of the upper-tail residuals from the log-normal probability distributions that govern most of the data range (e.g. Bommer at al., 2004; Bommer and Abrahamson, 2006). Such departures are usually found to be only marginal, and of low statistical significance, because of the small number of observations associated with extreme residuals.

These approaches tackle the issue of extreme residuals rather than extreme motions per se. The residual approach has the disadvantage that extreme upper-tail residuals may not be associated with the strongest motions found in the dataset or predicted by the model. High residuals are not necessarily associated with high accelerations. In an approach using analysis of residuals, a peak ground acceleration value of 0.1g associated with a median prediction of 0.01g produces a much more extreme residual for the logarithm of the acceleration than a value of 1.5g associated with a median prediction of 0.5g, while it is the latter that is an extreme motion in absolute terms.

The current paper reports a more direct approach, tackling the issue of extreme accelerations rather than extreme residuals. It analyses strong-motion attenuation models and the associated datasets from which they were derived, and compares the actual numbers of exceedances of various acceleration values (e.g. 0.01g, 0.1g, 0.2g, 0.5g, 0.75g, 1g….) with the expected numbers predicted by the derived attenuation model. The methodology is as presented by Rhoades et al. (2008).

2. METHOD

Consider a strong-motion dataset containing n records numbered $i=1,..n$, with observed values y_i of some ground-motion measure together with their vector of "explanatory variables" $\overline{x_i}$, such as magnitude, source-to-site distance, site class and source mechanism. A model of functional form f and coefficients $\overline{\theta}$ predicts a median value $\hat{y}_i$ associated with the observation y_i, where

$$\hat{y}_i = f\left(\overline{x}_i, \overline{\theta}\right) \tag{4.1}$$

In practice, the ground-motion values that are modelled by the attenuation expression are often the logarithms of either the peak ground acceleration or 5% damped response spectral accelerations. The model is associated with a probability distribution, usually a normal distribution for the ground-motion measure y, or log-normal distribution for the accelerations. The observation y_i is the value taken by a random variable Y_i with mean $\hat{y}_i$ and some standard deviation σ_i, with associated explanatory variables $\overline{x_i}$. Assuming a normal distribution, for a given value y of the strong-motion measure, the attenuation model then provides the probability that Y_i exceeds the value y

$$P(Y_i > y) = 1 - \Phi\left(\frac{y - \hat{y}_i}{\sigma_i}\right) \tag{4.2}$$

where Φ is the standard normal cumulative probability function. Repeating this evaluation for the particular value y of the ground-motion measure over the n data points and summing the probabilities gives the expected number $N(y)$ of exceedances of the strong-motion level y in the whole data set, given the attenuation model and the conditions under which the strong motion records ($y_i, i = 1, \cdots, n$) were obtained. That is

$$N(y) = \sum_{i=1}^{n} P(Y_i > y) \tag{4.3}$$

Let us denote the actual number of exceedances by k(y), i.e.

$$k(y) = \sum_{i=1}^{n} I(y_i > y) \tag{4.4}$$

where $I(e)$ = 1 if the expression e is true, and 0 otherwise. There would be evidence of inconsistency with the model for strong motion at level y if $k(y)$ was found to be significantly less than $N(y)$. The particular interest is in the situation where the actual number of exceedances $k(y)$ falls below the theoretical number $N(y)$ as the ground motion level becomes large, perhaps indicative of some physical limitation on very strong ground motions. If there is some physical constraint restricting the occurrence of very strong ground motion, the ratio $r(y) = k(y)/N(y)$ would become progressively smaller as y increases.

Under the model, $k(y)$ can be regarded as the realisation of a Poisson distribution with mean (and variance) $N(y)$. For sufficiently large $N(y)$, the normal approximation, justified by the central limit theorem, can be used to compute tolerance limits for $k(y)$. However, for the cases of most interest here, where y is large, $N(y)$ is likely to be small, so that the normal approximation is invalid. Assuming a Poisson distribution for $k(y)$, an upper 100(1-α)% confidence limit, u, for $r(y)$ can be calculated by solving

$$\sum_{j=k(y)+1}^{\infty} \frac{e^{-\lambda} \lambda^{j}}{j!} = \alpha \tag{4.5}$$

where $\lambda = uN(y)$.

3. RESULTS

The method is demonstrated by applying it to two attenuation models and their associated strong-motion datasets. One model is that of McVerry et al. (2006), based primarily on New Zealand data (the New Zealand model). The other model is that of Zhao et al. (2006), based primarily on Japanese data (the Japanese model). Both models were derived using random effects methodology (Abrahamson and Youngs, 1992), with the variances separated into a between-earthquake component τ^2 and a within-earthquake component σ^2, producing a total variance equal to ($\tau^2 + \sigma^2$). Both models give expressions for peak ground accelerations and 5% damped response spectral accelerations *SA(T)* for a range of spectral periods *T*.

The New Zealand model is based on records from 48 New Zealand earthquakes supplemented by near-source peak ground accelerations from overseas crustal earthquakes. A total of 526 New Zealand records and 64 overseas records were used to derive the attenuation model for pga, and 414 New Zealand records for the model for *SA(1s)*. The strongest peak ground accelerations in the dataset were associated with the overseas records, because they were selected to provide data at source-to-site distances of less than 10 km, a distance range that is lacking in the New Zealand data.

The Japanese dataset is much larger than the one for New Zealand. The Japanese model was derived from 4695 records from 269 earthquakes, including 208 near-source records from outside Japan. The specific model used is that corresponding to equations (1) and (2) and Tables 4 and 5 of Zhao et al. (2006).This is their model derived directly using the random effects methodology, without further modification from the addition of quadratic magnitude terms derived from further regression on the inter-earthquake residuals. The number of records affects the precision of the comparisons.

Standard residual comparisons are shown for both models. Figure 1 plots the actual deviate against the normal deviate for pga for both models. The actual deviate is the standardized residualof *ln(pga)* from the model. The normal deviate is the corresponding quantile of the standard normal distribution. For standardized residuals less than 2, both plots lie close to the identity line on which normally distributed residuals should fall. Neither the New Zealand nor Japanese *ln(pga)* residuals indicate over-estimation of the largest positive residuals by the assumed normal distribution, the indication sought in this type of analysis as evidence of over-prediction of the strongest accelerations. In fact, although the greatest divergence from the identity line in the New Zealand model is for the largest positive residuals, in the top right corner of Figure 1(a), the actual largest positive residuals are greater than expected from the normal distribution of residuals, i.e. it seems that the normal distribution under- rather than over-predicts the largest residuals. There is no indication of suppression of upper-tail residuals, such as would be expected if there were physical limitations on the strongest ground motions, in either model.

The same two sets of data are now used to compare actual and expected numbers of exceedances of various acceleration levels. In Figure 2(a), the solid line plots the expected number of exceedances as a function of peak ground acceleration for the New Zealand model and dataset, with the actual numbers shown as points. The dotted lines are the 95% tolerance limits, calculated using the Poisson expression of equation (4.5). Figure 2(b) shows the same information in terms of the ratio of actual to expected numbers of exceedances.

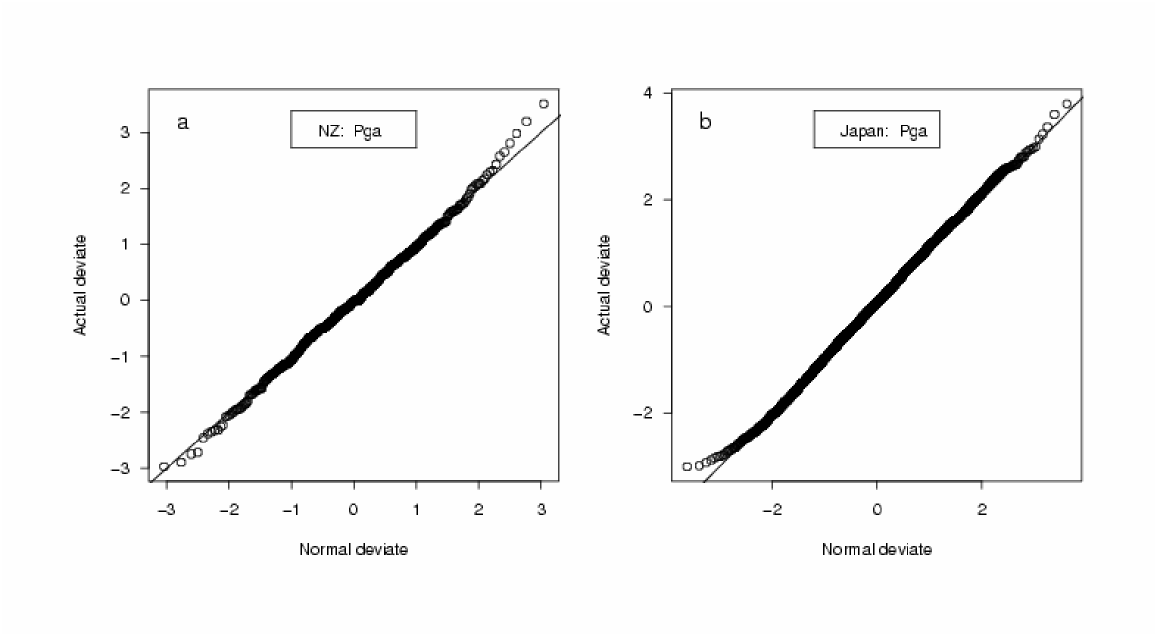

Figure 1 Examples of standard residual analysis.

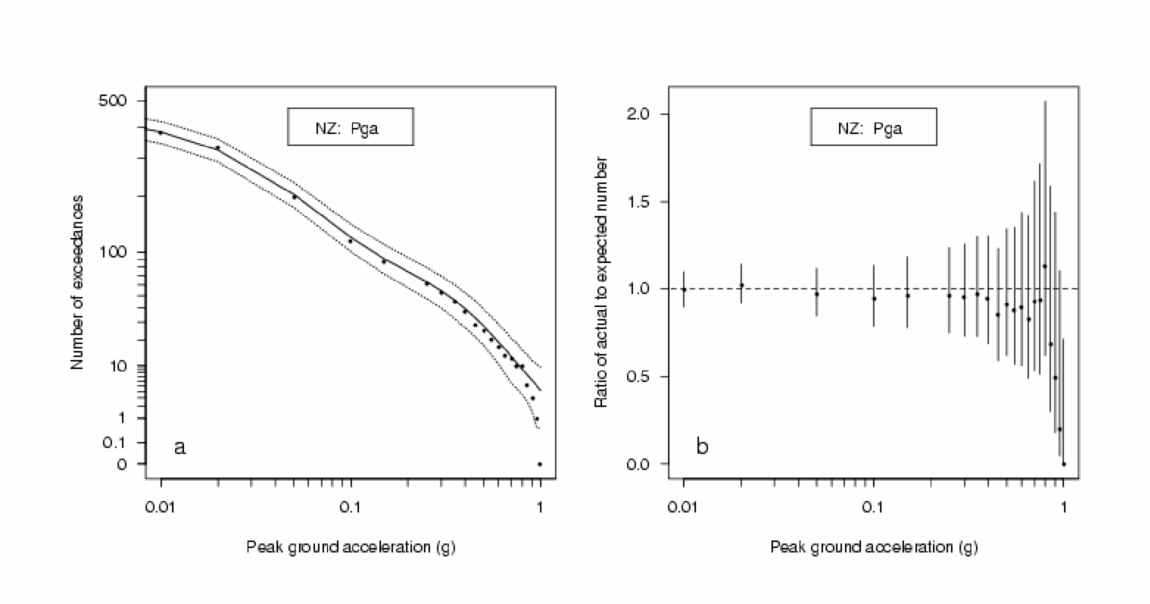

Figure 2 (a) Expected (solid line) and actual (points) number of exceedances of levels of peak ground acceleration in the New Zealand model. Dotted lines are 95% tolerance limits. (b) Ratio of actual to expected number of exceedances and its 95% tolerance limits.

In Figure 2a, there is a slight trend for the actual number of exceedances to drop below the expected number as peak ground acceleration increases, but only the value at 1.0g (which is zero, when the expected number is more than four) is outside the 95% tolerance limits. In Figure 2b, the ratio of the actual to expected number of exceedences is well constrained where the expected number of exceedances is large, but much less so when the expected numbers are less than about 20, i.e. for accelerations greater than about 0.7g. Consistent with Figure 2a, the entire confidence interval is less than 1 (i.e., the ratio is significantly less than 1) only at 1.0g.

A similar pair of plots for the Japanese pga data and Zhao et al. (2006) model is shown in Figure 3. There is a strong and statistically significant trend for the actual number of exceedances to progressively decline below the expected number as the value of acceleration increases. The decline is significant at 0.4g and the ratio of the actual number of exceedances to the expected number drops to about 0.15 at 1.0g. In contrast to the New Zealand data, the expected numbers of exceedances are high enough that the confidence limits on the ratio are quite narrow, even at values exceeding 1g. For example, we can say with high confidence that the ratio is less than 0.5 for pga values exceeding 0.7g (Figure 3b). Note that the decline is gradual. The strongest pga in the dataset exceeds 1.2g, three times the value at which the decline first becomes significant. The data at high accelerations deviate markedly from the model in a manner which is consistent with the inhibition of very strong motions. The probabilities of very strong ground motions occurring are much less than those predicted by the model. This behaviour was not apparent in the standard residual analysis, where the data showed more extreme positive residuals (i.e. data stronger than model) than indicated by the normal distribution of residuals. Apparently, these most extreme residuals are not associated with the largest peak ground accelerations, for which our new analysis shows that the rates of occurrence are over- rather than under-predicted by the model.

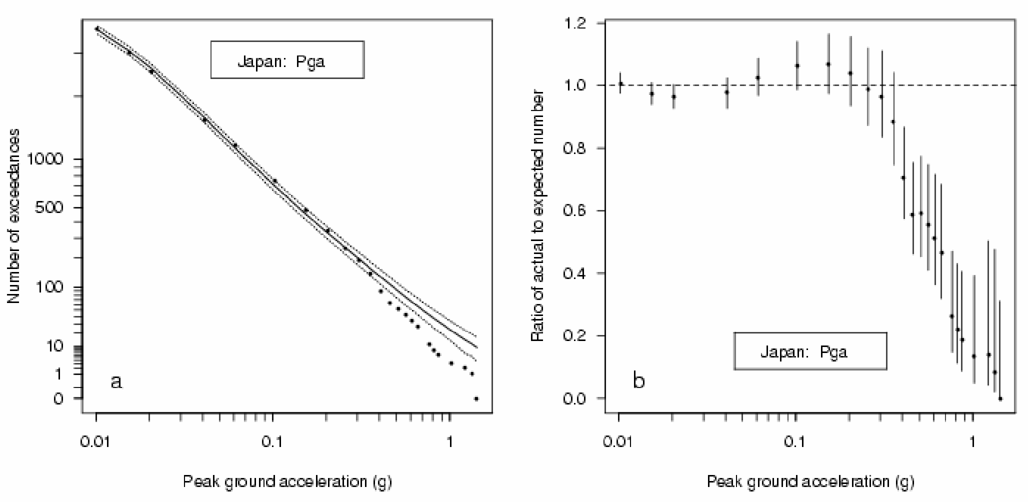

Figure 3 (a) Expected (solid line) and actual (points) number of exceedances of levels of peak ground acceleration in the Japanese model. (b) Ratio of actual to expected number of exceedences. The 95% confidence limits are shown as in Figure 2.

The over-prediction of the rates of occurrence of the strongest motions is not restricted to peak ground accelerations, but pervades response spectral accelerations for all spectral periods. Figure 4 shows a similar pair of plots related to actual and predicted numbers of exceedances for 1s response spectral accelerations, *SA(1s)*, for the Japanese data and model. In the case of *SA(1s)*, the noticeable decline from the expected numbers of exceedances begins at less than 0.1g but is more gradual than for *pga*, with the ratio reaching about 0.2 at 1.0g . We can say with high confidence that the ratio is less than 0.6 for SA(1s) values greater than 0.8g (Figure 4b).

A similar plot (not shown) of exceedance numbers for *SA(1s)* values for the New Zealand dataset is inconclusive. There are fewer *SA(1s)* values than *pga* values in the New Zealand dataset, and both the actual and expected numbers of exceedances are very small at accelerations large enough to be of interest. The few data lead to confidence limits that are simply too broad to be useful.

For the three cases presented in Figures 2, 3 and 4, the data at high accelerations deviate markedly from the associated model in a manner which is consistent with the ground being inhibited in producing very strong motions. The probabilities of very strong ground motions occurring are much less than those predicted by the models.

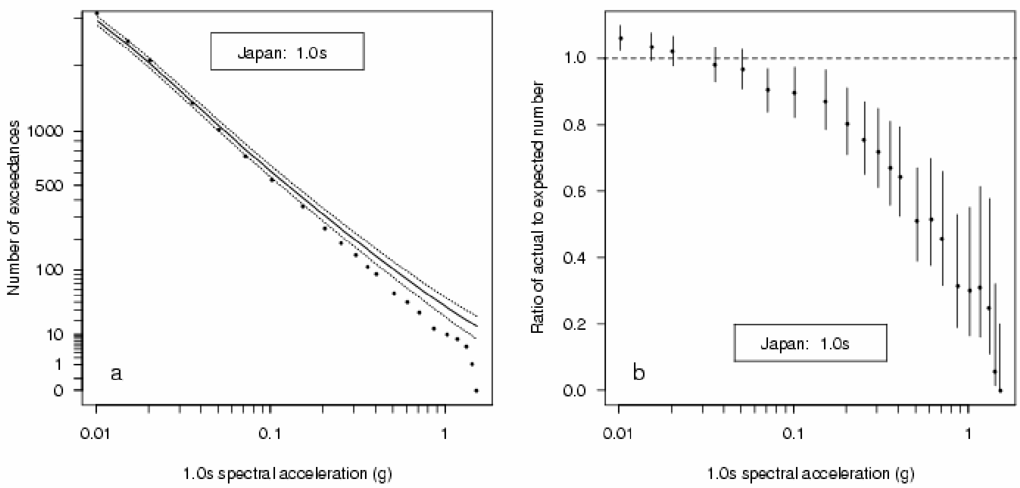

Figure 4 (a) Expected (solid line) and actual (points) number of exceedances of levels of 1.0s spectral acceleration in the Japanese model. (b) Ratio of actual to expected number of exceedences. The 95% confidence limits are shown as in Figure 2.

4. FUTURE WORK

Work is in progress using the large Japanese dataset to investigate whether the over-prediction is associated with certain classes of records (e.g. particular site classes, earthquake types or mechanisms). Some subsets of data show different behaviour of the ratios of actual to expected numbers of exceedances than the overall dataset, but explanation of the differences is proving difficult.

It is also intended to extend the study to better-known datasets and models, such as those of the NGA project in the USA.

Techniques are also being investigated to fit functions that diminish the predicted rates of extreme motions, to be included as a part of the standard regression procedure. The ratio $r(y)$, or an approximation to it through a function $\hat{r}(y)$ that remains non-zero beyond the largest value in the dataset, can be used to correct the results of a seismic hazard analysis based on a straightforward application of the attenuation model. Such a functional approximation might be obtained from an appropriate generalized linear model (McCullagh and Nelder, 1989), such as a suitably adapted logistic regression analysis. If such a function can be developed it could be included directly in probabilistic seismic hazard analyses. Suppose that a seismic hazard analysis yields a curve of return period $T(y)$ against y. Then a first-order correction to the curve is to replace $T(y)$ by $T(y)/\hat{r}(y)$. For a more rigorous correction, it would be necessary to consider the uncertainty in $\hat{r}(y)$ along with other uncertainties in the analysis.

5. DISCUSSION AND CONCLUSIONS

The examples show that the method proposed here can be effective in identifying deviations from expected numbers of occurrences of high accelerations, consistent with inhibition of very strong ground motions. The

Japanese attenuation model is supported by a very large data set. For the examples presented, analyses of the upper tail of the distribution of residuals show no significant departure from the normal distribution. If applied to sufficiently data-rich models, the method could be used to examine how the inhibition of strong motions varies with such factors as the period of the spectral response, the site conditions, magnitude or distance, such as we have begun for the Japanese model.

If the over-prediction of rates of the stronger motions such as we have found for the New Zealand and Japanese model and dataset combinations is shown to be a general feature of ground-motion models, the results are of profound importance for probabilistic seismic hazard estimates for critical facilities, for which annual probabilities of exceedance from about 10^{-3} to 10^{-6} or lower are often considered, leading to pga estimates often in excess of 1g in highly-seismic parts of the world.

6. ACKNOWLEDGEMENTS

This research was supported by research grants from the New Zealand Earthquake Commission, and by the Foundation for Research, Science and Technology, under programme CO5X0402.

REFERENCES

Abrahamson, N.A., and Youngs, R.R. (1992). A stable algorithm for regression analyses using the random effects model. *Bulletin of the Seismological. Society of America* **82:1** , 505-510.

Bommer, J.J., Abrahamson, N.A., Strasser, F.O., Pecker, A., Bard, P.Y., Bungun, H., Cotton, F., Fäh, D., Sabetta, F., Scherbaum, F. and Studer, J. (2004). The challenge of determining upper limits on earthquake ground motions. *Seismological Research Letters* **75:1**, 82-95.

Bommer, J. J., and N.A. Abrahamson (2006). Why do modern probabilistic seismic hazard analyses often lead to increased hazard estimates? *Bulletin of the Seismological Society of America* **96:6**, 1967-1977.

McCullagh, P. and Nelder, J.A. (1989) Generalized Linear Models (second edition), Chapman and Hall, London, U.K.

McVerry, G.H., Zhao, J.X., Abrahamson, N.A. and Somerville, P.G. (2006). New Zealand acceleration response spectrum attenuation relations for crustal and subduction zone earthquakes. *Bulletin of the New Zealand Society for Earthquake Engineering* **39:1**, 1-58.

Rhoades, D., Zhao, J., and McVerry, G. (2008). A simple test for inhibition of very strong shaking in ground-motion models. *Bulletin of the Seismological Society of America,* **88:1**, 448-453.

Zhao, J.X., Zhang, J., Asano, A., Ohno, Y., Oouchi, T., Takahashi, T., Ogawa, H., Irikura, K., Thio, H.K., Somerville, P.G., Fukushima, Yasuhiro, and Fukushima, Yoshimitsu. (2006). Attenuation relations of strong ground motion in Japan using site classification based on predominant period. *Bulletin of the Seismological. Society of America* **96:3** , 898-913.