Adaptive Fuzzy C-Mean Clustering of Ground Motion Prediction Equations

Alireza Azarbakht[1*] and Zahra Minaei [2]

1. Associate Professor, Department of Civil Engineering, Faculty of Engineering, Arak University, Iran
*Corresponding Author; email: a-azarbakht@araku.ac.ir
2. M.Sc. Student, Department of Civil Engineering, Faculty of Engineering, Arak University, Iran

Received: 17/02/2016
Accepted: 16/05/2017

ABSTRACT

Selection of Ground Motion Prediction Equations (GMPEs) within the Seismic Hazard Analysis (SHA) is an important and timely research line of inquiry. A set of 22 regional and worldwide GMPEs have been selected in this research for the purpose of classification. They are classified into clusters in which each cluster is defined to have the most dissimilarity with the other clusters as well as having the most similarity within the cluster. The C-mean clustering algorithm is modified and adapted in order to be applicable in the current study. In addition, six groups are defined for different focal mechanisms and soil types. Then, the GMPE clustering is performed for each group and the obtained clusters are proposed and discussed. The results confirm that the obtained spectral ordinated from GMPEs of different clusters can meaningfully differed from each other.

Keywords:
Earthquake;
GMPE;
Fuzzy C-mean clustering;
Un-supervised classification;
Validity index;
Next Generation Attenuation (NGA)

1. Introduction

Ground Motion Prediction Equation (GMPE) is the key element within any seismic hazard analysis (SHA). Different GMPEs have been revised by enrichment of earthquake catalogues in which a wide variety of GMPEs are now available [1-2]. However, any GMPE depends strongly on the selected ground motion database. In other words, employing non-local GMPEs may be a challenging task [3]. Additionally, selection of a suitable GMPE is usually performed without enough attention to the compatibility of the chosen GMPE on historical site database. Therefore, the selection of appropriate GMPEs for regions, which suffer from the lack of available GMPEs, is usually a serious challenge. Therefore, the fuzzy C-mean clustering approach [4] is adapted and employed in this study in order to classify the available well-known GMPEs into different clusters. A set of models can be divided into a number of subsets, i.e. clusters [4], in which each cluster members have the most dissimilarity with the other cluster members and having the most similarity within the given cluster [5].

Clustering approaches are categorized into the hard and soft clusters. In hard clustering approach, each data is assigned to only a specific cluster whereas in the soft (fuzzy) clustering approach each data has a certain membership value for each cluster [5]. Additionally, fuzzy clustering approach does not employ any category label that is a unique characteristic in this sort of approaches [6]. In other words, in contradiction of the classification in which the given data are classified into pre-defined classes, clusters are not defined in priori in fuzzy clustering approach. Therefore, fuzzy clustering approach is a sort of un-supervised

classification [5], which makes it suitable for the purpose of the current study. It is worth mentioning that the idea of GMPE clustering is a novel approach in which this study is the first attempt in this issue.

The fuzzy clustering approach is introduced in the upcoming section. Then, the fuzzy C-mean clustering approach is adapted in order to be applicable in the case of GMPE clustering. A set of 22 GMPE's are introduced later and employed within the adaptive fuzzy C-mean clustering in order to obtain different clusters. A validity measure for the number of clusters is discussed and the best number is chosen. Finally, the spectral accelerations based on different clusters are obtained and discussed.

2. Fuzzy Clustering

The concept of fuzzy clustering, in order to construct a set of unknown border clusters, has been proposed first by Zadeh [7]. Each data in clustering process is assigned to a cluster by a membership value based on the following criteria:

$$u_{ij} \in [0\ ,\ 1] \quad 1 \le i \le C, \quad 1 \le j \le N \tag{1}$$

$$\sum_{i=1}^{C} u_{ij} = 1 \quad 1 \le j \le N \tag{2}$$

$$\sum_{j=1}^{N} u_{ij} > 0 \quad 1 \le i \le C \tag{3}$$

where N, C and u_{ij} are, respectively, number of data, number of clusters and membership value of the j^{th} data for the i^{th} cluster. According to Eqs. (1) and (2), each data gets a membership value between 0 and 1 in the case of each cluster. Hence, the summation of all membership values for a specific data is equal to unity. Eq. (3) guarantees that there will not be a null cluster among all clusters [5].

The Fuzzy C-Mean (FCM) clustering is one of the most popular methods within the fuzzy clustering methods. As proposed by Bezdek in 1981 [8], a centre point is assigned to each cluster in which the distance of each data from this point is defined as the corresponding membership value. The membership value is closer to unity for the cases that the given data is close enough to the centre point [8]. It is worth noting that the final clustering depends only on the data distribution and is independent of the cluster centres. In other words, after defining the number of clusters, the cluster centres as well as the membership values are calculated without any influence of the user. This characteristic is the most important scheme of the FCM approach, which makes it suitable for the current study. On the other hand, some modifications of the original version of the FCM are necessary in order to be applicable in the GMPEs clustering. These modifications are mandatory since each GMPE has its own input variables that may be different with the other models. Additionally, for a specific physical phenomenon, e.g. distance, different definitions are available in the literature. Therefore, the next section is discussed on the details of the adapted FCM clustering approach in the current study.

3. Adapted Fuzzy C-Mean Clustering Algorithm

In order to define different GMPEs, SA matrix is introduced here as written in Eq. (4).

$$SA = \begin{bmatrix} Sa_1 \\ Sa_2 \\ \vdots \\ \vdots \\ Sa_N \end{bmatrix} = \begin{bmatrix} Sa_1(A) & \cdots & \cdots & Sa_1(B) \\ Sa_2(A) & & & Sa_2(B) \\ & & & \\ & & & \\ Sa_N(A) & & & Sa_N(B) \end{bmatrix} \tag{4}$$

$$A = {M_W}^1, T^1, R_{JB}^1, V_{S30}^1, \alpha^1, h^1, \lambda^1$$

$$B = {M_W}^n, T^n, R_{JB}^n, V_{S30}^n, \alpha^n, h^n, \lambda^n$$

where Sa_1 to Sa_N are spectral acceleration obtained from N (given) GMPEs. Mw, T, R_{JB}, V_{S30}, α, h, λ are, respectively, moment magnitude, period, horizontal distance to the surface projection of the rupture (Joyner-Boore distance), time-averaged shear-wave velocity over the top 30 meters of the subsurface, source-to-site azimuth, hypo-central depth, rake angle. Moreover, S_{aN} $({M_W}^n, T^n, R_{JB}^n, V_{S30}^n, \alpha^n, h^n, \lambda^n)$ is the spectral acceleration corresponding to N^{th} model which has

been calculated by using the following values: $M_W{}^n, T^n, R_{JB}^n, V_{S30}^n, \alpha^n, h^n$ and λ^n.

Now a *SA* matrix is in hand, by using Eq. (4), in which each row corresponds to a specific GMPE and each column corresponds to a specific input variable for different models. It is worth emphasizing that RJB is selected in all models for the distance variable. Other definitions for the distance variable are transformed to the RJB definition by employing the transformation relationships [9]. The transformation relationships are also used to transform all the models outputs to the geometric mean of the two horizontal components [10]. In the case that a specific variable value is not valid for the defined range of a given GMPE, Not a Number (NaN) is placed in the corresponding cell in the *SA* matrix. It means that in the case of that cell, there is no value for the specific model.

The adapted FCM clustering algorithm, which has been employed in the current study, is introduced as the following steps:

- After defining the number of clusters, the membership matrix is randomly calculated as written in Eq. (5).

$$U=\begin{bmatrix} u_{11} & \cdots & u_{1N} \\ \vdots & & \vdots \\ u_{C1} & \cdots & u_{CN} \end{bmatrix} \qquad 0 \le u_{CN} \le 1 \tag{5}$$

where u_{CN} is the membership value of Nth data for Cth cluster centre.

- Eq. (4) is re-written as Eq. (6) for the purpose of simplicity. The cluster centres is calculated by using Eq. (7). As some of the cells in the SA matrix is NaN, the corresponding is set to zero when using Eq. (7). Additionally, in the case of in Eq. (7), the membership value corresponding to NaN case is also set to zero.

$$SA=\begin{bmatrix} Sa_1 \\ Sa_2 \\ \vdots \\ \vdots \\ Sa_N \end{bmatrix}=\begin{bmatrix} Sa_1^1 & \cdots & \cdots & Sa_1^n \\ Sa_2^1 & \cdots & \cdots & Sa_2^n \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ Sa_N^1 & \cdots & \cdots & Sa_N^n \end{bmatrix} \tag{6}$$

$$C=\begin{bmatrix} \frac{\sum_{i=1}^{N} u_{1i}^m \cdot Sa_i^1}{\sum_{i=1}^{N} u_{1i}^m} & \cdots & \cdots & \frac{\sum_{i=1}^{N} u_{1i}^m \cdot Sa_i^n}{\sum_{i=1}^{N} u_{1i}^m} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \frac{\sum_{i=1}^{N} u_{ci}^m \cdot Sa_i^1}{\sum_{i=1}^{N} u_{ci}^m} & \cdots & \cdots & \frac{\sum_{i=1}^{N} u_{ci}^m \cdot Sa_i^n}{\sum_{i=1}^{N} u_{ci}^m} \end{bmatrix} 1<\mathrm{m}<\infty, \tag{7}$$

- The similarity is defined in this step as the Manhattan distance [5], which is written in Eq. (8).

$$D_{Manw}(Sa_N, C_c)=\frac{\sum_{i=1}^{n} w_i \left| Sa_N^i - C_C^i \right|}{\sum_{i=1}^{n} w_i} \tag{8}$$

here $D_{Manw}(Sa_N, C_c)$ is the distance of Nth model with Cth cluster. When either Sa_N^i or C_C^i in Eq. (8) is NaN, then, w_i is set to zero. Otherwise w_i is set to unity. Therefore, the similarity matrix is calculated as written in Eq. (9).

$$D=\begin{bmatrix} \frac{\sum_{i=1}^{n} w_i \left| Sa_1^i - C_1^i \right|}{\sum_{i=1}^{n} w_i} & \cdots & \cdots & \frac{\sum_{i=1}^{n} w_i \left| Sa_N^i - C_1^i \right|}{\sum_{i=1}^{n} w_i} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \frac{\sum_{i=1}^{n} w_i \left| Sa_1^i - C_C^i \right|}{\sum_{i=1}^{n} w_i} & \cdots & \cdots & \frac{\sum_{i=1}^{n} w_i \left| Sa_N^i - C_C^i \right|}{\sum_{i=1}^{n} w_i} \end{bmatrix}, \tag{9}$$

where $C \times N$ cell in the matrix corresponds to the distance between N^{th} model with Cth cluster centre. The objective function is defined as the sum of the distances as written in Eq. (10).

$$\begin{aligned} J_{FCM}(U,C) &= \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^m \cdot D_{Manw}^2(Sa_j, C_i) \\ &= \sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij}^m \cdot D_{ij}^2 \qquad 1<m, \end{aligned} \tag{10}$$

where D_{ij} corresponds to the ith row by jth column in *D* matrix.

As the *U* matrix is calculated randomly, it

should be corrected in an iterative process by the goal of minimization of the objective function. This is achieved by updating the *U* matrix by employing Eq. (11). The old *U* matrix, then, is replaced with the new *U* matrix and all the steps are repeated. This iteration process is repeated until the difference between the objective function with the previous objective function becomes less than a predefined tolerance, i.e. 10^{-5} in the current study. The mathematical form of the stopping criterion is written in Eq. (12).

$$U_{new} = \begin{bmatrix} \frac{D_{11}^{\frac{-2}{m-1}}}{\sum_{i=1}^{C} D_{i1}^{\frac{-2}{m-1}}} & \cdots & \cdots & \frac{D_{1N}^{\frac{-2}{m-1}}}{\sum_{i=1}^{C} D_{iN}^{\frac{-2}{m-1}}} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \frac{D_{C1}^{\frac{-2}{m-1}}}{\sum_{i=1}^{C} D_{i1}^{\frac{-2}{m-1}}} & \cdots & \cdots & \frac{D_{CN}^{\frac{-2}{m-1}}}{\sum_{i=1}^{C} D_{iN}^{\frac{-2}{m-1}}} \end{bmatrix} \quad (11)$$

$$\left|(J_{FCM})_J - (J_{FCM})_{J-1}\right| < 10^{-5} \quad (12)$$

where $(J_{FCM})_J$ and $(J_{FCM})_{J-1}$ are, respectively, corresponding to the objective functions of j[th] and $(J-1)^{th}$ iteration.

The GMPEs are clustered into a certain number based on their prediction capability of the spectral values. Each given GMPE has a membership value for each cluster. In addition, each cluster centre is accounted as the representative of the corresponding cluster. To clarify, three simple constant functions with different domains, as seen in Figure (1), are examined with the proposed clustering algorithm. The proposed FCM clustering algorithm has been applied to the given functions in order to produce two clusters. The results are shown in Figure (1) in which four different regions are distinguished. The following characteristics are discussed based on Figure (1).

- At the first region, only one function is defined. Hence, all the clusters centres are identical with the defined function.
- At the second region, the function $y_2=0$ has NaN values. Therefore, the clusters centres are identical with the other functions.
- At the third region, the $y_2=0$ function has numeric values, in contrast to the previous region. Hence, the clusters centres are closer to this function compared with the previous case.

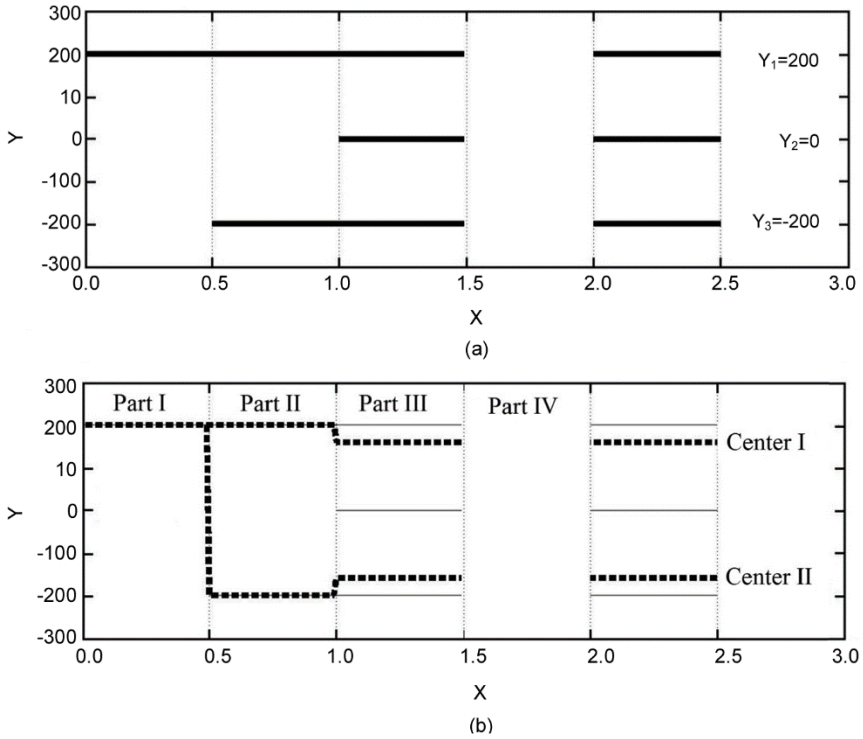

Figure 1. Performance of the proposed FCM clustering algorithm. (a) Introduction of three hypothetical functions y_1, y_2, y_3. (b) Clustering of the three hypothetical functions into two clusters using the proposed FCM clustering algorithm.

- At the forth region, all the functions have NaN values. This results in NaN in cluster centres.

The behaviour that is discussed in the four regions confirms that this adapted FCM algorithm is suitable to be applied to a set of GMPEs which is discussed in the following section.

4. Fuzzy Clustering of GMPEs

A wide variety of GMPEs are selected from 1997 to 2012 [2] that include 22 worldwide and regional models. Three different limitations are taken into consideration in the model selection as the followings:

1. The transformation relationships [9-10] should be applicable to any of the selected GMPE.
2. All the models use the moment magnitude as one of the inputs. Any model which uses other magnitude definition is eliminated. The only exception is Ghodrati et al [11] which is one of the Iranian well-known models. Ghodrati et al [11] model uses surface magnitude which was transformed into the moment magnitude by employing the transformation relationships [1].
3. Any GMPE corresponding to the near-field region is eliminated.

By considering the three above criteria, 22 models were selected which are shown in Table 1. Each model and the corresponding input variables are shown in Table (1). It is worth noting that NGA-WEST2 models were not available in the time of the current research. Therefore, these new models are left for future researches.

Table 1. Selected GMPEs with the range of input parameters.

No.	Model	M_w	Distance (km)	V_{S30} (m/s)	Depth (km)	λ	Definitions of Horizontal Component of Motion	Ref.
				Worldwide Shallow Crustal				
1	Boore & Atkinson 08 (Crustal)	5-8	0-200	180-1300	2-31	-90°, 0°, 90°	GMRotI50[a]	[12]
2	Campbell & Bozorgnia 08 (Crustal)	4-8.5	0-200	150-1500	0-30	-90°, 0°, 90°	GMRotI50	[13]
3	Abrahamson & Silva 08 (Crustal)	5-8.5	0-200	0-1000	0-30	-90°, 0°, 90°	GMRotI50	[14]
4	Chiou & Youngs 08 (Crustal)	4-8.5	0-200	150-1500	0-19	-90°, 0°, 90°	GMRotI50	[15]
5	Idriss 08 (Crustal)	4.5-7.7	0-200	450-900	0-30	-90°, 0°, 90°	AMxy[b]	[16]
				Worldwide				
6	Kuehn 09	5.6-7.9	0.1-200	200-1500	0-10.13	-90°, 0°, 90°	GM[c]	[17]
				European & Middle East				
7	Ambraseys 05	5-7.6	0-100	180-1000	1-30	-90°, 0°, 90°	Envxy[d]	[18]
8	Akkar & Bommer 10	5-7.6	0-100	0-1000	0-30	-90°, 0°, 90°	GM	[19]
				Italy				
9	Bindi 10	4-6.9	1-100	180-1000	1.5-30	-90°, 0°, 90°	Envxy	[20]
				Japan				
10	Zhao 06 (Crustal)	5-8.3	0-300	0-1000	0-25	-90°, 0°, 90°	GM	[21]
11	Zhao 06 (Interface)	5-8.3	0-300	0-1000	10-50	-90°, 0°, 90°	GM	[21]
				Turkey				
12	Kalkan & Gulkan 04	4-7.5	0-250	200,400,700	0-111	-90°, 0°, 90°	Envxy	[22]
13	Ozbey 04	5-7.4	5-300	0-1000	5-25	-90°, 0	GM	[23]
14	Akkar & Cagnan 10	5-7.6	0-200	180-1000	0-50	-90°, 0°, 90°	GM	[24]
				Iran				
15	Ghasemi 09	5-7.4	0.5-100	0-1000	5-22	-90°, 0°, 90°	GMRotI50	[25]
16	Ghodrati 10 (Alborz)	4-7.7	5-200	0-1000	5-59	-90°, 0°, 90°	Envxy	[11]
17	Ghodrati 10 (Zagros)	4-7.7	5-200	0-1000	5-59	-90°, 0°, 90°	Envxy	[11]
18	Saffari 12 (Central Iran)	5-7.3	15-135	175-1000	7-27	0°, 90°	GMRotI50	[26]
19	Saffari 12 (Zagros)	5-6.5	15-135	175-1000	7-27	0°, 90°	GMRotI50	[26]
20	Zafarani 12(Zagros)	4.4-7.5	1-200	175-1000	0-30	-90°, 0°, 90°	GM	[27]
				Himalaya				
21	Sharma 09	5-7	0-100	0-1000	5-50	0°, 90°	GM	[28]
				Eurasia				
22	Fukushima 03	5.5-7.4	0.5-235	0-1000	3-30	0°	Both[e]	[29]

[a]: The geometric mean determined from the 50th percentile values of the geometric means computed for all non-redundant rotation angles and all periods less than the maximum useable period.
[b]: Arithmetic mean of spectra of x and y components
[c]: Geometric mean of spectra of x and y components
[d]: Envelope of x and y spectra: At each period the maximum spectral ordinate from all possible orientations of the horizontal axis is chosen. This is the common understanding of the "larger component" definition.
[e]: Both horizontal components of a record are considered and treated as two independent realizations of a random process. This definition was used, in particular, when ground-motion data were still very sparse.

The input variables ranges are defined as shown in Table (2) for M_w, T, R_{JB}, V_{S30}, α, h and λ. The authors tried to create a balance between the ranges of the considered input variables with the computational efforts of the clustering algorithm. The input SA matrix dimension is 22 x 179928 that reveals the serious difficulties in the computational aspects of the current research.

Table 2. Introduction of the range of input variables.

Parameters	Range
Moment Magnitude	M=4:0.25:8
Period (s)	$T=\begin{bmatrix}0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 0.7\\ 0.8 & 0.9 & 1.0 & 1.25 & 1.5 & 1.75 & 2\end{bmatrix}$
Distance (km)	R_{jb}=25:25:150;
Shear-Wave Velocity (m/s)	Vs=[175 275 375]; Vs=[400 550 700];
Azimuth (Degree)	$\alpha=\{-180^\circ,-120^\circ,-60^\circ,0^\circ,60^\circ,120^\circ,180^\circ\}$
Focal Depth (Km)	h= 5:5:30
Rake Angle (Degree)	$\lambda=\{-90^\circ,0^\circ,90^\circ\}$

Two soil types are defined in this study, i.e. Soft soil is corresponding to 175<Vs30<375 and the firm soil is corresponding to 400<Vs30<700. Additionally, three different fault mechanisms are taken into consideration, i.e. normal ($\lambda=-90$), strike-slip ($\lambda=0$) and reverse ($\lambda=+90$). Therefore, six combinations of the soil type and fault mechanism are available in order to classify the considered GMPEs. The number of GMPEs in each combination (group) is illustrated in Table (3).

The next step in the clustering algorithm is the decision on the number of clusters which is a challenging task. Therefore, the number of clusters was varied between 2 to $\sqrt{N_{GMPE}}$ achieve the best choice in which N_{GMPE} is the number of GMPEs within a given group. For this purpose, the validity index (V_{XB}) has been used as written in Eq. (13) [30]. The result of this index is shown in Table (4). It is worth noting that high values, in Table (4), are corresponding to the better validity indices. Hence, it is concluded that three clusters in the case of Group No. 1, five clusters in the cases of Groups Nos. 2, 3, 4 and 5, and four clusters in the case of Group No. 6 are appropriate in the current study.

Table 3. The available GMPEs in each group.

No.	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
			Model			
1	Boore & Atkinson 08 (Crustal)	Boore & Atkinson 08 (Crustal)	Boore & Atkinson 08 (Crustal)	Boore & Atkinson 08 (Crustal)	Boore & Atkinson 08 (Crustal)	Boore & Atkinson 08 (Crustal)
2	Campbell & Bozorgnia 08 (Crustal)	Campbell & Bozorgnia 08 (Crustal)	Campbell & Bozorgnia 08 (Crustal)	Campbell & Bozorgnia 08 (Crustal)	Campbell & Bozorgnia 08 (Crustal)	Campbell & Bozorgnia 08 (Crustal)
3	Abrahamson & Silva 08 (Crustal)	Abrahamson & Silva 08 (Crustal)	Abrahamson & Silva 08 (Crustal)	Abrahamson & Silva 08 (Crustal)	Abrahamson & Silva 08 (Crustal)	Abrahamson & Silva 08 (Crustal)
4	Chiou & Youngs 08 (Crustal)	Chiou & Youngs 08 (Crustal)	Chiou & Youngs 08 (Crustal)	Chiou & Youngs 08 (Crustal)	Chiou & Youngs 08 (Crustal)	Chiou & Youngs 08 (Crustal)
5	-	-	-	Idriss 08 (Crustal)	Idriss 08 (Crustal)	Idriss 08 (Crustal)
6	Kuehn 09	Kuehn 09	Kuehn 09	Kuehn 09	Kuehn 09	Kuehn 09
7	Ambraseys 05	Ambraseys 05	Ambraseys 05	Ambraseys 05	Ambraseys 05	Ambraseys 05
8	Akkar & Bommer 10	Akkar & Bommer 10	Akkar & Bommer 10	Akkar & Bommer 10	Akkar & Bommer 10	Akkar & Bommer 10
9	Bindi 10	Bindi 10	Bindi 10	Bindi 10	Bindi 10	Bindi 10
10	Zhao 06 (Crustal)	Zhao 06 (Crustal)	Zhao 06 (Crustal)	Zhao 06 (Crustal)	Zhao 06 (Crustal)	Zhao 06 (Crustal)
11	Zhao 06 (interface)	Zhao 06 (interface)	Zhao 06 (interface)	Zhao 06 (interface)	Zhao 06 (interface)	Zhao 06 (interface)
12	Kalkan & Gulkan 04	Kalkan & Gulkan 04	Kalkan & Gulkan 04	Kalkan & Gulkan 04	Kalkan & Gulkan 04	Kalkan & Gulkan 04
13	Ozbey 04	Ozbey 04	-	Ozbey 04	Ozbey 04	-
14	Akkar & Cagnan 10	Akkar & Cagnan 10	Akkar & Cagnan 10	Akkar & Cagnan 10	Akkar & Cagnan 10	Akkar & Cagnan 10
15	Ghasemi 09	Ghasemi 09	Ghasemi 09	Ghasemi 09	Ghasemi 09	Ghasemi 09
16	Ghodrati 10 (Alborz)	Ghodrati 10 (Alborz)	Ghodrati 10 (Alborz)	Ghodrati 10 (Alborz)	Ghodrati 10 (Alborz)	Ghodrati 10 (Alborz)
17	Ghodrati 10 (Zagros)	Ghodrati 10 (Zagros)	Ghodrati 10 (Zagros)	Ghodrati 10 (Zagros)	Ghodrati 10 (Zagros)	Ghodrati 10 (Zagros)
18	-	Saffari 12 (Central Iran)	Saffari 12 (Central Iran)	-	Saffari 12 (Central Iran)	Saffari 12 (Central Iran)
19	-	Saffari 12 (Zagros)	Saffari 12 (Zagros)	-	Saffari 12 (Zagros)	Saffari 12 (Zagros)
20	Zafarani 12 (Zagros)	Zafarani 12 (Zagros)	Zafarani 12 (Zagros)	Zafarani 12 (Zagros)	Zafarani 12 (Zagros)	Zafarani 12 (Zagros)
21	-	Sharma 09	Sharma 09	-	Sharma 09	Sharma 09
22	-	Fukushima 03	-	-	Fukushima 03	-
			The total number of attenuation models in each case.			
	17	21	19	18	22	20

Table 4. Xie and Beni index to evaluate the validity of clusters.

The Number of Clusters (N)	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
N=2	0.2827	0.3260	0.3082	0.2662	0.2629	0.2761
N=3	0.1815	0.3636	0.4778	0.1739	0.2864	0.2595
N=4	0.2790	0.3759	0.2957	0.1651	0.2353	0.1346
N=5	0.2206	0.1255	0.2376	0.1576	0.1014	0.2546

The clustering results in each group are shown through Table (5) to Table (10) as the main result of the modified C-mean clustering algorithm. Each table illustrates the clusters GMPEs as well as the membership values for a given group.

$$V_{XB} = \frac{J_{FCM}}{N \min \left\| C_i - C_j \right\|^2} \qquad 1 \le i < j \le C \tag{13}$$

To more elaborate with the obtained clusters, each cluster centre is taken as the representative of that cluster. For example, different cluster centres in the case of group No. 1 are shown in Figure (2) versus period. The third cluster is remarkably different with the other cluster centres as seen in Figure (2). These clustering results are applicable to a given seismic data in a specific region. The observed data are comparable with the cluster centres in order to find the best fitted cluster to be used in PSHA.

Table 5. Clustering of 17 models in case 1.

Number of Clusters = 3		
No.	Attenuation Relationship	U
	Cluster 1	
1	Boore & Atkinson 08 (Crustal)	0.8992
2	Campbell & Bozorgnia 08 (crustal)	0.4080
4	Chiou & Youngs 08 (crustal)	0.8583
8	Akkar & Bommer 10	0.8829
9	Bindi 10	0.5701
`12	Kalkan & Gulkan 04	0.4878
13	Ozbey 04	0.6414
14	Akkar & Cagnan 10	0.8516
16	Ghodrati 10 (Alborz)	0.7734
17	Ghodrati 10 (Zagros)	0.8279
	Cluster 2	
7	Ambraseys 05	0.4625
	Cluster 3	
3	Abrahamson & Silva 08 (crustal)	0.4579
6	Kuehn 09	0.9037
20	Zafarani 12 (Zagros)	0.5903

Table 6. Clustering of 21 models in case 2.

Number of Clusters = 5		
No	Attenuation Relationship	U
	Cluster 1	
13	Ozbey 04	0.5269
14	Akkar & Cagnan 10	0.7068
16	Ghodrati 10 (Alborz)	0.6158
17	Ghodrati 10 (Zagros)	0.8788
21	Sharma 09	0.7106
	Cluster 2	
6	Kuehn 09	0.9705
18	Saffari 12 (Central Iran)	0.3945
19	Saffari 12(Zagros)	0.3756
20	Zafarani 12 (Zagros)	0.4616
	Cluster 3	
3	Abrahamson & Silva 08 (crustal)	0.9965
	Cluster 4	
7	Ambraseys 05	0.4769
9	Bindi 10	0.5530
10	Zhao 06 (crustal)	0.9274
11	Zhao 06 (interface)	0.9237
15	Ghasemi 09	0.4906
	Cluster 5	
1	Boore & Atkinson 08 (crustal)	0.9066
2	Campbell & Bozorgnia 08 (crustal)	0.8813
4	Chiou & Youngs 08 (crustal)	0.5716
8	Akkar & Bommer 10	0.5022
12	Kalkan & Gulkan 04	0.7805
22	Fukushima 03	0.6473

Table 7. Clustering of 19 models in case 3.

Number of Clusters = 5		
No	Attenuation Relationship	U
	Cluster 1	
14	Akkar & Cagnan 10	0.4018
16	Ghodrati 10 (Alborz)	0.7572
17	Ghodrati 10 (Zagros)	0.8748
	Cluster 2	
7	Ambraseys 05	0.6242
9	Bindi 10	0.5202
11	Zhao 06 (interface)	0.9211
15	Ghasemi 09	0.5975
	Cluster 3	
10	Zhao 06 (crustal)	0.9521
18	Saffari 12 (Central Iran)	0.4389
20	Zafarani 12 (Zagros)	0.6736
	Cluster 4	
3	Abrahamson & Silva 08 (crustal)	0.9830
6	Kuehn 09	0.3300

Table 7. Continue

	Cluster 5	
1	Boore & Atkinson 08 (crustal)	0.9024
2	Campbell & Bozorgnia 08 (crustal)	0.7637
4	Chiou & Youngs 08 (crustal)	0.7258
8	Akkar & Bommer 10	0.4586
12	Kalkan & Gulkan 04	0.6403
19	Saffari 12(Zagros)	0.3495
21	Sharma 09	0.3887

Table 8. Clustering of 18 models in case 4.

Number of Clusters = 5		
No	Attenuation Relationship	U
	Cluster 1	
6	Kuehn 09	0.9862
7	Ambraseys 05	0.3280
15	Ghasemi 09	0.2860
	Cluster 2	
3	Abrahamson & Silva 08 (crustal)	0.9817
20	Zafarani 12 (Zagros)	0.4140
	Cluster 3	
9	Bindi 10	0.5964
10	Zhao 06 (crustal)	0.9624
11	Zhao 06 (interface)	0.9538
	Cluster 4	
13	Ozbey 04	0.5411
14	Akkar & Cagnan 10	0.6499
16	Ghodrati 10 (Alborz)	0.6564
17	Ghodrati 10 (Zagros)	0.8738
	Cluster 5	
1	Boore & Atkinson 08 (crustal)	0.8759
2	Campbell & Bozorgnia 08 (crustal)	0.8199
4	Chiou & Youngs 08 (crustal)	0.6577
5	Idriss 08 (crustal)	0.4999
8	Akkar & Bommer 10	0.7499
12	Kalkan & Gulkan 04	0.4901

Table 9. Clustering of 22 models in case 5.

Number of Clusters = 5		
No	Attenuation Relationship	U
	Cluster 1	
6	Kuehn 09	0.9504
15	Ghasemi 09	0.3331
19	Saffari 12(Zagros)	0.5310
20	Zafarani 12 (Zagros)	0.3772
22	Fukushima 03	0.4314
	Cluster 2	
7	Ambraseys 05	0.3503
9	Bindi 10	0.7015

Table 9. Continue

10	Zhao 06 (crustal)	0.9463
11	Zhao 06 (interface)	0.9325
18	Saffari 12 (Central Iran)	0.3076
	Cluster 3	
3	Abrahamson & Silva 08 (crustal)	0.9869
	Cluster 4	
13	Ozbey 04	0.6434
14	Akkar & Cagnan 10	0.7039
16	Ghodrati 10 (Alborz)	0.7309
17	Ghodrati 10 (Zagros)	0.8549
21	Sharma 09	0.6033
	Cluster 5	
1	Boore & Atkinson 08 (crustal)	0.9220
2	Campbell & Bozorgnia 08 (crustal)	0.9143
4	Chiou & Youngs 08 (crustal)	0.5657
5	Idriss 08 (crustal)	0.5985
8	Akkar & Bommer 10	0.7258
12	Kalkan & Gulkan 04	0.7018

Table 10. Clustering of 20 models in case 6.

Number of Clusters = 5		
No	Attenuation Relationship	U
	Cluster 1	
1	Boore & Atkinson 08 (crustal)	0.9237
2	Campbell & Bozorgnia 08 (crustal)	0.9173
4	Chiou & Youngs 08 (crustal)	0.8236
5	Idriss 08 (crustal)	0.5280
8	Akkar & Bommer 10	0.5694
11	Zhao 06 (interface)	0.3637
12	Kalkan & Gulkan 04	0.7229
19	Saffari 12(Zagros)	0.3605
21	Sharma 09	0.3916
	Cluster 2	
6	Kuehn 09	0.8826
7	Ambraseys 05	0.6741
9	Bindi 10	0.3300
15	Ghasemi 09	0.4589
18	Saffari 12 (Central Iran)	0.4984
	Cluster 3	
3	Abrahamson & Silva 08 (crustal)	0.4235
10	Zhao 06 (crustal)	0.9097
20	Zafarani 12 (Zagros)	0.6570
	Cluster 4	
14	Akkar & Cagnan 10	0.4963
16	Ghodrati 10 (Alborz)	0.8794
17	Ghodrati 10 (Zagros)	0.9012
14	Akkar & Cagnan 10	0.4963

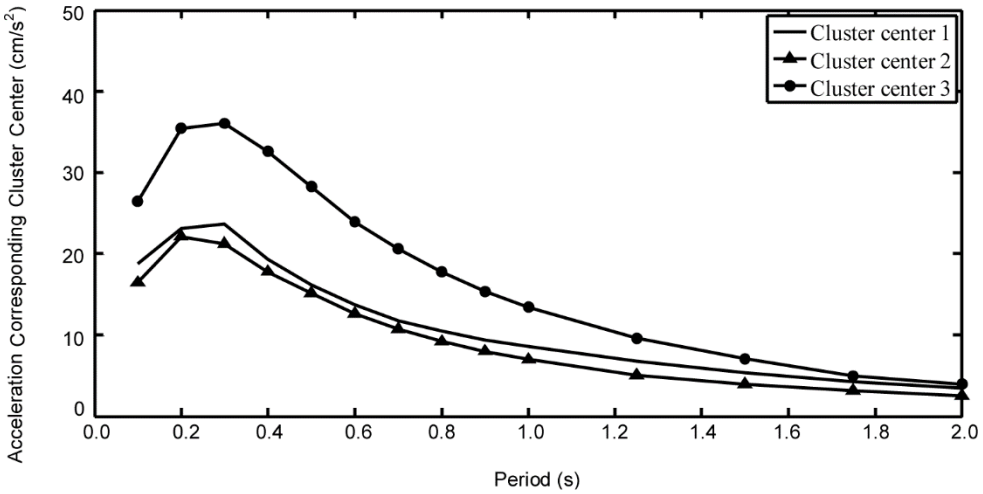

Figure 2. Different cluster centres in the case of group No. 1 and M_W = 6, R_{JB} = 150km, V_{S30} = 375m/s, α = 180°, h = 20m, λ = -90°.

5. Conclusions

The GMPEs are widely used within any seismic hazard analysis. On the other hand, local GMPEs are not available in many regions. That is, a systematic clustering algorithm is employed in this study in order to join similar models in a same cluster. The adapted fuzzy c-mean clustering algorithm is utilized to obtain the final clusters since this method is only depends on the models variation and not on the models centres. Each model is assigned to a cluster by a certain membership value that is between 0 and 1. A high membership value shows significant dependence of the selected model to the obtained cluster.

A set of 22 regional and worldwide GMPEs were selected in order to be clustered. Three different focal mechanisms and two soil types were defined in which six groups were organized. The clustering algorithm was performed for each group and the resulting clusters introduced. The Xie and Beni validity index was utilized in order to decide on the best number of clusters. The clusters can be employed in order to judge on applicability of GMPEs for a specific seismic region.

It is worth mentioning that the obtained results are limited to the given assumptions in this study and further investigations are necessary in order to shed light to this area of research.

Data and Resources

The ground motion records were provided by the Building and Housing Research Centre (BHRC), Iran (http://www.bhrc.ac.ir/portal/, last accessed May 2011).

Acknowledgements

The authors are also very grateful to three anonymous reviewers for their important and valuable comments which helped to improve the paper.

References

1. Elnashai, A.S. and Di Sarno, L. (2008) *Fundamentals of Earthquake Engineering*. John Wiley and Sons, Ltd, New York.
2. Douglas, J. (2011) *Ground-Motion Prediction Equations 1964–2010.* Pacific Earthquake Engineering Research Center, College of Engineering, University of California, Berkeley.
3. Mousavi, M., Ansari, A., Zafarani, H., and Azarbakht, A. (2012) Selection of ground motion prediction models for seismic hazard analysis in the Zagros region, Iran. *J. Earthq. Eng.*, **16**(8)**,** 1184-1207.
4. Hoeppner, F., Klawonn, F., Kruse, R., and Runkler, T. (1999) *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*. John Willey and Sons, Ltd, New York.

5. Gan, G., Ma, C., and Wu, J. (2007) *Data Clustering: Theory, Algorithms, and Applications*. Society for Industrial and Applied Mathematics, American Statistical Association, Philadelphia.

6. Jain, A. K. and Dubes, R.C. (1988) *Algorithms for Clustering Data*. Prentice Hall, Inc., New Jersey.

7. Zadeh, L.A. (1965) Fuzzy sets. *Inf. Control.*, **8**(3), 338-353.

8. Bezdek, J.C. (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms*. Springer Science and Business Media, Plenum Press, New York.

9. Kaklamanos, J., Baise, L.G. and Boore, D.M. (2011) Estimating unknown input parameters when implementing the NGA ground-motion prediction equations in engineering practice. *Earthquake Spectra.* **27**(4), 1219–1235.

10. Beyer, K. and Bommer, J.J. (2006) Relationships between median values and between aleatory variabilities for different definitions of the horizontal component of motion. *Bull. Seismol. Soc. Am.* **96**(4A), 1512-1522.

11. Ghodrati Amiri, G., Khorasani, M., Mirza Hessabi, R., and Razavian Amiri, S.A. (2009) Ground-motion prediction equations of spectral ordinates and arias intensity for Iran. *J. Earthq. Eng.*, **14**(1), 1-29.

12. Boore, D.M. and Atkinson, G.M. (2008) Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 sec and 10.0 sec. *Earthquake Spectra*, **24**(1), 99-138.

13. Campbell, K.W. and Bozorgnia, Y. (2008) 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 0.01sec to 10 sec. *Earthquake Spectra*, **24**(1), 139-171.

14. Abrahamson, N. and Silva, W. (2008) Summary of the Abrahamson and Silva NGA ground-motion relations. *Earthquake Spectra*, **24**(1), 67-97.

15. Chiou, B.S.J. and Youngs, R.R. (2008) An NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra*, **24**(1), 173-215.

16. Idriss, I.M. (2008) An NGA empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes. *Earthquake Spectra*, **24**(1), 217-242.

17. Kuehn, N.M., Scherbaum, F., and Riggelsen, C. (2009) Deriving empirical ground-motion models: balancing data constraints and physical assumptions to optimize prediction capability. *Bull. Seismol. Soc. Am.*, **99**(4), 2335–2347.

18. Ambraseys, N.N., Douglas, J., Sarma, S.K. and Smit, P.M. (2005) Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: horizontal peak ground acceleration and spectral acceleration. *Bull. Earthq. Eng.*, **3**(1), 1–53.

19. Akkar, S. and Bommer, J.J. (2010) Empirical equations for the prediction of PGA, PGV and spectral accelerations in Europe, the Mediterranean region and the Middle East. *Seismol. Res. Lett*., **81**(2), 195-206.

20. Bindi, D., Luzi, L., Massa, M. and Pacor, F. (2010) Horizontal and vertical ground motion prediction equations derived from the Italian Accelerometric Archive (ITACA). *Bull. Earthq. Eng.*, **8**(5), 1209-1230.

21. Zhao, J.X., Zhang, J., Asano, A., Ohno, Y., Oouch, T., Takahashi, T., Ogawa, H., Irikura, K., Thio, H.K., Somerville, P.G., Fukushima, Y., and Fukushima, Y. (2006) Attenuation relations of strong ground motion in Japan using site classification based on predominant period. *Bull. Seismol. Soc. Am.*, **96**(3), 898-913.

22. Kalkan, E. and Gulkan, P. (2004) Site-

dependent spectra derived from ground motion records in Turkey. *Earthquake Spectra*, **20**(4), 1111-1138.

23. Ozbey, C., Sari, A., Manuel, L., Erdik, M., and Fahjan, Y. (2004) An empirical attenuation relationship for Northwestern Turkey ground motion using a random effects approach. *Soil Dynam. Earthquake Eng*. **24**(2), 115-125.
24. Akkar, S. and Cagnan, Z. (2010) A local ground-motion predictive model for Turkey, and its comparison with other regional and global ground-motion models. *Bull. Seismol. Soc. Am.*, **100**(6), 2978–2995.
25. Ghasemi, H., Zare, M., Fukushima, Y., and Koketsu, K. (2009) An empirical spectral ground-motion model for Iran. *J. Seismol*., **13**(4), 499-515.
26. Saffari, H., Kuwata, Y., Takada, S., and Mahdavian, A. (2012) Updated PGA, PGV, and spectral acceleration attenuation relations for Iran. *Earthq. Spectra*, **28**(1), 257-276.
27. Zafarani, H. and Soghrat, M. (2012) Simulation of ground motion in the Zagros region, Iran using the specific barrier model and stochastic method. *Bull. Seismol. Soc. Am.*, **102**(5), 2031-2045.
28. Sharma, M.L., Douglas, J., Bungam, H., and Kotadia, J. (2009) Ground-motion prediction equations based on data from the Himalayan and Zagros regions. *J. Earthq. Eng*., **13**(8), 1191-1210.
29. Fukushima, Y., Berge-Thierry, C., Volant, P., Griot-Pommera, D.A., and Cotton, F. (2003) Attenuation relation for West Eurasia determined with recent near fault records from California, Japan and Turkey. *J. Earthq. Eng*., **7**(4), 573-598.
30. Xie, X.L. and Beni, G. (1991) A validity measure for fuzzy clustering. *IEEE Trans. Pattern Anal. Machine Intell.*, **13**(8), 841-847.