

# Quantifying the epistemic uncertainty in ground-motion maps

**R. Foulser-Piggott**

*Cambridge Architectural Research Ltd*

**P. J. Stafford, W. Y. Ochieng**

*Imperial College London, UK*



## SUMMARY:

This paper looks at the sources, quantification and impacts of epistemic uncertainty on ground-motion predictions for future, or retrospective, earthquake scenarios. The focus is specifically on the statistical epistemic uncertainty in the estimation of the ground motion model parameters. Current techniques, which use additive factors on the median model prediction, are valid for dealing with the epistemic uncertainty in the individual model, however the actual values of this factor are currently guessed. This paper presents a method by which this factor can be directly quantified for non-linear ground motion prediction equations using simulation techniques. The method is applied to existing non-linear ground motion equations and recommendations are made for the quantification of the factor in future studies. An important practical application of this work is also presented; the quantification of uncertainty in Hazard Maps.

*Keywords: Ground-motion maps, epistemic uncertainty, model-specific uncertainty, prediction.*

## 1. THE SOURCE OF STATISTICAL UNCERTAINTY IN PARAMETER ESTIMATION

In future predictions of earthquake ground-motion, the standard deviation of the prediction, characterised by  $\sigma$ , is commonly assumed to be aleatory (Strasser *et al.*, 2009). However the approaches taken to estimate this parameter dictate that it also represents a degree of epistemic uncertainty. The value of  $\sigma$  has a significant effect on seismic hazard analysis (Bommer and Abrahamson, 2006). Therefore, the portion of  $\sigma$  which is epistemic should be correctly characterised. Removing this epistemic uncertainty will result in a reduction of  $\sigma$ , which can have a significant impact on the generation of hazard maps or loss estimates which use the ground-motion prediction as an input (Bommer and Abrahamson, 2006).

In published GMPEs, the uncertainty in the parameters is usually not presented. In the few cases in which an estimate is provided it is represented by the standard error associated with each coefficient. This parameter uncertainty results in epistemic uncertainty in median model predictions made using a GMPE (Arroyo and Ordaz, 2011). The sources of this statistical uncertainty are now discussed, following the work presented in Arroyo and Ordaz (2011), to whom the reader is referred for a more detailed derivation. The linear regression model for a GMPE can be expressed as shown in Equation 1.

$$\mathbf{Y} = \mathbf{X}_k \mathbf{B} + \mathbf{e} \quad (1)$$

Where  $\mathbf{Y}$  is a known vector, with dimensions  $n \times 1$ , including  $n$  observations of a measure of ground motion,  $\mathbf{X}_k$ , is a known matrix of the input variables, of dimension  $n \times k$ ,  $\mathbf{B}$  is an unknown  $k \times 1$  vector of the parameters to be determined in the regression analysis and  $\mathbf{e}$  is an unknown  $n \times 1$  vector of the regression residuals. The elements of  $\mathbf{e}$  are the sum of an inter ( $\eta_i$ ) and intra-event ( $\varepsilon_{ij}$ ) residual. Following Joyner and Boore (1993), Arroyo and Ordaz (2011) assume that the correlation between  $\eta_i$  values for different events is 0 and the correlation between  $\varepsilon_{ij}$  at different sites for a given earthquake is  $\rho_\eta$ . The correlation between the values of  $\mathbf{e}$  can therefore be represented as a block diagonal matrix:  $\mathbf{C} = \sigma_\epsilon \mathbf{c}$  of

dimensions  $N_e \times N_e$ , where  $N_e$  is the number of earthquake events.  $\mathbf{c}_e$  is an  $N_{re} \times N_{re}$  matrix, where  $N_{re}$  is the number of records for the  $i^{th}$  earthquake, with diagonal elements equal to one and off-diagonal equal to  $\rho_\eta$ .

For a given value of  $\rho_\eta$ , the values of  $\sigma_T$  and  $\mathbf{B}$  which maximise the likelihood of  $\mathbf{Y}$  are the weighted least squared estimators  $\hat{\mathbf{B}}$  and  $\hat{\sigma}_T^2$  (Draper, 1998) shown in Equations 1.2 and 1.3.

For a given value of  $\rho_\eta$ , the values of  $\hat{\sigma}_T^2$  and  $\mathbf{B}$  which maximise the likelihood of  $\mathbf{Y}$  are the weighted least squared estimators  $\hat{\mathbf{B}}$  and  $\hat{\sigma}_T^2$  (Smith, 1998), shown in Equations 2 and 3.

$$\hat{\mathbf{B}} = (\mathbf{X}^T \mathbf{c}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{c}^{-1} \mathbf{Y} \quad (2)$$

$$\hat{\sigma}_T^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^T \mathbf{c}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})}{n} \quad (3)$$

The value of  $\rho_\eta$  that maximises the likelihood is found iteratively and therefore the values of  $\sigma_T^2$  and  $\mathbf{B}$  are related to  $\rho_\eta$  of the maximum likelihood. The covariance matrix ( $\text{COV}(\hat{\mathbf{B}})$ ) for a given  $\mathbf{c}$  is defined as follows (Smith, 1998; Arroyo, 2011):

$$\text{COV}(\hat{\mathbf{B}}) = \frac{1}{n - k - 2} [(\mathbf{X}^T \mathbf{c}^{-1} \mathbf{X})^{-1} (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^T \mathbf{c}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})] \quad (4)$$

A GMPE may be used to predict future values of  $\mathbf{Y}$ , represented by  $w$ , for a given scenario characterised by a set of values of  $\mathbf{X}$ , using  $\hat{\mathbf{B}}$  and  $\hat{\sigma}_T^2$ . However, as the coefficient values  $\mathbf{B}$  and the total model variance  $\hat{\sigma}_T^2$ , are conditioned on  $\mathbf{Y}$  and  $\mathbf{X}$  used in the analysis, it is not valid to predict  $w$  for values of  $\mathbf{X}$  outside the range of the original data (Arroyo and Ordaz, 2011). That is, if the model is extrapolated, the variance of the model prediction is not necessarily related to the variability of the data used to derive the model (i.e. the range of values of the observed ground-motion at a given magnitude and distance). Arroyo and Ordaz (2011) present Equation 5 to demonstrate how the variance of  $w$  can be represented by the predictive variance  $\sigma_p^2$ . This value depends on both the uncertainty in the regression coefficients ( $\mathbf{Z} \text{COV}(\hat{\mathbf{B}}) \mathbf{Z}^T$ ) and the variability of the dataset  $\sigma_T^2$ , where  $\mathbf{Z}$  is a vector of  $k$  parameters for which  $w$  is being predicted.

$$\sigma_p^2 = \sigma_T^2 + \mathbf{Z} \text{COV}(\hat{\mathbf{B}}) \mathbf{Z}^T \quad (5)$$

Arroyo and Ordaz (2011) discuss how the variance of  $w$  is affected by uncertainty in the coefficient values. Equation 5 demonstrates that  $\sigma_p^2$  will tend to  $\sigma_T^2$  if the  $\mathbf{Z}$ , as the variance of the regression coefficients will be small. However, for a dataset which has a great deal of variability in the ground-motion recordings for a particular magnitude-distance combination, the variance of the regression coefficients will be large and the variance of the predicted value increases. The uncertainty in parameter estimates can therefore clearly be classified as epistemic because estimates of the parameters become more precisely defined as the dataset becomes larger (Arroyo and Ordaz, 2011). In their paper, Arroyo and Ordaz (2011) examine the impacts of statistical uncertainty in the parameters only for linear models. To illustrate the impact of this uncertainty on future predictions made using a non-linear GMPE, Monte Carlo simulations are now used to quantify the influence that uncertainty in the parameters has on ground-motion predictions using the predictive model for Arias Intensity presented in Foulser-Piggott and Stafford (2011) (FPS). This is achieved by sampling from the covariance matrix,  $\text{COV}(\hat{\mathbf{B}})$ . The FPS model is shown here in Equations 6, 7 and 8 and the values of the coefficients are given in Table 1.1.

$$\ln I_a = \ln(\hat{I}_a^{ref}) + f_{site}(V_{S30}, \hat{I}_a^{ref}) \quad (6)$$

$$\ln(\hat{I}_a^{ref}) = c_1 + c_2(8.5 - M_w)^2 + (c_3 + c_4 M_w) \ln \sqrt{R_{rup}^2 + c_5^2} + c_6 F_{RV} \quad (7)$$

$$f_{site}(V_{S30}, \hat{I}_a^{ref}) = v_1 \ln\left(\frac{V_{S30}}{V_{ref}}\right) + v_2 \left[ e^{v_3(\min[V_{S30}, 4400] - V_1)} - e^{v_3(V_{ref} - V_1)} \right] \ln\left(\frac{\hat{I}_a^{ref} + v_4}{v_4}\right) \quad (8)$$

**Table 1.1.** The coefficient values for the FPS model. The model uncertainty, partitioned into an intra-event and inter-event component is also provided for both functional forms.

Coefficient	FN value	Coefficient	FN value
$c_1$	5.1961	$v_1$	-1.1335
$c_2$	-0.2371	$v_2$	-0.6519
$c_3$	-3.6561	$v_3$	-0.0022
$c_4$	0.2309	$v_4$	0.1327
$c_5$	5.4651	$\sigma_E$	0.6812
$c_6$	0.3186	$\sigma_A$	0.8975

## 2. THE IMPACTS OF STATISTICAL ERRORS IN THE PARAMETERS ON $I_a$ PREDICTIONS

The impact of statistical uncertainty in the coefficients on the median predictions of the equation for Arias Intensity  $I_a$ , referred to as FPS, is examined using Monte Carlo simulations.

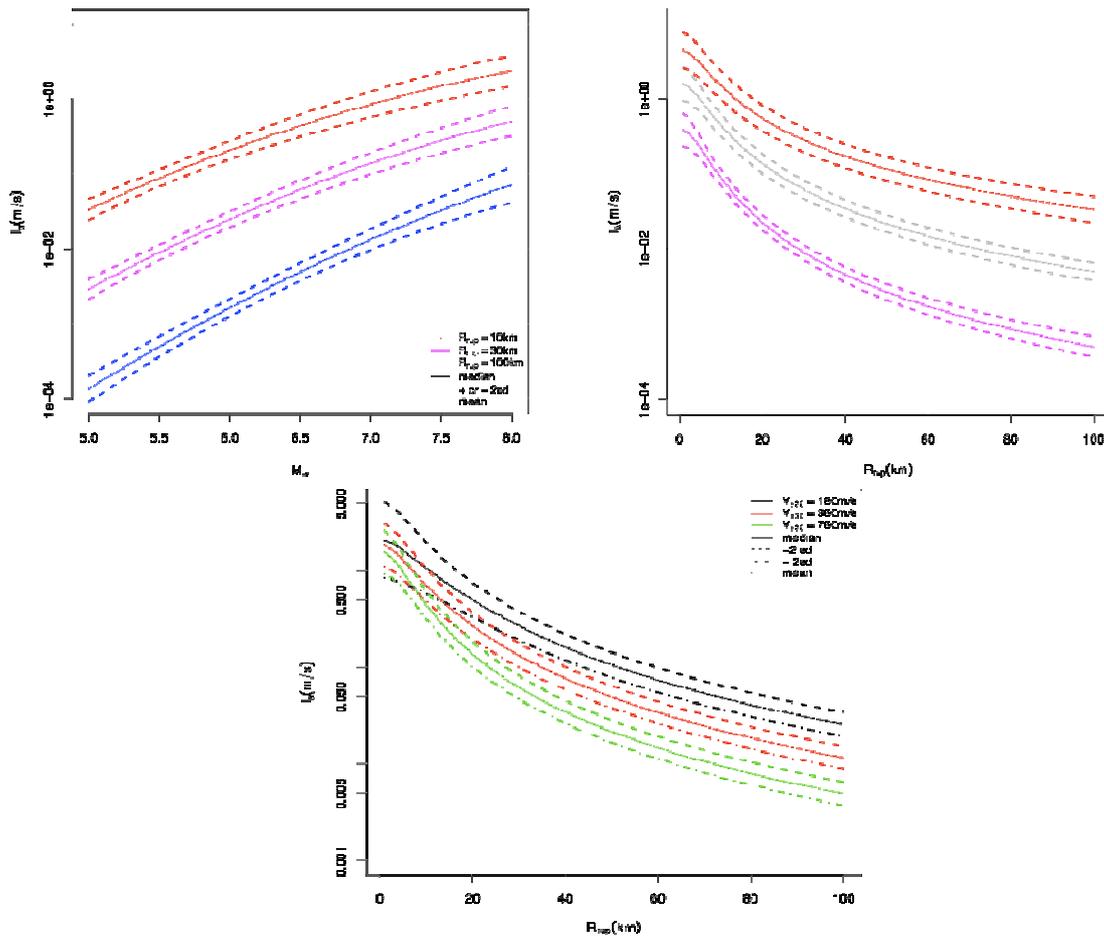
Firstly, a regression analysis is conducted in ‘‘R’’ using the FPS model form and the metadata used to derive this model. From this regression, the estimated coefficients of the model, the standard error associated with each coefficient and the correlation and covariance matrices of the coefficients are extracted. The correlation matrix is shown in Table 2.1.

There are very strong correlations between a number of the coefficients which would result in sensitivity of the coefficients to changes in the dataset. As the coefficients are correlated, the Monte Carlo simulations draw  $n$  possible coefficient values from a multivariate normal distribution defined by the coefficient values (presented in Table 1.1) and the extracted covariance matrix of these coefficients. From the simulations, there are two relevant issues relating to the coefficients of the model. The first is the assumption of normally distributed coefficients, which may be challenged as coefficients could in theory have any distribution. The second is that the values of the coefficients,  $v_3$  and  $v_4$ , have a large standard error which leads to a large range of simulated values and results in strong effects on the model predictions. This analysis suggests that including these terms in the model, whilst having a physical meaning, has a weak statistical basis. It can be questioned whether it is appropriate to select terms on the basis that they are *just* statistically significant due to the impact on the overall variability of the model.

**Table 2.1.** Correlation matrix of coefficients  $COR(B)$

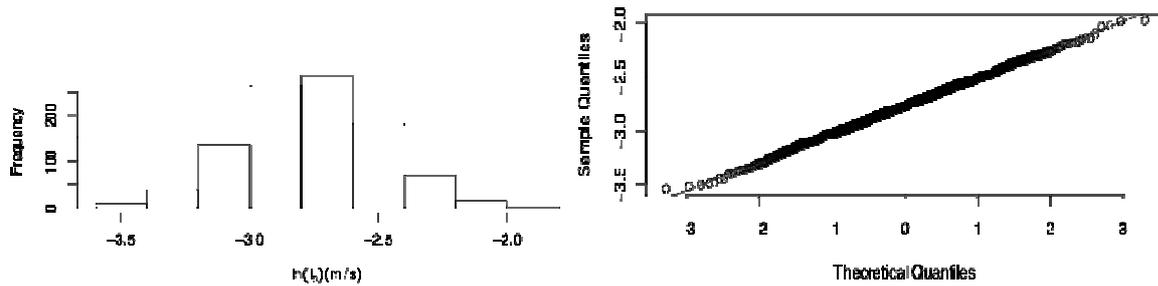
Coefficients (7 d.ps)	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$	$v_1$	$v_2$	$v_3$	$v_4$
$c_1$	1	-0.72142	0.312842	-0.51121	0.274256	-0.16021	-0.22075	-0.15275	0.650571	-0.4298
$c_2$	-0.72142	1	-0.78972	0.854061	0.163609	0.044718	0.101949	0.43004	-0.50278	0.048243
$c_3$	0.312842	-0.78972	1	-0.96978	-0.28961	-0.01064	0.098647	-0.49786	0.288127	0.242738
$c_4$	-0.51121	0.854061	-0.96978	1	0.174501	0.011588	-0.0009	0.482605	-0.42083	-0.10123
$c_5$	0.274256	0.163609	-0.28961	0.174501	1	-0.04882	-0.12915	0.542967	-0.14514	-0.34999
$c_6$	-0.16021	0.044718	-0.01064	0.011588	-0.04882	1	-0.01666	-0.01305	0.037428	-0.00574
$v_1$	-0.22075	0.101949	0.098647	-0.0009	-0.12915	-0.01666	1	-0.23918	-0.05023	0.54932
$v_2$	-0.15275	0.43004	-0.49786	0.482605	0.542967	-0.01305	-0.23918	1	-0.27861	-0.65543
$v_3$	0.650571	-0.50278	0.288127	-0.42083	-0.14514	0.037428	-0.05023	-0.27861	1	-0.40531
$v_4$	-0.4298	0.048243	0.242738	-0.10123	-0.34999	-0.00574	0.54932	-0.65543	-0.40531	1

The impact of statistical uncertainty in the parameters on both the mean and median of future predictions made using the GMPE is shown for three cases, in Figure 1. These are: the variation of the mean and median prediction of Arias Intensity ( $I_a$ ) with magnitude ( $M_w$ ) for distance ranges, the variation of  $I_a$  with  $R_{rup}$  for  $M_w$  values and the variation of  $I_a$  with  $R_{rup}$  for  $V_{s30}$  ranges. In each case, the original model predictions are shown and the  $n$  randomly drawn coefficients and the FPS functional form are used to obtain the mean or median prediction of Arias Intensity and mean model predictions plus or minus two standard deviations.



**Figure 1.** Variation of median and mean predictions of  $I_a$  with: *Left (a)*  $R_{rup}$  for  $V_{s30} = 760$  m/s,  $F_{IV} = 0$  and  $R_{rup}$  values: 10km, 30km and 100km. *Right (b)*  $R_{rup}$ , for  $V_{s30} = 760$  m/s,  $F_{IV} = 0$  and  $M_w$  values: 5.5, 6.5, 7.5. *Bottom: (c)*  $R_{rup}$  for  $M_w = 6.5$ ,  $F_{IV} = 0$  and  $V_{s30}$  values: 180m/s, 360m/s, 760m/s.

The first aspect to test is whether or not the distribution of the predictions of  $I_a$  are normal, i.e. if it is valid to represent the distribution of the predictions by plotting a mean or median prediction plus or minus two standard deviations. For a sample of all of the predictions in each range, a Kolmogorov-Smirnov normality test is conducted and a QQ-plot drawn. Examples of a histogram and QQ-plot are shown in Figure 2 for:  $R_{rup} = 30km$ ,  $F_{rv} = 0$  and  $V_{s30} = 180m/s$ . The results of the K-S tests for this example is:  $D = 0.328$ ,  $p\text{-value} = 0.01$  which is representative of the results for the other events and therefore indicates that the distributions are normal. However, Figure 2 shows that there may be some skew in the distributions of predictions and therefore, the skewness for each distribution is also calculated. The calculated skewness is extremely small and the value is not consistent within ranges, i.e. distributions may be skewed towards the maximum or minimum prediction in the same range. It is therefore assumed that the distributions  $I_a$  are normal and the variation in the median prediction may be well represented using a normal distribution and therefore standard deviations.



**Figure 2.** Histogram and QQ-plot for  $\ln I_a$  variation with  $M_w$  for  $R_{rup} = 30km$ ,  $F_{rv} = 0$  and  $V_{s30} = 180m/s$ .

The figures demonstrate that the variation of values of the median predictions, represented by the mean prediction plus and minus two standard deviations, is large and the variation is not constant across the predicted range. In particular there is a large variation at extreme values of the variables  $R_{rup}$  and  $M_w$ , i.e. for large events at close distances. Additionally Figure 1c demonstrates that the variation in values for sites experiencing non-linear effects (low  $V_{s30}$  values) is extremely large. Figure 1c shows a key result of the analysis, as the variation in predictions for sites with non-linear site response ( $V_{s30} = 180m/s$ ) at a close distance, exceeds the variation in model predictions obtained for sites with different shear-wave velocities  $V_{s30} = 360m/s$  and  $760m/s$ . This indicates that the non-linear site response terms have a large impact on the uncertainty in the model predictions. The difference in mean and median predictions can also be investigated using these figures in order to determine the uncertainty in the median prediction. The mean prediction appears to be consistent with the median prediction from the simulations, which indicates that the distribution of  $I_a$  values is not skewed.

This analysis has implications for future predictions made using a GMPE, particularly in the context of Earthquake Loss Estimation that assumes that the median prediction is exact. Instead, there is a large range of values of median predictions resulting from the epistemic uncertainty associated with the statistical estimation of the parameters. This also has implications for ground-motion models which are currently compared on the basis of their median predictions. Ground motion models which are assumed to give significantly different median predictions may actually give comparable ranges of predictions. An important application of this work is in the use of the results to quantify the epistemic uncertainty due to parameter estimation in hazard maps. This is discussed in further detail in the next section.

### 3. QUANTIFYING THE EPISTEMIC UNCERTAINTY IN GMPE PREDICTIONS FOR HAZARD MAPS

Seismic Hazard Maps display earthquake ground motions for various return periods and are used in a

number of applications including seismic provision for building codes, insurance calculations, risk assessments, and public policy decisions (Petersen *et al.*, 2008). For example, the 2008 U.S. Geological Survey (USGS) National Seismic Hazard Maps are derived from seismic hazard curves calculated at sites across the United States. These maps provide levels of ground motion that have the same frequency of being exceeded in a given period of time (Petersen *et al.*, 2008). The USGS maps are used as an example in this section, as they can be considered to be the current state of practice, having been constructed using expert opinion and then extensively reviewed (Petersen *et al.*, 2008). The maps also make an attempt to account for the model-specific epistemic uncertainty associated with scenario predictions of ground-motions (Stafford, 2008).

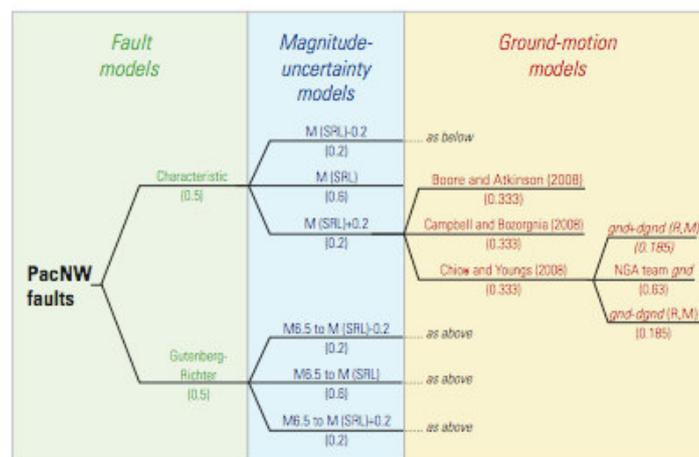
There are three sources of uncertainty of interest to the production of hazard maps that arise from the process of GMPE development:

- Inexact form of the model and selection of particular model formulation.
- Selection of a particular database.
- Statistical errors in the estimation of parameters.

The first two may be addressed by the use of multiple models (Youngs, 2006; Stafford, 2008). In this section, it is demonstrated that the third can be quantified using the method presented in Section 2.

### 3.1 State-of-practice

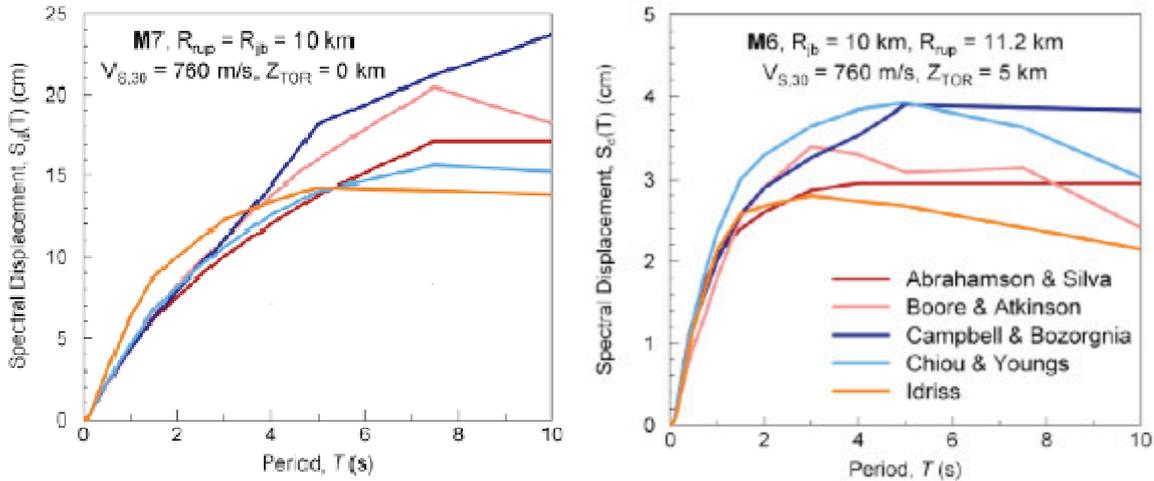
The current USGS approach for generating National Hazard Maps incorporating epistemic uncertainty is to use a logic tree formulation during the probabilistic seismic hazard analysis (PSHA), shown in Figure 3. The portion of interest in this diagram is the “Ground-motion models” section on the *right*.



**Figure 3.** Logic tree for fault sources in the compressional region of the Pacific Northwest (PacNW). The portion of interest in this research is the “Ground-motion models” section on the *right* of the diagram. *gnd* is the logarithm of median spectral acceleration or peak ground acceleration; *dgnd* is uncertainty in median spectral acceleration or peak ground acceleration at a given distance (*R*) and magnitude (*M*), (from Petersen *et al.* (2008).

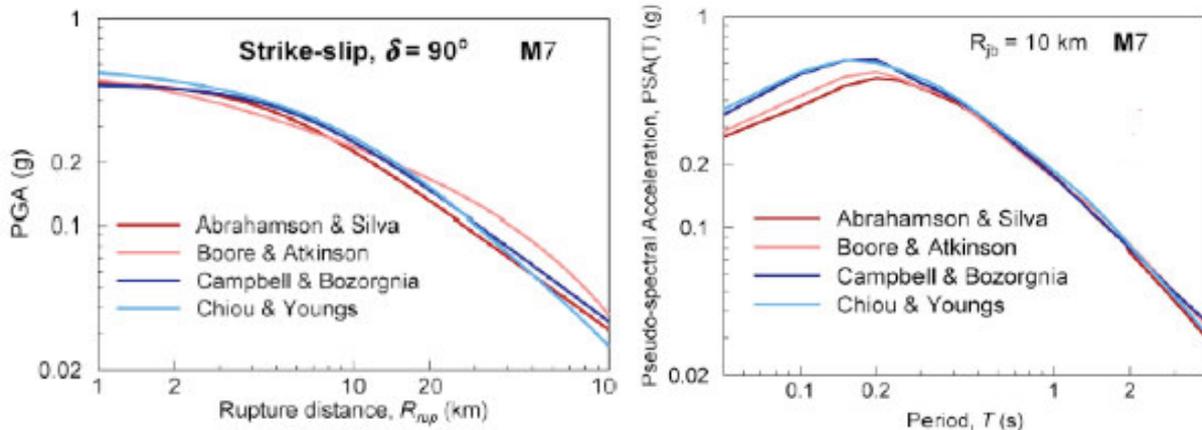
A logic tree theoretically allows the first two sources of uncertainty in the list above to be captured by allowing a suite of ground-motion models to be used to make scenario predictions. This is important because ground-motion models can have divergent predictions for many scenarios the result of which are loss estimates which are very sensitive to the ground-motion model used (Stafford, 2008a). An example of the difference in predictions obtained by different NGA ground-motion models at long periods is shown in Figure 4. PSHA commonly uses multiple GMPEs within a logic tree framework to account for these two sources of uncertainty (Arroyo and Ordaz, 2011). In Figure 3, the different ground-motion models used to predict ground motions for the USGS National Hazard Maps are shown in the “Ground-motion models” section and are the relationships of Boore and Atkinson (2008),

Campbell and Bozorgnia (2008) and Chiou and Youngs (2008). These three NGA models used to predict ground motions have equal weighting (0.333). For each individual model, there are another set of branches intended to account for epistemic uncertainty in the individual models. This “double branch” is intended to account for all components of epistemic uncertainty (Petersen *et al.*, 2008).



**Figure 4.** From Bommer *et al.* (2008), showing variation of predictions of ground-motion models particularly at long periods.

The purpose of using a suite of equations in a logic tree approach is to capture the first two sources of epistemic uncertainty. However, a portion of the uncertainty is still not identified as epistemic, as the suite of equations does not constitute the complete set of all possible models (Bommer and Scherbaum, 2008). This is compounded by the use of similar datasets and the same functional forms that are largely governed by the same theory to develop ground-motion models, which is likely to hide the true extent of the first two sources of the epistemic uncertainty. This is demonstrated by examining Figure 5 (Stafford, 2008), where the median predictions for a number of ground-motion models, three of which are used in the production of National Hazard Maps for the Western United states, are shown. It is evident from this figure that there is little variation in the median predictions from the different models which suggests that the true epistemic uncertainty may not be being captured.



**Figure 5.** The median predictions for four ground-motion models (where Boore and Atkinson (2008), Campbell and Bozorgnia (2008) and Chiou and Youngs (2008) are used in the production of National Hazard Maps) showing little variation in the median predictions from the different models which suggests that the true epistemic uncertainty may not be being captured, from Stafford (2008).

The USGS method deals with epistemic uncertainty for individual models by assuming, with no theoretical justification, that it is  $\pm 50\%$  on the model median prediction for scenarios with magnitudes ( $M_w$ ) greater than 7 and distances less than 10km (Petersen *et al.* 2008). When applied to models predicting spectral acceleration, this  $\pm 50\%$  relates to the addition ( $\alpha$ ) of 0.4 to the natural logarithm of the spectral acceleration. For different combinations of magnitude and distance values, or magnitude-distance bins, the value of an additive factor  $dgnd$  is calculated on the basis of the relative number of earthquakes that have been recorded in these bins, as shown in Equation 9 (Petersen *et al.* 2008).

$$dgnd_i = 0.4 \sqrt{\frac{n}{N_i}} \quad (9)$$

Where  $n$  is the number of earthquakes with recordings in the bin:  $M_w, R_{rup} < 10km$ , and  $N_i$  is number of earthquakes with recordings in the  $i^{th}$  magnitude-distance bin. The way in which the epistemic uncertainty resulting from parameter uncertainty can then be incorporated into the logic tree is seen on the far right of Figure 6. For each individual model, additional branches are added based on  $\pm dgnd$ . The first branch is the median value predicted by the GMPE (logarithm of the ground-motion measure) and has a weight  $w_1 = 0.63$ . The second and third branches are the median value predicted by the GMPE  $\pm dgnd$  with weights  $w_2 = w_3 = 0.185$  (Arroyo and Ordaz, 2011).

A shortcoming of the USGS National Hazard Maps approach to dealing with the epistemic uncertainty in ground-motion model predictions is that the source of uncertainty arising from the statistical errors in the estimation of the parameters is not directly quantified. Stafford (2008a) identifies the following specific problems with the USGS approach to the quantification of epistemic uncertainties in the individual GMPEs. Firstly, the  $dgnd$  terms cannot theoretically be the same for all models, as the  $dgnd_i$  value is calculated using characteristics of the different datasets used to derive the individual models. Secondly, the  $\pm 50\%$  uncertainty has no empirical basis. Thirdly, the factors to adjust this 50% uncertainty for other magnitude-distance bins do not account for the fact that this value is calculated based on the number of events in each bin and not the number of recordings. Arroyo and Ordaz (2011) look at the quantification of epistemic uncertainties for National Hazard Maps. However these authors focus only on linear ground-motion models and do not address the shortcomings in the way in which the variation in  $dgnd$  for different magnitude-distance bins is dealt with. Therefore, in the next section, an improved method for quantifying the epistemic uncertainties in future GMPE predictions for Hazard Maps is presented.

### 3.2. A new method for quantifying epistemic uncertainty in individual GMPE predictions

This section focusses on the quantification of the value of  $dgnd$  using the work presented in Section 6.1 and looks in more detail at the effects that the distribution of ground-motion predictions about the median value has on the quantification of epistemic uncertainty in ground-motion predictions. Here, a method is presented which can be used to obtain  $dgnd$  for Arias Intensity ( $I_a$ ) predictions. This method may easily be applied to any ground-motion measure of interest.

The value of  $dgnd$  is quantified for  $I_a$  predictions using the method presented in Section 2. The  $n$  sets of coefficient values created using Monte Carlo simulations are used to produce  $n$  model predictions for the different combinations of magnitude and distance values plotted in Figures 1a and 1b. In a logic tree framework analogous to that presented by Petersen *et al.* (2008), it is desirable to use a three point discrete approximation to represent the distribution of the ground-motion predictions for a given magnitude-distance combination. This allows a ‘‘triple branch’’ of predictions for a given GMPE, a median and an upper and lower value, where each branch has an assigned weight. The discrete approximations method of Miller and Rice (1983) is used to assign weights to logic tree branches.

The method presented by Miller and Rice (1983) is applied, by assuming that the predictions for  $I_a$  are normally distributed about a mean value. The normal distribution is then approximated using a 3 point discrete approximation method. Miller and Rice (1983) present 3-point discrete approximation values

for the normal distribution (with mean = 0 and standard deviation = 1) as follows, with weights shown in brackets: -1.732051 (0.166667), 0 (0.666667), 1.732051 (0.166667). The values with these weights which correspond to the distribution of  $I_a$  values for the relevant magnitude-distance combination, can then be calculated by multiplying by the standard deviation and adding the mean of the distribution of  $I_a$  values (for the magnitude-distance combination). The major assumption here that may be challenged is that the distribution of  $I_a$  values is normal.

The value of  $dgnd$  for each magnitude-distance combination may now be calculated as the percentage difference between the median  $I_a$  prediction and the  $I_a$  prediction with weight 0.166667. The  $dgnd$  values for the magnitude-distance combinations are then used to obtain average values of  $dgnd$  for the magnitude-distance bins shown in Table 4.1 (column 2). These bins are chosen to allow comparison of the results of this study with existing studies and column 3 of Table 4.1 gives the published values of Arroyo et al. (2011) who use a similar 3 point approximation of the distribution of the predictions with marginally different weights on the logic tree branches 0.185, 0.63, 0.185. The final column of Table 4.1 shows the results obtained for  $dgnd$  using the method of Petersen et al. (2008) on the dataset used in this study, i.e. defining weights on the basis of the number of earthquakes in the dataset for a given magnitude distance range. The results produced by the Petersen method are neither consistent with those obtained using the Monte Carlo approach nor those obtained by Arroyo et al. (2011). This gives further support to adopting the approach outlined in this paper for the quantification of epistemic uncertainty due to statistical uncertainty in the parameters.

**Table 4.1.** For different magnitude and distance ranges ( $V_{s30} = 760m/s$ ,  $F_{rv} = 0$ ),  $dgnd$ : column 2, results of this study, percentage difference between prediction with weight 0.17 (defined according to Miller and Rice (1983) and median prediction (weight 0.7); column 3,  $dgnd_A$  from Arroyo et al. (2011) (percentage difference between prediction with weight 0.18 and median prediction (weight 0.6); column 4,  $dgnd_p$  from the method of Petersen et al. (2008) and the dataset used in this study.

<b>Magnitude range</b>	<b>Distance range</b>	<b>dgnd</b>	<b>dgnd<sub>A</sub></b>	<b>dgnd<sub>p</sub></b>
<b>5 &lt; M &lt; 6</b>	<b>R<sub>rup</sub> &lt; 10</b>	0.372	0.375	0.133
<b>5 &lt; M &lt; 6</b>	<b>10 &lt; R<sub>rup</sub> &lt; 30</b>	0.229	0.21	0.082
<b>5 &lt; M &lt; 6</b>	<b>R<sub>rup</sub> &gt; 30</b>	0.262	0.245	0.141
<b>6 &lt; M &lt; 7</b>	<b>R<sub>rup</sub> &lt; 10</b>	0.396	0.23	0.092
<b>6 &lt; M &lt; 7</b>	<b>10 &lt; R<sub>rup</sub> &lt; 30</b>	0.272	0.225	0.078
<b>6 &lt; M &lt; 7</b>	<b>R<sub>rup</sub> &gt; 30</b>	0.246	0.23	0.082
<b>M &gt; 7</b>	<b>10 &lt; R<sub>rup</sub> &lt; 30</b>	0.468	0.4	0.4
<b>M &gt; 7</b>	<b>R<sub>rup</sub> &gt; 30</b>	0.377	0.36	0.283
<b>5 &lt; M &lt; 6</b>	<b>R<sub>rup</sub> &lt; 10</b>	0.372	0.31	0.4

#### 4. CONCLUSION

The results presented herein demonstrate that the epistemic uncertainty associated with the statistical uncertainty in the parameters can be quantified for any combination of magnitude and distance values of interest and predictions can also be made for magnitude-distance bins. For Arias Intensity ( $I_a$ ), the largest epistemic uncertainty is obtained for large magnitude events at close distances and there is a large amount of variation in the values of  $dgnd$ . This is expected, as the scenarios with the highest uncertainty, i.e. extreme values with less data and regions where non-linear site response occurs, correspond to the scenarios with the least empirical constraint. This method also allows the flexibility to define different positive and negative values for the increment  $dgnd$  if the distribution of predictions is skewed. The results of the analyses conducted are similar to those presented by Arroyo et al. (2011). However this study suggests that larger values of uncertainty are obtained for all bins, with a particularly significant difference for large magnitude and close distance scenarios. The work presented in this section provides a consistent approach to quantifying the epistemic uncertainty due to

the statistical errors in the parameters using Monte Carlo simulations and can be applied to hazard maps using a logic tree framework. The ability to quantify this component of epistemic uncertainty offers significant enhancements over methods currently used in the creation of hazard maps as it is both theoretically consistent and can be used for any magnitude-distance scenario.

## REFERENCES

- Arroyo, D. A. & Ordaz, M. (2011). On the forecasting of ground-motion parameters for probabilistic seismic hazard analysis. *Earthquake Spectra* **27:1**,1–21.
- Bommer, J. J. & Abrahamson, N. A. (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.
- Bommer, J. J. Scherbaum, F. (2008). The use and misuse of logic trees in probabilistic seismic hazard analysis. *Earthquake Spectra*, **24**,997-1009.
- Draper, N. R. Smith, H. (1998). Applied Regression Analysis, (3rd ed.). *Wiley Series in Probability and Statistics*. Wiley.
- Foulser-Piggott, R. & Stafford, P. J. (2012). A predictive model for Arias Intensity and consideration of spatial correlations. *Earthquake Engineering & Structural Dynamics* **41:3**,431-451.
- Joyner, W. B. & Boore, D. M. (1993). Methods for regression analysis of strong-motion data. *Bulletin of the Seismological Society of America* **83:2**,469–487.
- Miller, A.C. Rice, T.R. (1983). Discrete approximations of probability distributions. *Management Science*, **29:3**,352- 362
- Petersen, M.D., Frankel, A.D., Harmsen, S.C., Mueller, C.S., Haller, K.M., Wheeler, R.L., Wesson, R.L., Zeng, Y. Boyd, O.S., Perkins D.M., Luco, N., Field, E.H., Wills, C.J. & Rukstales, K.S. (2008) Documentation for the 2008 Update of the United States National Seismic Hazard Maps. *USGS Open File Report* (2008-1128)
- Stafford, P.J. (2008) Development and Implementation of the NGA models for ground-motion prediction. *Willis Research Network and Imperial College London Presentation*.
- Strasser, F. O., Abrahamson, N. A., & Bommer, J. J. (2009). Sigma: Issues, insights, and challenges. *Seismological Research Letters* **80:1**,40–56.
- Youngs, R. R. (2006). Epistemic uncertainty model for use of PEER-NGA ground motion models in national hazard mapping. *Powerpoint presentation*.