Dependence of Correlations between Spectral Accelerations at Multiple Periods on Magnitude and Distance

A. Azarbakht[*,&,1], M. Mousavi[*], M. Nourizadeh[*] and M. Shahri[*]
[*] *Department of Civil Engineering, Faculty of Engineering, Arak University, Arak, Iran.*
[&] *Lehrstuhl fur Baustatik, RWTH Aachen, Aachen, Germany.*

Abstract:

In this paper the dependence of correlations between spectral accelerations at multiple periods on Magnitude (M) and Distance (R) has been investigated. For this purpose, a relatively large dataset of ground motion records (GMRs), containing 1551 records with a wide range of seismic characteristics, was selected. It is shown that the difference in the correlation coefficient is statistically meaningful when the general GMR dataset is divided into two subsets based on an arbitrary M or R. The observed difference is more meaningful in the case of magnitude when compared with distance. The general dataset of GMRs was then divided into four separate subsets based on optimum values of M and R, so that the four obtained subsets were given the greatest dissimilarity in terms of the correlation coefficients. The correlation coefficients between spectral accelerations at multiple periods were calculated in the case of the four subsets, and compared with the available correlations in the literature. The conditional mean spectrum was also calculated by means of the conventional correlation coefficients, as well as by using the proposed M and R dependent correlation coefficients. The results show that, despite the commonly available findings in the literature, this dependence is significant and should not be neglected in the conditional spectra calculation process.

Keywords: Correlation coefficient; Magnitude; Distance; Epsilon; Z-Fisher test; Confidence interval; Ground motion record.

1. INTRODUCTION

Ground motion selection is an important element in the dynamic analysis of structures where different criteria are recommended in different regulations. The Uniform Hazard Spectrum (UHS) is a commonly used target in most design codes and guidelines [1], [2]. It is an elastic spectrum at a site with a given hazard level in terms of annual probability of exceedance. However, most previous research results have shown that the UHS is not a reasonable representation of a real earthquake event [3], and is thus considered to be a conservative target by researchers e.g. [4]. As in the low period range the UHS is affected by strong ground motions, and weak earthquakes make the greatest contribution to UHS values in the high period range, the UHS has not satisfied users as a suitable target spectrum for the purpose of GMR selection, but is considered as a conservative target by researchers e.g. [3]. The UHS for an ideal site, i.e. a spectrum with only one possible M and R scenario, can be defined, at any desired level of hazard, as is given in Equation (1).

---

[1] Correspondence to: Alireza Azarbakht, Department of Civil Engineering, Faculty of Engineering, Arak University, Iran, P.O. Box 38156-88359. E-mail: a-azarbakht@araku.ac.ir

$$\mu_{\ln Sa(T_i)} = \mu_{\ln Sa}(M, R, \theta, T_i) + \varepsilon(T^*)\, \sigma_{\ln Sa}(T_i) \tag{1}$$

where $\mu_{\ln Sa(Ti)}$ is the natural logarithm of the expected spectral acceleration at a given period T_i, and $\mu_{\ln Sa}$ and $\sigma_{\ln Sa}$ are, respectively, the median and standard deviation values obtained from a ground motion model, T* is the target period, and θ is representative of other seismic characteristics. The ε(T*) parameter is the target epsilon value obtained by disaggregation analysis [5]. Epsilon measures the deviation of the spectral acceleration for a recorded ground motion with the spectral acceleration computed from a ground motion model [6]. Based on Equation (1), each single ordinate of the UHS corresponds to the same target epsilon, i.e. ε(T*). As the target epsilon is an indicator of the hazard, all UHS amplitudes have the same hazard level, which makes UHS a conservative target spectrum. On the other hand, it is very rare to find an earthquake event spectrum with the same hazard level for all periods [7]. In order to clarify, in Figure 1 two selected record spectra based on a given target period are compared with the UHS, which shows that there is clearly observable difference at other periods between them that the target period. This figure is an illustration of the fact that why the UHS does not represent a realistic earthquake event [4], [8], [9].

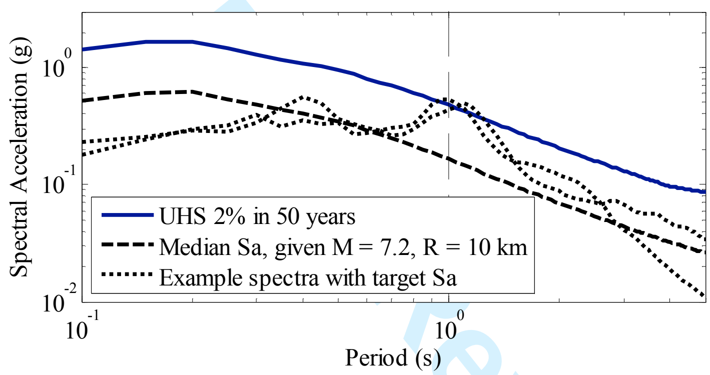

Figure 1: Median predicted spectrum using the CB08 ground motion model [10], where M = 7.2 and R = 10 km, UHS for 2% probability of exceedance in 50 years. The example record spectra are Colton-So Cal Edison and Anza Post Office, which were recorded during the Lytle Creek and San Fernando events respectively.

In order to deal with this problem the conditional mean spectrum (CMS) was introduced by Baker and Cornell [6, 11]), as a hazard based target response spectrum that could be used in structural analysis as an alternative to UHS. The correlation of epsilon values in the case of different periods was considered in the development of the CMS, so that expected values of epsilon could be predicted at different periods, given the value of the target epsilon at the period of interest. In fact, by predicting the epsilon values at other periods, a more realistic target spectrum (CMS) can be obtained. Based on probability calculations, Baker proposed that the expected epsilon values at any other period, $\varepsilon(T_i)$, should be taken as being equal to the target epsilon, ε(T*), multiplied by the correlation coefficient between the two epsilon values, as defined in Equation (2).

$$\varepsilon(T_i) = \rho\left(\varepsilon(T^*), \varepsilon(T_i)\right)\varepsilon(T^*) \tag{2}$$

The correlation coefficient between the epsilon values at two periods of interest can be obtained by using a maximum likelihood estimator, e.g. the Pearson product-moment correlation coefficient [12], as defined in Equation (3). The procedure can result in production of a matrix of correlation coefficients using a period range. The Pearson product-moment correlation coefficient can be employed since the epsilon values are normally distributed [13].

$$\rho(\varepsilon(T_1),\varepsilon(T_2)) = \frac{\sum_{i=1}^{m}(\varepsilon_i(T_1)-\mu_{\varepsilon(T_1)})(\varepsilon_i(T_2)-\mu_{\varepsilon(T_2)})}{\sqrt{\sum_{i=1}^{m}(\varepsilon_i(T_1)-\mu_{\varepsilon(T_1)})^2\sum_{i=1}^{m}(\varepsilon_i(T_2)-\mu_{\varepsilon(T_2)})^2}} \quad (3)$$

where m is the total number of observations (or GMRs); $\varepsilon_i(T_1)$ and $\varepsilon_i(T_2)$ are the epsilon values at T_1 and T_2 as the periods of interest, with respect to the record number i; $\mu_{\varepsilon(T_1)}$ and $\mu_{\varepsilon(T_2)}$ represent the sample means. Now, the CMS can be expressed mathematically as defined in Equation (4).

$$\mu_{\ln Sa(T_i)|\ln Sa(T^*)} = \mu_{\ln Sa}(M,R,\theta,T_i) + \rho\left(\varepsilon(T^*),\varepsilon(T_i)\right)\varepsilon * \sigma_{\ln Sa}(T_i) \quad (4)$$

Equation (4) indicates that the correlation coefficient plays a crucial role in the development of the CMS. This coefficient can be calculated using available closed-form models [14], [15] or by using a database of GMRs directly.

The selected dataset of GMRs, the chosen ground motion model, and the characteristics of the considered GMRs can have an effect on the correlation coefficient values. However, the conventional correlation coefficient is obtained based on a set of GMRs which cover a wide range of seismic characteristics. It is worth noting that the resulting CMS, which theoretically should correspond to a specific annual probability of exceedance, may not be adequately compatible with the employed correlation coefficients that are based on a general ground motion dataset. The effect of the ground motion model and ground motion dataset on the correlation coefficient values has shown to be negligible by a comparison of four Next Generation Attenuation (NGA) relationships [15]. The influence of some seismic parameters has also been investigated by researchers [16], [17]. Jayaram et al. (2008) showed that the effect of the ground motion model, the earthquake source mechanism, the seismic zone, the site conditions, and the source-to-site distance on estimated correlations was negligible in the case of a Japanese earthquake ground motion dataset [17]. In addition, Baker (2005) showed that, in the case of the classification of a dataset consisting of 534 GMRs into different subsets based on M and R, no trend could be seen, and therefore concluded that the correlation coefficient is not a function of M and R [16]. However, this conclusion was formed based on a relatively small ground motion database, as well as for some limited samples.

The current study focuses on a systematic investigation on the effect of magnitude and distance as two common key seismic characteristics of earthquakes on the correlation coefficients by means of using well-known statistical tests and a suitable ground motion dataset. The obtained results indicate that the mentioned seismic parameters have a meaningful influence on the correlation coefficients. This can be proved by investigating through all the target periods which can be followed in the next sections.

2. SELECTION OF THE GMRS DATASET, THE GROUND MOTION MODEL AND THE STATISTICAL TESTS

In the current study, the Campbell-Bozorgnia 2008 (CB08) model was employed as the attenuation model for the GMRs [10]. The GMRs dataset is the same as that employed in the development of the CB08 ground motion model, which contains 1551 orthogonal horizontal recordings from 64 earthquake events. In the definition of spectral acceleration, the 50^{th}-percentile was here selected, and denoted by "GMRotI50", as was the case when it was used in the NGA projects [18]. Note that 1551 available GMRs were obtained from the Pacific Earthquake Engineering Research (PEER) centre [19]. The CB08 model used a subset of the PEER NGA database, and excluded recordings that were believed to be inappropriate for estimating free-field ground motions from shallow earthquake main shocks in active tectonic regimes. The resulting equations are valid for magnitudes ranging between 4.0 to 7.5–8.5 (depending on the fault mechanism), and for distances ranging between 0 to 200 km. The model explicitly includes the effects of magnitude saturation, magnitude-dependent attenuation, style of faulting, rupture depth, hanging-wall geometry, linear and nonlinear site response, 3-D basin response, and inter-event and intra-event variability. Note that modern ground motion attenuation models such as CB08 are usually represented in the following form:

$$\ln Sa(T) = \mu_{\ln Sa}(M, R, \theta, T) + r_{ij} \tag{5}$$

where $\ln S_a(T)$ is the observed natural logarithmic spectral acceleration at a period of T, $\mu_{\ln Sa}$ is the predicted S_a, which is a function of magnitude (M), distance (R) and other seismic properties (θ), and the term r_{ij} is representative of the total residual between the observed logarithmic spectral acceleration and its predicted median value. It is worth noting that, in modern ground motion prediction models, r_{ij} is divided into two parts. The first part denotes the inter events residual and the second part represents the intra events residuals of the j^{th} recording for the i^{th} earthquake event.

As discussed in the following sections, hypothesis test criteria were used in this study which assume independence of the employed observations. Hence, employment of the total residuals might affect the final results, since they are not independent. Therefore, as the observations are not completely independent, as a consequence of common inter event residuals (or, in other words, the total residuals are correlated for a single earthquake), the current study was focused on the intra event residuals which are statistically independent. All of the employed statistical tests were hereafter based on the intra event residuals. The procedure for the calculation of intra-event residuals is given in [20].

As mentioned previously, it is necessary to use statistical tools to make logical comparisons between the obtained results in order to validate the findings. This study focuses on comparing the correlation coefficient values, and a powerful statistical test needs to be employed in order to determine whether the difference between the obtained coefficients is significant or not. In the given case the Z-Fisher test was used during the comparison process [21]. It is worth mentioning that the Z-Fisher test has been employed in several similar studies [16]. The null hypothesis (H_0) was defined as the equality of the two resulting correlation of coefficients, i.e. H_0: $\rho_1=\rho_2$. The value of ρ_1 resulted from subset number one, whereas ρ_2 was obtained from subset number two. Based on this procedure, the following statistics were calculated as shown below:

$$z = \frac{\rho_1' - \rho_2'}{\sqrt{\dfrac{1}{n_1 - 3} + \dfrac{1}{n_2 - 3}}} \quad , \quad \rho' = 0.5Ln\left|\frac{1+\rho}{1-\rho}\right| \tag{6}$$

where n_1, and n_2 represent the sample sizes of the two considered subsets, and ρ is the correlation coefficient value which can be calculated for each arbitrary subset by using

Equation (3). The null hypothesis was rejected for p-values lower than a significant level (i.e. 0.05). The p-value indicates the lowest level of significance that would lead to the rejection of the null hypothesis with the given data as written in Equation (7). Therefore, p-values lower than 0.05 indicates a significant difference between ρ_1 and ρ_2.

$$p-value = 2(1-\Phi(z)) \tag{7}$$

where Φ is the standard normal cumulative distribution function.

In order to illustrate the statistical uncertainty, due to the finite number of GMRs located within a subset, another statistical tool was also used, i.e. the Confidence Interval (CI) method [12]. CI, as written mathematically in Equation (8), is a type of interval that is estimated for a population parameter, and is used to indicate the reliability of an estimate. CI consists of a range of values (i.e. an interval) which act as good estimates of the unknown population parameter. If ρ_2 falls within the acceptance region of $\rho_{1,}$ this means that the difference in the estimated correlations, by using the two subsets, is statistically insignificant. On the other hand, if the confidence intervals of two datasets do not overlap, then, they are necessarily statistically significantly different. To calculate confidence interval, one needs to set confidence level as 90%, 95%, or 99% and etc. in which the most commonly used confidence level is 95%.

$$CI = X \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right) \tag{8}$$

where CI is representing lower and upper bounds (acceptance region), X is the sample mean, σ is the standard deviation of the samples and n is the sample size.

3. THE SIGNIFICANCE OF A STATISTICAL DIFFERENCE BETWEEN CORRELATION COEFFICIENTS

The general GMRs dataset was divided into seven subsets based on Magnitude. For example, the correlation coefficients of epsilon at T=1s with three arbitrary epsilons, i.e. at T=0.5, 2 and 4 s, are shown in Figure 2 versus the mean magnitude of the seven subsets. As can be seen in Figure 2, there is a clear trend between the mean magnitudes and the correlation coefficients. The obtained p-value for the fitted line is also an indicator of the significance of the correlations. Here, the p-value was defined as the likelihood of observing a slope coefficient equal to or greater than α, if the value of α is in fact zero. The α parameter was defined as the slope of the linear regression line. Lower p-values mean that the variation of the correlation coefficients within the subsets is meaningful. On the other hand higher p-values show that the variation of correlation coefficients, within the subset, regarding magnitude is not noticeable. This procedure was repeated for all combinations of periods within the dataset (The employed period range in the current study is 0.05 to 5 sec). The results show that 80 percent of the cases have p-values of less than 0.05. This means that the variation is significant in 80 percent of all the cases.

The above justifications make it clear that the claimed dependence of the correlation coefficients on the magnitude is reasonable. Increasing the number of subsets leads to a complicated problem since only a few GMRs will be available within some subsets. To better deal with this problem, and for the purpose of simplicity, the whole dataset was divided into two subsets based on an arbitrary magnitude i.e. say M=7. Then, the two obtained correlation matrixes based on the two divided subsets were compared by using an appropriate statistical test (the Z-Fisher test) as discussed in the former section. Figure 3 illustrates the variations of the correlation coefficient values by means of all the GMRs, and the two subsets. Note that each curve in Figure 3 is a row of the corresponding correlation coefficient matrix. As can be seen in Figure 3, it is clear that the difference between the correlation values is meaningful, which is also in agreement with the results based on Figure 2. It is worth noting that when computing these correlations, spectral values at all periods from all ground motions were used, without considering the potentially limited usable period range of individual ground motions due to filtering. The impact of this analysis decision has not yet been investigated.

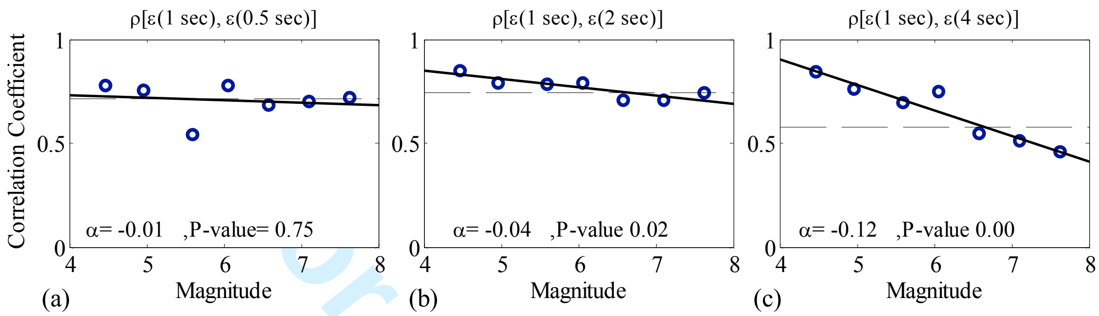

Figure 2: Correlation coefficients variations for the epsilon values at a period of 1 second versus magnitude with the epsilon values at (a) T=0.5sec, (b) T=2sec, (c) T=4sec.

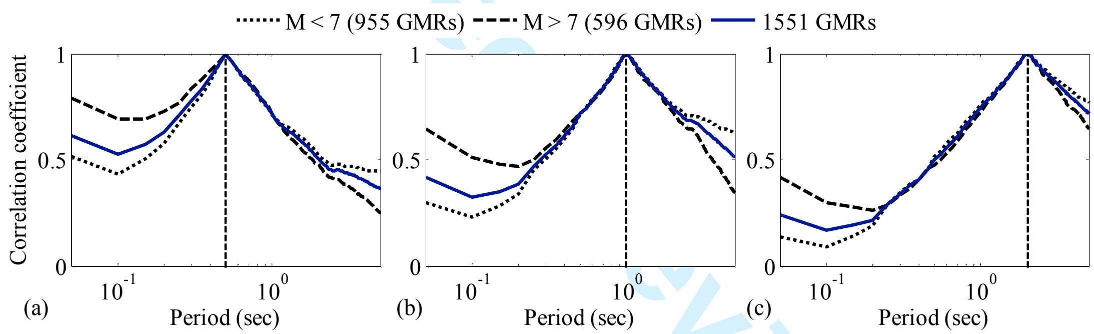

Figure 3: Variations of the correlation coefficients corresponding to (a) T* = 0.5 sec (b) T* = 1.0 sec (c) T* = 2.0 sec.

The observed difference can be quantified by means of a statistical test within the scope of a systematic procedure. For this purpose all the elements (i.e. the correlation coefficient values) in the correlation matrix corresponded to a subset were compared to another, one by one by using the Z-Fisher test. Each comparison of pairs of elements leads to a specified p-value that shows the significance of the difference. Here, p-values lower than 0.05 were defined as significant observations. By dividing the number of significant cases by the total number of all cases, the Significance Percentage (SP) was defined as a judgement index. The SP in the case of Figure 3 is equal to 75 percent which is obviously meaningful.

Different SP indices can clearly be obtained by assuming different arbitrary magnitude values for dividing the database. Additionally, the SP index can be calculated for different seismic parameters i.e. M, R and others. The variation of the SP index versus the different dividing values is shown in Figure 4 in the case of magnitude and distance. It is worth mentioning that the dividing value was chosen so that any of the two subsets contained at least 20 percent of the total number of GMRs, i.e. 300 GMRs in the current study. As can be seen in Figure 4, the SP index is much more significant in the case of magnitude in comparison with distance, which is also in agreement with the findings of [17].

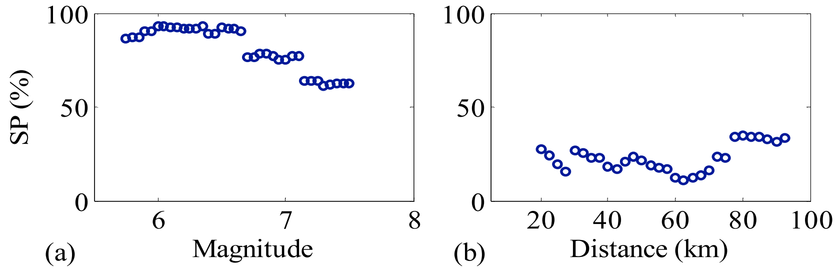

Figure 4: Variations of the SP value obtained by employing Z-fisher test for different boundary conditions versus (a) Magnitude (b) Distance.

Low numbers of GMRs in each subset against another subset might give rise to concerns about statistical uncertainty. Although the number of cases is accounted for within the Z-Fisher test algorithm, a complementary statistical test was also employed in this study. As discussed in the former section, the confidence interval is a statistical tool that illustrates the statistical uncertainty due to the finite number of GMRs.

Again the general dataset was divided into two subsets based on M=7. Then, the confidence intervals for the two subsets (M > 7 and M < 7) are shown in Figure 5 in the case of the three target periods (0.5, 1 and 2 sec). It can be seen that the confidence intervals do not overlap in some cases. To increase the accuracy of the investigation, the observed difference can again be quantified, as was done in the case of the Z-Fisher procedure. The number of non-overlap cases out of the total number of cases was defined as the Confidence Interval Significance Percentage (CISP) for a specified target period. This procedure can be developed to cover all the target periods, and it finally resulted in 68.5% as a CISP index. This value of CISP is noticeable enough, similar to the Z-Fisher test, as had been expected.

The CISP index was calculated for different values for the division of the global dataset. The variation of the CISP index versus the different dividing values is shown in Figures 6a and 6b, respectively, in the case of magnitude and distance. As can be seen in Figures 6a and 6b, the CISP index is much more significant in the case of magnitude when compared with the distance parameter, as was seen in the Z-Fisher test.

The results which were obtained based on the employed Z-Fisher and the confidence interval tests indicate that the differences between the correlation coefficients are considerable. In other words, at least the effect of magnitude is significant enough that this cannot be ignored. Thus, in the next section, a new approach is proposed in order to take this significance into account.

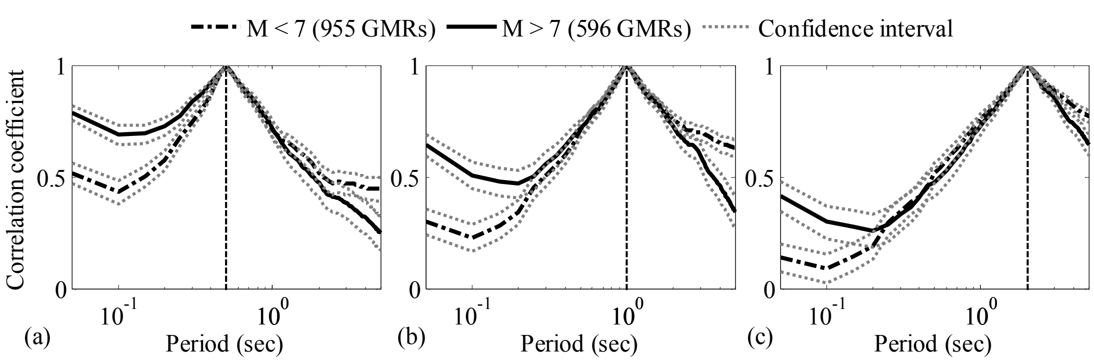

Figure 5: Variations of the correlation coefficients together with the confidence intervals corresponding to
(a) T* = 0.5 sec (b) T* = 1.0 sec (c) T* = 2.0 sec.

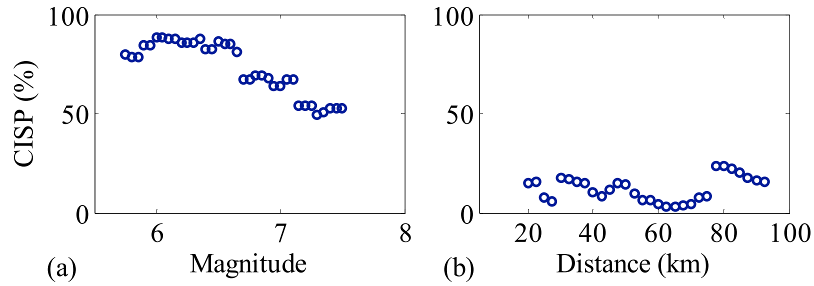

Figure 6: Variations of the CISP value obtained by employing the CI test for different boundary conditions versus (a) Magnitude (b) Distance.

4. DEVELOPMENT OF CORRELATION COEFFICIENTS DEPENDENT ON M AND R

It was shown in the previous section that the correlation of spectral acceleration depends on each of the individual M and R parameters. In other words, the correlation coefficient is statistically different when the general dataset is divided into two subsets, based on an arbitrary value of M or R. However distance was found to have less impact than magnitude. In this section M and R are taken into account simultaneously in order to account for any interaction between them. The statistical difference was quantified by a SP index and the Z-Fisher test. For this reason two arbitrary values were assumed for the parameters M and R, which result in four different subsets, (e.g. M < 6.5 and R < 50km, M < 6.5 and R > 50km, M > 6.5 and R < 50km, M > 6.5 and R > 50km). The SP index was also employed here in order to quantify the difference between each subset and the other three subsets. Thus, six independent possibilities are available for the purpose of comparison of the four subsets i.e. subset 1 with subsets 2, 3 and 4; subset 2 with subsets 3 and 4; and subset 3 with subset 4. The subdivision of a general dataset into four subsets is shown in Figure 7a, and the correlation coefficient versus period graph is shown in Figure 7b in the case of T*=0.5s. As can be clearly seen from Figure 7b, the correlation coefficients for the four subsets are significantly different, which again confirms the results of the previous section. The SP indices are also depicted in Figure 7b. The SP index for the subsets which only differ by distance is low, whereas it is remarkably high for the subsets having different magnitudes. The mean of six SP indices is 63.9%, which is again noticeable.

The results are clearly a function of the chosen values for M and R. A simple optimization procedure was therefore used in order to obtain the critical M and R values, so that the mean of the six SP indices was maximized. The obtained critical M and R values were appropriate values to divide the general set into four subsets that have maximum dissimilarity with each other in terms of obtaining the correlation coefficients i.e. M=6.33 and R=19.47 km in the current study. This division was able to produce a foundation in order to support the proposition of the correlation coefficients which are dependent on M and R. The correlation coefficients versus period are shown in Figures 8a, 8b and 8c, respectively, in the case of T* = 0.5s, 1s and 2s. As can be seen in Figure 8, the difference in the correlation coefficients is, as a rule, more important when the distance between two periods is higher. The mean of the SP in the optimized case is 70.5 %. The proposed correlation coefficients are shown in detail in Tables Ia to Id, for the four above-mentioned subsets.

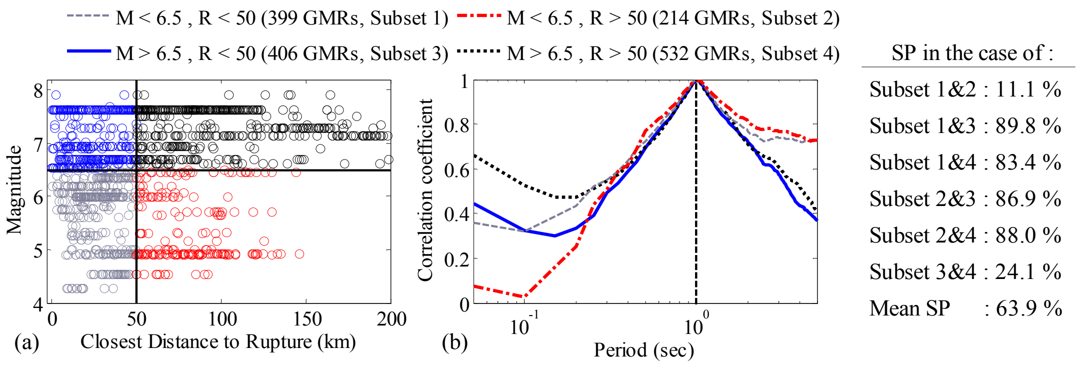

Figure 7: (a) Division of the general dataset into four subsets
(b) Correlation coefficient variation versus period for T*=1sec.

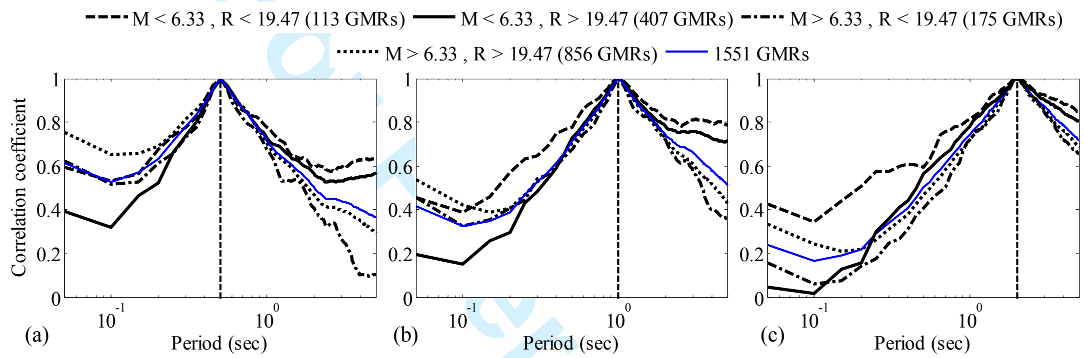

Figure 8: Variations of the correlation coefficients obtained from the different subsets corresponding to
(a) T* = 0.5 sec (b) T* = 1.0 sec (c) T* = 2.0 sec.

5. SENSITIVITY OF THE RESULTS TO THE GMRS DATASET

A well-known and suitable dataset was selected in this study in order to be compatible enough with the employed ground motion model (CB08). However, another concern that may arise is that the results could be sensitive to any changing of the chosen database. For this reason a sensitivity analysis was also performed in order to investigate how sensitive the results are to any changes in the database. For this purpose, the bootstrap method was used to analyze how significant the resulting correlation coefficient is [22]. The application of the bootstrap method involves sampling with replacements from one of the four subsets to generate an arbitrary number of alternate subsets of GMRs, which are hereinafter called bootstrap subsets, and to calculate the correlation coefficients for the thus generated bootstrap subsets. From such samples of subsets, it is easy to calculate the standard error and the confidence intervals for the correlation coefficients. To clarify, the M < 6.33 R > 19.47 subset was taken into consideration, which contains 407 GMRs. 1000 samples were generated as bootstrap subsets within this given subset. Thus, 1000 different correlation coefficients were calculated in which the upper and lower bounds are shown in Figure 9. As can be seen from Figure 9, the coefficient correlation based on the general dataset, at least within a wide range of interest, is not located within the upper and lower bounds region. This conclusion is still more meaningful when the two assumed periods in the correlation confident calculation procedure are farther apart from one other.

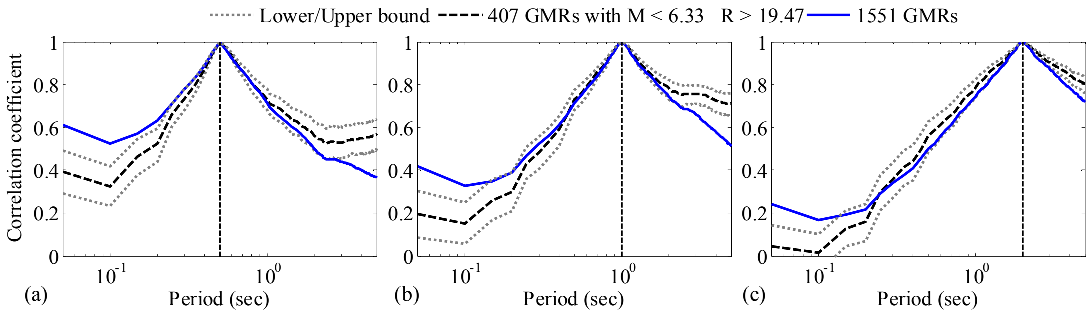

Figure 9: Variations of the correlation coefficients and the corresponding confidence interval based on the general dataset and the subset with M<6.33 and R>19.47,
(a) T* = 0.5 sec (b) T* = 1.0 sec (c) T* = 2.0 sec.

6. APPLICATION OF THE PROPOSED CORRELATION COEFFICIENTS TO THE CONDITIONAL MEAN SPECTRUM

The correlation coefficients, which have been discussed in the former sections, are commonly employed in the development of ground motion attenuation equations, as well as in obtaining target spectra e.g. CMS [11]. As illustrated in the previous sections, CMS is usually obtained by means of seismic hazard disaggregation and a chosen correlation model between spectral accelerations, e.g. that proposed by Baker and Jayaram 2008 [15]. Although seismic hazard disaggregation outputs are strongly tied to a specific seismic hazard level, i.e. a given return period. Nevertheless the employed correlation coefficient is not commonly obtained based on a pre-defined hazard level. On the other hand, it has been shown in the previous sections that the correlation coefficient can be affected by M and R. So it is more reasonable to propose and employ a new correlation coefficient. The proposed correlation coefficients, obtained in the four pre-mentioned cases, are shown, respectively, in Figures 10c to 10f. The Baker and Jayaram 2008 correlation coefficient is also shown in Figure 10b for comparison purposes. Additionally, CMS cases based on the conventional correlation coefficient as well as on the proposed correlation coefficient (see Table I) are shown in Figure 11 in the case of T*=0.5s, 1s and 2s. As can be seen in Figure 11, the difference between the conventional CMS and the proposed CMS is significant in some cases, especially when the target period is high. The conventional CMS uses the correlations obtained by employing all GMRs, whereas the enhanced CMS uses the correlation coefficients obtained from a GMRs subset containing seismic parameters which are compatible with a given scenario.

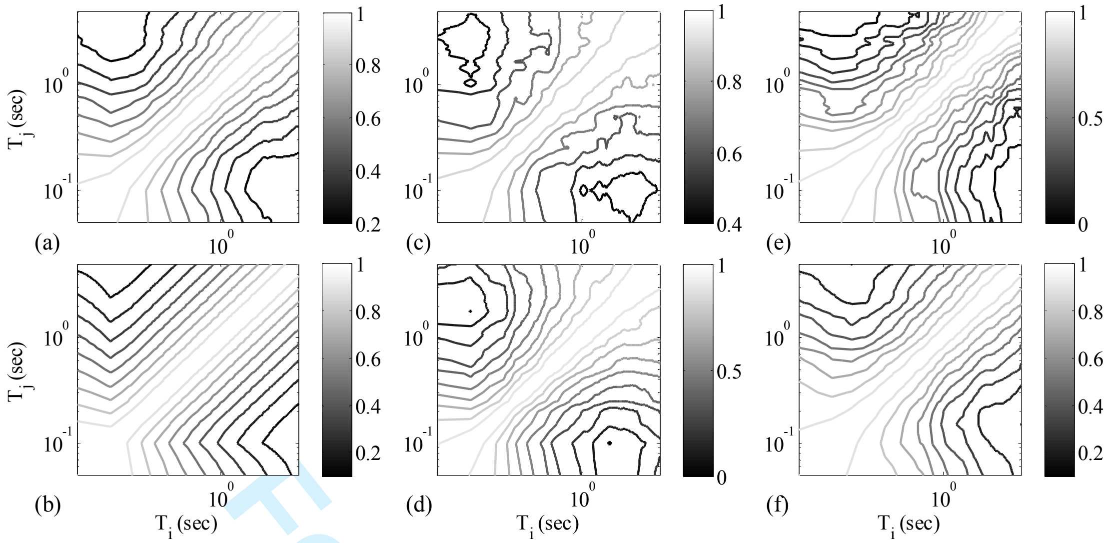

Figure 10: Contour plot of the correlation coefficients corresponding to (a) All the 1551 GMRs (b) The Baker and Jayaram 2008 proposed model (c) GMRs with M < 6.33 and R < 19.47 (d) GMRs with M < 6.33 and R > 19.47 (e) GMRs with M > 6.33 and R < 19.47 (f) GMRs with M > 6.33 and R > 19.47.

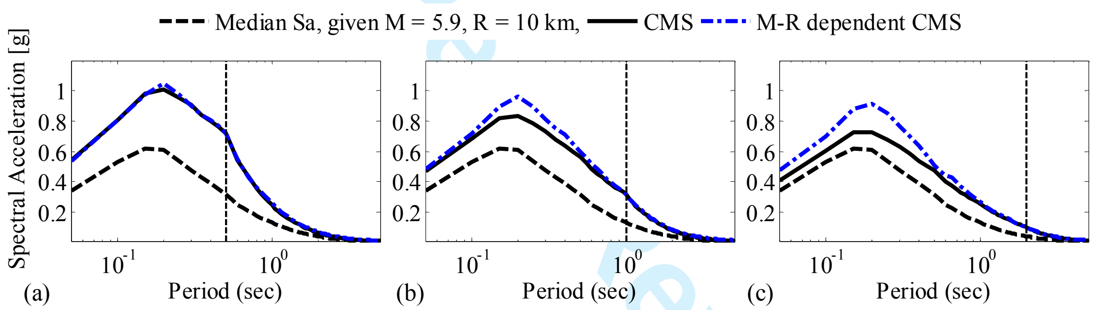

Figure 11: Conditional mean spectrum obtained by employing the coefficient correlations based on the general dataset and the proposed approach; the hazard level corresponds to M = 5.9 and R = 10, (a) T* = 0.5 sec (b) T* = 1.0 sec (c) T* = 2.0 sec.

Table I-a: The correlation coefficients obtained from GMRs with M < 6.33 and R < 19.47.

T_j ↓ \ T_i →	0.05	0.1	0.15	0.2	0.3	0.4	0.5	0.75	1	1.5	2	2.5	3	4	5
0.05	1	0.93	0.88	0.83	0.71	0.63	0.60	0.55	0.45	0.45	0.42	0.41	0.41	0.47	0.49
0.1	0.93	1	0.91	0.83	0.68	0.59	0.53	0.51	0.39	0.38	0.34	0.33	0.33	0.38	0.40
0.15	0.88	0.91	1	0.89	0.75	0.62	0.57	0.55	0.46	0.45	0.44	0.41	0.41	0.43	0.44
0.2	0.83	0.83	0.89	1	0.83	0.71	0.68	0.65	0.57	0.53	0.51	0.48	0.48	0.51	0.52
0.3	0.71	0.68	0.75	0.83	1	0.87	0.77	0.69	0.65	0.59	0.58	0.55	0.57	0.57	0.57
0.4	0.63	0.59	0.62	0.71	0.87	1	0.90	0.77	0.75	0.67	0.61	0.59	0.60	0.63	0.65
0.5	0.60	0.53	0.57	0.68	0.77	0.90	1	0.84	0.78	0.64	0.60	0.58	0.59	0.63	0.63
0.75	0.55	0.51	0.55	0.65	0.69	0.77	0.84	1	0.91	0.82	0.77	0.74	0.74	0.73	0.74
1	0.45	0.39	0.46	0.57	0.65	0.75	0.78	0.91	1	0.88	0.82	0.78	0.81	0.80	0.79
1.5	0.45	0.38	0.45	0.53	0.59	0.67	0.64	0.82	0.88	1	0.93	0.88	0.87	0.85	0.83
2	0.42	0.34	0.44	0.51	0.58	0.61	0.60	0.77	0.82	0.93	1	0.95	0.92	0.88	0.84
2.5	0.41	0.33	0.41	0.48	0.55	0.59	0.58	0.74	0.78	0.88	0.95	1	0.97	0.93	0.90
3	0.41	0.33	0.41	0.48	0.57	0.60	0.59	0.74	0.81	0.87	0.92	0.97	1	0.96	0.93

4	0.47	0.38	0.43	0.51	0.57	0.63	0.63	0.73	0.80	0.85	0.88	0.93	0.96	1	0.98
5	0.49	0.40	0.44	0.52	0.57	0.65	0.63	0.74	0.79	0.83	0.84	0.90	0.93	0.98	1

Table I-b: The correlation coefficients obtained from GMRs with M < 6.33 and R > 19.47.

T_i →

T_j ↓	0.05	0.1	0.15	0.2	0.3	0.4	0.5	0.75	1	1.5	2	2.5	3	4	5
0.05	1	0.89	0.83	0.74	0.58	0.47	0.39	0.27	0.19	0.07	0.04	0.04	0.08	0.15	0.22
0.1	0.89	1	0.84	0.73	0.53	0.42	0.32	0.23	0.15	0.04	0.02	0.02	0.05	0.12	0.18
0.15	0.83	0.84	1	0.87	0.65	0.54	0.46	0.34	0.26	0.16	0.13	0.14	0.17	0.23	0.29
0.2	0.74	0.73	0.87	1	0.76	0.60	0.52	0.38	0.30	0.19	0.16	0.17	0.19	0.25	0.30
0.3	0.58	0.53	0.65	0.76	1	0.85	0.74	0.57	0.48	0.39	0.36	0.35	0.36	0.40	0.45
0.4	0.47	0.42	0.54	0.60	0.85	1	0.87	0.70	0.59	0.50	0.44	0.42	0.42	0.45	0.48
0.5	0.39	0.32	0.46	0.52	0.74	0.87	1	0.84	0.73	0.63	0.56	0.53	0.53	0.55	0.57
0.75	0.27	0.23	0.34	0.38	0.57	0.70	0.84	1	0.88	0.76	0.68	0.65	0.65	0.65	0.65
1	0.19	0.15	0.26	0.30	0.48	0.59	0.73	0.88	1	0.86	0.78	0.75	0.76	0.72	0.71
1.5	0.07	0.04	0.16	0.19	0.39	0.50	0.63	0.76	0.86	1	0.91	0.85	0.82	0.76	0.73
2	0.04	0.02	0.13	0.16	0.36	0.44	0.56	0.68	0.78	0.91	1	0.95	0.90	0.84	0.80
2.5	0.04	0.02	0.14	0.17	0.35	0.42	0.53	0.65	0.75	0.85	0.95	1	0.96	0.90	0.87
3	0.08	0.05	0.17	0.19	0.36	0.42	0.53	0.65	0.76	0.82	0.90	0.96	1	0.94	0.91
4	0.15	0.12	0.23	0.25	0.40	0.45	0.55	0.65	0.72	0.76	0.84	0.90	0.94	1	0.97
5	0.22	0.18	0.29	0.30	0.45	0.48	0.57	0.65	0.71	0.73	0.80	0.87	0.91	0.97	1

Table I-c: The correlation coefficients obtained from GMRs with M > 6.33 and R < 19.47.

T_i →

T_j ↓	0.05	0.1	0.15	0.2	0.3	0.4	0.5	0.75	1	1.5	2	2.5	3	4	5
0.05	1	0.92	0.87	0.83	0.76	0.72	0.62	0.58	0.46	0.25	0.16	0.10	0.04	-0.08	-0.10
0.1	0.92	1	0.92	0.85	0.73	0.66	0.52	0.46	0.33	0.15	0.06	0.02	-0.02	-0.07	-0.09
0.15	0.87	0.92	1	0.93	0.79	0.69	0.53	0.46	0.35	0.18	0.07	0.03	-0.01	-0.05	-0.08
0.2	0.83	0.85	0.93	1	0.85	0.74	0.61	0.50	0.41	0.23	0.14	0.09	0.04	-0.01	-0.03
0.3	0.76	0.73	0.79	0.85	1	0.87	0.73	0.62	0.54	0.37	0.25	0.19	0.16	0.06	0.05
0.4	0.72	0.66	0.69	0.74	0.87	1	0.86	0.71	0.60	0.46	0.31	0.23	0.17	0.02	0.03
0.5	0.62	0.52	0.53	0.61	0.73	0.86	1	0.79	0.68	0.53	0.40	0.33	0.27	0.10	0.10
0.75	0.58	0.46	0.46	0.50	0.62	0.71	0.79	1	0.84	0.67	0.57	0.48	0.42	0.24	0.22
1	0.46	0.33	0.35	0.41	0.54	0.60	0.68	0.84	1	0.83	0.72	0.61	0.58	0.40	0.36
1.5	0.25	0.15	0.18	0.23	0.37	0.46	0.53	0.67	0.83	1	0.85	0.73	0.69	0.53	0.46
2	0.16	0.06	0.07	0.14	0.25	0.31	0.40	0.57	0.72	0.85	1	0.92	0.84	0.72	0.66
2.5	0.10	0.02	0.03	0.09	0.19	0.23	0.33	0.48	0.61	0.73	0.92	1	0.94	0.84	0.80
3	0.04	-0.02	-0.01	0.04	0.16	0.17	0.27	0.42	0.58	0.69	0.84	0.94	1	0.91	0.86
4	-0.08	-0.07	-0.05	-0.01	0.06	0.02	0.10	0.24	0.40	0.53	0.72	0.84	0.91	1	0.94
5	-0.10	-0.09	-0.08	-0.03	0.05	0.03	0.10	0.22	0.36	0.46	0.66	0.80	0.86	0.94	1

Table I-d: The correlation coefficients obtained from GMRs with M > 6.33 and R > 19.47.

T_i →

T_j ↓	0.05	0.1	0.15	0.2	0.3	0.4	0.5	0.75	1	1.5	2	2.5	3	4	5
0.05	1	0.93	0.89	0.86	0.82	0.79	0.75	0.64	0.54	0.44	0.33	0.28	0.26	0.21	0.15
0.1	0.93	1	0.92	0.86	0.77	0.71	0.65	0.53	0.42	0.32	0.24	0.20	0.17	0.15	0.11

0.15	0.89	0.92	1	0.92	0.81	0.73	0.65	0.52	0.38	0.29	0.21	0.17	0.15	0.14	0.10
0.2	0.86	0.86	0.92	1	0.86	0.78	0.69	0.54	0.41	0.31	0.22	0.18	0.17	0.15	0.11
0.3	0.82	0.77	0.81	0.86	1	0.88	0.81	0.67	0.53	0.41	0.31	0.27	0.25	0.23	0.18
0.4	0.79	0.71	0.73	0.78	0.88	1	0.90	0.74	0.61	0.49	0.38	0.32	0.31	0.27	0.22
0.5	0.75	0.65	0.65	0.69	0.81	0.90	1	0.84	0.71	0.58	0.47	0.41	0.40	0.35	0.29
0.75	0.64	0.53	0.52	0.54	0.67	0.74	0.84	1	0.86	0.72	0.61	0.55	0.51	0.44	0.38
1	0.54	0.42	0.38	0.41	0.53	0.61	0.71	0.86	1	0.83	0.72	0.65	0.61	0.52	0.43
1.5	0.44	0.32	0.29	0.31	0.41	0.49	0.58	0.72	0.83	1	0.88	0.79	0.74	0.66	0.57
2	0.33	0.24	0.21	0.22	0.31	0.38	0.47	0.61	0.72	0.88	1	0.91	0.86	0.76	0.69
2.5	0.28	0.20	0.17	0.18	0.27	0.32	0.41	0.55	0.65	0.79	0.91	1	0.93	0.84	0.76
3	0.26	0.17	0.15	0.17	0.25	0.31	0.40	0.51	0.61	0.74	0.86	0.93	1	0.90	0.81
4	0.21	0.15	0.14	0.15	0.23	0.27	0.35	0.44	0.52	0.66	0.76	0.84	0.90	1	0.93
5	0.15	0.11	0.10	0.11	0.18	0.22	0.29	0.38	0.43	0.57	0.69	0.76	0.81	0.93	1

CONCLUSIONS

The significance of magnitude and distance on the spectral acceleration correlation coefficients has been comprehensively investigated in this paper. The Z-Fisher test was applied as a powerful statistical tool in order to make a judgement on this issue. The results based on a relatively large ground motion database revealed that this significance is meaningful, despite the limited findings which were available in the literature. This difference is more meaningful in the case of magnitude when compared with distance.
A new set of correlation coefficients is also proposed, which are dependent on the magnitude and distance. The CMS is then calculated in the case of both the newly proposed correlation coefficient and the conventional correlation coefficient. The results show that the CMS is sensitive to the correlation coefficients in the range of low period values and high target period values. These results indicate some magnitude and distance dependence, and more work is needed to identify the cause of this apparent dependence and refine methods for incorporating it.

ACKNOWLEDGEMENTS

The authors are very grateful to J.W. Baker, J. Douglas and another anonymous reviewer for their important and valuable comments which helped to improve the paper.

REFERENCES

1. ASCE7-5, *Minimum Design Loads for Buildings and Other Structures*. 2005, American Society of Civil Engineers/Structural Engineering Institute, Reston ,VA.
2. CEN, *Eurocode 8 Design of structures for earthquake resistance. Part 1: general rules, seismic actions and rules for buildings. Final draft prEN 1998*. 2003, European Committee for Standardization, Brussels.
3. McGuire, R., *Probabilistic seismic hazard analysis and design earthquakes: closing the loop.* Bulletin of the Seismological Society of America, 1995. **85**(5): p. 1275-1284.
4. Bommer, J.J., Scott, S.G., and Sarma, S.K., *Hazard-consistent earthquake scenarios.* Soil Dynamics and Earthquake Engineering, 2000. **19**: p. 219-231.
5. Bazzurro, P., Cornell, C.A, *Disaggregation of seismic hazard.* Bulletin of the Seismological Society of America, 1999. **89**(2): p. 501-520.

6. Baker, J.W. and C.A. Cornell, *Spectral shape, epsilon and record selection.* Earthquake Engineering & Structural Dynamics, 2006. **35**(9): p. 1077-1095.
7. McGuire, R., *Seismic Hazard and Risk Analysis*. 2004: EERI Monograph, Earthquake Engineering Research Institute, Oakland, California.
8. Naeim, F., Lew, M., *On the use of design spectrum compatible time histories.* Earthquake Spectra, 1995. **11**: p. 111-127.
9. Reiter, L., *Earthquake Hazard Analysis: Issues and insights*. 1990: Columbia University Press, New York.
10. Campbell, K.W. and Y. Bozorgnia, *NGA Ground Motion Model for the Geometric Mean Horizontal Component of PGA, PGV, PGD and 5% Damped Linear Elastic Response Spectra for Periods Ranging from 0.01 to 10 s.* Earthquake Spectra, 2008. **24**(1): p. 139-171.
11. Baker, J.W., *Conditional Mean Spectrum: Tool for Ground-Motion Selection.* Journal of Structural Engineering, 2011. **137**(3): p. 322-331.
12. Kutner, M.H., C. Nachtsheim, and J. Neter, *Applied Linear Regression Models*. 2004: McGraw-Hill/Irwin, New York. 701.
13. Jayaram, N., Baker, J.W, *Statistical tests of the joint distribution of spectral acceleration values.* Bulletin of the Seismological Society of America, 2008. **98**(5): p. 2231-2243.
14. Baker, J.W. and C.A. Cornell, *Correlation of Response Spectral Values for Multicomponent Ground Motions.* Bulletin of the Seismological Society of America, 2006. **96**(1): p. 215-227.
15. Baker, J.W. and N. Jayaram, *Correlation of spectral acceleration values from NGA ground motion models.* Earthquake Spectra, 2008. **24(1)**: p. 299-317.
16. Baker, J.W., *Vector-valued ground motion intensity measures for probabilistic seismic demand analysis*. 2005: Report #150, John A.Blume Earthquake Engineering Center, Stanford, CA. p. 321.
17. Jayaram, N., Baker, J.W, Okano, H., Ishida, H., McCann, M. W., and Mihara, Y, *Correlation of response spectral values in Japanese ground motions.* Earthquakes and Structures, 2011. **2**(4): p. 357-376.
18. Boore, D.M., Watson-Lamprey, J. and Norman A. Abrahamson, *Orientation-independent measures of ground motion.* Bulletin of the Seismological Society of America, 2006. **96**: p. 1502–11.
19. PEER. *Pacific Earthquake Engineering Research Center, NGA Database, University of California, Berkeley, accessed 8/3/2012*. Available from: http://peer.berkeley.edu/ngawest/nga.
20. Abrahamson, N.A., Youngs, R. R., *A stable algorithm for regression analyses using the random effects model.* Bulletin of the Seismological Society of America, 1992. **82**(1): p. 505-510.
21. Cohen, J. and P. Cohen, *Applied multiple regression/correlation analysis for the behavioral sciences*: New York: Wiley & Sons, 1975.
22. Efron, B. and R. Tibshirani, *An Introduction to the Bootstrap*. 1993, New York: Chapman & Hall/CRC Monographs on Statistics & Applied Probability.