A GENERALISED CONDITIONAL INTENSITY MEASURE APPROACH AND HOLISTIC GROUND MOTION SELECTION

Brendon A Bradley, [1]GNS Science, Box 30368, Lower Hutt, New Zealand

Ph +64 4 570 4803 x 8244, b.bradley@gns.cri.nz

ABSTRACT

A generalised conditional intensity measure (GCIM) approach is proposed for use in the holistic selection of ground motions for any form of seismic response analysis. The essence of the method is the construction of the multivariate distribution of any set of ground motion intensity measures conditioned on the occurrence of a specific ground motion intensity measure (commonly obtained from probabilistic seismic hazard analysis). The approach therefore allows any number of ground motion intensity measures identified as important in a particular seismic response problem to be considered. A holistic method of ground motion selection is also proposed based on the statistical comparison, for each intensity measure, of the empirical distribution of the ground motion suite with the 'target' GCIM distribution. A simple procedure to estimate the magnitude of potential bias in the results of seismic response analyses when the ground motion suite does not conform to the GCIM distribution is also demonstrated. The combination of these three features of the approach make it entirely holistic in that: any level of complexity in ground motion selection for any seismic response analysis can be exercised; users explicitly understand the simplifications made in the selected suite of ground motions; and an approximate estimate of any bias associated with such simplifications is obtained.

KEYWORDS

Generalised conditional intensity measure (GCIM); Ground motion selection; conditional mean spectrum (CMS), statistical goodness-of-fit tests; seismic response bias.

INTRODUCTION

The rigorous selection of ground motions is an important consideration in the seismic assessment of an engineered system as it provides the link between seismic hazard (seismology) and seismic response (earthquake engineering). Rigorous and consistent ground motion selection requires both the determination of a 'target' to compare the appropriateness of different ground motions, as well as an objective method for the selection, simulation and/or modification of ground motions to 'match' this 'target'.

Probabilistic seismic hazard analysis (PSHA) [1] has become the almost unanimously adopted method by which seismic hazards are quantitatively assessed. One of the many outputs of a PSHA is a uniform hazard spectrum (UHS). The UHS, as the name implies represents a locus of the spectral accelerations at various periods which have an equal (i.e. uniform) probability/rate of exceedance. The UHS was quickly recognised as an excellent tool in force-based seismic design since for a structure, of any initial vibration period, the elastic spectral acceleration for a given rate of exceedance could be obtained. This elastic spectral acceleration once modified for damping and ductility could be used to obtain the peak base shear for the design of structures (e.g. NZS 1170.5 [2]).

Despite the fact that many studies (e.g. McGuire [3], Naeim and Lew [4], Bommer *et al*. [5], among others) have highlighted the differences between the UHS and individual earthquake scenarios, the UHS is still the primary method by which ground motion records are selected and scaled. Of the many alternatives to the UHS for ground motion selection is the conditional mean spectrum (CMS), as first presented by Baker and Cornell [6], and more recently discussed in Baker [7]. The CMS, as the name implies, provides the mean response spectral ordinates conditioned commonly on the occurrence of a specific value of a single spectral period, and is therefore directly linked to PSHA. As a result of its simplicity yet theoretical robustness, Baker [7] proposes the CMS as a tool for ground motion selection. There are however several limitations in the use of the CMS for ground motion selection, which primarily stem from the fact that spectral accelerations provide only a partial picture of the true character of a ground motion.

This paper begins by briefly discussing the CMS approach of Baker and Cornell [6], and identifying its limitations. Based on these stated limitations, a generalised conditional intensity measure (GCIM) approach for ground motion selection is presented, in which any number of ground motion intensity measures can be rigorously considered. Detailed discussion is given to the construction of the GCIM distribution; the possibilities and repercussions for selection, modification and/or simulation of ground motion records to 'fit' the GCIM distribution; and the examination of seismic response analysis results for biases due to inappropriate ground motion selection.

THE CONDITIONAL MEAN SPECTRUM

Based on the examination of the effect of the ground motion parameter ε on spectral shape, and its acknowledged correlation at multiple vibration periods, Baker and Cornell [6] proposed the concept of a conditional mean spectrum (CMS) considering ε. The basis of the CMS is that spectral accelerations at multiple vibration periods can be assumed to have a multivariate lognormal distribution (recently validated by Jayaram and Baker [8]). Based on this assumption, the conditional distribution of spectral acceleration ordinates, for a single causal earthquake, given the occurrence of a specific value of the spectral acceleration at some period, $Sa(T_j)=sa_j$, also has a lognormal distribution with mean and standard deviation given by:

$$\mu_{\ln Sa_i|Rup,Sa_j}(rup_k, sa_j) = \mu_{\ln Sa_i|Rup}(rup_k) + \sigma_{\ln Sa_i|Rup}(rup_k)\rho_{\ln Sa_i,\ln Sa_j}\varepsilon_{\ln Sa_j} \quad (1)$$

$$\sigma_{\ln Sa_i|Rup,Sa_j}(rup_k, sa_j) = \sigma_{\ln Sa_i|Rup}(rup_k)\sqrt{1-\rho^2_{\ln Sa_i,\ln Sa_j}} \quad (2)$$

where $\mu_{\ln Sa_i|Rup,Sa_j}(rup_k, sa_j)$ and $\sigma_{\ln Sa_i|Rup,Sa_j}(rup_k, sa_j)$ are the mean and standard deviation of $lnSa_i$ given $Sa_j = sa_j$ and other details related to the earthquake rupture scenario, $Rup = rup_k$ (i.e. magnitude, source-to-site distance, local soil properties, among others); $\mu_{\ln Sa_i|Rup}(rup_k)$ and $\sigma_{\ln Sa_i|Rup}(rup_k)$ are the mean and standard deviation of $lnSa_i$ given $Rup = rup_k$; $\rho_{\ln Sa_i,\ln Sa_j}$ is the Pearson correlation coefficient between $lnSa_i$ and $lnSa_j$ (assumed independent of Rup); and $\varepsilon_{\ln Sa_j} = \left(\ln sa_j - \mu_{\ln Sa_j|Rup}(rup_k)\right) / \sigma_{\ln Sa_j|Rup}(rup_k)$ is the number of standard deviations $Sa_j = sa_j$ is from that predicted by ground motion prediction equations.

Based on several assumptions, Baker [9, Appendix E], then demonstrates how considering all the causal earthquakes which contribute toward the total rate of $Sa_j = sa_j$ the mean and standard deviation of $lnSa_i$ given $Sa_j = sa_j$ can be approximated by:

$$\mu_{\ln Sa_i|Sa_j}(sa_j) \approx \mu_{\ln Sa_i|Rup}(\overline{Rup}) + \sigma_{\ln Sa_i|Rup}(\overline{Rup})\rho_{\ln Sa_i,\ln Sa_j}\varepsilon_{\ln Sa_j} \tag{3}$$

$$\sigma_{\ln Sa_i|Sa_j}(sa_j) \approx \sigma_{\ln Sa_i|Rup}(\overline{Rup})\sqrt{1-\rho^2_{\ln Sa_i,\ln Sa_j}} \tag{4}$$

where $\overline{Rup}$ represents the mean values of all the parameters which affect the ground motion at the site of interest (e.g., earthquake magnitude, source-to-site distance etc).

The mathematical simplicty and justified assumptions of the CMS approach clearly make it appealing for ground motion selection (as exemplified by Baker [7] and references therein) in relation to alternatives such as the UHS. It is noted that while equivalent versions of both Equations (3) and (4) are given in Baker [9, Appendix E], only Equation (3) appears in Baker and Cornell [6] and Baker [7]. Presumably this is done to demonstrate the simplicity of the approach and to enable an equivalent comparison with the UHS.

Limitations of the conditional response spectrum for ground motion selection

The primary limitation of the CMS stems from the fact that only characteristics of the ground motion represented in terms spectral accelerations are considered, while it is well acknowledged that the severity of a ground motion, in general, depends on its intensity, frequency content and duration. Spectral accelerations, by definition represent the peak response of a single-degree-of-freedom oscillator of a specific period, and therefore do not explicitly account for many other important features of ground motions, such as the duration, and energy (both total and its temporal accumulation) of ground shaking, among others.

The neglection of ground motion characteristics other than those reflected in a ground motions response spectrum has been somewhat justified out of both the convenience of the response spectrum as a tool for (predominately structural earthquake) engineers to understand ground motion's, and that several studies have concluded that, regarding the seismic response of structures, other features of a ground motion are of secondary concern to that of the ground motion's response spectrum. Examples of such studies include Shome *et al*. [10], Iervolino and Cornell [11], Baker [9], Baker and Cornell [12], and Luco and Bazzurro [13]. Baker [9] found that when predicting the peak interstorey drift over all floors for ductile structures scaling ground motions to the CMS produced seismic response estimates with lower uncertainty and less bias than: (i) randomly selecting records, (ii) selecting records based solely on magnitude and distance, or (iii) selecting records based on their epsilon value.

Despite the findings of the aforementioned studies, they must be kept in context with what was being analysed and what was being measured. In all these studies, either single-degree-of-freedom or multi-degree-of-freedom representations of multi-storey structures were analysed, and the response parameter of interest was the peak interstorey drift ratio over all floors. The obvious question is therefore: does the relative unimportance of aspects other than the response spectrum of a ground motion change when measuring an arbitrary response measure for any arbitrary system? It is argued that the answer to such a question is most certainly yes, which is easily justified using the following two examples. Firstly, even for such multi-storey structural systems, the importance of the ground motion response spectrum is somewhat an artefact of only measuring peak seismic responses. Given that the damage to engineering materials is a function of the number, magnitude, and sequence of plastic material strains [14] (i.e. not simply the peak strain), then in the near future more appropriate measures of seismic response, which account for more of these influential factors are likely to be adopted [15]. Shome *et al*. [10] find, for example, that the normalised hysteretic energy dissipated for a single-degree-of-freedom representation of a multi-storey structure is strongly dependent on ground motion duration. Secondly, there is no question that the seismic response of many systems where cumulative effects are important (e.g. liquefaction and other highly plastic phenomena in soils, peak displacements of unstable slopes, among others) are dependent on ground motion characteristics other than spectral ordinates.

Clearly, existing literature on the dependence of the results of seismic response analyses on particular ground motion intensity measures all have limitations which may or may not be applicable for the problem under an analyst's consideration. It is therefore desirable to have a holistic method for the selection of ground motions for any seismic response problem. To cater for the inevitably-wide variety of complexity in engineering design (and consequently in ground motion selection and seismic response analyses) such a method should allow for simplifications in the selection of ground motions, but analysts should have an explicit appreciation for the simplifications they make and be able to determine (in a simple manner) if such simplifications significantly affect the seismic response analysis results for the problem at hand. The remainder of this manuscript is dedicated to the presentation of a generalised conditional intensity measure (GCIM) approach, and a holistic method of ground motion selection which aims to meet the above stated objectives.

THE GENERALISED CONDITIONAL INTENSITY MEASURE (GCIM) APPROACH

Theoretical details

As was previously mentioned the conditional response spectrum proposed by Baker and Cornell [6] has its fundamental basis in the fact that spectral accelerations can be assumed to have a multivariate lognormal distribution. It is proposed here that for a given earthquake scenario (i.e. given a causal earthquake magnitude, source-to-site distance, local site properties etc.) any arbitrary vector of ground motion intensity measures, $\boldsymbol{IM}$, has a multivariate lognormal distribution (i.e. $\boldsymbol{IM}|Rup$ has a multivariate lognormal distribution, where "$|Rup$" indicates conditioning on a specific earthquake rupture scenario). $\boldsymbol{IM}$ may include any scalar intensity measures of a ground motion, e.g. spectral acceleration of various vibration periods, Sa_i, spectrum intensity, SI [16], acceleration spectrum intensity, ASI [17], Arias intensity, I_a [18], significant duration, SD [19], and vertical spectral acceleration at any vibration period, $Sa_{V,i}$, among others. Based on the properties of a multivariate lognormal distribution it then follows that the conditional distribution of $\boldsymbol{IM}|Rup$ given the occurrence of $IM_j = im_j$ is also a multivariate lognormal distribution.

How valid is the assumption of multivariate lognormality of $\boldsymbol{IM}|Rup$? It is almost unanimously accepted that the majority of ground motion intensity measures are marginally lognormally distributed as evident from regression on *lnIM* in empirical ground motion prediction equations (Boore and Atkinson [20], Bradley *et al*. [21], Bradley [22], Travasarou *et al*. [23], Abrahamson and Silva [24], and Abrahamson and Silva [25] are examples for Sa_i, SI, ASI, I_a, SD, and $Sa_{V,i}$, respectively). The accuracy of the assumption of multivariate lognormality however has, to the authors knowledge, only been scrutinized for horizontal spectral accelerations at various vibration periods [8]. It is therefore necessarily assumed here, without proof, that all intensity measure vectors, $\boldsymbol{IM}|Rup$, can be adequately represented by a multivariate lognormal distribution. Such an assumption (which can be relaxed as later discussed) can be validated in future studies.

Given that $\boldsymbol{IM}|Rup$ is assumed to be given by a multivariate lognormal distribution, and that the marginal distributions of all of the scalar intensity measures in $\boldsymbol{IM}|Rup$, $IM_i|Rup$, can be estimated via existing ground motion prediction equations, then only the correlation coefficient matrix of $\boldsymbol{IM}$ is needed to uniquely specify the multivariate lognormal distribution for $\boldsymbol{IM}|Rup$. Empirical prediction equations for such correlation coefficients are presently available for various intensity measure combinations. For example, Baker and Jayaram [26], Goda and Hong [27], and Goda and Atkinson [28] provide prediction equations for the correlation between horizontal Sa_i at different periods; Baker and Cornell [29] provide the correlation between both horizontal and vertical Sa_i's; Baker [30] provides the correlation between I_a and Sa_i; and Bradley [22] provides the necessary mathematical details for determining the correlation of SI and ASI with Sa_i, $Sa_{V,i}$, and I_a. The author is not aware of any empirical correlation equations relating duration intensity measures, such as SD, with other intensity measures. It is noted that when no empirical correlation coefficient prediction equation is available for a particular IM_i (i.e. the correlation between IM_i and IM_j is unknown), a distribution of IM_i can still be obtained conditional on all of the causal ruptures contributing to $IM_j = im_j$, but not specifically conditioning on $IM_j = im_j$ itself (i.e. neglecting the correlation between IM_i and IM_j). This is referred to informally here as an 'unconditional distribution' and is further discussed later in the manuscript.

In the most general sense, the result of a conventional probabilistic seismic hazard analysis (PSHA), is the annual frequency of IM_j exceeding im_j, which is given by:

$$\lambda_{IM_j}(im_j) = \sum_{k=1}^{N_{Rup}} P_{IM_j|Rup}(im_j|rup_k)\lambda_{Rup}(rup_k) \tag{5}$$

where $\lambda_{IM_j}(im_j)$ is the annual frequency of $IM_j > im_j$; $P_{IM_j|Rup}(im_j|rup_k)$ is the probability of $IM_j > im_j$ given $Rup = rup_k$; $\lambda_{Rup}(rup_k)$ is the annual frequency of earthquake rupture $Rup = rup_k$, and N_{Rup} is the number of different (assumed independent) possible earthquake ruptures. Empirical ground motion prediction equations are most commonly used to compute $P_{IM_j|Rup}(im_j|rup_k)$, while an earthquake rupture forecast (ERF) specifies $\lambda_{Rup}(rup_k)$ for all N_{rup} possible ruptures.

Given $IM_j > im_j$, the application of Bayes' Theorem [31, p 63] can be used to determine the relative contribution (so-called seismic hazard 'deaggregation' [3] or 'disaggregation' [32]) of $Rup = rup_k$ toward $IM_j > im_j$ from:

$$P_{Rup|IM_j}\left(rup_k \middle| IM_j \geq im_j\right) = \frac{P_{IM_j|Rup}\left(IM_j \geq im_j \middle| rup_k\right)\lambda_{Rup}\left(rup_k\right)}{\lambda_{IM_j}\left(im_j\right)} \tag{6}$$

Here it is important to note that Equation (6) gives the probability that if a ground motion with $IM_j > im_j$ is observed, it was caused by rupture rup_k. For the purposes of ground motion selection however, one is more interested in that probability of rupture rup_k given $IM_j = im_j$. Such a conditional probability can be obtained from the total probability theorem [31, p 57-62] as given by Equation (7):

$$\begin{aligned} P_{Rup|IM_j}\left(rup_k \middle| IM_j = im_j\right) \approx \frac{1}{\Delta\lambda_{IM_j}\left(im_j\right)} \Big[& P_{Rup|IM_j}\left(rup_k \middle| IM_j \geq im_j\right)\lambda_{IM_j}\left(im_j\right) - \\ & P_{Rup|IM_j}\left(rup_k \middle| IM_j \geq im_j + \Delta im_j\right)\lambda_{IM_j}\left(im_j + \Delta im_j\right)\Big] \end{aligned} \tag{7}$$

where $\Delta\lambda_{IM_j}\left(im_j\right) = \lambda_{IM_j}\left(im_j\right) - \lambda_{IM_j}\left(im_j + \Delta im_j\right)$; and Δim_j is a small increment (relative to im_j) of IM_j. It is noted that Equation (7) becomes exact in the limit as $\Delta im_j \to 0$. As herein we will deal solely with ground motion selection for the case of $IM_j = im_j$ and not $IM_j > im_j$, then for brevity $P_{Rup|IM_j}\left(rup_k \middle| IM_j = im_j\right)$ is referred to simply as $P_{Rup|IM_j}\left(rup_k \middle| im_j\right)$.

Since $P_{Rup|IM_j}\left(rup_k \middle| im_j\right)$ for all N_{rup} form a mutually exclusive and collectively exhaustive set, then the conditional distribution of IM_i given $IM_j = im_j$, is obtained via the total probability theorem from:

$$f_{IM_i|IM_j}\left(im_i \middle| im_j\right) = \sum_{k=1}^{N_{Rup}} f_{IM_i|Rup,IM_j}\left(im_i \middle| rup_k, im_j\right) P_{Rup|IM_j}\left(rup_k \middle| im_j\right) \tag{8}$$

where $f_{IM_i|IM_j}\left(im_i \middle| im_j\right)$ is the probability density function (pdf) of IM_i given $IM_j = im_j$; and $f_{IM_i|Rup,IM_j}\left(im_i \middle| rup_k, im_j\right)$ is the pdf of IM_i given $Rup = rup_k$ and $IM_j = im_j$. From the assumption that $\boldsymbol{IM}|Rup$ has a multivariate lognormal distribution, it follows that $\boldsymbol{IM}|Rup,IM_j$ also has a multivariate lognormal distribution [33], and in particular, that for each IM_i in $\boldsymbol{IM}$, $IM_i|Rup,IM_j$ has a univariate conditional lognormal distribution which can be expressed as:

$$f_{IM_i|Rup,IM_j}\left(im_i \middle| rup_k, im_j\right) \sim LN\left(\mu_{\ln IM_i|Rup,IM_j}\left(rup_k, im_j\right), \sigma^2_{\ln IM_i|Rup,IM_j}\left(rup_k, im_j\right)\right) \tag{9}$$

where $f_X(x) \sim LN\left(\mu_{\ln X}, \sigma^2_{\ln X}\right)$ is short-hand notation for X having a lognormal distribution with mean $\mu_{\ln X}$ and variance $\sigma^2_{\ln X}$. The conditional mean and standard deviation (square root of the variance) in Equation (9) can then be obtained from Equations (10) and (11), respectively [31, p 137]:

$$\begin{aligned} \mu_{\ln IM_i|Rup,IM_j}\left(rup_k, im_j\right) = & \ \mu_{\ln IM_i|Rup}\left(rup_k\right) \\ & + \sigma_{\ln IM_i|Rup}\left(rup_k\right)\rho_{\ln IM_i,\ln IM_j}\varepsilon_{\ln IM_j} \end{aligned} \tag{10}$$

$$\sigma_{\ln IM_i|Rup,IM_j}\left(rup_k, im_j\right) = \sigma_{\ln IM_i|Rup}\left(rup_k\right)\sqrt{1 - \rho^2_{\ln IM_i,\ln IM_j}} \tag{11}$$

where the parameter, $\varepsilon_{\ln IM_j}$, in Equation(10) is given by:

$$\varepsilon_{\ln IM_j} = \frac{\ln IM_j - \mu_{\ln IM_j|Rup}(rup_k)}{\sigma_{\ln IM_j|Rup}(rup_k)} \tag{12}$$

Thus, Equations (5)-(12) provide the necessary mathematical details to compute the conditional distribution of IM_i given $IM_j = im_j$ for all IM_i in $\boldsymbol{IM}$. Note that only Equations (9)-(12) are a result of the assumption of multivariate lognormality of $\boldsymbol{IM}$, and can be substituted for alternative relationships if such an assumption is shown to inappropriate for specific IM combinations. It is also important to note that the GCIM approach is not constrained to be applied simply as an extension of PSHA, but can be also used in a deterministic seismic hazard analysis. In such a case, $P_{Rup|IM_j}(rup_k|im_j)$ will simply equal 1.0 for the deterministic scenario considered.

Example GCIM distributions

Figure 1 illustrates the seismic hazard disaggregation for a rock site (Vs_{30} = 760 m/s) in Christchurch, New Zealand for $Sa(1.0) = 0.165$g (using the Boore and Atkinson [20] ground motion prediction equation), which has an annual exceedance rate of $4.04\text{x}10^{-4}$ (i.e. a 2% exceedance probability in 50 years). It can be seen that the seismic hazard is contributed to by a range of different potential casual earthquake ruptures.

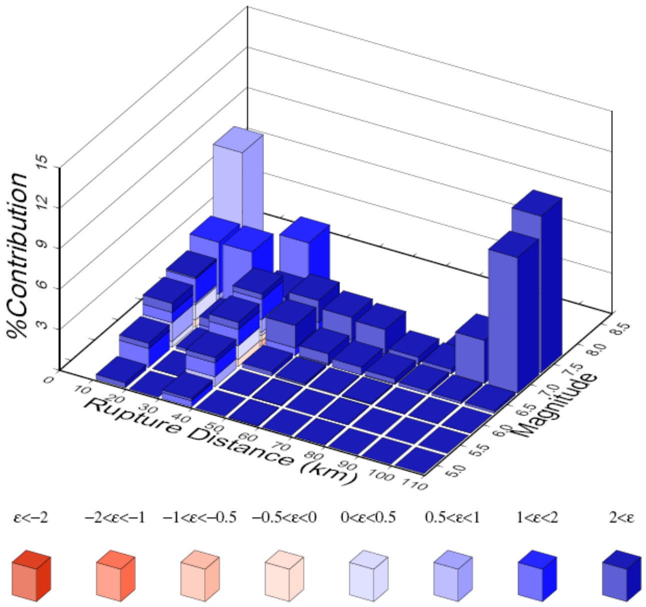

Figure 1: Magnitude-distance-epsilon disaggregation of PSHA for Christchurch, New Zealand for $Sa(1.0) = 0.165$g which has a 2% in 50 year probability of exceedance.

Figure 2 illustrates the mean (of *lnSa*), 16^{th} and 84^{th} percentiles of the conditional response spectrum obtained based on Equations (5)-(12) for the site in question. Also shown is the 'unconditional' distribution of spectral accelerations, which has been computed by replacing $f_{IM_i|Rup,IM_j}$ in Equation (8) with $f_{IM_i|Rup}$. That is, the 'unconditional' distribution of IM_i is obtained by neglecting the correlation between IM_i and IM_j. It can be seen that the conditional mean spectrum is largest relative to the 'unconditional' mean spectrum at $T = 1.0$s (as a result of most potential ruptures have an associated ε value greater than zero (Figure 1)) and tends toward the unconditional spectrum as the period of vibration tends away from $T = 1.0$s. The uncertainty in the response spectrum distribution also increases as the period of vibration tends away from $T = 1.0$s. Both the above two observations are the result of the correlation of spectral accelerations generally decreasing as the difference between the periods of interest increases [26].

Table 1: Correlations of various intensity measures with one-second spectral acceleration

IM	*Sa*(0.05)	*Sa*(0.5)	*SI*[1]	*ASI*[1]	I_a
$\rho_{\ln IM \mid \ln Sa(1.0)}$	0.42	0.75	0.92	0.61	0.7

[1]The correlation of *SI* and *ASI* with *Sa* strictly speaking is a function of earthquake scenario, but as noted in Bradley [22] the variation is negligible.

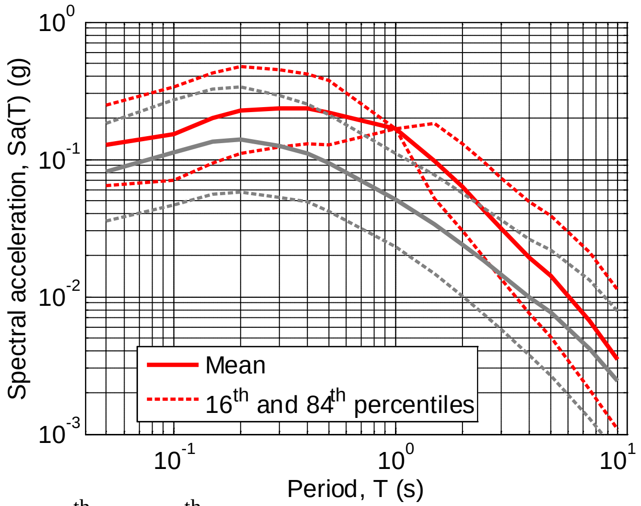

Figure 2: The mean, 16th and 84th percentiles of the conditional distribution of Sa given Sa(1.0) = 0.165g (red) and the 'unconditional' distribution of Sa (grey).

Figure 3 illustrates the conditional distributions for six different ground motion intensity measures, *Sa*(0.05), *Sa*(0.5), *SI*, *ASI*, I_a, and *SD*, which can be computed using Equations (5)-(12) for the site in question. The ground motion prediction equations of Boore and Atkinson [20], Bradley *et al*. [21], Bradley [22], Travasarou [23], and Abrahamson and Silva [24], were used for computing *Sa*, *SI*, *ASI*, I_a, and *SD*, respectively, while the correlation equations in Baker and Jayaram [26], Bradley [22], and Baker [30] were also adopted. Inspection of Figure 3 reveals that for some IM_i, the conditional and 'unconditional' distributions are relatively similar, yet significantly different for others. This difference, or lack thereof, between the conditional and 'unconditional' distributions is obviously a function of the correlation between IM_i and IM_j, as illustrated in Table 1. For example, *Sa*(0.05), which has a relatively weak correlation with *Sa*(1.0) ($\rho_{\ln Sa(0.05),\ln Sa(1.0)}$ =0.42 from Table 1) has a conditional distribution which is relatively similar to the 'unconditional' distribution, while for *SI*, which is highly correlated to *Sa*(1.0) ($\rho_{\ln SI,\ln Sa(1.0)}$ =0.92 from Table 1) the two distributions are notably different. As no empirical correlation coefficient prediction equation exists for significant duration and spectral accelerations at present, Figure 3f merely shows the 'unconditional' distribution.

Only six conditional distributions are shown in Figure 3, although as previously mentioned the conditional distribution for any intensity measure can be constructed. Since the conditional distributions will be used for ground motion selection (as discussed in the remainder of the manuscript), it is most advantageous to select a vector of intensity measures which are nearly orthogonal (i.e. measure different properties of the ground motion). The 'quality' of such a constructed conditional distribution is of course a function of the 'quality' of the ground motion and correlation prediction equations used to develop it (both the rigour with which the equations are developed and their applicability to the subject region in question). As is conventional in contemporary PSHA, epistemic uncertainties in the ground motion and correlation prediction equations can be easily considered within the framework given by Equations (5)-(12), although to avoid detraction from the pertinent issues of discussion, epistemic uncertainty consideration is neglected in the examples in this manuscript.

GROUND MOTION SELECTION USING THE GCIM DISTRIBUTIONS

The benefit of the GCIM approach is that $\boldsymbol{IM}|IM_j$ provides the exact distribution (for the given inputs in a PSHA) of intensity measures of potential ground motions with $IM_j = im_j$ which may be observed at the site. The GCIM distributions are therefore the 'target' which should be used in selecting a suite of ground motions for seismic response analysis. The aim of this section is to describe a holistic method by which ground motions can be selected to match $\boldsymbol{IM}|IM_j$ for any seismic response analysis problem.

Methods for selecting ground motions against a 'target'

Given the previous focus in literature of selecting ground motions based on their response spectrum, their are many alternative methods which provide an objective approach to select, scale and/or modify ground motions to match a target response spectrum [e.g. 34, 35, 36]. Almost all such methods (Kottke and Rathje [36] being an exception) attempt to scale ground motions to a deterministic target response spectrum (i.e. a single value of spectral acceleration per period), and those records which have the best (typically least-squares) fit to the target deterministic spectrum are generally used in seismic response analysis. This is exactly the approach taken by most seismic design guidelines throughout the world [e.g. 2, 37, 38].

Here, it is desired to perform ground motion selection which is completely consistent with the results of a PSHA. Therefore when selecting ground motions for seismic response analyses based on $\boldsymbol{IM}|IM_j = im_j$, the first step is that all potential ground motions must be scaled to have $IM_j = im_j$. Although this requirement may seem restrictive, it is also convenient in that it uniquely specifies the scaling of the (either as-recorded, modified or simulated) ground motion, and therefore the only task left to do is select a suite of such ground motions which are representative of $\boldsymbol{IM}|IM_j = im_j$.

Another key difference adopted here from many other studies which attempt to only match some deterministic response spectrum, is that the focus in selecting ground motions is not specifically on the characteristics of the individual ground motions on their own, but more on the collective characteristics of the suite of ground motions. This focus on the collective characteristics of a suite of ground motions, similar to Kottke and Rathje [36], does present some complexities in determination of the final suite for use in seismic response analysis.

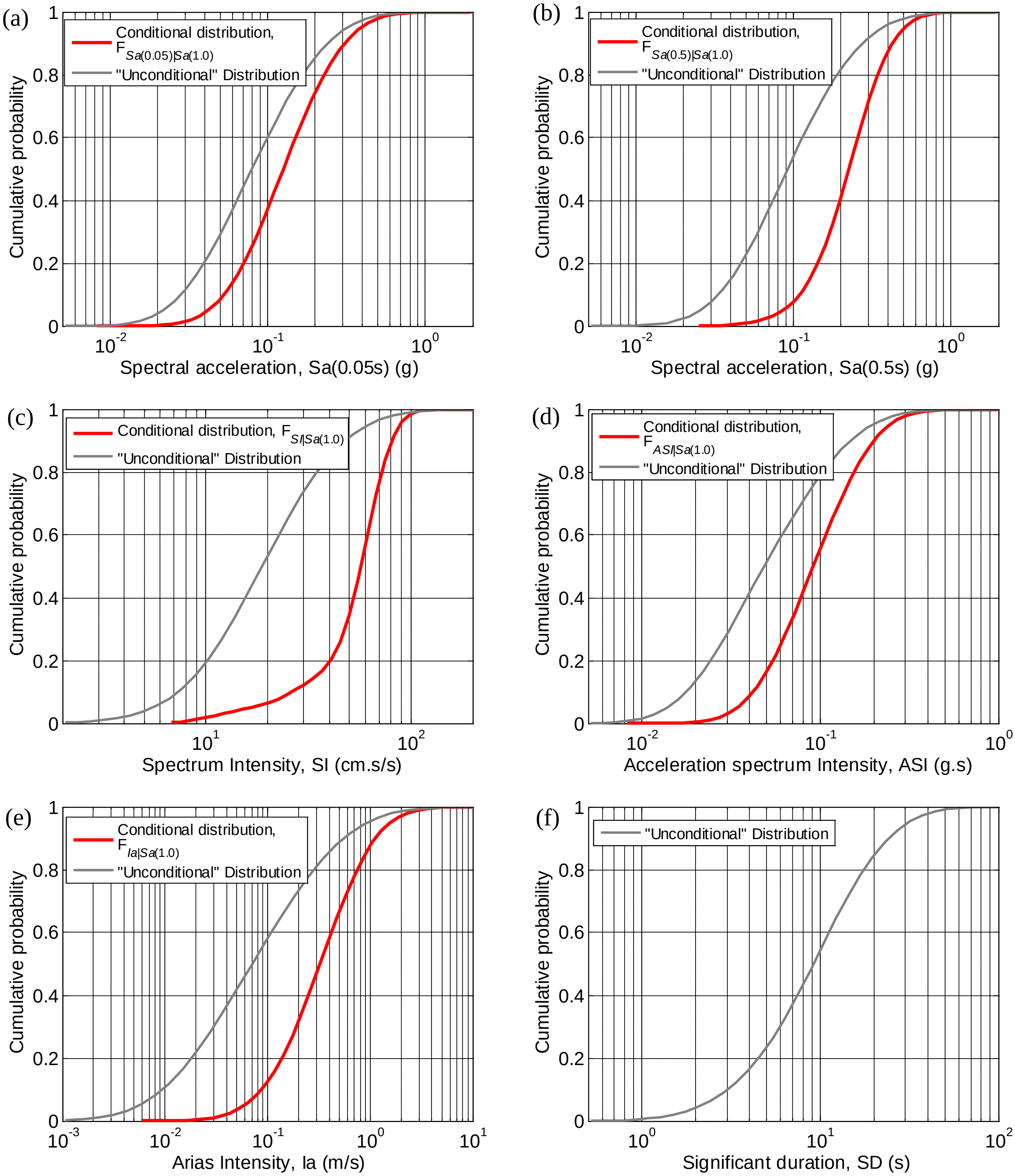

Figure 3: Conditional and 'unconditional' distributions of various IMs obtained from the conditional IM approach for Sa(1.0) = 0.165g at a site in Christchurch, NZ: (a) 0.05-second spectral acceleration, *Sa*(0.05); (b) 0.5-second spectral acceleration, *Sa*(0.5); (c) Spectrum Intensity, *SI*; (d) Acceleration Spectrum Intensity, *ASI*; (e) Arias Intensity, I_a; and (f) Significant duration, *SD*.

In order to (in a somewhat automated fashion) obtain a suite of ground motions for a specific seismic response problem, one needs an objective function for determining the relative hierarchy of possible ground motions so that a set of the 'best' N_{gm} motions can be chosen (e.g. in the aforementioned studies the least-squares residual of a ground motion compared to a target response spectrum is used as the objective function). It is argued that for a general seismic response problem, in which a multitude of seismic response measures are of interest, (which depend on several ground motion intensity measures), such an objective function is not known *a priori* (e.g. it may be unknown prior to conducting seismic response analyses how sensitive the seismic demands of interest are to the duration of the input ground motion). In such a case it is not possible to determine a 'best' suite of N_{gm} motions in an automated fashion. It is therefore proposed that a suite ground motions is selected which is consistent with the distribution of those intensity measures which are likely to influence the results of the seismic response analysis. While specific details in this initial selection may be problem dependent (and are therefore not discussed here) a starting point may be to loosely follow seismic hazard disaggregation results. From the resulting seismic response analyses the importance of various ground motion intensity measures can be determined, and the selected ground motion suite scrutinized based on the criteria they were selected upon. The remainder of this section provides an objective means by which the consistency of a ground motion suite with respect to various IM_i distributions can be obtained, while the subsequent section examines the dependence of the seismic response analysis results on the selected ground motion suite.

Ground motion selection based on $IM_i|IM_j$

As previously mentioned, $\boldsymbol{IM}|IM_j$, as obtained from the GCIM approach provides the exact distribution of intensity measures of ground motions with $IM_j = im_j$ which may be observed at the site. That is, the aim of the ground motion selection is to select a ground motions suite which matches $\boldsymbol{IM}|IM_j$.

A primary relaxation made here is that ground motions are selected to match all of the univariate distributions of $\boldsymbol{IM}|IM_j$ (i.e. $IM_i|IM_j$ for all i), but not the complete multivariate distribution, $\boldsymbol{IM}|IM_j$, itself. This relaxation is considered pragmatic because of the significant reduction in complexity that it entails. Such a relaxation is unlikely to be of consequence when 'as-recorded' ground motions are being selected (upon which the empirical ground motion prediction equations used to determine $\boldsymbol{IM}|IM_j$ are based). However, such a relaxation could possibly be more of an issue in the case of stochastically simulated ground motions, which have less of an underlying physical basis [39]. Such issues are left for future research.

Because ground motion selection is desired for a finite number of N_{gm} ground motions, then a comparison of the appropriateness of the N_{gm} ground motions as representative of $IM_i|IM_j$ (for all i) must be done so using statistical goodness-of-fit tests. Two such tests are outlined here, for continuous and discrete IM_i's of interest.

Continuous IM_i variables: The Kolmogorov-Smirnov goodness-of-fit test

The majority of ground motion intensity measures of engineering interest are continuous variables (e.g. *Sa, SI, ASI,* I_a, *SD,* in Figure 3). In such cases, the adequacy of a particular suite of ground motions with respect to a pre-defined theoretical distribution of a single IM_i can be verified by the Kolmogorov-Smirnov (KS) goodness-of-fit test [e.g. 31, p. 293-296]. The KS test measures the absolute difference between the theoretical cumulative distribution function (CDF) and the empirical distribution function (EDF) of the sample, which is mathematically given by Equation (13).

$$D_{N_{gm}} = \max_{IM_i} \left| F_{IM_i|IM_j}\left(im_i | im_j\right) - S_{N_{gm}}\left(im_i\right) \right| \qquad (13)$$

where $D_{N_{gm}}$ is the KS test-statistic for N_{gm} ground motions; $F_{IM_i|IM_j}\left(im_i|im_j\right)$ is the theoretical CDF (obtained from the GCIM approach); and $S_{N_{gm}}\left(im_i\right)$ is the EDF (of IM_i) of the suite of ground motions selected. The null hypothesis that the distribution of IM_i of the suite of ground motions is representative of the theoretical distribution is rejected if $D_{N_{gm}}$ is greater than the critical KS test-statistic for a given confidence level, α.

Figure 4a illustrates graphically the KS test. The critical KS test-statistic for $\alpha = 0.1$ is shown in the figure. If the EDF intersects the KS test-statistic 'bounds', then it indicates that the null hypothesis should be rejected. Thus, based on Figure 4a it can be stated that (at the $\alpha = 0.1$ significance level) the *Sa*(0.05) values of ground motion Suite 2 (as given in Table 2) are not representative of the theoretical distribution of *Sa*(0.05) obtained from the GCIM approach.

Although it is likely sufficient to compare the adequacy of a ground motion suite of the 'body' of the cumulative distributions, if the distribution 'tails' are of particular interest, the Anderson-Darling goodness-of-fit test should also be considered [e.g. 31, p. 296-300].

Discrete IM_i variables: The Chi-Square goodness-of-fit test

Although the majority of ground motion intensity measures of interest are continuous variables, there are also discrete variables used to classify ground motions which may be of interest. Such variables include, for example: the focal mechanism and tectonic type of the ground motion's causal earthquake, and the site class classification on which the seismograph recording the ground motion is founded. For notational consistency, such variables are still referred to here as intensity measures, although one may argue that they are variables which influence ground motions characteristics, but are not a measure of the ground motion itself.

The adequacy of a particular suite of ground motions with respect to a pre-defined discrete theoretical distribution of a single IM can be verified by the Chi-Square goodness-of-fit test [e.g. 31, p. 289-293]. The Chi-Square test-statistic, χ^2, is given by:

$$\chi^2 = \sum_{m=1}^{k} \frac{\left(n_m - e_m\right)^2}{e_m} \qquad (14)$$

where n_m are the number of observations of $IM_i = m$, e_m is the expected (predicted) number of observations of $IM_i = m$, and k is the number of discrete values IM_i can take. Similar to the KS test, the null hypothesis that the distribution of IM_i of the ground motion suite is representative of the theoretical distribution is rejected if χ^2 is larger than the critical Chi-Square test-statistic for a given confidence level, α.

Figure 4b illustrates a comparison of the discrete distribution of the focal mechanism of the ground motions comprising Suite 1 (as given in Table 2), as compared to the theoretical distribution of focal mechanism obtained from the GCIM approach. Unlike the KS test, the summative nature of the χ^2 test-statistic (Equation (14)) does not make it possible to plot 'rejection bounds' on Figure 4b. Despite this, the p-value of the χ^2 test of 0.2009 indicates that the null hypothesis would not be rejected at the 10% significance level.

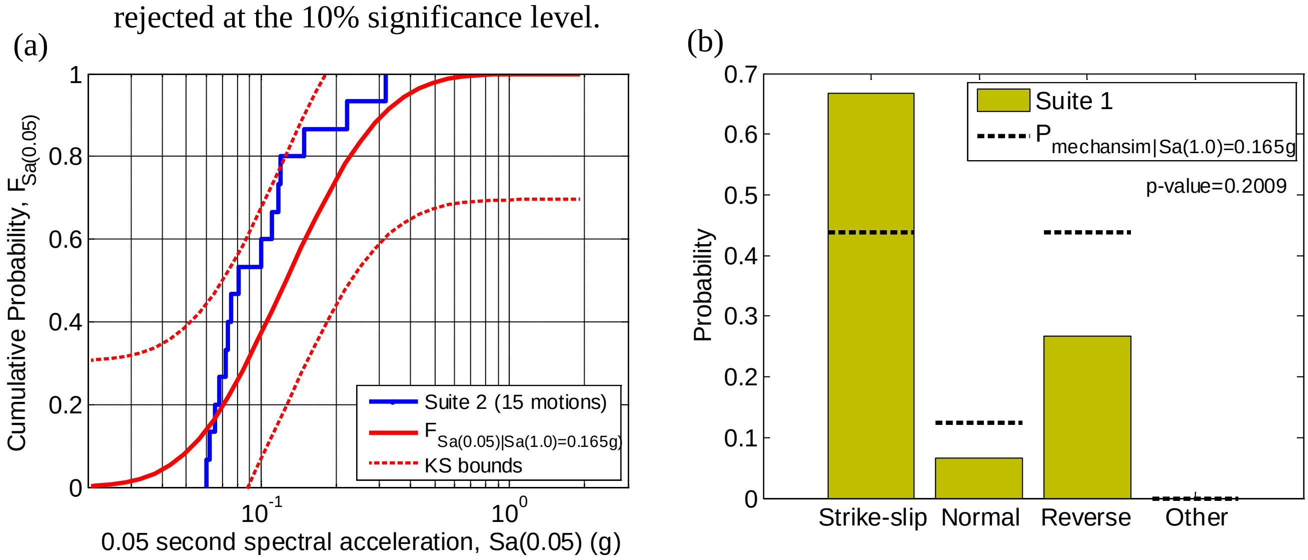

Figure 4: Examples of ground motion selection to match the 'target' $IM_i|IM_j$ distributions from the GCIM approach: (a) IM_i = Sa(0.05); and (b) IM_i = focal mechansim.

Examples and compromises in obtaining a suite of ground motions compatible with $IM|IM_j$

In order to illustrate the aforementioned concepts two suites of ground motions were selected for the site in Christchurch previously discussed. In order to elucidate the features of the proposed procedure the two ground motion suites (shown in Table 2) were selected based on different aspects of the seismic hazard disaggregation at the site (Figure 1). Suite 1 was selected on the basis of ground motions having causal earthquake magnitudes less than 6.0 and source-to-site distances less than 20 km, while Suite 2 was selected on the basis of ground motions having causal earthquake magnitudes greater than 7.0 and source-to-site distances greater than 50 km. Hence comparison of the basis for selecting these suites with the disaggregation given in Figure 1, indicates that, when scaled to have $Sa(1.0) = 0.165$g, the two suites may not provide an appropriate representation of all the ground motions expected at the site. This is however a weak statement because of the fact that many features in addition to magnitude and source-to-site distance affect ground motion characteristics. For example, based on the previously mentioned studies of Shome *et al.* [10], Iervolino and Cornell [11], Baker [9], Baker and Cornell [12], and Luco and Bazzurro [13], magnitude and source-to-site distance are of secondary importance to the spectral shape of the ground motion.

Figure 5a illustrates the un-scaled response spectra of the two ground motion suites compared to the mean (of $lnSa$), 16th and 84th percentiles of the conditional response spectrum obtained from the GCIM approach. Figure 5b shows the corresponding response spectra of the two ground motion suites once they have been scaled in amplitude to have $Sa(1.0) = 0.165$g. It can be seen from Figure 5b, that the Suite 1 ground motions (which have relatively low causal magnitudes) have notably larger response spectral ordinates at periods less than one second, and weaker spectral ordinates at periods greater than one second, relative to the conditional response spectrum. Conversely, the Suite 2 ground motions have slightly lower response spectra for periods less than one second, but notably larger response spectra for periods greater than one second. Despite the fact that, in their current form, Suites 1 and 2 show several departures from the theoretical response spectrum distribution at various periods, it would not be overly difficult to selectively choose other suites of ground motions (or combine these two) within these magnitude-distance ranges which provide a better fit. However, as stated previously, in the general seismic response problem, the appropriateness of a suite of ground motions is dependent on more than just their response spectrum characteristics, which is where a significant benefit of the GCIM approach can be obtained.

Table 2: Ground motion suites used in the examples

Suite 1				Suite 2			
ID[2]	Earthquake (year)	M	R_{jb} (km)	ID[2]	Earthquake (year)	M	R_{jb} (km)
0145	Coyote Lake (1979)	5.74	5.30	0012	Kern County (1952)	7.36	114.62
0146	Coyote Lake (1979)	5.74	10.21	0013	Kern County (1952)	7.36	122.65
0147	Coyote Lake (1979)	5.74	8.47	0014	Kern County (1952)	7.36	81.30
0148	Coyote Lake (1979)	5.74	6.75	0832	Landers (1992)	7.28	69.21
0149	Coyote Lake (1979)	5.74	4.79	0833[1]	Landers (1992)	7.28	144.90
0233	Mammoth Lakes (1980)	5.69	2.91	0834	Landers (1992)	7.28	137.25
0234	Mammoth Lakes (1980)	5.69	14.28	0835[1]	Landers (1992)	7.28	135.22
0235	Mammoth Lakes (1980)	5.69	1.44	0836[1]	Landers (1992)	7.28	87.94
0545	Chalfant Valley (1986)	5.77	14.99	0837	Landers (1992)	7.28	131.92
0547	Chalfant Valley (1986)	5.77	6.07	1759	Hector Mine (1999)	7.13	176.59
1641	Sierra Madre (1991)	5.61	8.57	1760[1]	Hector Mine (1999)	7.13	174.90
1642	Sierra Madre (1991)	5.61	17.79	1761[1]	Hector Mine (1999)	7.13	166.11
1645	Sierra Madre (1991)	5.61	2.64	1763[1]	Hector Mine (1999)	7.13	89.98
1646	Sierra Madre (1991)	5.61	13.91	1764	Hector Mine (1999)	7.13	102.40
1740	Little Skull Mtn (1992)	5.65	14.12	1765[1]	Hector Mine (1999)	7.13	193.80

[1]Used in a modified set of 7 ground motions in Figure 6.
[2]As given in the NGA database http://peer.berkeley.edu/nga/earthquakes.html

In addition to examination of the appropriateness of the ground motion suites in terms of their response spectra, using the GCIM approach comparisons can also be made with respect to any other arbitrary ground motion intensity measure. For example, Figure 5c-Figure 5f illustrate the comparisons between the EDF's of Suite's 1 and 2 and the theoretical distributions from the GCIM approach for Acceleration Spectrum Intensity, *ASI*; Spectrum Intensity, *SI*; Arias Intensity, I_a; and Significant Duration, *SD*, respectively. It can be seen that the ground motions of Suite 2 provide a good representation of the theoretical distribution of *ASI*, *SI* and I_a but not for *SD*, while the *ASI* and I_a distributions of Suite 1 are significantly larger (at the 10% significance level) than the theoretical distributions. Based on the significant correlation of the response of pile foundations in liquefying and non-liquefying soils with *SI* [40], one may argue that both Suites 1 and 2 are appropriate for seismic response analysis of such a seismic response analysis problem. However, as it is also well recognised that soil liquefaction is a cumulative phenomena, then the fact that Suite 2 has a significantly larger distribution of *SD* may mean it will give a biased prediction of seismic response (as demonstrated in the following section).

Clearly, the conditional distributions obtained for any arbitrary IM parameter from the GCIM approach can provide a significant constraint on the selection of ground motions. In fact, in many applications where as-recorded ground motions (with only amplitude scaling) are desired, it may not be possible to find a suite of N_{gm} ground motions which are not statistically different from the GCIM distributions. In such cases two options are available. The first option is to reduce the number of ground motions which are desired. The second option is to explicitly neglect the statistically significant differences between the ground motion suite and theoretical distribution for one of more IM_i.

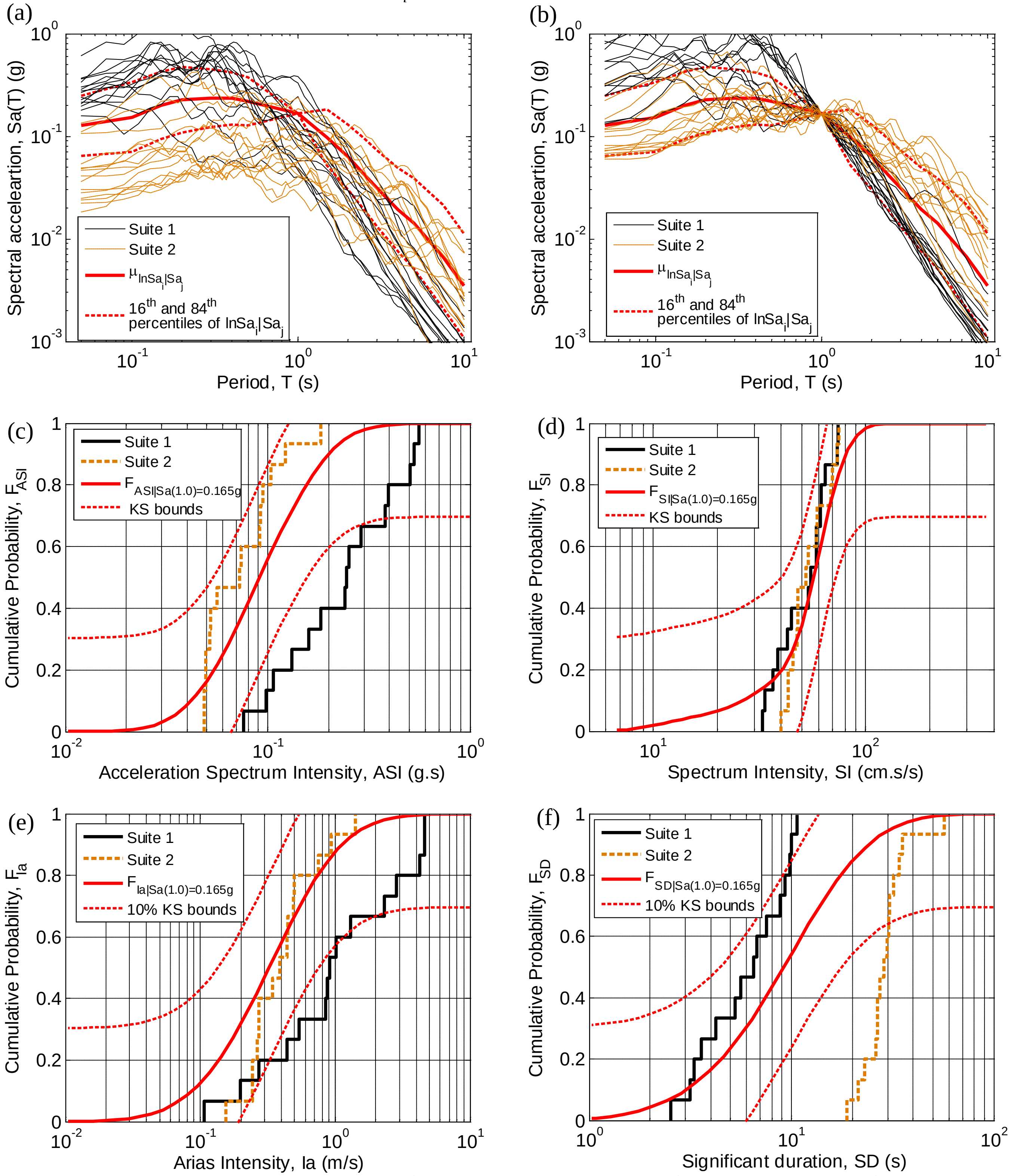

Figure 5: The appropriateness of two ground motion suites for the distribution of $\boldsymbol{IM}|Sa(1.0) = 0.165$g: (a) comparison of the unscaled suite spectra; (b) comparison of the scaled response spectra of the suites; and the distributions of (c) *ASI*; (d) *SI*; (e) I_a; and (f) *SD*.

Reducing the number of ground motions required in a suite provides two benefits in attempting to find a statistically consistent suite of ground motions. Firstly, given the finite number of possible as-recorded ground motions, reducing the number of desired ground motions allows the analyst to be more aggressive in the rejection of particular ground motions. Secondly, as the number of ground motions in the suite reduces the value of the critical test-statistic increases. Figure 6 illustrates the effect of reducing Suite 2 from 15 ground motions down to 7 ground motions (by removing those closest to the KS bounds) on the statistical significance of the distribution of *Sa*(0.05). It can be seen that originally Suite 2 was statistically significantly different than the theoretical *Sa*(0.05) distribution (i.e. the KS bounds and the EDF intersect in Figure 6). However, by reducing the required number of ground motions from 15 to 7, those ground motions causing the statistically significant result can be removed, as well as the critical KS test statistic value increasing (as indicated by the fact that the KS bounds for 7 ground motions are 'farther' from the theoretical distribution, than the KS bounds for 15 ground motions).

Neglecting certain differences between the characteristics of a ground motion suite and the theoretical characteristics of ground motions expected at the site is universally adopted in earthquake engineering research and practice. For example, selecting ground motions on the basis of their response spectra alone, as implemented in seismic design guidelines [e.g. 2, 37, 38], implicitly neglects all other characteristics of a ground motion that are not directly represented through its response spectrum. The main feature of this common approach to ground motion selection is that such neglection is implicit. Two key unknowns therefore are: (i) whether the distribution of these other characteristics of the selected ground motion suite are consistent with the theoretical distribution of such characteristics; and (ii) whether the seismic response analysis problem for which such ground motions are being selected for is dependent on any of these other characteristics. The GCIM approach previously outlined allows an analyst to determine the theoretical conditional distribution of any intensity measure, and therefore allows an explicit answer to (i) above. That is, for any arbitrary intensity measure, the EDF of the potential ground motions suite can be compared with the theoretical distribution (as in Figure 5). The analyst therefore can see explicitly when a potential ground motion suite provides a misrepresentation of a particular intensity measure. The answer to (ii) is discussed later in the manuscript.

It should be made clear that the two options of: reducing the number of ground motions which are desired; and explicitly neglecting the statistically significant differences between the ground motion suite and theoretical distribution for one of more IM_i, are not mutually exclusive. In fact, in most practical applications a combination of the two will most likely be employed. However, both of these options cause adverse effects. A fewer number of ground motions leads to a larger uncertainty in the estimated seismic response statistics (e.g. mean, standard deviation) obtained based on the seismic response analyses conducted, while neglection of statistically significant differences between the ground motion suite and the theoretical distribution for one or more intensity measures can lead to bias in the estimated seismic response statistics.

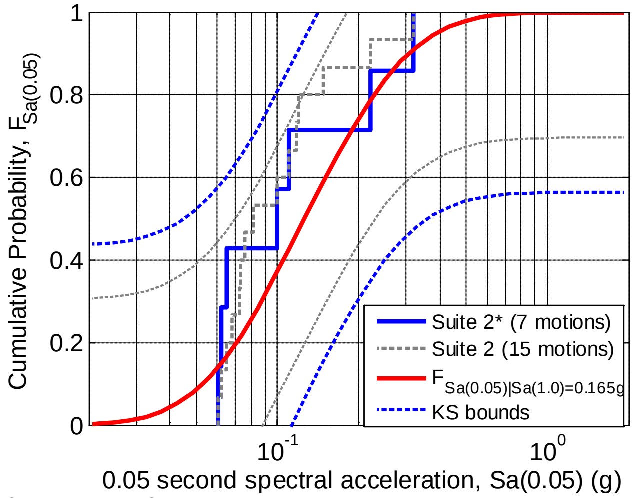

Figure 6: Effect of the size of the ground motion suite on its statistical significance against the GCIM distributions.

Appropriateness of modified and stochastically simulated motions and amplitude scaling

The additional constraint on the characteristics of ground motions expected at the site provided by the GCIM approach, is of great benefit in the examination of the appropriateness of modified (in both the time and frequency domain) and stochastically simulated ground motions. Ground motion modification and simulation procedures have long been criticised for potentially producing ground motions which are physically unrealisable. For example, Bommer *et al*. [39] note that there is no guarantee that stochastic-method [41] derived ground motions are physically realistic, since while the amplitude spectrum is quasi-realistic, the phase spectrum is often assumed entirely random. Frequency domain modification of ground motions are also known to significantly alter the velocity and displacement time histories of a ground motion and lead to unrealistically high energy content [4], something which time-domain approaches appear to circumvent [42]. Comparison of a potential suite of modified or stochastically simulated ground motions with the conditional distributions obtained from the GCIM approach allows an explicit examination of the physical appropriateness of the ground motion suite (e.g. a realistic energy content), and therefore offers the potential to reduce some of the scepticism of using such ground motion modification and simulation procedures.

It is also timely to discuss the issue of excessive (amplitude) scaling of as-recorded ground motions. Because of the conventional use of response spectra for examination of ground motions, and the fact that (elastic) response spectral ordinates scale linearly with (amplitude) scale factor, amplitude scaling of ground motions is an easy method to 'match' ground motions to a target response spectrum. Not surprisingly, the excessively scaling of ground motions can produce bias in the results of seismic response analysis, a topic which has received quantative attention recently [13, 43]. Such bias is primarily the result of the different ways in which ground motion intensity measures scale with seismological parameters such as earthquake magnitude and source-to-site distance, compared to the scaling of such intensity measures with amplitude scale factor. For example, *Sa*, *SI*, and *ASI* all scale linearly with amplitude scale factor, while I_a scales with the square of scale factor, and *SD* is independent of scale factor. Since *SD* is empirically observed to scale linearly with source-to-site distance (i.e. $SD \propto R$) and exponentially with earthquake magnitude (i.e. $SD \propto \exp(0.87M)$) [24], then clearly amplitude scaling of ground motions cannot reproduce the seismological scaling of the duration characteristics of ground motions (as evident from Figure 5f). Similar arguments can be made regarding the seismological scaling of response spectral ordinates and other ground motion measures.

Thus use of the GCIM approach can explicitly identify inappropriate ground motions which have been artificially generated, or excessively amplitude scaled based on comparisons with the GCIM distribution of various ground motion intensity measures.

BIAS IN SEISMIC RESPONSE ESTIMATES FROM INCOMPATIBLE GROUND MOTIONS

Once seismic response analyses are preformed using the final selected suite of ground motions, the resulting seismic responses of interest should be examined for bias. Bias can potentially occur if the distribution of one or more intensity measures of the ground motion suite differs from the theoretical distribution. This can be explained by considering Equation (15), an application of the total probability theorem, which gives the seismic demand distribution of an arbitrary engineering demand parameter (EDP) of interest:

$$\hat{f}_{EDP|IM_j}(edp|im_j) = \int f_{EDP|\boldsymbol{IM},IM_j}(edp|\boldsymbol{im},im_j)\hat{f}_{\boldsymbol{IM}|\boldsymbol{IM_j}}(\boldsymbol{im}|im_j)d\boldsymbol{IM} \tag{15}$$

where $\hat{f}_{\boldsymbol{IM}|\boldsymbol{IM_j}}(\boldsymbol{im}|im_j)$ is the distribution of $\boldsymbol{IM}$ of the ground motion suite when scaled to $IM_j = im_j$; $f_{EDP|\boldsymbol{IM},IM_j}(edp|\boldsymbol{im},im_j)$ is the distribution of $EDP|IM_j = im_j$ as a function of $\boldsymbol{IM}$; $\hat{f}_{EDP|IM_j}(edp|im_j)$ is the obtained distribution of $EDP|IM_j = im_j$; and the 'hat' symbols (e.g. $\hat{f}$) indicate an approximation. Thus if the distribution of the ground motion records, $\hat{f}_{\boldsymbol{IM}|\boldsymbol{IM_j}}(\boldsymbol{im}|im_j)$, does not match the 'true' distribution, $\boldsymbol{IM}|IM_j = im_j$, and $f_{EDP|\boldsymbol{IM},IM_j}(edp|\boldsymbol{im},im_j)$ is in fact dependent on $\boldsymbol{IM}$, then $\hat{f}_{EDP|IM_j}(edp|im_j)$ will be a biased estimate of the 'true' response distribution, $f_{EDP|IM_j}(edp|im_j)$. Those IM_i's in $\boldsymbol{IM}$ for which $\hat{f}_{IM_i|\boldsymbol{IM_j}}(im_i|im_j)$ is not consistent with $f_{IM_i|\boldsymbol{IM_j}}(im_i|im_j)$ was determined in the last section when ground motions are selected. For such IM_i's, it is therefore necessary to quantify the dependence of EDP on IM_i in order to determine if $\hat{f}_{EDP|IM_j}(edp|im_j)$ is biased. The following paragraphs explain a simple method of determining the dependence of EDP on IM_i.

The statistical dependence of EDP on IM_i can be obtained from regression analyses using the results of the seismic response analyses obtained with the adopted suite of ground motions. Given that EDPs are generally assumed to be lognormally distributed [e.g. 44, 45] a simple model is to use linear regression on $\ln EDP$ vs. $\ln IM_i$ (i.e. $\ln EDP = a + b\ln IM_i$). In such a relationship the parameter b represents the dependence of IM_i on EDP, and hence statistical tests should be used to determine its significance, and then consequently estimate the potential bias in

$\hat{f}_{EDP|IM_j}\left(edp|im_j\right)$.

If $\hat{f}_{IM_i|\boldsymbol{IM}_j}\left(im_i|im_j\right)$ does not equal $f_{IM_i|\boldsymbol{IM}_j}\left(im_i|im_j\right)$ and it is found that in the regression $\ln EDP = a + b\ln IM_i$, b is statistically rejected as having a value of zero, then it indicates that $\hat{f}_{EDP|IM_j}\left(edp|im_j\right)$ is a biased predictor of $f_{EDP|IM_j}\left(edp|im_j\right)$. In such cases, it is insightful to determine approximately the bias induced by this dependence of *EDP* on IM_i. The theoretical distribution of *EDP* given IM_j can be approximately computed from:

$$f_{EDP|IM_j}\left(edp|im_j\right) \approx \int f_{EDP|IM_i,IM_j}\left(edp|im_i, im_j\right) f_{IM_i|IM_j}\left(im_i|im_j\right) dIM_i \tag{16}$$

where $f_{IM_i|IM_j}\left(im_i|im_j\right)$ is the theoretical distribution of $IM_i|IM_j$ and $f_{EDP|IM_i,IM_j}\left(edp|im_i, im_j\right)$ is obtained from aforementioned regression analysis. Equation (16) is approximate as it assumes that $EDP|IM_j = im_j$ is a function of only IM_i (and not of any other terms in ***IM***). In the case of multiple IMs for which *EDP* is dependent on, the correlation between the different IMs needs to be considered [46]. Given that the purpose of Equation (16) in this application is merely to observe what possible bias exists in $\hat{f}_{EDP|IM_j}\left(edp|im_j\right)$ due to IM_i alone, then Equation (16) is considered sufficient. If it is further assumed that both $f_{EDP|IM_i,IM_j}\left(edp|im_i, im_j\right)$ and $f_{IM_i|IM_j}\left(im_i|im_j\right)$ have a lognormal distribution, i.e. $EDP|IM_i, IM_j \sim LN\left(a + b\ln IM_i, \sigma^2_{\ln EDP|IM_i,IM_j}\right)$, and $IM_i|IM_j \sim LN\left(\mu_{\ln IM_i|IM_j}, \sigma^2_{\ln IM_i|IM_j}\right)$ then Equation (16) further simplifies to:

$$EDP|IM_j \sim LN\left(a + b\mu_{\ln IM_i|IM_j}, \sigma^2_{\ln EDP|IM_i,IM_j} + b^2\sigma^2_{\ln IM_i|IM_j}\right) \tag{17}$$

Figure 7a illustrates, for a hypothetical seismic response problem, the dependence of peak free-field ground displacement, U_g, on significant duration, *SD*, based on the Suite 2 ground motions in Table 2. The p-value of $6.9\text{x}10^{-10}$ for *b* indicates that there is a statistically significant dependence of U_g on *SD*. Therefore, given that the Suite 2 ground motions have a distribution of *SD* which is different from the theoretical distribution obtained from the GCIM approach (Figure 5f), the distribution of *EDP* (i.e. U_g) obtained from the seismic response analysis is potentially biased. Figure 7b illustrates the EDF of U_g, its lognormal approximation, and the corrected distribution obtained from Equation (17). It can be seen that the corrected and uncorrected distributions are notably different. Firstly, because of the positive correlation of U_g and *SD*, and that the Suite 2 distribution of *SD* is 'larger' than the theoretical distribution of SD (i.e. Figure 5f), then the corrected distribution of U_g is generally "to the left" of the uncorrected distribution. Secondly, because the standard deviation in the theoretical distribution of *SD* is larger than the Suite 2 distribution of *SD* (Figure 5f), the corrected distribution of U_g also has a larger standard deviation than the uncorrected distribution.

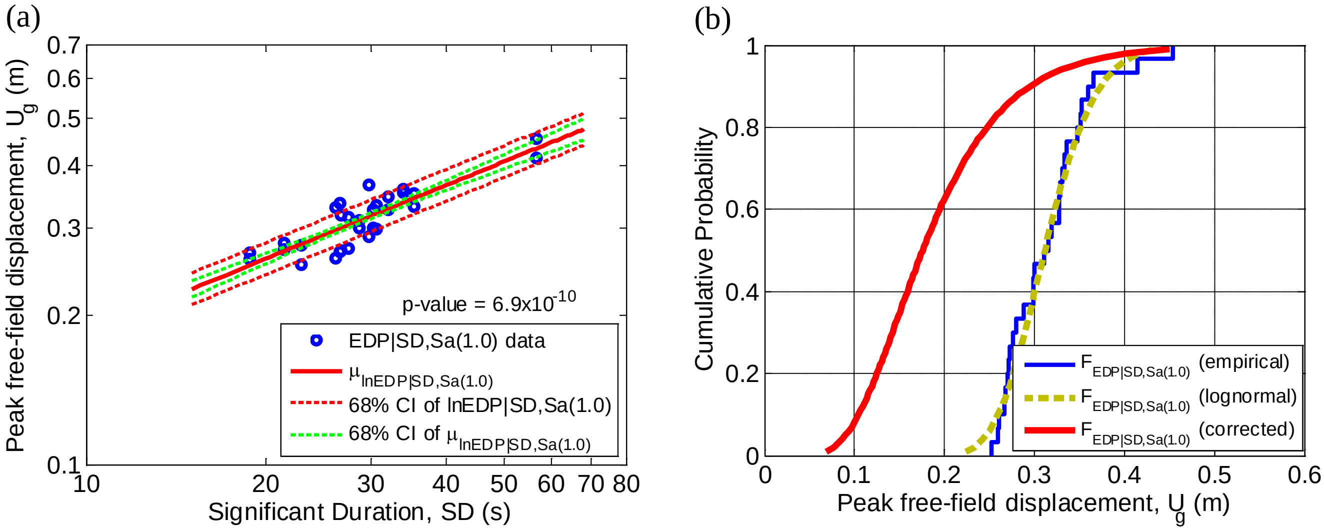

Figure 7: Hypothetical results (*EDP*)of a seismic response analysis and their dependence with IM_i: (a) Regression of peak free-field displacement, U_g, with *SD*; and (b) comparison of observed and corrected Ug distributions due to the dependence of U_g on *SD* and the difference between the *SD* statistics of the ground motion suite and theoretical distribution.

Note that Figure 7 and the related discussion consider the continuous IM, *SD*. Although not discussed here because of space limitations, in the case of discrete IMs, such as fault mechanism, equivalent discrete forms of Equations (15)-(17) can be used to determine the magnitude of any potential bias.

The estimation of seismic response bias due to improper ground motion selection, as given by Equations (15)-(17) and illustrated in Figure 7, therefore provides a simple method to examine the final consequences (in terms of bias in the estimated response) of the selected ground motion suite. Based on such information, an analyst could then decide whether the estimated bias is acceptable or whether a more suitable suite of ground motions is necessary.

CONCLUSIONS

A generalised conditional intensity measure (GCIM) approach was presented which allows the construction of the conditional distribution of any arbitrary ground motion intensity measure which may be considered as important in a particular seismic response analysis problem. Based on the obtained GCIM distribution, a holistic method of ground motion selection was presented. The holistic ground motion selection is based on statistical goodness-of-fit tests between the empirical distribution of the ground motion suite and theoretical distribution of a particular intensity measure obtained from the GCIM approach. A simple method in which the magnitude of any potential bias in the results of seismic response analyses when the ground motion suite does not conform to the GCIM distribution was also presented. The combination of the above features of the approach make it entirely holistic in that: any level of complexity in ground motion selection for any seismic response analysis can be exercised; users explicitly understand the simplifications made in the selected suite of ground motions; and an approximate estimate of any bias associated with such simplifications is obtained.

REFERENCES

[1] Cornell CA. Engineering seismic risk analysis. Bulletin of the Seismological Society of America 1968; 58(5): 1583–1606.

[2] NZS 1170.5. Structural design actions, Part 5: Earthquake actions - New Zealand. Standards New Zealand, Wellington, New Zealand, 2004.

[3] McGuire RK. Probabilistic seismic hazard analysis and design earthquakes: closing the loop. Bulletin of the Seismological Society of America 1995; 85(5): 1275-1284.

[4] Naeim F and Lew M. On the use of design spectrum compatible time histories. Earthquake Spectra 1995; 11(1): 111-127.

[5] Bommer JJ, Scott SG, and Sarma SK. Hazard-consistent earthquake scenarios. Soil Dynamics and Earthquake Engineering 2000; 19: 219-231.

[6] Baker JW and Cornell CA. Spectral shape, record selection and epsilon. Earthquake Engineering and Structural Dynamics 2006; 35(9): 1077-1095.

[7] Baker JW. The conditional mean spectrum: A tool for ground motion selection. Journal of Structural Engineering 2009: (in press).

[8] Jayaram N and Baker JW. Statistical tests of the joint distribution of spectral acceleration values. Bulletin of the Seismological Society of America 2008; 98(5): 2231-2243, DOI: 10.1785/0120070208.

[9] Baker JW. Vector-valued ground motion intensity measures for probabilistic seismic demand analysis. Ph.D Thesis, Department of Civil and Environmental Engineering Stanford University, 2005, 347pp.

[10] Shome N, Cornell CA, Bazzurro P, and Carballo JE. Earthquakes, records, and nonlinear responses. Earthquake Spectra 1998; 14(3): 469-500.

[11] Iervolino I and Cornell CA. Record selection for nonlinear seismic analysis of structures. Earthquake Spectra 2005; 21(3): 685-713, DOI: 10.1193/1.1990199.

[12] Baker JW and Cornell CA. A vector-valued ground motion intensity measure consisting of spectral acceleration and eplison. Earthquake Engineering and Structural Dynamics 2005; 34(10): 1193-1217.

[13] Luco N and Bazzurro P. Does amplitude scaling of ground motion records result in biased nonlinear structural drift responses? Earthquake Engineering and Structural Dynamics 2007; 36(13): 1813-1835, DOI: 10.1002/eqe.695.

[14] Krawinkler H. Cyclic loading histories for seismic experimentation on structural components. Earthquake Spectra 1996; 12(1): 1-12.

[15] Bradley BA. Epistemic uncertainty in component fragility functions. Earthquake Spectra 2009: (in press).

[16] Housner GW. Spectrum intensities of strong-motion earthquakes, in Symposium on earthquakes and blast effects on structures, Los Angeles, CA, 1952.

[17] Von Thun J, Roehm L, Scott G, and Wilson J. Earthquake ground motions for design and analysis of dams. Earthquake Engineering and Soil Dynamics II - Recent Advances in Ground-Motion Evaluation, Geotechnical Special Publication 1988; 20: 463-481.

[18] Arias A. A measure of earthquake intensity, in Seismic Design for Nuclear Power Plants, R. J. Hansen, Ed. Cambridge, MA: MIT Press, 1970, 438–483.

[19] Trifunac MD and Brady AG. A study on the duration of earthquake strong motion. Bulletin of the Seismological Society of America 1975; 65: 581-626.

[20] Boore DM and Atkinson GM. Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01s and 10.0s. Earthquake Spectra 2008; 24(1): 99-138.

[21] Bradley BA, Cubrinovski M, MacRae GA, and Dhakal RP. Ground motion prediction equation for spectrum intensity from spectral acceleration relationships. Bulletin of the Seismological Society of America 2009; 99(1): 277-285, DOI: doi: 10.1785/0120080044.

[22] Bradley BA. Site specific and spatially distributed prediction of acceleration spectrum intensity. Bulletin of the Seismological Society of America 2009: (in press).

[23] Travasarou T, Bray JD, and Abrahamson NA. Empirical attenuation relationship for Arias Intensity. Earthquake Engineering and Structural Dynamics 2003; 32: 1133-1155, DOI: 10.1002/eqe.270.

[24] Abrahamson NA and Silva WJ. Empirical ground motion models. Report to Brookhaven National Laboratory, 1996.

[25] Abrahamson NA and Silva WJ. Empirical response spectral attenuation relations for shallow crustal earthquakes. Seismological Research Letters 1997; 68(1): 94–126.

[26] Baker JW and Jayaram N. Correlation of spectral acceleration values from NGA ground motion models. Earthquake Spectra 2008; 24(1): 299-317.

[27] Goda K and Hong HP. Spatial correlation of peak ground motions and response spectra. Bulletin of the Seismological Society of America 2008; 98(1): 354-465, DOI: 10.1785/0120070078.

[28] Goda K and Atkinson GM. Probabilistic characterization of spatially correlated response spectra for earthquakes in Japan. Bulletin of the Seismological Society of America 2009; 99(5): (in press), DOI: 10.1785/0120090007.

[29] Baker JW and Cornell CA. Correlation of response spectral values for multi-component ground motions. Bulletin of the Seismological Society of America 2006; 96(1): 215-227.
[30] Baker JW. Correlation of ground motion intensity parameters used for predicting structural and geotechnical response, in 10th International Conference on Application of Statistics and Probability in Civil Engineering, Tokyo, Japan, 2007, 8pp.
[31] Ang AHS and Tang WH. Probability concepts in engineering: Emphasis on applications in civil and environmental engineering. John Wiley & Sons, 2007;
[32] Bazzurro P and Cornell CA. Disaggregation of seismic hazard. Bulletin of the Seismological Society of America 1999; 89(2): 501-520.
[33] Johnson NL and Kotz S. Distributions in statistics: Continuous multivariate distributions vol. 4. John Wiley & Sons, Inc., 1972;
[34] Naeim F, Alimoradi A, and Pezeshk S. Selection and scaling of ground motions time histories for structural design using genetic algorithms. Earthquake Spectra 2004; 20(2): 413–426, DOI: 10.1193/1.1719028.
[35] Gasparini D and Vanmarcke EH. SIMQKE: A program for artificial motion generation. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA, 1976.
[36] Kottke AR and Rathje EM. A semi-automated procedure for selecting and scaling recorded earthquake motions for dynamic analysis. Earthquake Spectra 2008; 24(4): 911-932, DOI: 10.1193/1.2985772.
[37] American Society of Civil Engineers. Minimum design loads for buildings and other structures. American Society of Civil Engineers/Structural Engineering Institute, Reston, VA, 2005.
[38] CEN. Eurocode 8: Design of structures for earthquake resistance. Part 1: General rules, seismisc actions and rules for buildings. Final Draft prEN 1998, European Committee for Standardization, Brussels, 2003.
[39] Bommer JJ, Stafford PJ, and Akkar S. Current empirical ground-motion prediction equations for Europe and their application to Eurocode 8. Bulletin of Earthquake Engineering 2009: (in press), DOI: 10.1007/s10518-009-9122-9.
[40] Bradley BA, Cubrinovski M, Dhakal RP, and MacRae GA. Intensity measures for the seismic response of pile foundations. Soil Dynamics and Earthquake Engineering 2009; 29(6): 1046-1058, DOI: 10.1016/j.soildyn.2008.12.002.
[41] Boore DM. Stochastic simulation of high-frequency ground motions based on seismological models of the radiated spectra. Bulletin of the Seismological Society of America 1983; 73(6): 1865-1894.
[42] Hancock J, Watson-Lamprey J, Abrahamson NA, Bommer JJ, Markatis A, McCoy E, and Mendis R. An improved method of matching response spectra of recorded earthquake ground motion using wavelets. Journal of Earthquake Engineering 2006; 10(S1): 67-89.
[43] Baker JW. Measuring bias in structural response caused by ground motion scaling, in 8th Pacific Conference on Earthquake Engineering, Singapore, 2007, 8pp.
[44] Aslani H and Miranda E. Probability-based Seismic Response Analysis. Engineering Structures 2005; 27(8): 1151-1163.
[45] Bradley BA, Dhakal RP, Cubrinovski M, and MacRae GA. Prediction of spatially distributed seismic demands in structures: ground motion and structural response. Earthquake Engineering and Structural Dynamics 2009: (avaliable online), DOI: 10.1002/eqe.954.
[46] Bradley BA, Dhakal RP, Cubrinovski M, and MacRae GA. Prediction of spatially distributed seismic demands in structures: from structural response to loss estimation. Earthquake Engineering and Structural Dynamics 2009: (in press).