Makara Journal of Science

Volume 25
Issue 1 *March*

Article 4

3-30-2021

Cluster Analysis of Lithology Grouping Trends using Principal Component Spectral Analysis and Complex Seismic Attributes

Isfan Isfan
PT. Elnusa Tbk, Jakarta Selatan 12560, Indonesia, isfan@elnusa.co.id

Agustinus Harsono
Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok 16424, Indonesia

Abdul Haris
Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok 16424, Indonesia

Recommended Citation

Isfan, Isfan; Harsono, Agustinus; and Haris, Abdul (2021) "Cluster Analysis of Lithology Grouping Trends using Principal Component Spectral Analysis and Complex Seismic Attributes," *Makara Journal of Science*: Vol. 25 : Iss. 1 , Article 4.
DOI: 10.7454/mss.v25i1.1227
Available at: https://scholarhub.ui.ac.id/science/vol25/iss1/4

Makara Journal of Science, 25/1 (2021), 21–27
doi: 10.7454/mss.v25i1.1227

Cluster Analysis of Lithology Grouping Trends using Principal Component Spectral Analysis and Complex Seismic Attributes

Isfan[1,2*], Agustinus Harsono[1], and Abdul Haris[1]

1. Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia,
Depok 16424, Indonesia
2. PT. Elnusa Tbk, Jakarta Selatan 12560, Indonesia

**E-mail: isfan@elnusa.co.id*

Received March 25, 2019 | Accepted May 11, 2020

Abstract

Cluster analysis is used to determine possible lithology groupings on the basis of information from seismic data. Specifically, k-means is used in the cluster analysis of different lithologies. The data center is determined randomly and updated through an iterative process (unsupervised). The cluster analysis process involves combinations of complex seismic attributes and spectral decomposition as inputs. The complex seismic attributes are reflection strength and cosine phase. Reflection strength clearly describes the lithology boundary while the cosine phase describes the lithologies. Spectral decomposition is used to detect the presence of channels. The resolution of seismic data generally reaches 90 Hz. Spectral decomposition can produce outputs with up to 1 Hz intervals. The spectral components are correlated and repeated. To reduce the repetition of spectral data and increase the trend within the data, we use principal component spectral analysis. We apply and validate the workflow using the seismic data volume acquired over Boonsville, Texas, USA. The results of the cluster analysis method show good consistency with existing lithological maps interpreted from well data correlations.

Keywords: cluster analysis, principal component analysis, unsupervised, k-means clustering

Introduction

In the oil and gas industries, geophysical methods are used to obtain subsurface imaging during exploration. Geophysical methods can also be used for reservoir characterization. The seismic reflection method remains the primary tool for hydrocarbon exploration. This method captures the geological conditions below the subsurface, including hydrocarbon traps, and provides accurate information on structural and stratigraphic features.

Seismic data are often used to derive geological properties by using a deterministic approach. An example is the use of compressional and shear wave velocities to identify Poisson’s ratio anomalies. However, deterministic relationships are general and might not be appropriate for certain areas. Consequently, the optimal relationship between the seismic data and geological properties of the subsurface cannot be produced. To overcome these problems, we employ the statistical method called cluster analysis in the current study. This technique has been used in different fields of study, such as engineering, medicine, marketing, and seismology. In such fields, cluster analysis involves the application of data partitioning to disjointed subsets. At present, this technique is widely used for the classification of seismic data.

Numerous researchers have explored the use of cluster analysis in seismology. A related study shown the used of cluster analysis in structural interpretation [1]. Another study determined the strike and average depth of geological sources that can then be used as inputs for 3D geological modeling algorithms [2]. Cluster analysis also used in channel delineation [3], and in earthquake clustering on the basis of the k-means technique [4]. In the present study, we implemented cluster analysis based on k-means to classify the lithologies obtained from recorded and processed seismic data. In principle, cluster analysis projects N seismic attributes to an N-dimension coordinate system, resulting in K groups of clouds representing possible lithologies. The identification of the center of the clouds and the related samples can be conducted through an iterative process (unsupervised) or by fixing the centers from known information (supervised). The information may come, for example, from attributes at well locations [5]. Seismic attributes and spectral decomposition are used as inputs in cluster analysis.

Spectral decomposition is usually used to identify hydrocarbons, classify facies, and calibrate of thin-bed thickness. Different frequency sections display different geological features. Given the nature of geological characteristics, such as thickness and fluid content, they can only be clearly observed at the appropriate frequency level. Generally, spectral decomposition generates single-frequency volume data for the analysis of one seismic volume. With the large volume of data available, the key issue in clustering is the development of an effective approach to data representation and reduction among single-frequency data.

Principal component analysis (PCA) is a method used to identify patterns of data distribution and to reduce the dimensions of data without losing meaningful information. In principle, scattered data have maximum and minimum axes. PCA projects data toward their corresponding trend (maximum). The direction can be determined by determining the eigenvector value of the covariance matrix with the largest eigenvalue value. The eigenvalue value is related to data variations in the eigenvector direction. The eigenvector value is also called the principal component (PC). Once the value is obtained, the data are projected to the value of the PC so that a new matrix can be obtained with the direction of the data toward the trend. In other words, data are projected toward the PC so that the data distribution pattern is obtained in the form of a straight or linear line [6].

PCA has been widely used in various fields, such as mechanical engineering, linear algebra, and seismic processing. A study showed that PCA can be used to reduce redundant spectral components into fewer, more manageable bands that capture most of the statistical variance of the original spectral response; the process is called principle component spectral analysis (PCSA) [7]. The use of PSCA in seismic thickness delineation from seismic data was also demonstrated [8]. PCSA has also been applied to single-frequency data to obtain the most important components from overabundant data.

The primary purpose of the present study is to generate a lithology map on the basis of seismic data and in the absence of any well log information. For our case study, we employ unsupervised cluster analysis to determine the lithology grouping of the Caddo formation in the Boonsville field by using the information extracted from 3D seismic data; the result is then compared with the trend log map based on log data [9]. We use the following attributes for the cluster analysis: reflection strength, cosine phase, and spectral decomposition. In particular, we use reflection strength to image the boundary lithology, cosine phase to image the lateral lithology variation, and spectral decomposition to image thin layers (e.g., channels).

Materials and Methods

Study location. An analysis was performed using the seismic data from the Boonsville 3D survey located in the Fort Worth basin, Central Texas (Figure 1). Previous seismic studies showed that the field consists of the Caddo, Davis, Runaway, and Vineyard formations. The field is one of the largest natural gas fields in the United States. Natural gas and oil production is found in the Caddo formation [10]. On the basis of this information, we chose the Caddo formation as the zone of interest.

Seismic and well data set. The seismic data have a record length of 2 s and a sample rate of 1 ms. The number of inlines is 97, and the number of crosslines is 133. In general, the seismic data used herein have good quality and good continuity of the seismic reflection characteristics. The seismic data are onshore seismic data with 0 datum on true vertical depth subsea (TVDSS). The area has 38 well log data sets with 19 well SP and gamma ray as the lithology indicator. The well basemap is shown in Figure 2.

Geology of real seismic data. The trend log data from the Caddo formation are illustrated in Figure 3. As shown in this figure, the study area consists of two delta systems. The main delta system is located in the eastern quadrant, and the second delta system is located in the northwestern quadrant. The main delta system is divided into four delta subfacies: the distal channel located at the eastern boundary, the proximal delta front in the east, the distal delta extending from the south and north to the east, and the interdelta that lies to the south and southeast. In terms of the cross sections of the main delta system, the distal channel is characterized by

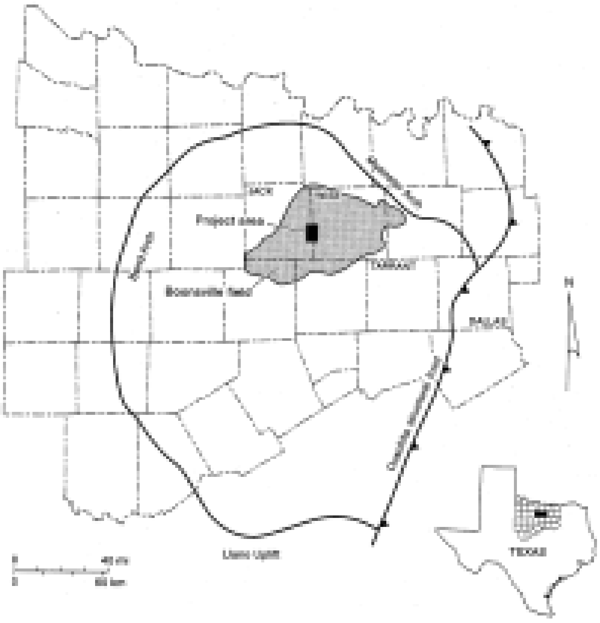


Figure 1. Location of Boonsville Field [10]

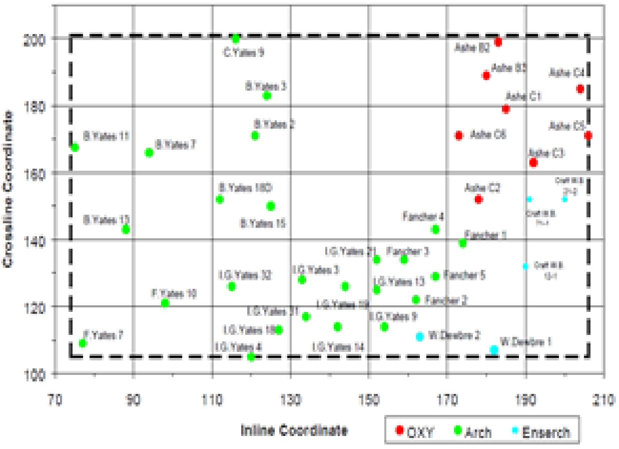


Figure 2. Well Log Data [10]

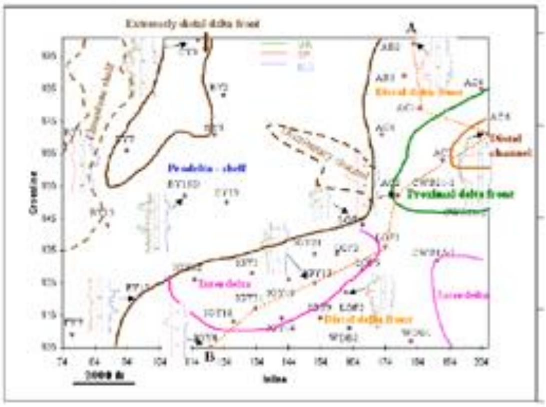


Figure 3. Representative Trend Log and Deposition Subfacies of Caddo Formation based on Log Data [9]

deposition subfacies and reservoir distribution from the area producing oil in the east, and the proximal delta front is characterized by deposition subfacies, which contain a thin layer of sandstone; most of the sandstone reservoirs are scattered in the delta front subfacies [9].

Horizon Interpretation. The results from this step are crucial for the initial detection and interpretation of lithology. Therefore, the horizon interpretation should be tied to the well log data. In this work, the horizon interpretation was applied to 3D seismic data, and the interpreted horizon was displayed in a time section to help define the area of interest. Geological trend analysis was conducted on the basis of three wells, namely, BY11, IGY31, and AC3 that represent the prodelta, interdelta, and main delta depositional subfacies, respectively.

Seismic Attributes. The seismic attributes were extracted from the 3D seismic data. They were calculated from the amplitude and complex seismic trace analysis. We then computed the root mean square of the attributes over a certain time window. The reflection strength and cosine phase are attributes extracted from the area. Reflection strength is associated with a reflection event. Therefore, it can be associated with major lithologic changes between adjacent rock layers. The displays of the cosine phase are effective in showing the discontinuities, faults, pinchouts, angularities, and events of different dip attitudes that interfere with one another.

Spectral Decomposition. Spectral decomposition can extract stratigraphic patterns that help refine the geologic interpretation of seismic data. A conventional method for computing time–frequency spectra using short-time Fourier transform requires a predefined time window and therefore has a fixed time–frequency resolution. However, to analyze a nonstationary signal whose frequency changes with time, we should have a time varying window. Thus, we implemented spectral decomposition via continuous wavelet transform (CWT). CWT utilizes the dilation and compression of wavelets to decompose the data from the time–space domain into the time–frequency domain [11].

PCSA. In the present study, PCA was used to obtain the spectral decomposition results by finding new data with a perpendicular axis and rotating it so that the data variant value is maximum. The perpendicular axis is called the eigenvector and represents the spectrum in the frequency domain. The results of the projection to the axes from the original spectrum are called the PC bands. The total number of variants can be represented by each PC band that correlates with the eigenvalue. The first PC band represents the data with the largest variant, and the second PC band represents the data with the second largest variant, and so forth. The last PC band represents that parts that are not correlated with the original spectrum; these parts are considered to be random noise.

The PCA of spectral components consists of three steps [7]. The first step is to compute the covariance matrix from the raw data for every frequency component with itself and all other frequency components.

$$C_{j,k} = \sum_{n=1}^{N} \sum_{m=1}^{M} d_{mn}^{(j)} d_{mn}^{(k)} \tag{1}$$

where $C_{j,k}$ is the *j*kth element of the covariance matrix C; *N* is the number of seismic lines in the survey; *M* is the number of seismic crosslines in the survey; and $d_{mn}^{(j)}$ and $d_{mn}^{(k)}$ are the spectral magnitudes of the *j*th and *k*th frequencies at line *n* and crossline *m*, respectively.

The second step is to decompose the covariance matrix into eigenvector v_p and eigenvalue λ_p.

$$Cv_p = \lambda_p v_p \tag{2}$$

The third step is to project the spectrum at each trace z onto eigenvector v_p by using the eigenvector with the highest eigenvalue to obtain the final data.

$$z = v_p d \tag{3}$$

In general, the first spectrum PC usually represents 50% of the data variant, the second PC represents 15%–25% of the data variant, and the third PC represents 5% of the data variant [7].

K-means Clustering. Clustering is a statistical method used to classify objects by dividing them into groups. Data are grouped together if they share similar characteristics so that each group has homogenous characteristics that differ from those of other groups. K-means clustering can be described as a method for classifying objects by dividing data into k-clusters (groups). In the present study, we employed a squared Euclidean method based on distance.

The k-means method divides data into groups according to the closest distance. In other words, data with close distances are combined into one group while data that are far apart are grouped into other groups. Each cluster in a given group is defined by the number of objects and centroids or center positions.

The k-means method calculates the distance between the center and the data grouped according to that center position and that between different positions for each measurement to minimize the distance. In the present case, we used an iterative process to minimize the distance from each point to the center position of each group. We also employed an algorithm that makes the center position move until the minimum distance is obtained. In this way, the data groups were effectively separated. The parameters used in the k-means algorithm can be determined beforehand; they include the maximum number of iteration processes that we want and the starting position of the center that we want. This process of determining the initial center is called semisupervized clustering.

In the present study, we used squared Euclidean distance as our k-means clustering algorithm; it is based on minimizing the function of the distance of points to the position of the center. In theory, if a data point is included in a given group, that data point cannot be included in other groups. However, in reality, data tend to overlap.

We used a silhouette plot with input from the results of the k-means algorithm to determine how good the results of the data grouping process were. The silhouette image displays the measurements of the distance between the points in one cluster and the points in the neighboring cluster. This measurement ranges from +1, which indicates points that are far from the neighboring cluster to 0, which indicates points that cannot be divided into one cluster to −1, which indicates points that may have been clustered incorrectly.

Results and Discussion

Horizon Interpretation. The depth of the Caddo formation ranges from 4500 feet to 4570 feet [10]. The focus at this seismic depth is the seismic–well tie. The horizon picking is based on the event marker on the well. The seismic–well tie results show that the top Caddo formation is located in approximately 870 ms (Figure 4). Seismic flattening is applied to the Caddo formation using a 15 ms window above and below the Caddo formation.

Seismic Attributes. Reflection strength (Figure 5a) shows the depositional environment of the Caddo horizon directly. The depositional environments are the distal channel in the east margin, the proximal delta front in the east, and the prodelta or limestone shelf in the west. This attribute also shows a good lithology boundary for limestone and sandstone, but it cannot differentiate between limestone and sandstone in terms of phase. The cosine phase (Figure 5b) shows the different phases between limestone and sandstone, with sandstone having negative phase values and limestone having positive phase values.

Spectral Decomposition. Using CWT, we compute 86 spectral components ranging from 5 to 90 Hz for the Boonsville seismic data. Figure 3 shows that the sandstone lithology boundary is best delineated by the 30 Hz components displayed in Figure 6a, whereas the distributary channel is best delineated by the 50 Hz components displayed in Figure 6b. The lithology boundary in Figure 6c is not clearly delineated. This result is due to the existence of high-frequency noise. The seismic data show the good depositional environment of the Caddo horizon at a frequency of 50 Hz. However, the frequency scanning method reveals that only a small portion of the spectral variation displays 1 out of 86 possible spectral

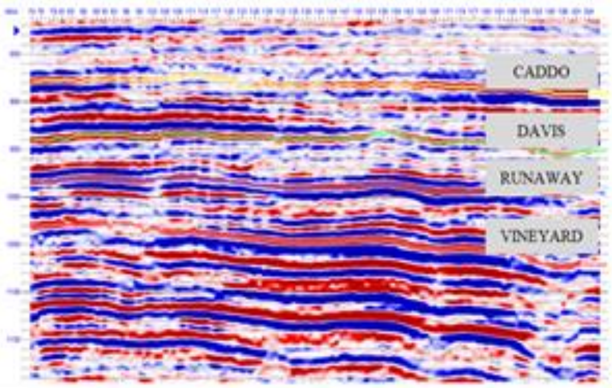
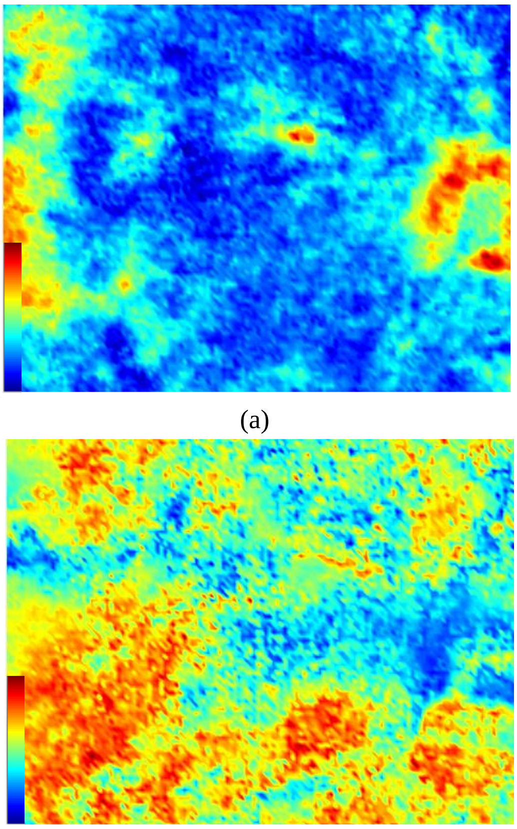


Figure 4. Horizon Picking


(a)

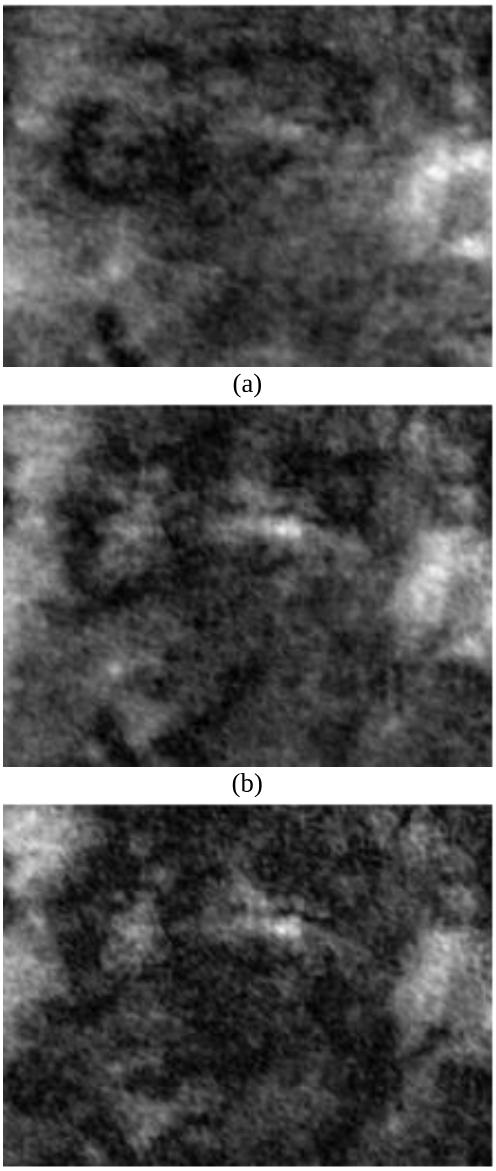
(b)

Figure 5. (a) Reflection Strength; (b) Cosine Phase

components. We use PCSA to determine the best frequency for multiple spectral component volumes.

PCSA. The result of spectral decomposition (i.e., 86 seismic frequency volumes) serves as the input of the covariance matrix for computing the eigenvector and eigenvalue. The highest eigenvalue is 8.319 × 10^7. Then, we project the original data to obtain the final data (Figure 7). The first PCSA band shows a good lithology boundary for limestone and sandstone, and this result resembles that with a frequency slice of 50 Hz. We use the combination of PCSA, reflection strength, and cosine phase as inputs for clustering.

K-means Clustering. The clustering result based on the silhouette plot in Figure 8 shows that 8 is the proper number of clusters. However, the silhouette plot does not strongly support this value. The silhouette plot shows negative silhouette values that indicate the presence of data points that are assigned to the wrong cluster. The negative values could be attributed to the existence of noise or subclusters in a cluster. The best number of clusters is 8 because the lithology map consisting of 8 clusters provides appropriate results when correlated with the well log trends and depositional subfacies of the interest area in the Caddo formation.


(a)

(b)

(c)

Figure 6. Horizon Slice: (a) CWT at 30 Hz; (b) CWT at 50 Hz; (c) CWT at 70 Hz

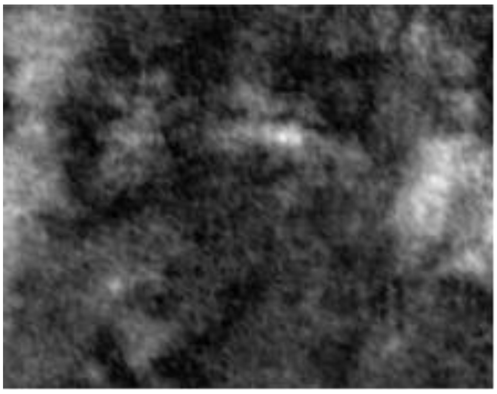


Figure 7. Spectra Projected onto First Principal Component

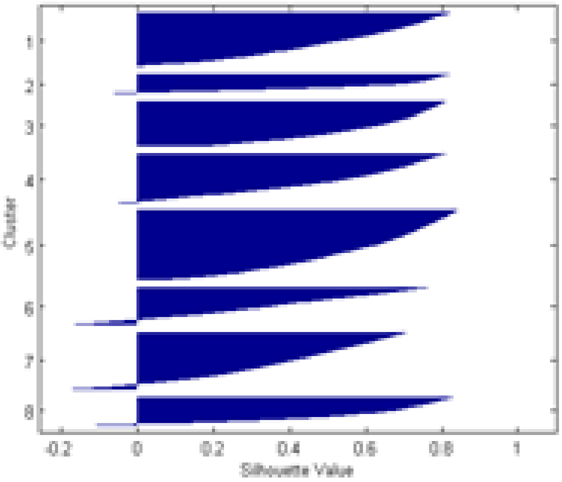


Figure 8. Silhouette Plot of Interest Area in Caddo formation

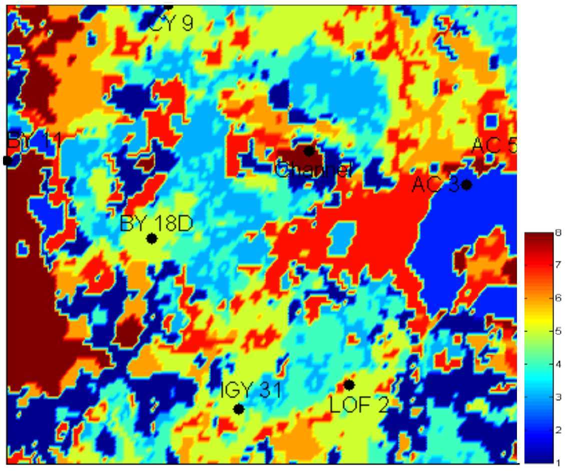


Figure 9. Cluster Analysis Result from Caddo Formation

The clustering result (Figure 9) is correlated with the well log trends and depositional subfacies of the interest area in the Caddo formation (Figure 3). Thus, the lithology types can be interpreted. Figure 9 shows that limestone and sandstone can now be differentiated. In the west, the brown area represents the limestone zone in the prodelta subfacies. In the east, the blue area represents the sandstone zone in the main delta subfacies. The pattern of the interdelta subfacies indicates the mixture of limestone–sandstone in the south and a dark blue area in the southeast. The cluster analysis result (Figure 9) shows a good correlation with the well log trends (Figure 3).

Conclusion

In the present work, we demonstrate that the cluster analysis method can be used to determine lithology groupings in an area on the basis of 3D seismic data. We use this method as a quick method for generating an initial lithology map in the absence of any well log information. The application of cluster analysis to real seismic data (3D Boonsville seismic data) reveals 8 as the appropriate number of clusters. The selected number of clusters is illustrated as cluster models called the lithology map. The lithology map patterns show that they match the well log trends and depositional subfacies of the interest area in the Caddo formation. Seismic attributes such as reflection strength, cosine phase, and PCSA can provide optimum results in the Boonsville area.

Further research should be carried out to improve the results of this study. One drawback of the k-means clustering method is that the initial center location yields different results due to varying central locations. Hence, a supervised cluster study should be conducted.

Acknowledgements

We acknowledge and grateful to the the Bureau of Economic Geology, University of Texas and Geophysics Department, University of Indonesia for The Boonsville 3-D seismic data set.

References

[1] Krehbiel. 1993. Cluster analysis applied to low relief structural interpretations. The Leading Edge. 831–836.

[2] Ugalde, H., Morris, W.A. 2003. Cluster analysis of Euler deconvolution solutions: New filtering techniques and geologic strike determination. The Leading Edge. 942–953.

[3] Matos, M.C., Yenugu, M., Angelo, S.M., Marfurt, K.J. 2011. Channel delineation and chert reservoir characterization by integrated seismic texture segmentation and cluster analysis. SBGf. 801–806.

[4] Kamath, R.S. 2017. Earthquake Cluster Analysis: K-Means Approach. JCPS.2017: 250–253.

[5] Coleou, T., Poupon, M., Azbel, K. 2003. Unsupervised Seismic Facies Classification: A Review and Comparison of Techniques and Implementation. The Leading Edge. 942–953.

[6] Smith, L. 2002. A Tutorial on Principal Component Analysis. https://ourarchive.otago.ac.nz/handle/10523/7534.

[7] Guo, H., Marfurt, K.J., Liu, J. 2009. Principal component spectral analysis. Geophys. (74): 35.

[8] Zhou, J. and Castagna, J.P. 2017. Seismic thickness delineation using spectral principal component analysis: theory and a synthetic turbidite example. 87th Annual International Meeting, SEG, Expanded Abstracts. pp. 3158–3162.

[9] Penington, W.D., Acevedo, H., Green, A., Haataja, J., Len, S., Minaeva, A., Xie, D. 2001. Calibration of Seismic Attributes for Reservoir Characterization.

Annual Technical Progress Report Michigan Technological University.

[10] Tanakov, M.Y., Kelkar, M. 2000. Integrated Reservoir Description for Boonsville, Texas Field Using 3D Seismic Well and Production Data. Soc. Pet. Eng.

[11] Sinha, S., Routh, P.S., Anno, P.D., Castagna, J.P. 2005. Spectral decomposition of seismic data with continuous-wavelet transform. Geophys. (70): 19–25.