# The conditional mean spectrum based on the robust regression analysis

M. Nourizadeh[1], M. Mousavi[2], A. Azarbakht[3]

**Abstract** – *The conditional mean spectrum (CMS) has been recently proposed as an alternative to the uniform hazard spectrum (UHS) to be employed as a target spectrum in ground motion record selection. The CMS provides the expected response spectrum, conditioned on occurrence of a target spectral acceleration value at the period of interest. A robust regression analysis is proposed in this manuscript to improve the current CMS which is based on a conventional regression analysis. The results show that the proposed robust CMS significantly differs from the conventional CMS, especially for higher periods of interest. The shape of the robust CMS represents the rare ground motions in a more reliable manner, comparing with the conventional CMS.*

***Keywords**: Conditional Mean Spectrum (CMS), Epsilon, Probabilistic Seismic Hazard Analysis, Robust regression analysis, Seismic Hazard, Uniform Hazard Spectrum.*

## I. Introduction

The assessment of structural seismic response is often done by selecting ground motion records that conforms to the seismic hazard conditions of the objective site which can be obtained based on the probabilistic seismic hazard analysis (PSHA). A common record-selection practice (e.g. ACSE7-05) suggests selecting seven records which are compatible with the dominant earthquake scenario in a given site. This dominant scenario is represented with two key parameters, magnitude (Mw) and distance (R), which can be obtained by disaggregation analysis [1]. The selected records are then scaled (if necessary) to match the design level of the uniform hazard spectrum (UHS). Both the UHS and the disaggregation analysis are outputs of PSHA and can be determined for any desired level of hazard such as 10% or 2% probability of exceedance in 50 years. For clarify of exposition, consider a hypothetical site where the only possible nearby earthquake event is a magnitude 7 earthquake at a distance of 10 km, occurring at a mean rate of once every 75 years. The soil shear wave velocity is assumed equal to 760 m/sec for this site. By using an appropriate ground motion prediction model, here CB08 [2], the median (μ) and the median plus one standard deviation (μ + σ) spectra are plotted in Fig 1.

By using basic probability theory, these two spectra correspond to ground motions with 150 and 475 years return period, respectively (i.e. $\Phi^{-1}(1-\frac{75}{475})=1$ where Φ is the standard Gaussian cumulative distribution function). Also, the expected ground motion spectrum for 2475 years return period corresponds to μ+1.88σ. As a result, it is rational to say that the different levels of hazard for this site are represented by vector <M, R, ε>. The third component of this vector specifies the level of increase/decrease of mean spectrum for different hazard levels.

Table I signifies the corresponding ε values for different hazard levels for the prescribed assumed site. The exceedance probability in 50 years is also indicated in this Table for different hazard levels. Therefore, the UHS for any desired level of hazard at the mentioned site is defined as:

$$\mu_{\ln Sa(T_i)} = \mu_{\ln Sa}(M,R,T_i) + \varepsilon\sigma_{\ln Sa}(T_i) \qquad (1)$$

where $\mu_{\ln Sa(Ti)}$ is the natural logarithm of the expected spectral acceleration at the given period $T_i$ and $\mu_{\ln Sa}$ and $\sigma_{\ln Sa}$ are, respectively, the predicted median and standard deviation values obtained from the ground motion prediction model for the dominant event.

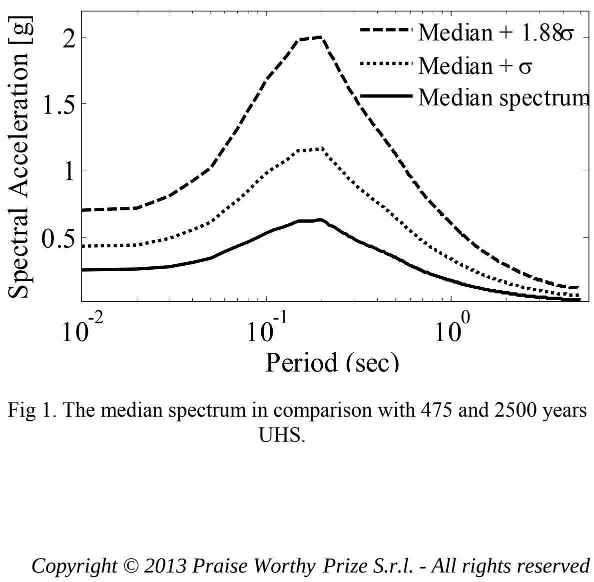


Fig 1. The median spectrum in comparison with 475 and 2500 years UHS.

TABLE I
THE VALUE OF ε FOR DIFFERENT HAZARD LEVELS IN THE CONSIDERED SITE.

| Return period (years) | Exceedance Probability in 50 years | ε |
|---|---|---|
| 100 | 39% | -0.67 |
| 150 | 28% | 0 |
| 250 | 18% | 0.52 |
| 475 | 10% | 1 |
| 2475 | 2% | 1.88 |

As shown in Equation (1), the UHS remains at a constant ε standard deviation from the median response spectrum at all periods. Considering a high positive ε value, the UHS represents an expected ground motion spectrum that all of its points take simultaneously rare values. This issue is in a significant conflict with the nature of real ground motion records [3]. The rate of observing a high positive ε at all periods is much lower than the rate of observing a high ε at any single period. Thus it can be concluded that the UHS represents a nearly impossible earthquake scenario, especially in higher levels of hazard [4]. The conservatism of the UHS has been addressed also by other researchers [5],[6].

To deal with this problem the conditional mean spectrum (CMS) has been recently introduced by Baker to be used in structural analysis instead of using UHS [7]. The correlation of ε values in different periods is considered in CMS development; hence, the mentioned conservatism is taken into account.

In order to explain this spectrum, consider a structural system with fundamental period, $T^*$=1.0 sec, located at the above mentioned site. Suppose that the level of hazard of interest is 2475 years return period which corresponds to ε=1.88. Since the Sa value at the fundamental period of structure has been accepted as a common efficient intensity measures in literature [8], we assign the target ε value to this period e.g. ε(1.0s)=1.88. Considering that assignment of this ε value to other periods leads to a conservative spectrum, the question is then, what are the associated ε values at other periods, given that we know ε(1.0s)=1.88? This question can be responded by considering Fig 2. By using a large set of ground motions (as defined in Section II.2), the correlation between ε of T=1.0s and ε of other periods has been revealed in Fig 2. It is obvious that the different ε values are correlated, however with different coefficients of correlation (ρ). For more clarification, ε(1.0s) is plotted against ε(0.25s), ε(0.50s), ε(2.0s), and ε(4.0s), respectively, in Fig 2a, 2b, 2c and 2d.

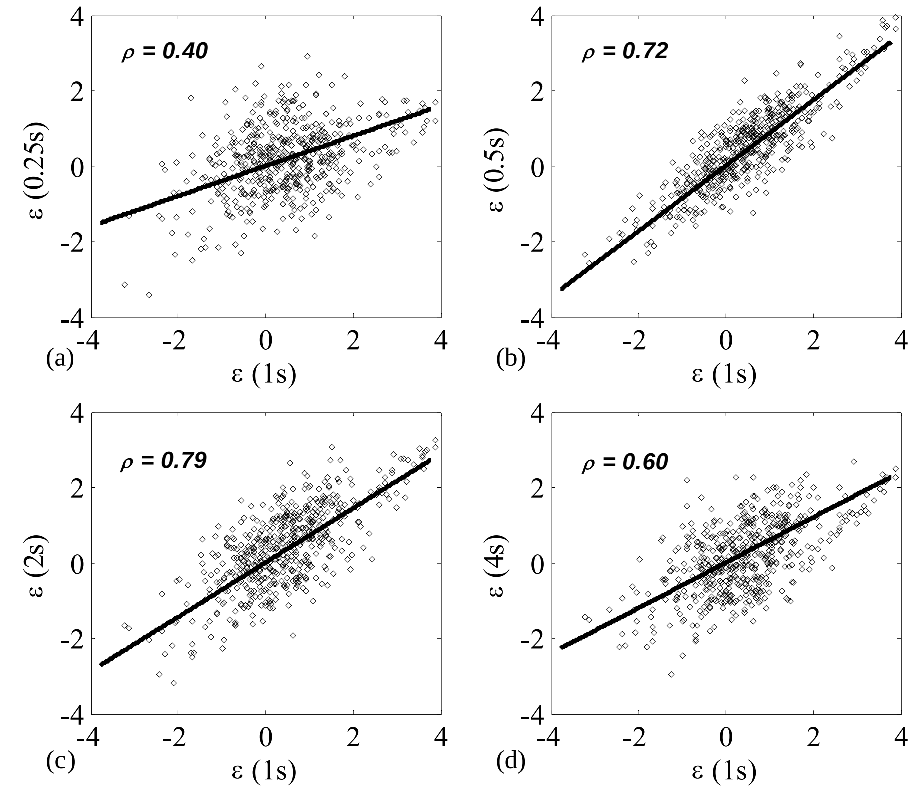


Fig 2. The Scatter plot of ε values from a large set of ground motions.
ε(1.0s) versus (a) ε(0.25s), (b) ε(0.5s), (c) ε(2.0s), and (d) ε(4.0s).

By using this information, it is possible to predict the expected value of ε at different periods, given that we know the value of target ε at the period of interest (here, T=1.0 s). Based on probability calculations, Baker proposed to take the expected ε values at any other period equal to the target ε multiplied by the correlation coefficient between the two ε values as mathematically written in Equation (2) [7]:

$$\varepsilon(T_i) = \rho\left(\varepsilon(T^*), \varepsilon(T_i)\right)\varepsilon(T^*) \quad (2)$$

The correlation coefficient between two sets of observed ε values can be estimated as:

$$\rho\left(\varepsilon(T_1), \varepsilon(T_2)\right) = \frac{\sum_{k=1}^{n}(\varepsilon_k(T_1) - \overline{\varepsilon(T_1)})(\varepsilon_k(T_2) - \overline{\varepsilon(T_2)})}{\sqrt{\sum_{k=1}^{n}(\varepsilon_k(T_1) - \overline{\varepsilon(T_1)})^2 \sum_{k=1}^{n}(\varepsilon_k(T_2) - \overline{\varepsilon(T_2)})^2}} \quad (3)$$

where $\varepsilon_k(T_1)$ and $\varepsilon_k(T_2)$ are the $k^{th}$ observations of $\varepsilon(T_1)$ and $\varepsilon(T_2)$, and are their sample means, and n is the total number of observations (records). The expected ε values for periods 0.25, 0.5, 2.0, and 4.0 s are predicted 0.75, 1.35, 1.49, and 1.13, respectively, given ε(1.0sec)=1.88. This procedure has been repeated for an entire range of periods and the resulted correlations are used in Equation (4) to find the conditional mean spectrum (CMS).

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

Fig 3 compares the CMS given ε(1.0sec)=1.88 in comparison with the UHS and the mean spectrum. This figure also includes CMS at a few other periods, having equal ε=1.88.

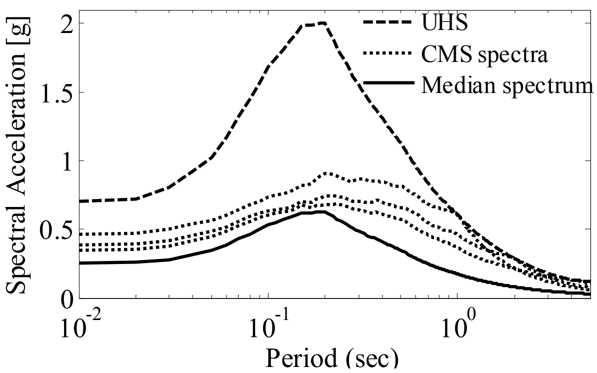


Fig 3. Conditional Mean Spectra, conditioned on Sa values at a few periods, having equal ε=1.88.

A few closed form empirical formulas have been proposed in literature to find the correlation coefficient for different periods [9],[10]. Each of these closed form formulas in conjunction with Equation (2) build a complete framework to construct the CMS and use it instead of UHS in structural dynamic analyzes. For more details, see [7].

The main objective of this paper is to reveal an important drawback in the above mentioned procedure for calculation of CMS. The authors think that the developed procedure for CMS leads to an anomaly spectral shape that is not enough consistent with real ground motions. This guess arises from comparison of the CMS and mean spectrum in Fig 3. The mean spectrum, as shown in Fig 3, has a peak value near the period T=0.15 sec and then this spectrum decays with a relatively rapid rate. However, this pattern is not observed in the shape of different CMS curves. The peak shape is not clear and a relatively flat region is observed near the mentioned period in all CMS curves. According to these observations, this hypothesis was formed that the developed CMS procedure suffers from a systematic bias. A conventional regression analysis has been applied in the original paper of CMS [7] to measure the degree of correlation of ε values in different periods and the influence of outlier data has not been studied. The authors think that this overlook significantly affects the resulted correlation coefficients and it shall be taken into account in the CMS development procedure. This issue is considered in the following sections.

# II. Introducing a Procedure to Find Robust CMS

### *II.1. About the Robust Regression Analysis*

Robust regression is an important tool for analyzing data that are contaminated with outliers. Outliers are sample values that cause surprise in relation to the majority of the sample [11]. Many robust methods have been developed since 1960 to detect outliers and to provide resistant (stable) results in the presence of outliers [11]. In order to achieve this stability, robust regression limits the influence of outliers.

Robust regression analysis works by assigning a weight to each data point. Weighting is done automatically and iteratively using a process called iteratively reweighted least squares [12]. In the first iteration, equal weight is assigned to each point and model parameters are estimated using conventional least squares. At subsequent iterations, weights are recomputed so that points farther from model predictions in the previous iteration are given lower weight. The model parameters are then recomputed using weighted least squares. The process continues until the values of the estimated parameters converge within a specified tolerance. It is needed to mention that except the iteratively reweighted least squares process; many other techniques have been also developed for robust regression problems which can be found in [11].

In the following sections, the iteratively reweighted least squares technique has been applied to find the robust correlation coefficient of ε values. Before that, the employed ground motion set is defined.

### *II.2. The Ground Motion Data Set*

A strong ground motion records data set based on worldwide recordings of shallow crustal earthquakes is selected which was also used by Baker and Cornell [9] to analyze the correlation of response spectral values. This set includes 267 pairs of horizontal ground motion records with magnitudes greater than 5.5 and source-to-site distances of less than 100 km. The other selection criteria, as well as the detailed documentation about this set, are given in [13].

### *II.3. The Robust CMS*

Using the above mentioned data set, a conventional regression analysis has been done by Baker and Cornell [9] to build a linear model between ε values in different periods. In that study, the influence of outlier ground motions was not taken into account. Here, it is supposed that the results obtained from that study are not sufficiently stable and application of a robust regression is needed to find more strength correlation coefficients.

For convenience, assume that for the interest period $T^*$=1.0 sec, the correlation analysis is done just for four periods: T=0.25, 0.50, 1.0, 2.0, and 4.0 sec. On the other hand, the objective CMS is constructed just based on five period values ($T^*$, and T). Fig 4 shows the scatter plot of ε(1.0sec) versus ε(0.25s), ε(0.50s), ε(2.0s), and ε(4.0s). The slope of solid line corresponds to the correlation of coefficient, as mentioned in section I.

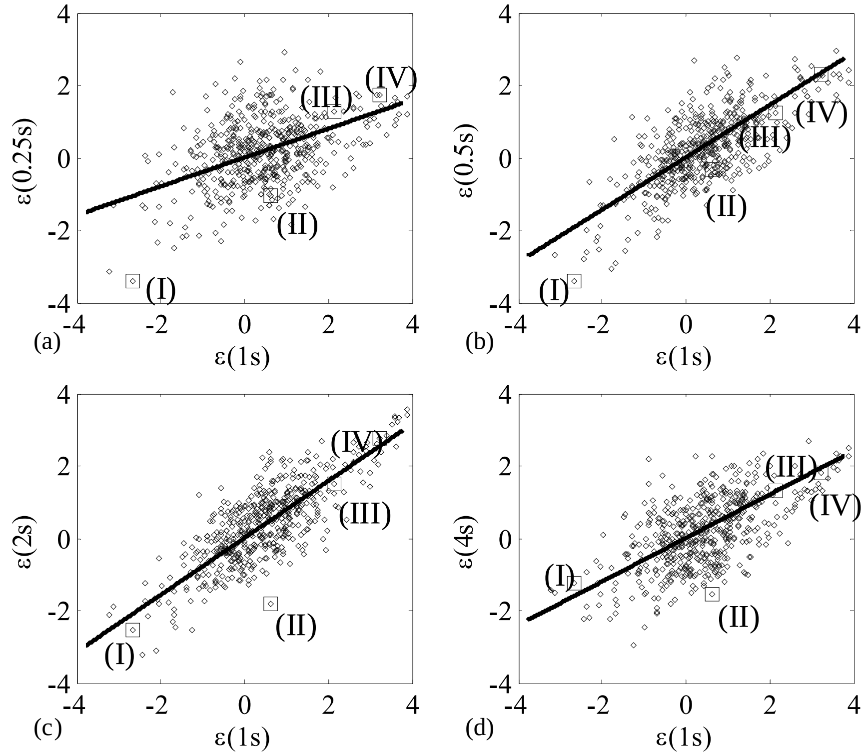


Fig 4. The Scatter plot of ε values from the introduced data set,
ε(1.0s) versus (a) ε(0.25s), (b) ε(0.5s), (c) ε(2.0s), and (d) ε(4.0s). Four random records are marked.

Four random ground motions are marked in Fig 4. Since the selected records do not match exactly to the regression line, a residual parameter is defined for each record corresponding to the considered period:

$$r(T^*,T_i) = \varepsilon(T_i) - \rho\left(T^*,T_i\right)\varepsilon(T^*) \qquad (5)$$

where $r(T^*,T_i)$ symbolizes the observed residual in prediction of $\varepsilon(T_i)$ from $\varepsilon(T^*)$. Investigation of residual values for these four ground motions illuminates that why it is necessary to assign different weight to each record and then re-calculate the correlation coefficient.

- The record, marked as (I), shows significant negative residual at periods 0.25 and 0.50 sec, slight negative residual at T=2.0 sec, and slight positive residual at T=4.0 sec. The residual vector [-2.35, -1.49, -0.42, 0.37] quantifies the residuals corresponding to this record. The mean of residual vector is μ= -0.97 and its standard deviation is σ=1.19. Based on the mentioned periods, this record can be accounted as a highly outlier ground motion.

 

- The residual vector for the record (II) is [-1.27, -0.76, -2.32, -1.92] with mean value μ=-1.57 and σ=0.69. This record has negative residual values in all periods. Comparing with record (I), it has greater negative μ and it should be accounted as a more outlier data. As a direct result, a smaller weight shall be assigned to this record in comparison with the record (I).

- The residual vector for the record (III) is [0.44, -0.29, -0.16, 0.034], with μ=-0.01 and σ=0.32. Comparing with two former records, this record is well-matched to the regression line in all periods and a higher weight value should be assigned to it.

- The residual vector for the record (IV) is [0.44, -0.01, 0.21, -0.12], with μ=-0.13 and σ=0.25. A higher weight should be assigned to this record in comparison with records (I) and (II). However, due to greater μ and smaller σ value, the judgment about the weight of this record is not straightforward in comparison with the record (III).

The resulted residual vectors can be integrated in a residual matrix,

$$R = \begin{bmatrix} -2.35 & -1.49 & -0.42 & 0.37 \\ -1.27 & -0.76 & -2.32 & -1.92 \\ 0.44 & -0.29 & -0.16 & 0.03 \\ 0.44 & -0.01 & 0.21 & -0.12 \end{bmatrix}$$

Also, the vectors M and Σ indicates the mean and standard deviation of records:

$$M = \begin{bmatrix} -0.97 \\ -1.57 \\ 0.01 \\ 0.13 \end{bmatrix}, \text{ and } \Sigma = \begin{bmatrix} 1.19 \\ 0.69 \\ 0.32 \\ 0.25 \end{bmatrix}$$

Fig 5 shows the box-plot of residual vectors for the considered ground motion records. The median and standard deviation of each residual vector was explained by using this figure. A snapshot to the degree of outlier of each ground motion record is achievable through this figure.

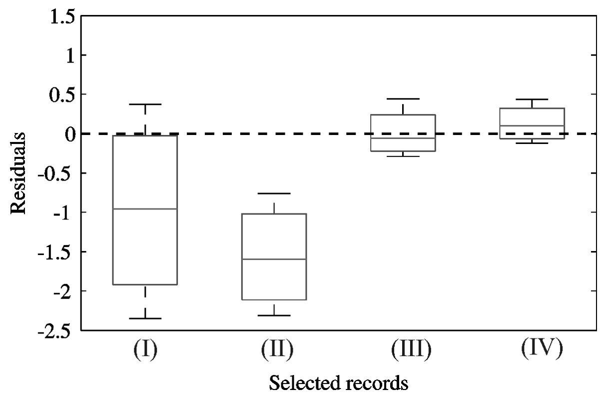


Fig 5. The box-plot of residual vectors for the selected ground motion records.

Rationally, the greater weight should be assigned to the ground motions with lower absolute mean and standard deviation. Here, a Gaussian function has been selected to meet the mentioned concern:

$$w_i = \frac{1}{\sigma_i \sqrt{2}} e^{-\frac{-\mu_i^2}{2\sigma_i^2}} \quad (6)$$

where $w_i$ is the objective weight, and $\mu_i$, and $\sigma_i$ correspond to the mean and standard deviation of the residual vector of the considered record.

Now, the iteratively reweighted least squares procedure is applied to find the robust correlation coefficient values. The used algorithm can be summarized as:

1-Take the initial weighting vector as unity, $w = [1 \quad 1 \quad 1 \quad 1]$

2-Calculate the ρ values for regression model using Equation (3).

3- Find the residual mean vector (M) as well as the residual standard deviation vector (Σ).

4-Determine the weighting vector based on Equation (6).

5-Multiply ε values of each record by the corresponding weight value.

6-Repeat steps 2 to 5 until the ρ values converge to a stable rate.

Table II comprises the resulted values during the iterations. As indicated in this table, the iterative procedure converges to stable ρ and w values after a few iterations.

 

TABLE II
THE ITERATIVE PROCEDURE TO FIND THE ROBUST COEFFICIENT CORRELATION

| Iteration | Weights (w) | | | | $\rho(T^*=1s, T_i)$ | | | |
|---|---|---|---|---|---|---|---|---|
| | Selected Records | | | | $T_i$ | | | |
| | I | II | III | IV | 0.25s | 0.50s | 2.0s | 4.0s |
| 1 | 1.000 | 1.000 | 1.000 | 1.000 | 0.40 | 0.72 | 0.79 | 0.60 |
| 2 | 0.240 | 0.043 | 1.253 | 1.413 | 0.49 | 0.82 | 0.86 | 0.73 |
| 3 | 0.33 | 0.026 | 0.158 | 0.176 | 0.54 | 0.83 | 0.89 | 0.80 |
| 4 | 0.318 | 0.028 | 0.411 | 0.469 | 0.56 | 0.84 | 0.89 | 0.81 |
| 5 | 0.317 | 0.028 | 0.433 | 0.455 | 0.57 | 0.84 | 0.88 | 0.80 |
| 6 | 0.317 | 0.028 | 0.436 | 0.455 | 0.58 | 0.84 | 0.88 | 0.80 |
| 7 | 0.317 | 0.028 | 0.436 | 0.454 | 0.59 | 0.84 | 0.88 | 0.79 |
| 8 | 0.317 | 0.028 | 0.436 | 0.454 | 0.60 | 0.84 | 0.88 | 0.79 |
| 9 | 0.317 | 0.028 | 0.436 | 0.454 | 0.61 | 0.84 | 0.88 | 0.79 |
| 10 | 0.317 | 0.028 | 0.436 | 0.454 | 0.61 | 0.84 | 0.88 | 0.79 |

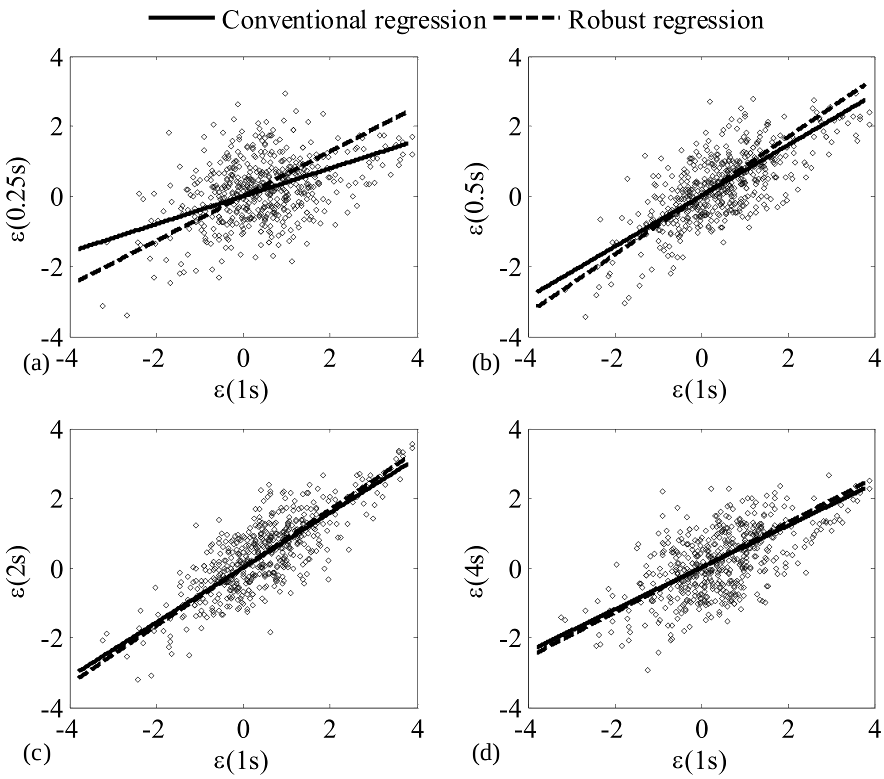


Fig 6. The comparison of the robust regression with conventional regression model.

Fig 6 compares the resulted robust regression with the conventional regression model. As it is obviously clear, the robust regression increases the values of correlation coefficient in all periods. This increase is more significant in the lowest period case i.e. T=0.25 sec.

By replacement of the robust correlation coefficients in Equation (4), the robust CMS is achievable. Fig 7 compares the resulted robust and the conventional CMS. Note that the CMS are formed here by using only five spectral acceleration values in five periods.

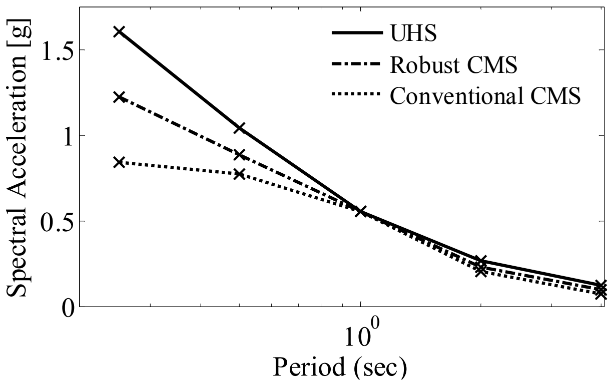


Fig 7. The comparison of the robust CMS with the conventional CMS.

Manuscript received January 2013, revised February 2013

As shown in Fig 7, the shape of obtained robust CMS is significantly different with the conventional CMS, especially in low periods. This issue will be more discussed in the next section, where the robust CMS procedure is extended for an entire range of periods.

# III. Final Results of the Robust CMS

## *III.1. Development of the Robust CMS Procedure for The Entire Period Range.*

In the former simple example, the ground motions were weighted just for interest period T*=1.0sec. This weighting procedure was done based on analysis of residuals in four other periods 0.25, 0.50, 2, and 4 sec. Here, this procedure is extended to several interest periods and the analysis of residuals for each interest period is completed for an entire range of other periods.

The conventional and robust correlation coefficients, at a variety of period pairs, are plotted in Fig 8. This figure shows correlation coefficients for a selected set of periods Ti, plotted versus T* values between 0.01 and 5 seconds. The higher value of ρ in lower periods in robust analysis is clear in this figure, comparing to the conventional analysis results. Fig 9 shows the same results, plotted using contours of correlation coefficients as a function of both Ti and T*.

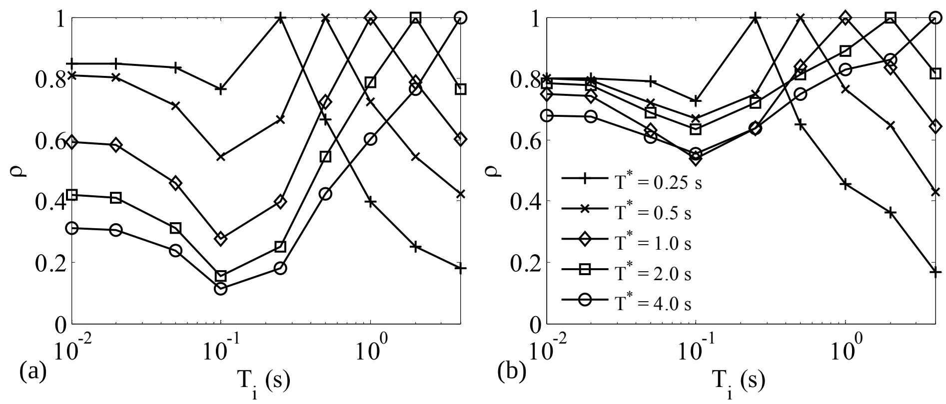


Fig8. Plot of correlation coefficients versus T*, for several Ti values; based on
(a) Conventional analysis (b) Robust correlation analysis.

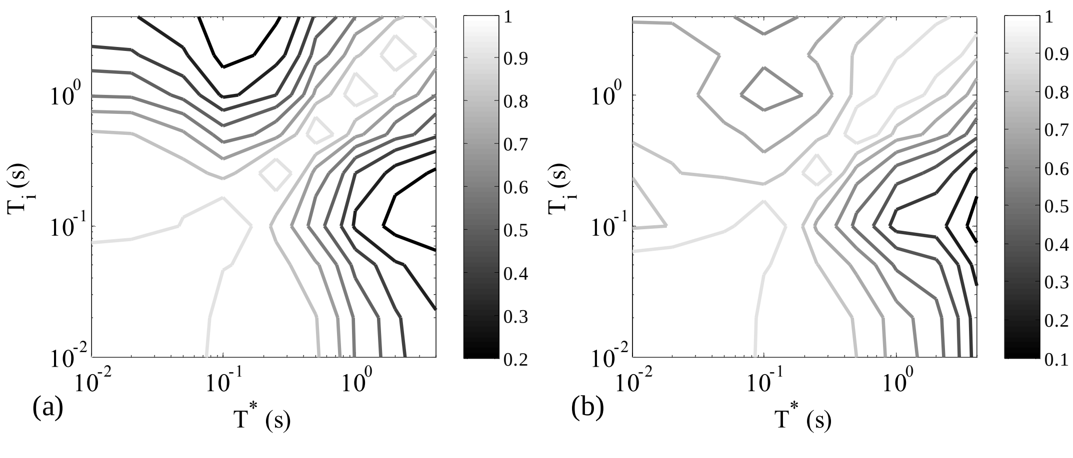


Fig9. Contours of correlation coefficients versus T* and $T_i$; based on
(a) Conventional analysis (b) Robust correlation analysis.

The numerical values from Fig 9 are provided in Table III. This table can be used when correlation coefficients for a specific ground motion scenario are needed. For comparison, this table includes both conventional and robust analysis results.

TABLE III
Correlation coefficients of ε($T^*$) versus ε($T_i$) obtained using CB08 ground motion prediction model;
(a) Conventional analysis approach, and (b) Robust approach.

**(a)**

| | | $T_i$ | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **0.02** | **0.05** | **0.1** | **0.2** | **0.3** | **0.4** | **0.5** | **0.75** | **1** | **1.5** | **2** | **2.5** | **3** | **4** | **5** |
| | **0.02** | 1.00 | 0.97 | 0.85 | 0.84 | 0.85 | 0.83 | 0.80 | 0.69 | 0.58 | 0.50 | 0.41 | 0.36 | 0.34 | 0.31 | 0.28 |
| | **0.05** | 0.97 | 1.00 | 0.91 | 0.85 | 0.81 | 0.76 | 0.71 | 0.58 | 0.46 | 0.39 | 0.31 | 0.27 | 0.26 | 0.24 | 0.23 |
| | **0.1** | 0.85 | 0.91 | 1.00 | 0.85 | 0.71 | 0.61 | 0.54 | 0.40 | 0.27 | 0.22 | 0.15 | 0.13 | 0.10 | 0.11 | 0.12 |
| | **0.2** | 0.84 | 0.85 | 0.85 | 1.00 | 0.78 | 0.67 | 0.59 | 0.45 | 0.31 | 0.25 | 0.15 | 0.11 | 0.08 | 0.08 | 0.08 |
| | **0.3** | 0.85 | 0.81 | 0.71 | 0.78 | 1.00 | 0.83 | 0.75 | 0.61 | 0.50 | 0.43 | 0.32 | 0.28 | 0.27 | 0.24 | 0.22 |
| | **0.4** | 0.83 | 0.76 | 0.61 | 0.67 | 0.83 | 1.00 | 0.88 | 0.73 | 0.62 | 0.56 | 0.46 | 0.41 | 0.40 | 0.34 | 0.29 |
| | **0.5** | 0.80 | 0.71 | 0.54 | 0.59 | 0.75 | 0.88 | 1.00 | 0.83 | 0.72 | 0.64 | 0.54 | 0.49 | 0.48 | 0.42 | 0.37 |
| $T^*$ | **0.75** | 0.69 | 0.58 | 0.40 | 0.45 | 0.61 | 0.73 | 0.83 | 1.00 | 0.87 | 0.76 | 0.67 | 0.62 | 0.59 | 0.53 | 0.46 |
| | **1** | 0.58 | 0.46 | 0.27 | 0.31 | 0.50 | 0.62 | 0.72 | 0.87 | 1.00 | 0.87 | 0.79 | 0.72 | 0.69 | 0.60 | 0.52 |
| | **1.5** | 0.50 | 0.39 | 0.22 | 0.25 | 0.43 | 0.56 | 0.64 | 0.76 | 0.87 | 1.00 | 0.89 | 0.81 | 0.76 | 0.68 | 0.61 |
| | **2** | 0.41 | 0.31 | 0.15 | 0.15 | 0.32 | 0.46 | 0.54 | 0.67 | 0.79 | 0.89 | 1.00 | 0.92 | 0.86 | 0.77 | 0.67 |
| | **2.5** | 0.36 | 0.27 | 0.13 | 0.11 | 0.28 | 0.41 | 0.49 | 0.62 | 0.72 | 0.81 | 0.92 | 1.00 | 0.94 | 0.82 | 0.72 |
| | **3** | 0.34 | 0.26 | 0.10 | 0.08 | 0.27 | 0.40 | 0.48 | 0.59 | 0.69 | 0.76 | 0.86 | 0.94 | 1.00 | 0.88 | 0.77 |
| | **4** | 0.31 | 0.24 | 0.11 | 0.08 | 0.24 | 0.34 | 0.42 | 0.53 | 0.60 | 0.68 | 0.77 | 0.82 | 0.88 | 1.00 | 0.90 |
| | **5** | 0.28 | 0.23 | 0.12 | 0.08 | 0.22 | 0.29 | 0.37 | 0.46 | 0.52 | 0.61 | 0.67 | 0.72 | 0.77 | 0.90 | 1.00 |

**(b)**

| | | $T_i$ | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **0.02** | **0.05** | **0.1** | **0.2** | **0.3** | **0.4** | **0.5** | **0.75** | **1** | **1.5** | **2** | **2.5** | **3** | **4** | **5** |
| | **0.02** | 1.00 | 0.97 | 0.88 | 0.88 | 0.82 | 0.78 | 0.79 | 0.74 | 0.63 | 0.59 | 0.53 | 0.45 | 0.42 | 0.25 | 0.27 |
| | **0.05** | 0.96 | 1.00 | 0.91 | 0.85 | 0.75 | 0.67 | 0.67 | 0.61 | 0.48 | 0.47 | 0.42 | 0.34 | 0.31 | 0.15 | 0.20 |
| | **0.1** | 0.80 | 0.88 | 1.00 | 0.79 | 0.60 | 0.47 | 0.47 | 0.38 | 0.25 | 0.28 | 0.23 | 0.16 | 0.12 | 0.04 | 0.12 |
| | **0.2** | 0.80 | 0.82 | 0.80 | 1.00 | 0.71 | 0.55 | 0.52 | 0.40 | 0.31 | 0.30 | 0.23 | 0.14 | 0.11 | 0.01 | 0.07 |
| | **0.3** | 0.81 | 0.78 | 0.73 | 0.81 | 1.00 | 0.78 | 0.75 | 0.65 | 0.57 | 0.53 | 0.44 | 0.36 | 0.32 | 0.18 | 0.20 |
| | **0.4** | 0.83 | 0.77 | 0.70 | 0.74 | 0.83 | 1.00 | 0.88 | 0.77 | 0.68 | 0.65 | 0.59 | 0.51 | 0.48 | 0.33 | 0.33 |
| | **0.5** | 0.80 | 0.72 | 0.67 | 0.69 | 0.79 | 0.87 | 1.00 | 0.85 | 0.76 | 0.72 | 0.65 | 0.59 | 0.56 | 0.43 | 0.40 |
| $T^*$ | **0.75** | 0.78 | 0.69 | 0.58 | 0.63 | 0.74 | 0.78 | 0.86 | 1.00 | 0.88 | 0.81 | 0.75 | 0.71 | 0.68 | 0.55 | 0.48 |
| | **1** | 0.74 | 0.63 | 0.54 | 0.56 | 0.68 | 0.75 | 0.84 | 0.91 | 1.00 | 0.89 | 0.84 | 0.79 | 0.75 | 0.64 | 0.58 |
| | **1.5** | 0.79 | 0.69 | 0.65 | 0.65 | 0.74 | 0.78 | 0.83 | 0.88 | 0.91 | 1.00 | 0.91 | 0.84 | 0.81 | 0.74 | 0.70 |
| | **2** | 0.78 | 0.69 | 0.63 | 0.62 | 0.73 | 0.76 | 0.81 | 0.87 | 0.89 | 0.92 | 1.00 | 0.93 | 0.88 | 0.82 | 0.76 |
| | **2.5** | 0.74 | 0.67 | 0.63 | 0.61 | 0.71 | 0.74 | 0.79 | 0.85 | 0.86 | 0.88 | 0.94 | 1.00 | 0.94 | 0.85 | 0.82 |
| | **3** | 0.72 | 0.64 | 0.61 | 0.58 | 0.71 | 0.75 | 0.80 | 0.87 | 0.86 | 0.87 | 0.90 | 0.95 | 1.00 | 0.90 | 0.86 |
| | **4** | 0.67 | 0.61 | 0.55 | 0.54 | 0.68 | 0.69 | 0.75 | 0.82 | 0.83 | 0.83 | 0.86 | 0.88 | 0.90 | 1.00 | 0.93 |
| | **5** | 0.54 | 0.48 | 0.40 | 0.40 | 0.57 | 0.53 | 0.66 | 0.73 | 0.74 | 0.75 | 0.76 | 0.78 | 0.81 | 0.89 | 1.00 |

Investigation of Fig 9 shows that despite the conventional analysis, the robust analysis has been yielded to asymmetric ρ results, i.e. $\rho(T_1, T_2) \neq \rho(T_1 = T_2)$. It means that the assigned weight to a single record is not constant and depends on the interest period (T*).

The robust CMS is achievable through the robust correlation coefficients, as stated in the former section. Fig 10 shows the robust CMS in a few interest periods for ε=1.88. This figure also comprises the conventional CMS and UHS curves.

The difference between the robust and the conventional CMS curves is significant, especially for higher interest periods which clarify the need to take the outliers into account.

### *III.2. Discussion on the Results*

By focusing on the case $T^*$=4.0 sec in Fig 10, the conventional CMS approximately touches the mean spectrum in lower periods. This issue is resulted from a weak correlation between ε(4.0s) and ε values at low periods. In the conventional CMS procedure, the correlation between two periods is independently analyzed, without any consideration to the other periods. This memory-less correlation calculation approach is questionable from the author's viewpoint. Let us take a simple example for more clarification. Two given records with equal ε(4.0sec)=1.88 are plotted in Fig 11. The record A has a relatively high positive ε values in the range of period 0.01 to 4.0 sec. On the other hand, the record B has negative value of ε in the majority of remained periods except in T=4 sec, as shown in Fig 11.

 

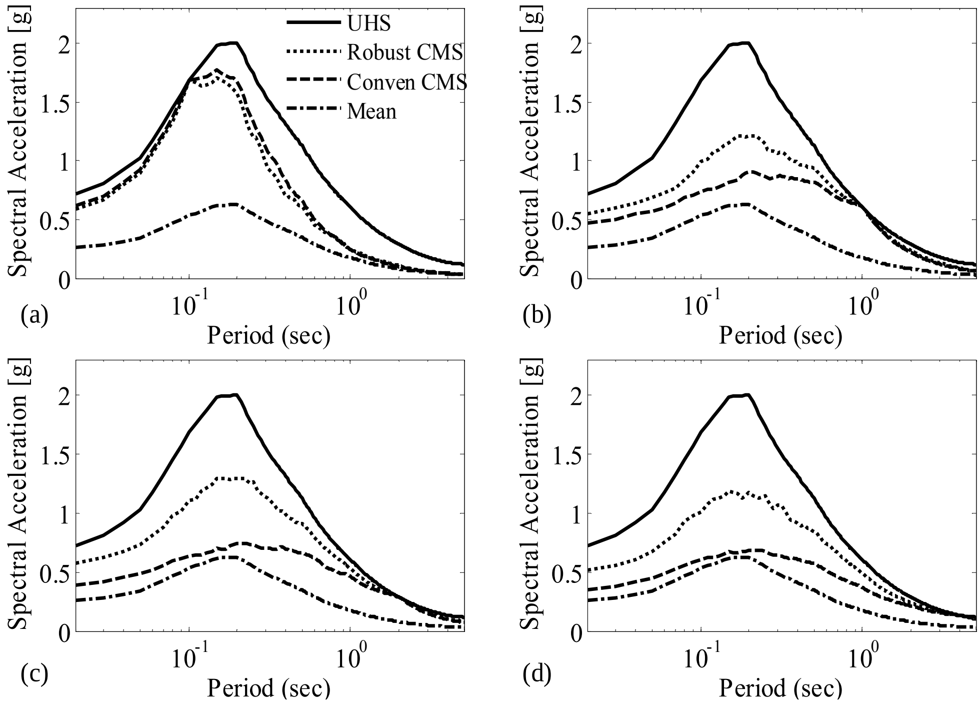


Fig 10. The robust CMS, conditioned on Sa values at the interest period.
(a) $T^*$=0.50 sec, (b) $T^*$=1.0 sec, (c) $T^*$=2.0 sec, and (d) $T^*$=4.0 sec, having equal ε=1.88.

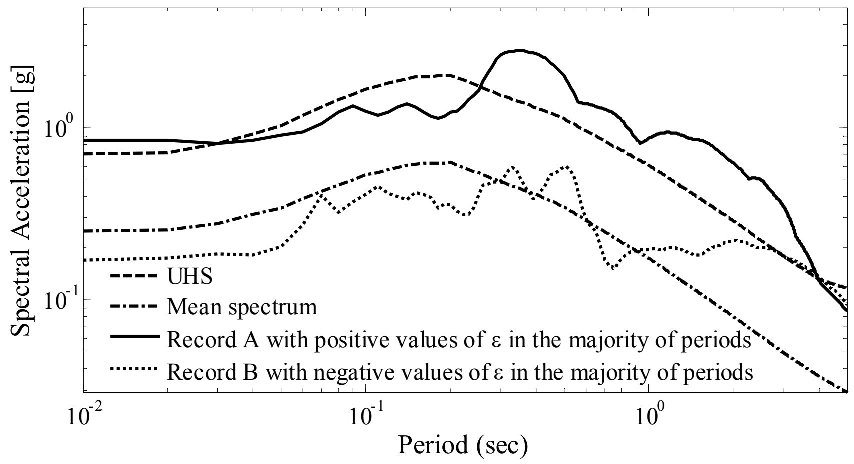


Fig 11. Comparison of two single records that have equal ε(4.0s)=+1.88, but different distribution of ε in the other periods.

Now the problem is to predict the value of ε(0.25s), given the ε(4.0s)=1.88, based on the given two records. According to the conventional regression analysis, the correlation of ε(0.25s) and ε(4.0s) is analyzed without any consideration to the shape of records, and a linear regression model is constructed to solve the problem. So, both records are accounted similar and a unit weight is assigned to them. According to Fig 11, it is obvious that a weak correlation is calculated between ε(4.0s) and ε(0.25s) based on two given records. Now coming back to the philosophy of hazard consistent spectral shape, it is absolutely obvious that the record A represent a rare record, despite the record B that barely shows a rare event. At least, it can be said that each of these two records may represent rare events but with different degrees. The assigned weight to each of records in robust regression analysis corresponds to the mentioned degree. It means that the weight assigned to the record A is higher than the weight of the record B because the ε values in all periods are in agreement with ε(4.0s) for the record A. On the other hand, since the ε(4.0s) value is not compatible with the ε in the other periods, a lower weight should be assigned to the record B in comparison with the first record. As a direct consequence of this weighting procedure, the correlation between ε(4.0s) and ε(0.25s) arises comparing with the conventional regression.

As the concluding result, the robust CMS shown in Fig 10 characterizes the rare 2500 years ground motions in a more reliable way comparing with the conventional CMS procedure. It worth emphasizing that the conventional CMS is non-conservative, specifically for the mid-rise and high-rise structures in higher levels of seismic hazard.

## IV. Conclusions

The uniform hazard spectrum (UHS), as a common target spectrum for structural dynamic analysis, does not represent a spectrum caused by a single earthquake at a given site and leads to a conservative spectrum in higher hazard levels. As an alternative, the conditional mean spectrum (CMS) has been introduced in recent years. The CMS uses the advantages of correlation between spectral accelerations at different periods. This correlation is calculated based on a conventional regression analysis between different pairs of periods and the correlation analysis for each pair of periods is not influenced by result of correlation analyzes in other pairs. In this paper, it was demonstrated that using the mentioned conventional regression analysis may leads to a systematic bias in results and a robust regression analysis was proposed to calculate the robust correlation coefficients. In the robust CMS, the ground motion records are contributed with different weights in regression analysis, depending on a rational residual analysis. The final results show that the robust CMS significantly differs from the conventional CMS, especially for higher interest periods. As a final point, the shape of robust CMS represents the rare events in a more reliable manner, comparing with the conventional CMS.

## References


1. Bazzurro, P., Cornell, C.A, *Disaggregation of seismic hazard.* Bulletin of the Seismological Society of America, 1999. **89**(2): p. 501-520.
2. 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.
3. Haselton, C.B., et al., *Accounting for Ground-Motion Spectral Shape Characteristics in Structural Collapse Assessment through an Adjustment for Epsilon.* Journal of Structural Engineering, 2011. **137**(3): p. 332-344.
4. Baker, J.W. and C.A. Cornell, *Spectral shape, epsilon and record selection.* Earthquake Engineering & Structural Dynamics, 2006. **35**(9): p. 1077-1095.
5. Naeim, F., Lew, M., *On the use of design spectrum compatible time histories.* Earthquake Spectra, 1995. **11**: p. 111-127.
6. Reiter, L., *Earthquake Hazard Analysis: Issues and Insights*. 1990, Columbia University Press: New York;254
7. Baker, J.W., *Conditional Mean Spectrum: Tool for Ground-Motion Selection.* Journal of Structural Engineering, 2011. **137**(3): p. 322-331.
8. Tothong, P., Luco, N, *Probabilistic seismic demand analysis using advanced ground motion intensity measures.* Earthquake Engineering & Structural Dynamics, 2007. **36**: p. 1837-1860.
9. 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.
10. Baker, J.W. and N. Jayaram, *Correlation of spectral acceleration values from NGA ground motion models.* Earthquake Spectra, 2008. **24(1)**: p. 299-317.
11. Rousseeuw, R.J., and Leroy, A. M., *Robust Regression and Outlier Detection*. 1987, New York: Wiley.
12. Holland, P., and Welsch, R., *Robust regression using interatively reweighted least-squares.* Communications in Statistics-Theory and Methods, 1977. **6**(9): p. 813–827.
13. Baker, J.W. and C.A. Cornell, *A vector-valued ground motion intensity measure consisting of spectral acceleration and epsilon.* Earthquake Engineering & Structural Dynamics, 2005. **34**(10): p. 1193-1217.


## Authors' information


1 Department of Civil Engineering, Arak University, Arak, IRAN

2 Department of Civil Engineering, Arak University, Arak, IRAN

3 Department of Civil Engineering, Arak University, Arak, IRAN


**M. Nourizadeh** was born in Arak, Iran, in 1987. He received his M.S. in structural engineering in 2012 from Arak University. His major interest is earthquake engineering, and conditional spectrums in particular. He is also working on some professional issues on seismic hazard along with Dr. Azarbakht and Dr. Mousavi at present (the co-authors in this article).

Dr. Azarbakht is an associated professor and Dr.Mousavi is an assistant professor at Arak university. Their area of interests include earthquake engineering and related issues and their main research area addresses the structural dynamic analysis, seismic hazard analysis, and earthquake ground motion selection.

 