The robust conditional mean spectrum

M. Mousavi, A. Azarbakht & M. Noorizadeh
Arak University, Iran

SUMMARY:
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 interest periods. The shape of the robust CMS represents the rare ground motions in a more reliable manner, comparing with the conventional CMS.

Keywords: epsilon, conditional mean spectrum, robust regression, seismic hazard

1. 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., ASCE7-05 2005) 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 the distance (R) which can be obtained by disaggregation analysis (Bazzurro and cornell 1999). 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.

The UHS for any desired level of hazard at the mentioned site is defined as

$$\mu_{\ln Sa(T_i)} = \mu_{\ln Sa}(M, R, T_i) + \varepsilon\,\sigma_{\ln Sa}(T_i) \tag{1.1}$$

where $\mu_{lnSa(Ti)}$ is the natural logarithm of the expected spectral acceleration at the given period Ti and μ_{lnSA} and σ_{lnSa} are the predicted median and standard deviation values obtained from the ground motion prediction model for the dominant event. 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 (e.g., Haselton et al. 2011). 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 (Baker and Cornell 2006a). The conservatism of the UHS has been addressed also by other researchers (e.g., Reiter 1990, Naeim and Lew 1995).

To deal with this problem the conditional mean spectrum (CMS) has been recently introduced by

Baker (Baker 2011) to be used in structural analysis instead of using UHS. The correlation of ε values in different periods is considered in CMS development; hence, the mentioned conservatism is improved. In order to explain this spectrum, consider a structural system with fundamental period, T*=1.0sec, 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 more common efficient intensity measures in literature (Luco and Bazzurro 2007, Tothong and Luco 2007), 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 Figure 1. By using a large set of ground motions (as will be defined in Section 2.2), the correlation between ε of T=1.0s and ε of other periods has been revealed in Figure 1. It is obvious that the different ε values are correlated, however with different coefficients of correlation (ρ). In Figures 1a, 1b, 1c and 1d, ε(1.0s) is plotted against ε(0.25s), ε(0.50s), ε(2.0s), and ε(4.0s), respectively.

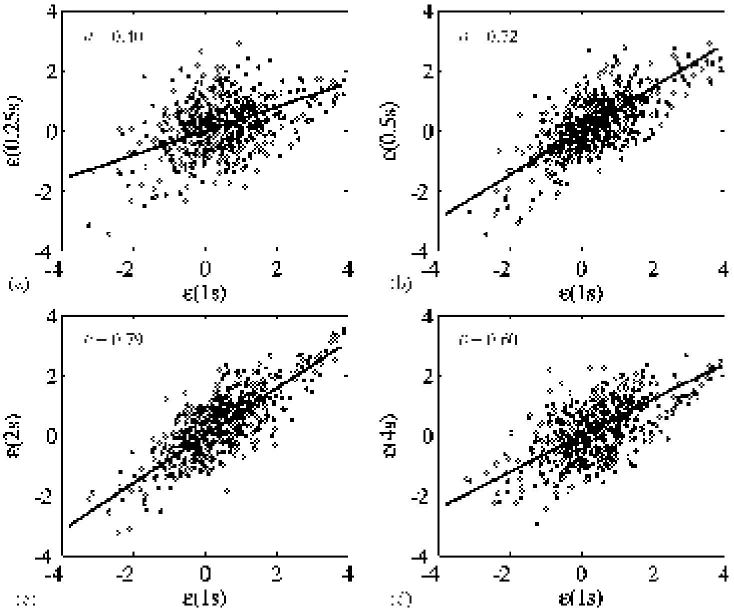

Figure 1. The Scatter plot of ε values from a large set of ground motions.
ε(1.0s) versus (a) ε(0.25s), (b) ε(0.5s), (c) ε(2.0s), and (d) ε(4.0s)

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 periods equal to the target ε multiplied by the correlation coefficient between the two ε values (Baker 2011):

$$e(T_i) = r\left(e(T^*), e(T_i)\right) e(T^*) \tag{1.2}$$

Here, 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 (1.3) to find the conditional mean spectrum (CMS).

$$m_{\ln Sa(T_i)|\ln Sa(T^*)} = m_{\ln Sa}(M, R, T_i) + r\left(T^*, T_i\right) e\, s_{\ln Sa}(T_i) \tag{1.3}$$

Figure 2 compares the CMS given ε(1.0sec)=1.88 in comparison with the UHS and the median spectrum. This Figure also includes CMS at a few other 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(Baker and Cornell 2006b, Baker and Jayaram 2008). Each of these closed form formulas in conjunction with Equation (1.2) build a complete framework to construct the CMS and use it instead of UHS in structural dynamic analyzes. For more details, see (Baker 2011).

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 median spectrum in Figure 2. The median spectrum, as shown in Figure 2, 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 (Baker 2011) 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.

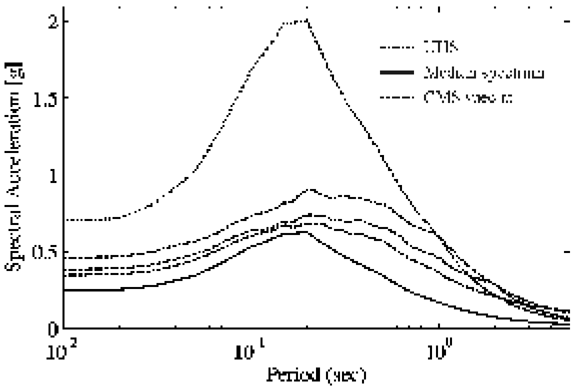

Figure 2. Conditional Mean Spectra, conditioned on Sa values at a few periods, having equal ε=1.88.

2. INTRODUCING A PROCEDURE TO FIND ROBUST CMS

2.1. 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 (Rousseeuw and Leroy 1987). Many robust methods have been developed since 1960 to detect outliers and to provide resistant (stable) results in the presence of outliers (Rousseeuw and Leroy 1987). 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 (Holland and Welsch 1977). In the first iteration, each point is assigned equal weight 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 (Rousseeuw and Leroy 1987).

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.

2.2. The Ground Motion Data set the Robust Regression Analysis

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 to analyze the correlation of response spectral values (Baker and Cornell 2006b). 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 (Baker 2005).

2.3. The Robust CMS

Using the above mentioned data set, a conventional regression analysis has been done by Baker and Cornell (Baker and Cornell 2006b) 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, 2.0, and 4.0 sec. On the other hand, the objective CMS is constructed just based on five period values (T*, and T). Figure 3 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 1.

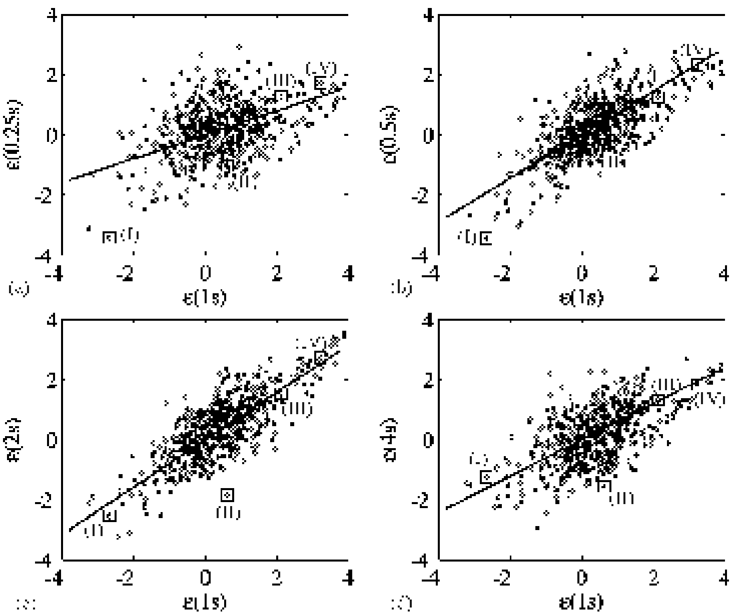

Figure 3. The Scatter plot of ε values from the introduced data set, ε(1.0s) versus (a) ε(0.25s), (b) ε(0.5s), (c) ε(2.0s), and (d) ε(4.0s). Four random records are marked.

Four random ground motions are marked in Figure 3. 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:

$$r(T^*,T_i) = e(T_i) - r\left(T^*,T_i\right)e(T^*) \tag{2.1}$$

where r(T*,Ti) symbolizes the observed residual in prediction of ε(Ti) from ε(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.

- Record marked as (I), this record 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], value 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 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 record (III).

Figure 4 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.

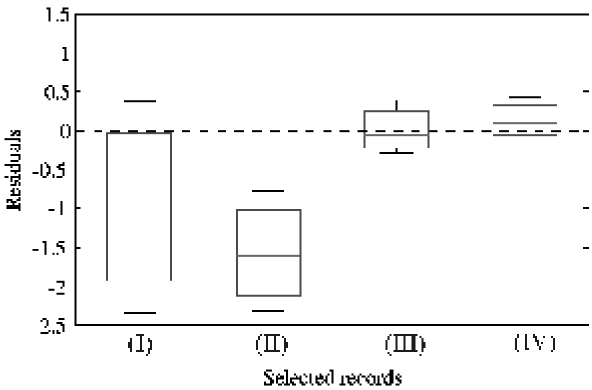

Figure 4. The box-plot of residual vectors for the selected ground motion records.

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:

$$w_i = \frac{1}{s_i\sqrt{2}} e^{\frac{-m_i^2}{2s_i^2}} \tag{2.2}$$

where w_i is the objective weight, and μi, and σi 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,
2- Calculate the ρ values for regression model.
3- Find the residual mean vector (M) as well as the residual standard deviation vector (Σ).
4- Determine the weighting vector based on Equation (2.2).
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 1 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 1. 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

Figure 5 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 lower period T=0.25 sec. By replacement of the robust correlation coefficients in Equation (1.3), the robust CMS is achievable. Figure 6 compares the resulted robust and the conventional CMS. Note that the CMS are formed by just five spectral acceleration values in five periods, here.

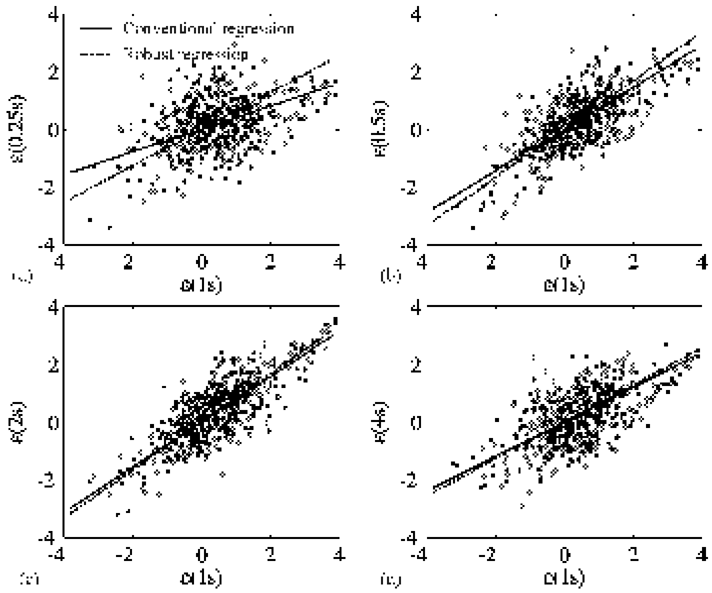

Figure 5. The comparison of the robust regression with conventional regression model

As shown in Figure 6, the shape of obtained robust CMS is significantly different with the conventional CMS, especially in lower periods. This issue will be more discussed in the next section, where the robust CMS procedure is extended for an entire range of periods.

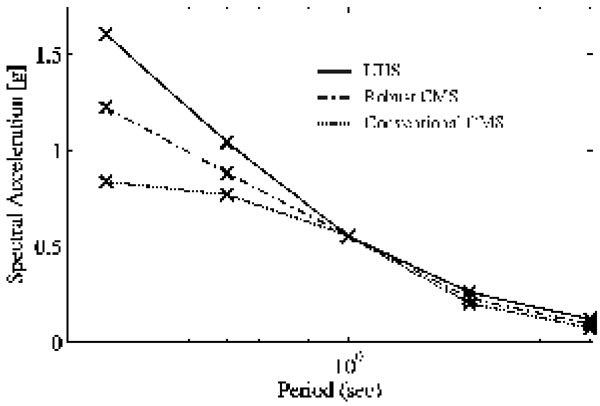

Figure 6. The comparison of the robust CMS with the conventional CMS

3. FINAL RESULTS OF THE ROBUST CMS

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 Figure 7. 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.

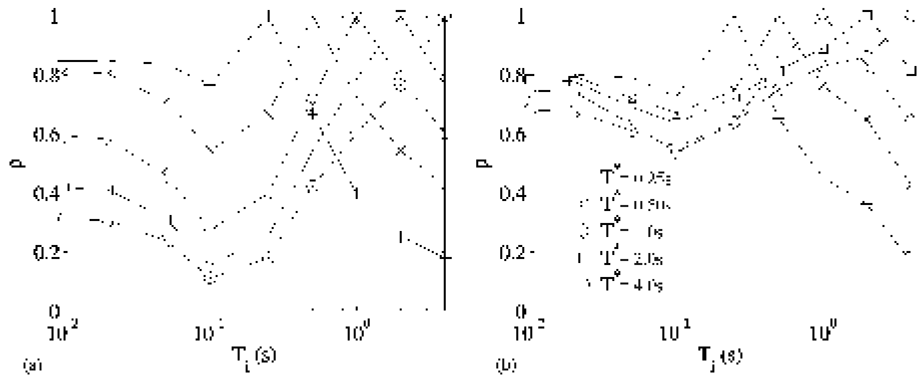

Figure 7. Plot of correlation coefficients versus T*, for several Ti values; based on (a) Conventional analysis (b) Robust correlation analysis

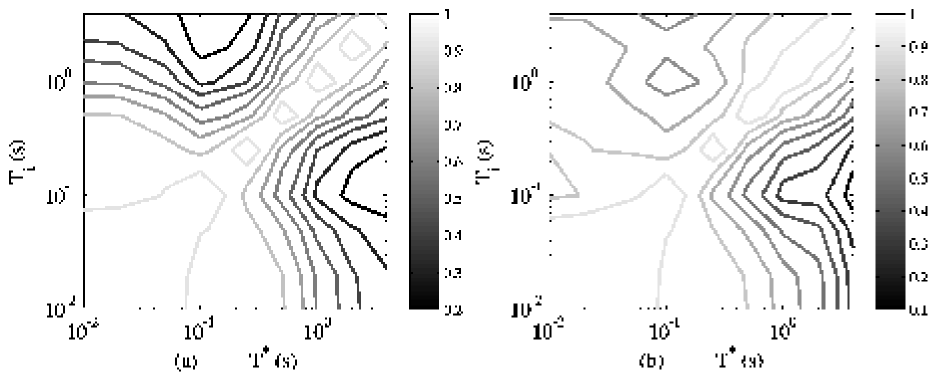

Figure 8. Contours of correlation coefficients versus T* and Ti; based on (a) Conventional analysis (b) Robust correlation analysis

Figure 8 shows the same results, plotted using contours of correlation coefficients as a function of both Ti and T*. The numerical values from Figure 8 are provided in Table 2. 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 2. Correlation coefficients of ε(T*) versus ε(Ti) obtained using CB08 ground motion prediction model; (a) Conventional analysis approach, and (b) Robust approach

(a)

	Ti														
T*	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
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)

	Ti														
T*	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
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 Figure 8 shows that despite the conventional analysis, the robust analysis has been yielded to asymmetric ρ results, i.e. $r(T_1, T_2) \neq r(T_1 = T_2)$. It means that the assigned weight to a single record is not fix and depends on the interest period (T*).

The robust CMS is achievable through the robust correlation coefficients, as stated in the former section. Figure 9 shows the robust CMS in a few interest periods for ε=1.88.

This Figure also comprises the conventional CMS and UHS curves.

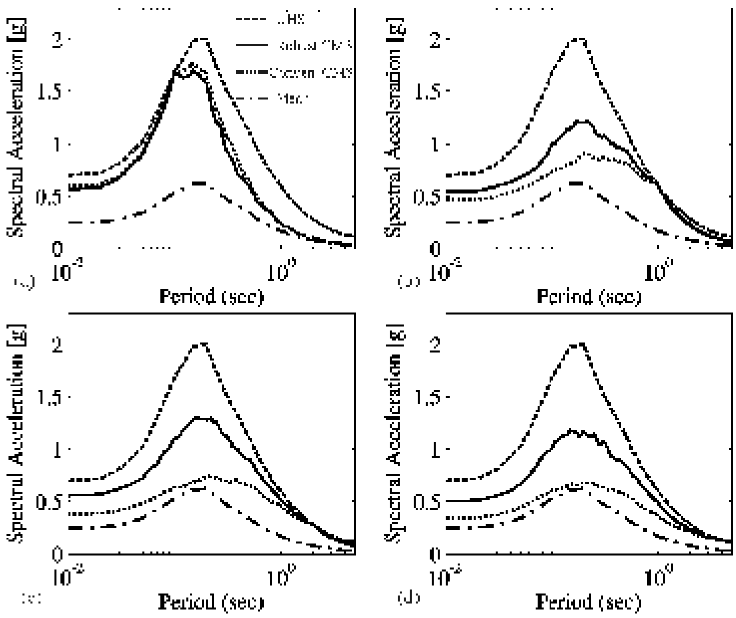

Figure 9. The robust CMS, conditioned on Sa values at the interest period.
(a)T*=0.50 sec, (b) T*=1.0 sec, (c) T*=2.0 sec, and (d) T*=4.0 sec, having equal ε=1.88.

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.

As the concluding result, the robust CMS shown in Figure 9 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 midrise structures in higher levels of seismic hazard.

4. CONCLUSIONS

The uniform hazard spectrum (UHS), as the more 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 the recent years. The CMS uses the advantages of correlation between spectral accelerations at different periods. This correlation is calculated based on an 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 conventional CMS.

REFERENCES

ASCE7-05 (2005). Minimum Design Loads for Buildings and Other Structures, American Society of Civil Engineers/Structural Engineering Institute, Reston ,VA.
Baker, J. W. (2005). Vector-valued ground motion intensity measures for probabilistic seismic demand analysis.

Report #150, John A.Blume Earthquake Engineering Center, Stanford.
Baker, J. W. (2011). Conditional Mean Spectrum: Tool for Ground-Motion Selection. *Journal of Structural Engineering* **137:3**,322-331.
Baker, J. W., Cornell, C.A (2006b). Correlation of Response Spectral Values for Multicomponent Ground Motions. *Bulletin of the Seismological Society of America* **96:1**, 215-227.
Baker, J. W., Cornell, C.A (2006a). Spectral shape, epsilon and record selection. *Earthquake Engineering & Structural Dynamics* **35:9**, 1077-1095.
Baker, J. W., Jayaram, N (2008). Correlation of spectral acceleration values from NGA ground motion models. *Earthquake Spectra* **24:1**, 299-317.
Bazzurro, P., Cornell, C.A. (1999). Disaggregation of seismic hazard. *Bulletin of the Seismological Society of America* **89:2**, 501-520.
Haselton, C. B., Baker, J. W., Liel, A. B., Deierlein, G. G. (2011). Accounting for Ground-Motion Spectral Shape Characteristics in Structural Collapse Assessment through an Adjustment for Epsilon. *Journal of Structural Engineering* **137:3**, 332-344.
Holland, P., and Welsch, R. (1977). Robust regression using iteratively reweighted least-squares. *communications in statistics-Theory and methods* **6:9**, 813-827.
Luco, N., Bazzurro, P. (2007). Does amplitude scaling of GMRs result in biased nonlinear structural drift responses? *Earthquake Engineering and Structural Dynamics* **36:13,** 1813-1835.
Naeim, F., Lew, M. (1995). On the Use of Design Spectrum Compatible Time Histories. *Earthquake Spectra* **11:1**, 111-127.
Reiter, L. (1990). Earthquake Hazard Analysis: Issues and Insights. Columbia University Press: New York;254
Rousseeuw, R. J., and leroy, A. M. (1987). Robust Regression and outlier Detection. New York, Wiley.
Tothong, P., Luco, N. (2007). Probabilistic seismic demand analysis using advanced ground motion intensity measures. *Earthquake Engineering and Structural Dynamics* **36:13**, 1837-1860.