Application of Gustafson-Kessel clustering algorithm for detecting fault through seismic attributes

Sima Shakiba[1], Nasser Keshavarz Farajkhah[2]

1- Simulation and Processing Lab. School of Mining Engineering, University of Tehran, Iran.
2- Geophysics Research Group, Geoscience Division, Research Institute of Petroleum Industry (RIPI).

* Corresponding Author: Sima_shakiba@ut.ac.ir

Received: 26 January 2015 / Accepted: 05 March 2016 / Published online: 10 March 2016

Abstract

In this paper an application of Gustafson-Kessel clustering algorithm is presented to create a fault detection map (FDM). Five post-stack seismic attributes are extracted from a desired seismic time slice related to 3D seismic data of a gas field located in southwest of Iran. To find the optimal cluster numbers, two frequently used clustering validity measures, i.e. SC and XB, are used and then the studied area were divided into regions which have high possibility for exploring faults. The proposed method helps expert geologists to enhance their interpretations in identifying subtle faults and provides an effective FDM generating approach.

Keywords: Gustafson-Kessel clustering algorithm, Fault Detection Map (FDM), Gas field, Seismic attributes, Clustering validity.

1- Introduction

Detecting fault and fractures is a significant step in geological studies (e.g. structural and stratigraphic interpretation) in both exploration and development phases. In general, fault structures are categorized into seismically resolvable and sub-seismic scale (subtle) faults that may be inferred more successfully with the aid of seismic attributes.

Although the seismically resolvable faults can be detected through conventional analytical criteria such as abrupt reflector cut off or kinks, but the subtle faults are not visibly imaged by the usual seismic sections and time slices displays (Odoh *et al.*, 2014).

Among the various geophysical methods available for identifying subtle faults and/or other discontinuities, employing seismic data are certainly the most powerful approach (Neves *et al.*, 2004). The most common approach for identifying faults and fractures is utilizing seismic attributes (Klein *et al.*, 2008). Utilizing seismic attributes can help interpreters in getting a better insight into the fault and fracture systems. Seismic attributes are useful tools for interpreting seismic sections and generating quantitative/qualitative maps (Shakiba *et al.*, 2015; Mahdavi Basir *et al.*, 2013; Chopra and Marfurt, 2007; Marfurt, 2006; Gresztenkorn and Marfurt, 1999; Bahorich and Farmer, 1995; Rijks and Jauffred, 1991).

Faults can be characterized more effectively by the use of seismic attributes, most particularly the ones that emphasize discontinuities in the seismic traces.

Regarding the importance of discovering fault regions, generating a Fault Detection Map (FDM) seems to be a hopeful idea for reducing the risk of placing wells within productive zones and increasing production success. Different techniques may be used for generating FDM, which can be divided into either, data-driven or knowledge-driven methods (Dai *et al*., 2008; Guo *et al*., 2005; Vicente *et al*., 2007).

The process of selecting the suitable area for exploring fault zones is similar to unsupervised classification problem. Consequently, different clustering algorithms can be employed to generate FDM.

The main objective of this study is to build a suitable FDM by employing Gustafson-Kessel clustering methods. In this regard, five post-stack seismic attributes are obtained from a time slice extracted from 3D seismic data related to a gas field in southwest of Iran.

2- Fuzzy clustering

Data clustering is the process of partitioning data elements from original dataset into different clusters based on the similarity between the elements by measuring a criterion. In traditional form of data clustering algorithms, the data which are close to each other and far from data in other clusters belong to a same cluster. But in fuzzy clustering (known as soft clustering), data can belong to more than one cluster using the concept of membership functions.

During the past years, fuzzy logic is extremely used in a number of different areas of geosciences (Grekousis and Hatzichristos, 2013; Grekousis, 2013; Grekousis and Photis, 2011; Mollajan, 2014). The fuzzy clustering algorithm is a novel extended data clustering method can enhance the process of data partitioning.

2.1- Gustafson-Kessel clustering algorithm

Gustafson and Kessel broaden the original fuzzy c-means clustering algorithm by employing an adaptive distance measurement in order to recognize clusters of different geometrical forms in dataset. While the fuzzy c-means presupposes that clusters are spherical shape, the Gustafson-Kessel is not subject to this limitation and can identify ellipsoidal clusters. The objective function J_m of GK algorithm can be defined as follows (Gustafson and Kessel, 1979):

$$J_m = \sum_{j=1}^{n} \sum_{i=1}^{k} \mu_{ij}^{m} \, d_{ij}^{2} \qquad (1$$

Where the distance between data and cluster centers, defined as:

$$d_{ir} = (x_i - w_r)^T A_r^{-1} (x_i - w_r) \qquad (2$$

In this algorithm, any cluster associate with cluster center and its covariant.

Suppose Fir is the influence of point *i* on cluster *r*, the cluster center is computed as a weighted means of the data (Serir *et al*., 2012):

$$W_r = \frac{\sum_{i=1}^{n} F_{ir}^{m} x_i}{\sum_{i=1}^{n} F_{ir}^{m}} \qquad (3$$

where m is a user-defined parameter named fuzzifier. Also, the covariance matrix is defined as a fuzzy comparable to classic covariance matrix is given by:

$$A_r = \sqrt[p]{\det(S_r)} \, S_r^{-1} \qquad (4$$

and S_r is defined as:

$$S_r = \sum_{i=1}^{n} F_{ir}^{m} (x_i - w_r)(x_i - w_r)^T \qquad (5$$

As Eq.4 indicates, a size constraint is imposed on the covariance matrix, so the determinant of covariance matrix must be 1. Therefore the algorithm can identify ellipsoidal clusters having more or less the same size.

When the clusters are ranged to a significant degree along the length of the greatest vector, the covariance matrix is able to determine the original data distribution. In this situation, a scaled identity matrix (g) can be added to the covariance matrix which changes between 0 and 1 (Grekousis and Hatzichristos, 2013).

2.2- Validity Measures

In this paper, two validity measures are use to find the optimal cluster numbers which are described below:

a) Partition Index (PI): is the ratio of the sum of compactness and separation of the clusters (Bensaid *et al*., 1996):

$$SC(c) = \sum_{i=1}^{c} \frac{\sum_{j=1}^{N} (\mu_{ij})^m \left\| x_j - v_i \right\|^2}{N_i \sum_{k=1}^{c} \| v_k - v_i \|^2} \qquad (6$$

b) The Xie-Beni (XB) Separation Index: this index utilizes a minimum-distance separation for partition validity (Xie and Beni, 1991):

$$XB = \frac{\sum_{i=1}^{c} \sum_{j=1}^{N} (\mu_{ij})^m \left\| x_j - v_i \right\|^2}{N min_{i,j} \left\| x_j - v_i \right\|^2} \qquad (7$$

3- Application of G-K algorithm to fault detection mapping

3.1- The study area and dataset

The investigated area of interest is a fractured gas field situated in southwest of Iran. The main focus of this study is on Gotnia Formation which mainly consists of Anhydrite depositions (Setudehnia, 1978). This formation represents the Upper Jurassic sequence of the studied area and can be correlated with evaporates of Upper Jurassic in Fars, Alborz and Central part of Iran (Setudehnia, 1978; Narin and Alsharhan, 1997; Alavi, 2004). 3D seismic survey was previously carried out in the studied area that covers the area of 500 Km^2. Data were sampled at an interval of 4ms and stored in SEG-Y format. The figure 1 shows a time slice at t= 2900 ms. As can be seen in this figure, two groups of faults can be identified at angles of about 10^o and 90^o with respect to the horizon.

3.2- Input feature selection

To commence this study, five post-stacked seismic attributes are extracted from a seismic time slice at t=2900 ms and used to generate FDM including instantaneous amplitude, similarity, energy, frequency and fault enhancement filter (FEF) (Fig. 2).

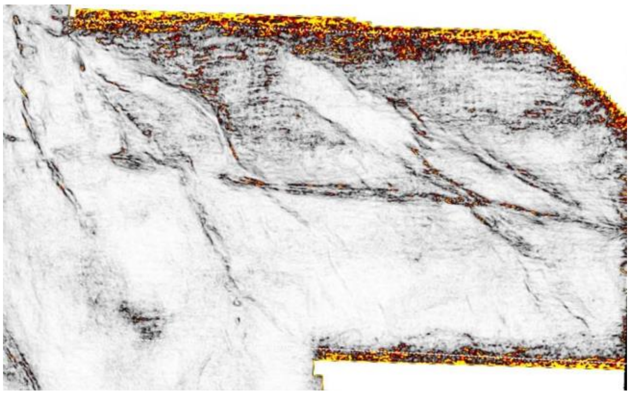

Figure 1) Identified fault direction in desired time slice at t = 2900 ms.

The similarity determines the similarities between seismic traces and highlights the connectedness and continuity of seismic horizons. High similarity specifies fault and fracture zones (Mahdavi Basir *et al*., 2013). The FEF is the combination of diffusion and median filter. This filter sharpens the faults and suppresses non-fault discontinuities using a cut-off value (Marfurt *et al*., 1998). The filter uses median filter for fault zones to sharpen the edges where the similarity is high and employs diffusion filtering where the quality of the seismic data is poor.

The energy is a seismic attribute which returns the energy of trace segment and is a measure of reflection in a time interval. The filter can effectively enhance the laterally discontinues events. The value of this attribute varies between zero and maximum amplitude, therefore the low energy values can specify the faults. Another sensitive attribute that derived from seismic data is spectral decomposition (frequency).

This attribute can effectively be used for evaluating sequence stratigraphy, illuminating shear faults and determining fracture properties (Neves *et al*., 2004). The latest attribute used in this study is instantaneous amplitude. This attribute can highlight the variation of acoustic impedance and improve detecting micro-faults and channels (Radovich and Oliveros, 1998).

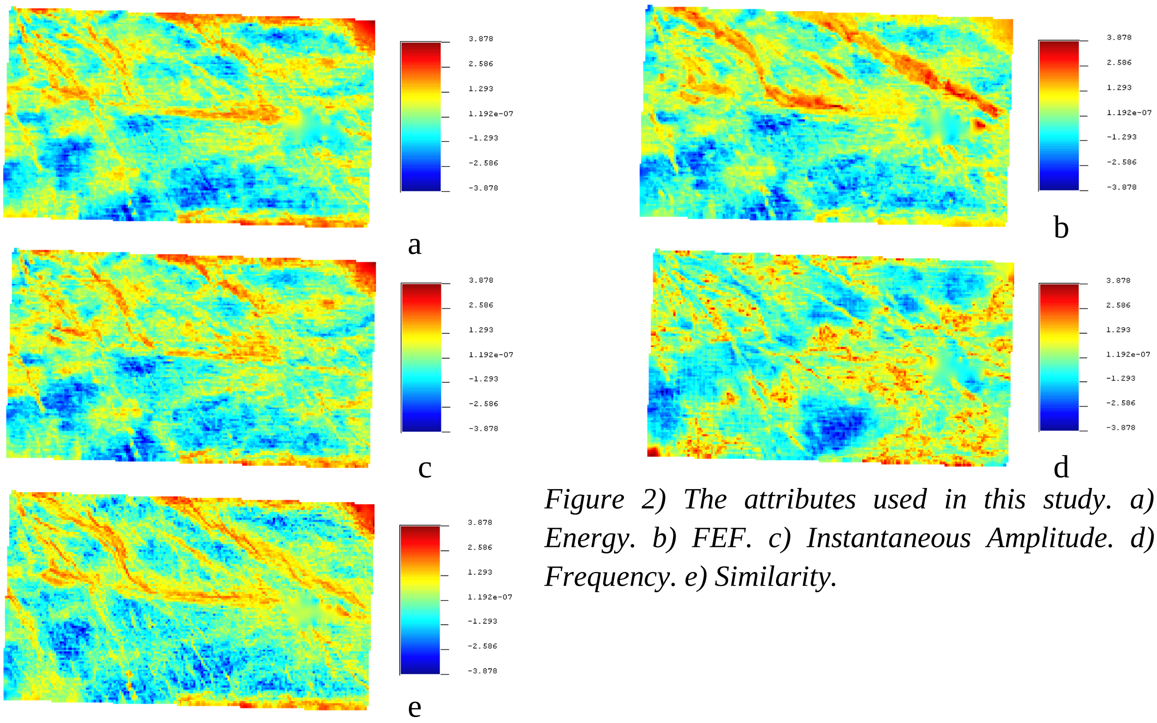

Figure 2) The attributes used in this study. a) Energy. b) FEF. c) Instantaneous Amplitude. d) Frequency. e) Similarity.

3.3- Finding optimal cluster number

As it is stated before, we have used two validity measures to find optimal cluster number in our experiments. The main purpose of employing cluster validity is to know which partition best describes the unknown cluster structure in a given dataset and then extract the optimal number of clusters.

Prior to using cluster validity measures, different fuzziness values (m) were examined (Table 1). Next, several runs of the algorithm were carried out to find the optimal number of cluster between 2 and 15 for ideal m. The results of the two validity indices applied to the selected data are shown in figure 3. As seen, the minimum of partition index (PI) and absolute maximum of separation index (SI) is reached by cluster number C=3 where m=2, which is somehow consistent with derived seismic time slice shown in figure 1.

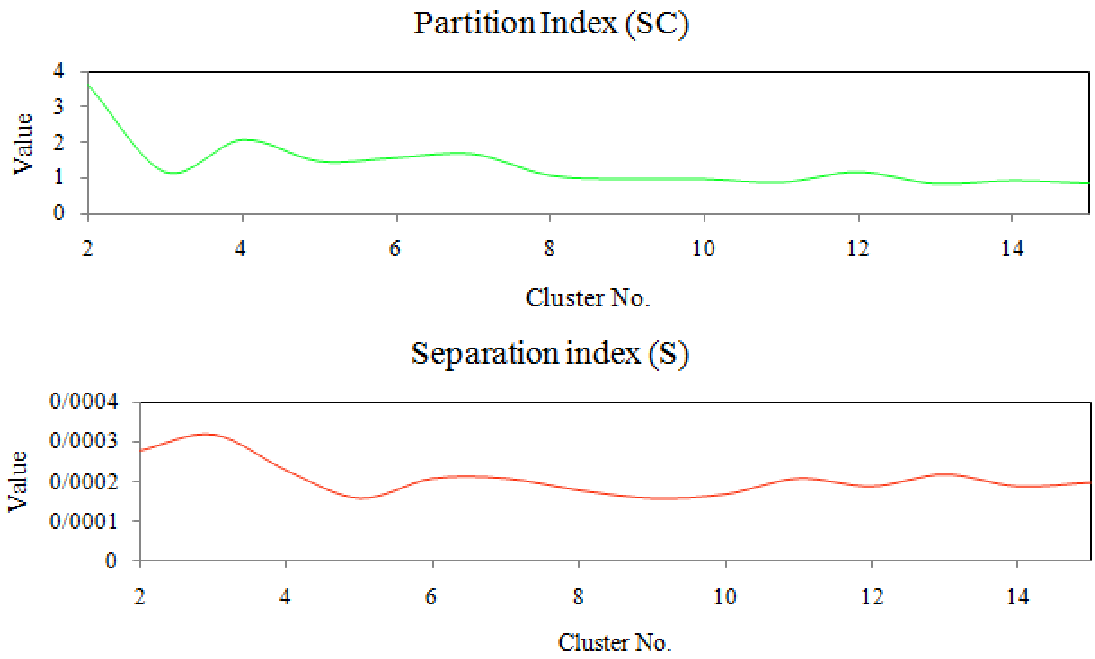

Figure 3) Finding optimal cluster number through two cluster validity index.

Moreover, different values of g are tested to choose suitable one. The g value was chosen to be equal to 0.5 by trial and error approach for a satisfactory fuzziness and ellipsoid clusters.

Table 1) Ideal fuzziness value.

Fuzziness value (m)	SC	XB
1.2	8.46E06	5.21
1.4	7.52E06	4.3
1.8	6.21E04	3.8
1.6	4.08E04	2.6
2	5.92E04	2.1
2.2	7.21E05	2.5

The energy and instantaneous amplitude attributes have low values in clusters 1 and 2, whereas the remaining attributes have their high values in these regions. Therefore, these clusters indicate fault zones. Similarly, cluster 3 corresponds to the non-fault zones as it shows a good match with regions having high values in energy and instantaneous amplitude attributes and low values in the three remaining attributes.

3.4- Validation of the results

The result of employing the proposed algorithm is a map indicating the most likely regions for exploring fault zones (Fig. 4). From figure 4, the most likely region for exploring faults corresponds to class 1, i.e., exact fault zones. Moreover, if the study area belongs to class 2 or 3, the possibility of the presence of fault is reduced.

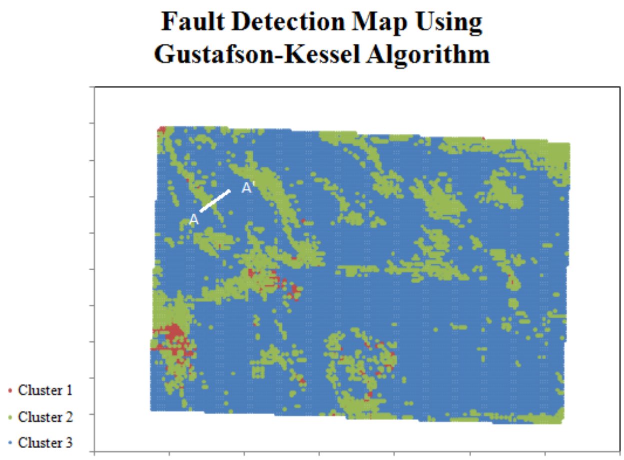

Figure 4) The final FDM generated by G-K algorithm.

In order to assess the accuracy of the obtained FDM, variography in identified fault direction was carried out. The variogram is a function reflecting the degree of spatial dependence of a regional variable and provides a description of how the data are related (correlated) with distance. It is suitable tool to characterize the spatial continuity or roughness of a data set.

In this study, we used variogram to identify and bold all discontinuity on seismic sections. It is expected to see a high correlation between traces in discontinuities which indicates the fault zones. As more discontinuity, more correlation is occurred and consequently sill of variogram shows less value.

To do so, section A-A' was considered on final FDM (Fig. 4), and variogram of all attributes along with output of G-K algorithm were drawn in this direction.

The results are shown in figure 5. As can be seen, the concluded variograms show the most correlation in direction of existing discontinuities.

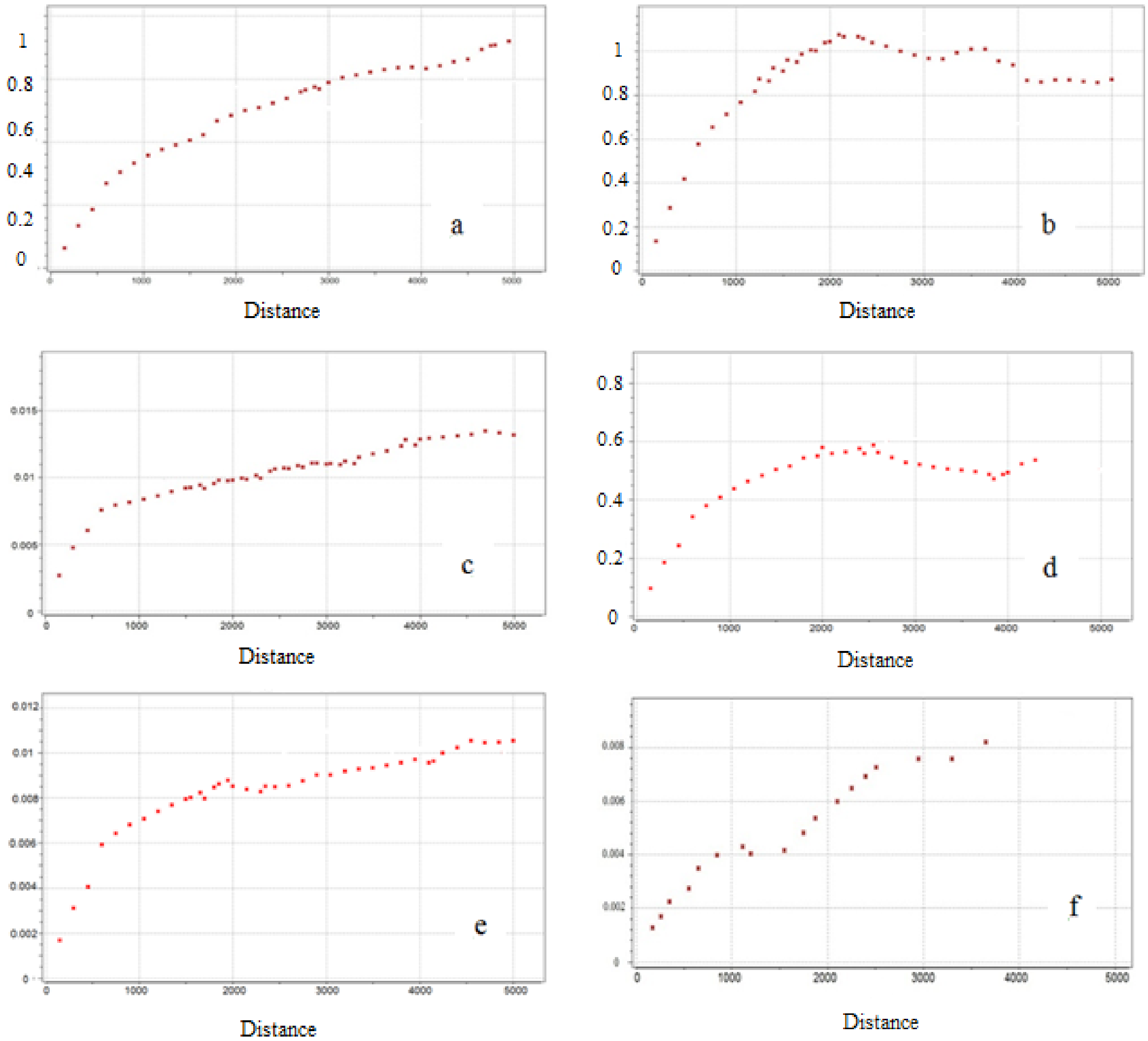

Figure 5) Plotting variograms to quantify fault: a) Energy. b) FEF. c) Similarity. d) Instantaneous Amplitude. e) Spectral decomposition. f) FDM from proposed algorithm.

4- Discussion

In order to find optimal fuzzy value (g), different values were tested. Choosing small values for g will result in limited changes in covariance table. Moreover, in a constant g, as fuzziness value (m) increases, fuzziness also increases.

On the other hand, for a constant c value, the more g reduces, the more the clusters become divided. For example, when the g value approaches 0, data in the corresponding clusters shows large membership values which is similar to a non-fuzzy clustering method. Based on the obtained results, the value g = 0.5 were chosen to have a satisfactory fuzziness and ellipsoid clusters. Finally, the Gustafson-Kessel algorithm with values m= 2, g=0.5 and C=3 was selected for creating a suitable FDM.

Table 2 shows the centers of each cluster. As it is expected, data which are close to cluster centers have high membership values reflecting the fact that they belong to an exact cluster. While data locate far from the centers have low membership values indicating they belong to more than one cluster.

Table 2) Cluster centers.

Cluster No.	Cluster 1	Cluster 2	Cluster 3
Centers	[0.15 0.02]	[0.06-0.11]	[0.05 - 0.09]
Center error	0.11	0.82	1.21

5- Conclusion

In this paper, an application of Gustafson-Kessel clustering algorithm for creating Fault Detection Map (FDM) was presented. The clustering process was applied on five post-stack seismic attributes to explore high potential regions for detecting fault zones.

Both Partition Index (PI) and Separation Index (SI) as clustering validity measures identified that using selected attributes three different clusters can be considered in studied are. The clusters generated by this approach were mapped to reveal most suitable regions for finding subtle faults. The identified clusters were then used to discriminate fault and non-fault zones in the studied area.

The accuracy of the obtained results was evaluated by plotting the variograms in identified fault direction. Through this approach, all discontinuities can quantitatively be assessed and the efficiency of the proposed algorithm in detecting fault zones may be examined.

According to the results of this study, it can be concluded that clustering of seismic attributes can effectively be considered as powerful tool for creating FDM and help the interpolators to detect faults more accurately in the studied carbonate reservoir.

References:

Alavi, M. 2004. Regional stratigraphy of the Zagros fold-thrust belt of Iran and its proforeland evaluation. American Journal of Science: 304, 1–20.

Bahorich, M. S., Farmer, S. L. 1995. 3D seismic discontinuity for faults and stratigraphic features: the coherence cube. The Leading Edge: 14, 1053–1058.

Bensaid, A. M., Hall, L. O., Bezdek, J. C., Clarke, L. P., Silbiger, M. L., Arrington, J. A. 1996. Validity-guided (Re) Clustering with applications to image segmentation. IEEE Transactions on Fuzzy Systems: 4, 112–123.

Chopra, S., Marfurt, K. 2007. Curvature attribute applications to 3D surface seismic data. The Leading Edge: 26, 404–414, doi:10.1190/1.2723201

Dai, X., Liu, G., Long, Zh. 2008. Discrete-time Robust Fault Detection Observer design: A Genetic algorithm approach. IEEE Proceedings of the 7th World Congress on Intelligent Control and Automation.

Guo, Q. J, Yu, H.b., Xu, A. 2005. Modified Morlet wavelet neural networks for fault detection. International Conference on Control and Automation (ICCA 2005), 2, 1209–1214.

Grekousis G., Hatzichristos T. 2013. Fuzzy clustering analysis in geomarketing research. Environment and Planning B. Planning and Design: 40: 95–116. doi: 10.1068/b37137.

Grekousis G, 2013. Giving fuzziness to spatial clusters: A simple index for choosing the optimal number of clusters. International Journal on Artificial Intelligence Tools: 22. doi: 10.1142/S0218213013500097.

Grekousis G., Photis, Y. 2011. A fuzzy index for detecting spatiotemporal outliers. Geoinformatica: doi: 10.1007/s10707-011-0145-4.

Gresztenkorn, A., Marfurt, K. J. 1999. Eigenstructure-based coherence computations as an aid to 3D structural and stratigraphic mapping. Geophysics: 64, 1468–79.

Gustafson, E., Kessel, W. 1979. Fuzzy clustering with a fuzzy covariance matrix. In Proc. of IEEE CDC.

Hwang, C. -L., Yoon, K. 1981. Multiple attribute decision making: methods and applications. In Lecture notes in economics and mathematical systems. Now York: Springer.

Klein, P., Richard, L., James, H. 2008. 3D curvature attributes: a new approach for seismic interpretation. First Break: 26, 105–111.

Mahdavi Basir, H., Javaherian, A., Tavakoli Yaraki, M. 2013. Multi-attribute ant-tracking and neural network for fault detection: a case study of an Iranian oilfield. Journal of Geophysics and Engineering 10, doi: 10.1088/1742-2132/10/1/015009.

Marfurt, K. J., Kirlin, R. L., Farmer, S. H., Bahorich, M. S. 1998. 3D seismic attributes using a running window semblance-based algorithm. Geophysics: 63, 1150–65.

Marfurt, K. J. 2006. Robust estimates of 3D reflector dip and azimuth. Geophysics: 71, 29–40.

Mollajan, A. 2014, Application of local linear neuro-fuzzy model in estimating reservoir water saturation from well logs. Arabian Journal of Geoscience: 8, 4863-4872.

Narin, A. E. M., Alshahran, A. S. 1997. Sedimentary Basins And Petroleum Geology of the Middle East. Elsevier, 942 pp.

Neves, F. A., Zahrani, M. S., Bremkamp, S. 2004. Detection of potential fractures and small faults using seismic attributes. The Leading Edge: 23, 903–906.

Radovich, B. J., Oliveros, R. B. 1998. 3D sequence interpretation of seismic instantaneous attributes from the Gorgon field. The Leading Edge: 17, 1286–1293.

Rijks, E. J. H., Jauffred, J. C. E. M. 1991. Attribute extraction, an important application in any detailed 3D interpretation study. The Leading Edge: 10, 11–19.

Setudehnia, A. 1978. The Mesozoic sequence in south-west Iran and adjacent areas. Journal of Petroleum Geology: 1, 3–42.

Taheri, K., Gutiérrez, F., Mohseni, H., Raeis, E., Taheri, M. 2015. Sinkhole susceptibility mapping using the analytical hierarchy process (AHP) and magnitude–frequency relationships: A case study in Hamadan province, Iran. Geomorphology: 234, 64–79.

Xie, X. L., Beni, G. A. 1991.Validity measure for fuzzy clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence: 3, 841–846.

Vicente, P., Rodriguez, J., Arkkio, A. 2007. Detection of stator winding fault in induction motor using fuzzy logic. Applied Soft Computing: 8, 1112–1120.