INTERNATIONAL JOURNAL OF GEOLOGY
DOI: 10.46300/9105.2020.14.9 Volume 14, 2020

Pattern Recognition on Seismic Data for Earthquake Prediction Purpose

Adel Moatti
Industrial Engineering
Tarbiat Modares University
Tehran, Iran
Adel.moatti@modares.ac.ir

Mohammad Reza Amin-Nasseri
Industrial Engineering
Tarbiat Modares University
Tehran, Iran
Amin_nas@modares.ac.ir

Hamid Zafarani
International Institute of Earthquake
Engineering and Seismology
Tehran, Iran
H.zafarani@iiees.ac.ir

Abstract— **Earthquakes has been known as a destructive natural disaster. Due to high human casualties and economical losses, earthquake prediction appears critical. The b-value of Gutenberg Richter law has been considered as precursor to earthquake prediction. Temporal variation of b-value before earthquakes equal or greater than $M_w = 6.0$ has been examined in the south of Iran, the Qeshm island and around of this from 1995 to 2012. Clustering method by the k-means algorithm has been performed to find pattern of variation of b-value. Three clusters are obtained as optimum number of clusters by the Silhouette Index. Before all mentioned earthquakes greater than $M_w = 6.0$, cluster 1, which is known as a decrease in b-value has been seen. so decreasing b-value before main shocks as distinctive pattern has been considered. Also an approximate time of decrease has been determined.**

Keywords— **earthquake prediction, long-term seismic hazard analysis, pattern recognition, clustering, seismicity rate, b-value.**

I. INTRODUCTION

The Earthquake prediction as a promising solution to reduce the toll number of victims has been performed since 70 years ago by Ishimoto and Idia[1]. Efforts in this field are divided into long-term and short-term prediction. The short-term predictions is based on precursors such as foreshock, seismic quiescence, decrease in radon concentrations and other geochemical phenomenon[1,2]. In long-term prediction the historical earthquake data has been used along with some empirical equations like Gutenberg-Richter low to discover seismic pattern. In fact, in many earthquake prone areas in the world, the time and the location of earthquake sequences and also the magnitude of major main shocks follow distinct patterns. So extracting the seismic patterns from earthquake parameters (e.g. times, locations and magnitudes) may be useful to long-term predictions [3-5]. One of the empirical relationships which have been used frequently in long-term prediction is the Gutenberg-Richter low. This equation expresses the relationship between earthquake magnitude and total number of events as follow:

$$\log N = a - bM \tag{1}$$

The a and b parameters are constants, M is magnitude of earthquake and N is the total number of earthquakes equal to or greater than M[6]. Space and temporal variations of the b-value have been known as an indicator to predict strong main shocks, because it presents the tectonic setting and geophysical characteristics of an area. The b-value over long time and large areas is usually reported around 1, but it can vary from 0.5 to 1.5 with the decrease of exploring area[7]. The study of temporal and spatial variations of b-value has been started by Mogi and Scholz in 1968 [8,9] and many researchers have used this parameter in order to find the pattern of medium-large earthquakes. By careful inspection of 15 large earthquake in the west of Indonesia Nuannin et al.[10] have reported a significant reduction of b-value before happening all of these events. In the southern Iran, temporal b-value variations from 2005 to 2011 show that, before two earthquake with magnitude greater than M=6.0 the significant reduction of b-value has been occurred [11].

Using three different approaches to study seismicity variations within a radius of 30 km around the epicenter of the largest shock (M_W=6.4), Tsukakoshi and Shimazaki [12] found the reduction of b-value from 1.2 to 0.7. Applying the sliding time and space windows method, temporal and spatial variations of b-value in the Andaman-Nicobar islands before two major shock in 2002 (M_s = 7.0) and 2004 (M_w = 9.0) shown that two significant drop on b-value in time and low b-value in space[13].

All of the previous researches, only the variations of b-value without specify the approximate time of these changes have been conducted. It is very important to know how the sequence of b-value changes has been achieved[14]. Clustering method In this paper, has been used to investigate the sequence and time of b-value variation in Qeshm island of Iran before earthquakes equal to or greater than M_w = 6.0 from 1995 to 2012by the use of performed method in the [14].

In the second part of this paper, the seismic catalog will be introduced. In the third part, the methodology to b-value temporal estimation, k-means algorithm and method of selection optimal number of clusters is explained. Finally, by represent every cluster before major earthquakes the pattern of these events is presented.

DOI: 10.46300/9105.2020.14.9

II. Tectonic Setting

The Zagros Mountains in southwestern of Iran are the largest mountain range in this country. The Zagros fold and thrust belt is bounded to the NE by the Main Zagros Thrust and to the southeast (SE) by the Zagros Frontal Fault[15]. Base on geomorphology and topography the Zagros is divided into two difference zone. The High Zagros zone in the north-eastern and bordering by Iranian plateau and the Simply Folded Belt in south-western zone that is bordering the Persian Gulf[16]. It is spread for about 1500 kilometers from southwestern Iranian plateau to the Strait of Hormoz and is formed by collision of the Eurasian and Arabian Plates[17]. The Zagros fold and thrust belt is one of the most rapidly deforming and seismically active in the world. The most active zone of the Zagros is the Simply Folded Belt[11,16], Fig. 1. In south-eastern of the Zagros, The Qeshm Island, at 110 km in length, and between 10 km and 35 km in width is the greatest island in the Persian Gulf.

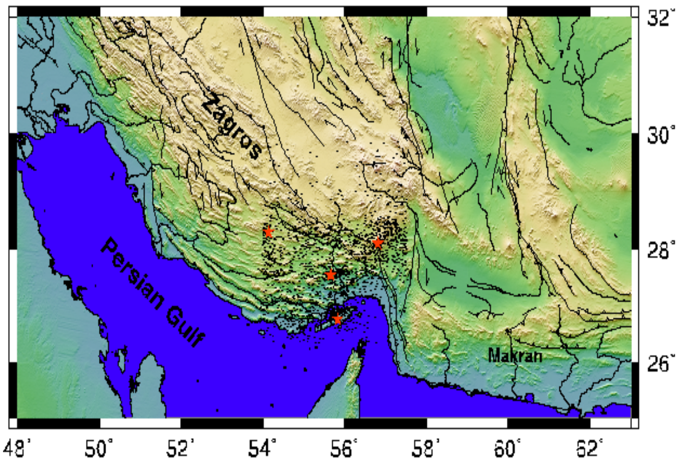

Fig. 1. Topographic map of southern, the Zagros Mountain in Iran showing major faults and epicenters of earthquakes with magnitudes equal or greater than Mw = 6.0 , between 1995 and 2012 extracted from ISC catalog.

It separated from mainland of Iran by the Strait of Khoran by trends ENE along the northern Strait of Hormoz[16]. However there are no major fault trace in Qeshm island, it has experience high seismic activities[11], Fig. 2.

III. Data Catalog

In this study, the seismic data from the International Seismological Center (ISC) catalog has been used. The examined region is limited by latitudes 26.5° to 30°N and longitudes 54° to 57.5°E and includes the Qeshm Island and spanning the period 2005.01.01 to 2012.06.19. During the period of study and for the studied region, 2046 earthquakes have been reported in ISC catalog. Magnitude and depth of the earthquakes range from 1.8 to 6.5 Mw and 1 to 256 km, consequently.

The minimum detectable magnitude in every region is known as threshold magnitude or completeness magnitude and is shown by Mc[18]. The minimum magnitude of completeness, Mc, in most seismicity studies considered as an important parameter. In seismicity study, to receive more high quality results, it is necessary to use the maximum number of event, so as much as Mc be lower, is better[19]. By Maximum likelihood estimation and 90% probability, Mc is calculated. Mc = 3.7 has been considered as threshold magnitude and all events less than M=3.7 has been deleted. At last 1065 events has remained. Also the overall b-value and a-value are estimated 0.89+/-0.04 and 6.16, respectively, by Maximum likelihood method, Fig. 3.

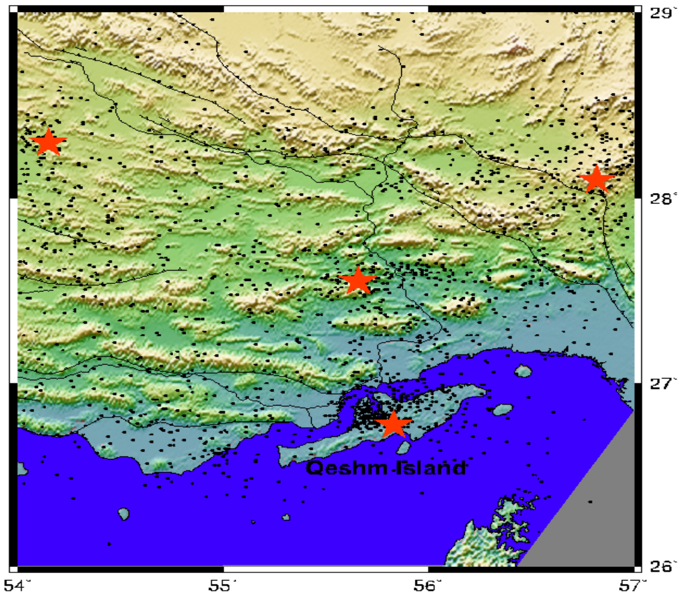

Fig. 2. Topographic map of the study area with major fault and earthquakes with magnitude equal or greater than Mw = 3.0 , of the Qeshm island and around of this region between 1995 and 2012. The earthquakes, with Mw > 6.0, marked with red stars.

In many seismicity study, declustering methods is performed to eliminate foreshocks and aftershocks to received independent data[13,17]. In this paper, the declustering method doesn't performed, due to lack of data[20,21].

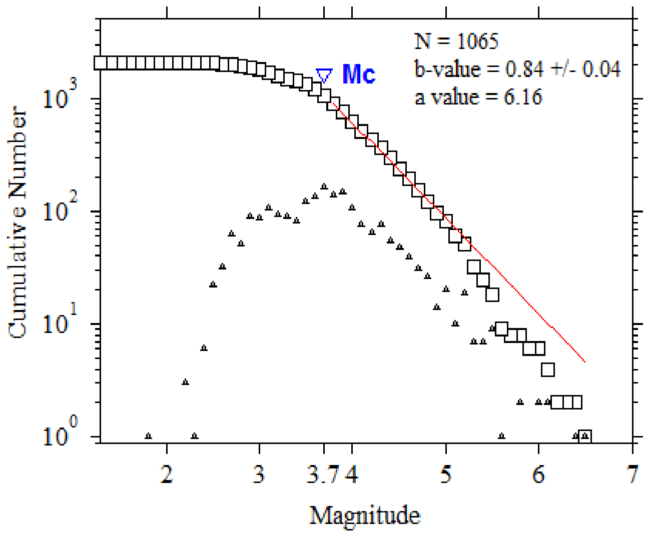

Fig. 3. Frequency magnitude distribution with respect to Mw. The strait line is the best fit by (1).

ISSN: 1998-4499

IV. METHOD

A. Calculate temporally b-value

Temporal variation in b-value of the Gutenberg-Richter relationship is calculated by sliding time windows method. The constant size window and constant number of events in each window are two options can be employed. Due to high different between numbers of events in each windows may be lead to uncertainty, constant number of events in each windows has been considered[9]. The b-value has been calculated using the Maximum likelihood method witch represent by Aki in 1965[22] in (2):

$$\log e = b / (M_{mean} - M_{min}) \quad (2)$$

The M_{Mean}, denotes the mean magnitude in each window and M_{Min} is the minimum magnitudes of sample earthquakes and is determined as $M_{Mean} = M_c - \Delta M / 2$, where ΔM is magnitude bin and here has been selected $\Delta M = 0.1$ [13], and M_C in each windows calculated separately. Different number of events in each sliding window has been tried, that is, 40, 50, 60, 70, 80, 90 and 100. At least to achieve best time resolution, 70 events in windows with an overlap 1 event have been selected. According to the above descriptions M_C and the b-value has been calculated by 1995 to 2012. The Fig. 4 show temporally variation of completeness of magnitudes, M_C in each window. The standard deviation also is calculated with bootstrapping method and is demonstrated by black color curve.

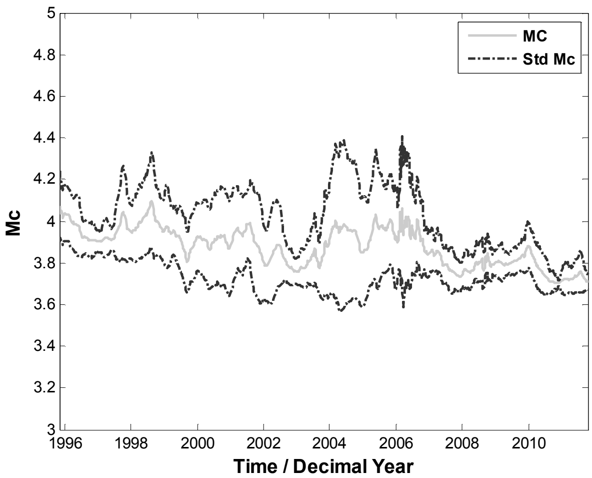

Fig. 4. Temporally variation of the Magnitude Completeness, M_c, by 1995 to 2012 with the sliding time window method between 1/1/1995 and 19/6/2012.

Also temporally variations of the b-values have been shown in Fig. 5. It can be observed, the b-value change between 0.8 and 1.4. According to Scholz 1968 and Gibowicz 1974 the high b-value shows that the low stress in the seismogenic zone and high b-value is related with high stress conditions[8,23].

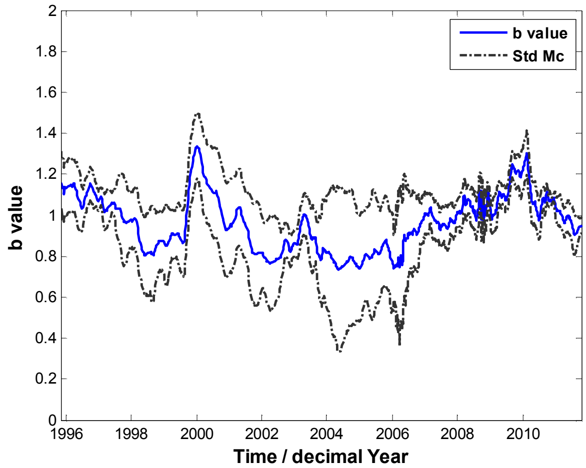

Fig. 5. Temporal variation of the b-value of earthquakes with $M_C = 3.7$ from 1/1/1995 to 19/6/2012.

B. The new dataset

Temporally b-value variation was calculated in pervious section, has been used to make new dataset according to Morales-Esteban et al. (2010)[14] for clustering purpose. Each earthquake in seismic catalog represent by three features, the b-value, b_i, the date of occurrence, T_i, and magnitude M_i. Each earthquake has been shown as:

$$e_i = (b_i, T_i, M_i) \quad (3)$$

After this, every five earthquakes has been grouped chronologically and the following calculations has been performed on each category[14]. Each group A_i containing the differences of two b-values at first and at the end of each group, Δb_i, the mean of five earthquakes in each group, $\overline{M}_i$, and the time elapsed of five earthquakes, ΔT_i, has determined. Thus,

$$A_i = (\Delta M_i, \Delta b_i, \Delta T_i) \qquad i = 1, ..., [N/5] \quad (4)$$

So that,

$$\overline{M}_i = \sum_{k=j-4}^{j} M_k \qquad j = 5i \quad (5)$$

$$\Delta b_i = \Delta b_j - \Delta b_{j-4} \qquad j = 5i \quad (6)$$

$$\Delta T_i = \Delta T_j - \Delta T_{j-4} \qquad j = 5i\,; \tag{7}$$

Where N, is the number of earthquakes in seismic catalog; As a final point, the new dataset, ND, is formed by all chronologic A_i, which is determined by defined (5), (6) and (7) as:

$$ND = \{A_1, A_2, A_3, \ldots, A_{[N/5]}\} \tag{8}$$

At last the ND has been performed to clustering for pattern recognition of earthquakes with M_w equal or greater than 6.0.

C. Clustering

1) The k-means algorithm

Cluster analysis is the task of grouping some objects in such a way that objects in the same group are more similar to each other than to those in other clusters. One of the most popular clustering method is the k-means algorithm that was introduced at first by Macqueen (1968)[24]. At first the algorithm selects k points as the initial centroid. After this, the algorithm collects rest of objects into k groups with the aim of increasing intra-clusters similarity at the same time. Each object is assigned to the cluster with the closest centroid. Actually, this similarity is measured according to centroid of each clusters and the aim is to reduce intra-cluster distance. Indeed to reduce intra-cluster distances the squared error function is used as follow:

$$SSE = \sum_{i=1}^{k} \sum_{X_j \in C_j} \left| X_j - m_j \right|^2; \tag{9}$$

Where k is the number of clusters, X_j, is the j-th object, m_j, is the centroid of j-th cluster and C_j, is the j-th cluster. In this paper, the k-means has been used multiple times to escape from entrapment in local minimum.

2) The silhouette index

In the most unsupervised clustering i.e. k-means, select the optimum number of clusters is a crucial challenge. Also it is very important to evaluate how much the result is accurate. There are many various quality measures to evaluate clustering result. The Silhouette index as one of the common index has been used in this paper. This validity index computes silhouette width for each object. Also this index calculate average silhouette width for each clusters and overall silhouette width for all dataset[25]. The following formula is used to measure silhouette width for each data point:

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \tag{10}$$

Where a_i is the average dissimilarity of i-th object to all other object in the same cluster; b_i is the minimum of average dissimilarity of i-th object to all objects in the other clusters. The S_i, is the value between -1 and 1. If S_i was closed to 1, it means that the object is assigned to proper cluster. If S_i was closed to zero, it means that, the object could be assigned to another closest cluster and if the S_i was -1, it means that the object assigned to improper cluster. The average of all S_i is the overall average silhouette width for all objects in dataset. Finally, the largest overall average, indicates clustering with high accuracy[26].

In this paper, different number of clusters has been performed to achieve optimum cluster. Figure 6, shows that the maximum overall silhouette width averages is 0.5938 and is related to 3 clusters. So the optimum cluster has been chosen three.

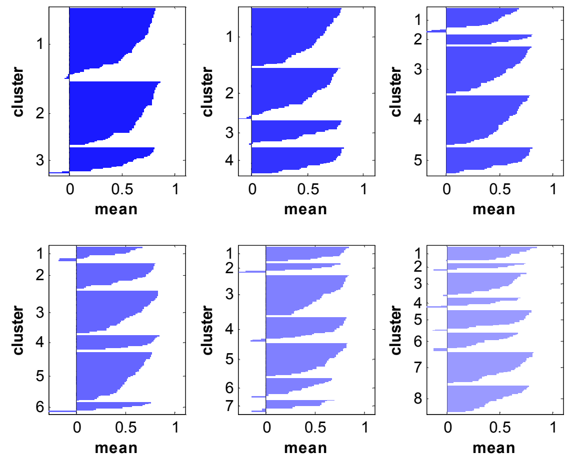

Fig. 6. The mean silhouette index values versus number of clusters for 3 to 8 clusters

According to the silhouette index result, the ND has been clustered by k-means algorithm by repeating 500 times and Table 1 shows obtained centriods of clusters.

TABLE I. CENTROIDS OF OBTAINED CLUSTERS WITCH IS OBTAINED BY K-MEANS

Cluster	$\overline{M}$	ΔT	Δb
1	3.987	0.062	- 0.023
2	4.235	0.067	0.003
3	4.531	0.062	+ 0.039

It can be seen cluster 1 represenet decrease in b-value and earthquakes with low magnitudes. The time interval in this cluster is 22 day approximatly. Cluster 2, show that there are no any changes in b-value. Cluster 3 demonstrate increase in b-value with large magnitude earthquakes in time interval similar cluster 1.

V. Pattern Recognition

According to clusters has been obtained from k-means algorithm, each five grouped seismic data in new dataset has been plotted chronologically with respect to cluster labels versus time. As can be seen from Fig. 2, four major earthquakes have been occurred in study area with magnitude $M_w \geq 6.0$ between 1995 and 2012. Table 2 has shown four main shocks.

After clustering, it can be seen that all earthquakes with magnitude $M_w \geq 6.0$ are presented by five-earthquake groups belongs to cluster 3. According to cluster 3 represent increases in b-value. Accordingly, Fig. 7 shows the patterns are extracted from b-value variation. The blue circles have identified five-earthquake groups and stars have specified main shocks with $M_w \geq 6.0$.

TABLE II. FOUR MAIN SHOCKS WITH MAGNITUDE EQUAL OR GREATER THAN M_W FROM 1/1/1995 TO 19/6/2012

Earthquake	Magnitude	Date	Longitude	Latitude
1	6.0	2003/7/10	28.30	54.15
2	6.1	2006/2/28	28.10	56.82
3	6.0	2006/3/25	27.55	55.66
4	6.1	2008/9/10	26.77	55.83

Figure 7, represent how five-earthquake group's clusters have changes between 2000 and 2005 interval. It can be seen that, before earthquake with $M_w = 6.0$, sequence of cluster 1, has been occurred. According to table 1, cluster 1 represent decrease in b-value by 22 days interval approximately. Indeed, the main shock that shows by star is preceded by cluster 3-2-1. This means that, significant decrease in b-value before main shock has been occurred. Also, time interval of this reduction is 3 month approximately.

As the Fig. 7, trend of clusters changes has been reported from 1/1/2005 to 29/12/2008 in Fig. 8. In this period, two large earthquakes with magnitudes, $M_w = 6.1$ and $M_w = 6.0$, have occurred within less than 1 month. For both events, clusters changes represent decrease in b-value. As can be seen in Fig. 8, first event precede by cluster 1. Also second event is preceded by five-earthquake groups classified to cluster 1. In fact clusters changes from 3 to 1 for both main shocks. This means that, significant reduction in b-value before main shocks has been occurred.

Finally, the interval between 2008 and 2012 has been investigated. In this period, an earthquake with magnitude $M_w = 6.1$ is occurred. According to Fig. 9, it can be seen, similar patterns have been occurred like previous results. A sequence of cluster 1 has been occurred before main shock at 10/9/2008. This trend is happened in 3 month roughly and reports significant decrease in b-value.

VI. CONCLUSIONS

In this paper, clustering method according to Morales-Esteban et.al.(2010)[1414], has been performed to pattern recognition of earthquake with magnitude equal or greater than $M_w = 6.0$, in south of Iran, the Qeshm island and around of this area. With respect to clusters that are obtained from the k-means, reduction of b-value before all events with $M_w \geq 6.0$ has been observed. This result is similar to past studies that reports b-value decrease as large earthquakes precursor.

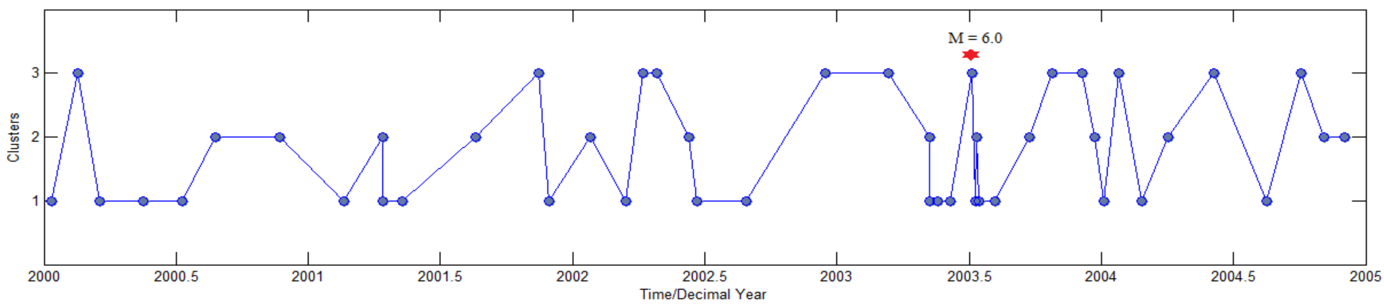

Fig. 7. changes of clusters between 2000 and 2005

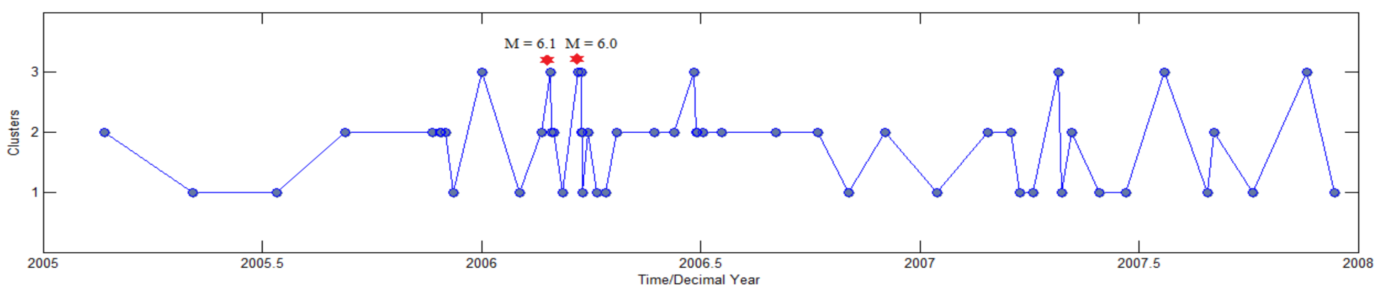

Fig. 8. Changes of clusters between 2005 and 2008

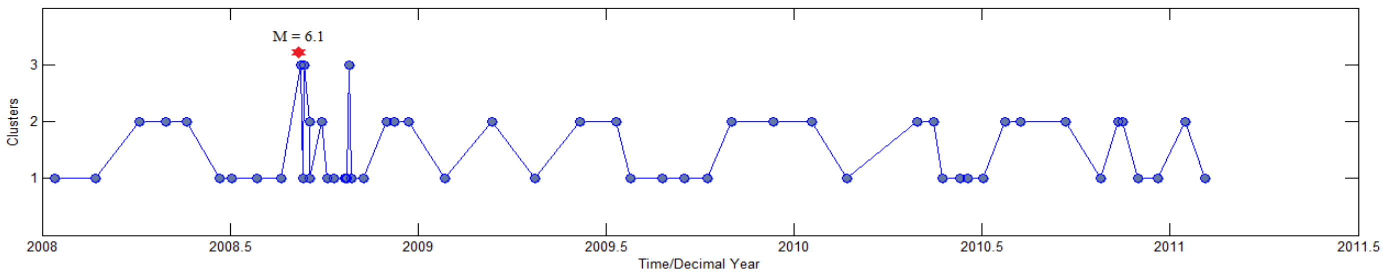

Fig. 9. Changes of clusters between 2008 and 2012

REFERENCES

[1] A. Panakkat and H. Adeli, "Recent efforts in earthquake prediction (1990–2007)," *Natural Hazards Review,* vol. 9, 2008, pp. 70-80.

[2] M. Wyss and D. C. Booth, "The IASPEI procedure for the evaluation of earthquake precursors," *Geophysical Journal International,* vol. 131, 1997, pp. 423-424.

[3] M. Wyss, F. Pacchiani, A. Deschamps, and G. Patau, "Mean magnitude variations of earthquakes as a function of depth: Different crustal stress distribution depending on tectonic setting," *Geophysical research letters,* vol. 35, 2008, p. L01307.

[4] B. CHEN, P. BAI, and T. LI, "THE b-VALUE AND EARTHQUAKE OCCURRENCE PERIOD," *Chinese Journal of Geophysics,* vol. 46, 2003, pp. 736-749.

[5] V. S. Rani, K. Srivastava, D. Srinagesh, and V. Dimri, "Spatial and temporal variations of b-value and fractal analysis for the Makran Region," *Marine Geodesy,* vol. 34, 2011, pp. 77-82.

[6] B. Gutenberg and C. F. Richter, *Seismicity of the earth and associated phenomena*: Hafner New York, 1965.

[7] C.-H. Chan, Y.-M. Wu, T.-L. Tseng, T.-L. Lin, and C.-C. Chen, "Spatial and temporal evolution of b-values before large earthquakes in Taiwan," *Tectonophysics,* vol. 532, 2012, pp. 215-222.

[8] C. Scholz, "The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes," *Bulletin of the Seismological Society of America,* vol. 58, 1968, pp. 399-415.

[9] O. Kulhanek, "Seminar on b-value," *Dept. of Geophysics, Charles University, Prague,* 2005.

[10] P. Nuannin, O. Kulhánek, and L. Persson, "Variations of b-values preceding large earthquakes in the Andaman–Sumatra subduction zone," *Journal of Asian Earth Sciences,* vol. 61, 2012, pp. 237-242.

[11] M. R. Sorbi, F. Nilfouroushan, and A. Zamani, "Seismicity patterns associated with the September 10th, 2008 Qeshm earthquake, South Iran," *International Journal of Earth Sciences,* 2012, pp. 1-9.

[12] Y. Tsukakoshi and K. Shimazaki, "Decreased b-value prior to the M 6.2 Northern Miyagi, Japan, earthquake of 26 July 2003," *Earth Planets and Space (EPS),* vol. 60, 2008, p. 915.

[13] P. Nuannin, O. Kulhanek, and L. Persson, "Spatial and temporal b value anomalies preceding the devastating off coast of NW Sumatra earthquake of December 26, 2004," *Geophysical research letters,* vol. 32, 2005, p. L11307.

[14] A. Morales-Esteban, F. Martínez-Álvarez, A. Troncoso, J. Justo, and C. Rubio-Escudero, "Pattern recognition to forecast seismic time series," *Expert Systems with Applications,* vol. 37, 2010, pp. 8333-8342.

[15] K. Hessami, F. Nilforoushan, and C. J. Talbot, "Active deformation within the Zagros Mountains deduced from GPS measurements," *Journal of the Geological Society,* vol. 163, 2006, pp. 143-148.

[16] E. Nissen, F. Yamini-Fard, M. Tatar, A. Gholamzadeh, E. Bergman, J. Elliott, J. Jackson, and B. Parsons, "The vertical separation of mainshock rupture and microseismicity at Qeshm island in the Zagros fold-and-thrust belt, Iran," *Earth and Planetary Science Letters,* vol. 296, 2010, pp. 181-194.

[17] A. Zamani and M. Agh-Atabai, "Temporal characteristics of seismicity in the Alborz and Zagros regions of Iran, using a multifractal approach," *Journal of Geodynamics,* vol. 47, 2009, pp. 271-279.

[18] P. A. Rydelek and I. S. Sacks, "Testing the completeness of earthquake catalogues and the hypothesis of self-similarity," *Nature,* vol. 337, 1989, pp. 251-253.

[19] S. Wiemer and M. Wyss, "Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the western United States, and Japan," *Bulletin of the Seismological Society of America,* vol. 90, 2000, pp. 859-869.

[20] T. Parsons, "Forecast experiment: Do temporal and spatial b value variations along the Calaveras fault portend M≥ 4.0 earthquakes?," *Journal of geophysical research,* vol. 112, 2007, p. B03308.

[21] Y.-M. Wu, C.-C. Chen, L. Zhao, and C.-H. Chang, "Seismicity characteristics before the 2003 Chengkung, Taiwan, earthquake," *Tectonophysics,* vol. 457, 2008, pp. 177-182.

[22] K. Aki, "17. Maximum Likelihood Estimate of b in the Formula logN= a-bM and its Confidence Limits," 1965.

[23] S. J. Gibowicz, "Frequency-magnitude, depth, and time relations for earthquakes in an island arc: North Island, New Zealand," *Tectonophysics,* vol. 23, 1974, pp. 283-297.

[24] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1967, p. 14.

[25] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics,* vol. 20, 1987, pp. 53-65.

[26] Z. Ansari, M. Azeem, W. Ahmed, and A. V. Babu, "Quantitative Evaluation of Performance and Validity Indices for Clustering the Web Navigational Sessions," *World of Computer Science and Information Technology Journal,* vol. 1, 2011, pp. 217-226.