International Journal of GEOMATE, June, 2022, Vol.22, Issue 94, pp.83-92
ISSN: 2186-2982 (P), 2186-2990 (O), Japan, DOI: https://doi.org/10.21660/2022.94.3246
Geotechnique, Construction Materials and Environment

SANDSTONE RESERVOIR DELINEATION USING MACHINE LEARNING-BASED SPECTRAL ATTRIBUTE ANALYSIS IN "G" FIELD JAMBI SUB-BASIN, SOUTH SUMATRA BASIN

Gabriella Eka Putri[2], *Abdul Haris[1,] and Muhammad Rizqy Septyandy[1]

[1]Geology and Geophysics Study Program, FMIPA Universitas Indonesia, Indonesia
[2]Reservoir Geophysics Study Program, Department of Physics, FMIPA Universitas Indonesia, Indonesia

*Corresponding Author, Received: 20 July 2021, Revised: 07 Jan. 2022, Accepted: 02 April 2022

ABSTRACT: A lot of reservoirs have thick and thin sand bodies at the same intervals, while the amplitude values of seismic data frequently highlight sand bodies near the ¼ wavelength for the tuning phenomena. These machine learning methods aim to link seismic attributes for the qualitative prediction of facies classification and compare the results obtained with the common seismic attributes visualizations controlled by the gamma-ray log data as the lithofacies guidance. This study performed an extraction of Seismic Spectral Attributes (SSAs) in the area of interest for the spectral decomposition RGB Blending visualization. Furthermore, numerical values were applied for several seismic attributes in the clustering step, while a principal component analysis (PCA) was proposed towards lowering the computational time and storage space on these values. Subsequently, a subsurface depositional facies map was obtained using the frequency cube from the red-green-blue (RGB) compounding technique, while the facies classification map, useful for the reservoir delineation, was obtained using the final facies map from the combination of principal component analysis of the original numerical seismic attributes value followed by unsupervised classification through clustering.

Keywords: Seismic Spectral Attributes, Principal Component Analysis, and Unsupervised Learning

1. INTRODUCTION

The seismic data method's low dominant frequency, narrow bandwidth, and low S/N ratio, owing to deficient contrast of impedance and powerful attenuation, necessitates the creation of other suitable methods for delineating subsurface depositional facies. The seismic spectral attribute has been used with some approaches, for instance, seismic spectral decomposition, while other approaches, including continuous wavelet, transforms (CWT) and the short-time Fourier transform (STFT), have been proposed. CWT results in time-frequency representative with various resolutions from the seismic signal in cases where a wavelet has been previously selected [1, 2, 3].

This study, therefore, used CWT to detect sandstone frequency thickness, combined with RMS amplitude and machine learning clustering map, due to the method's ability to result in localized time-frequency representative and the seismic signal from Ricker wavelet. Using the red-green-blue (RGB) color blending image shows three frequency cubes to be compared with the depth and thickness map, as well as other seismic attributes visualization, and this enables detection of sandstone facies boundary, as well as qualitative estimation of the sand's thickness.

Subsequently, a bridging was performed from the commonly used seismic attributes to the use of seismic attributes parameters in clustering algorithm application in facies classification by initially applying a principal component analysis (PCA), a broadly applied technique to reduce dimensionality in machine learning [4]. PCA orthogonally converts a big set of seismic spectral attributes into a smaller set, while retaining the same information [4]. The resultant of the brand-new variable, principal component (PC), is a linear combination from SSA with maximum variance and is not inter-correlated.

After the seismic attribute's value has been well packaged in the form of principal components, the unsupervised machine learning step is applied using several clustering algorithms to determine whether the algorithm can map the seismic attributes data into several clusters interpreted as facies in this case. The clustering algorithm used in this study are K-Means Clustering, Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH), and Gaussian Mixture. The interpretation of these clusters is then compared with the morphology forms found in spectral decomposition RGB visualization, RMS amplitude, and other seismic attributes visualization.

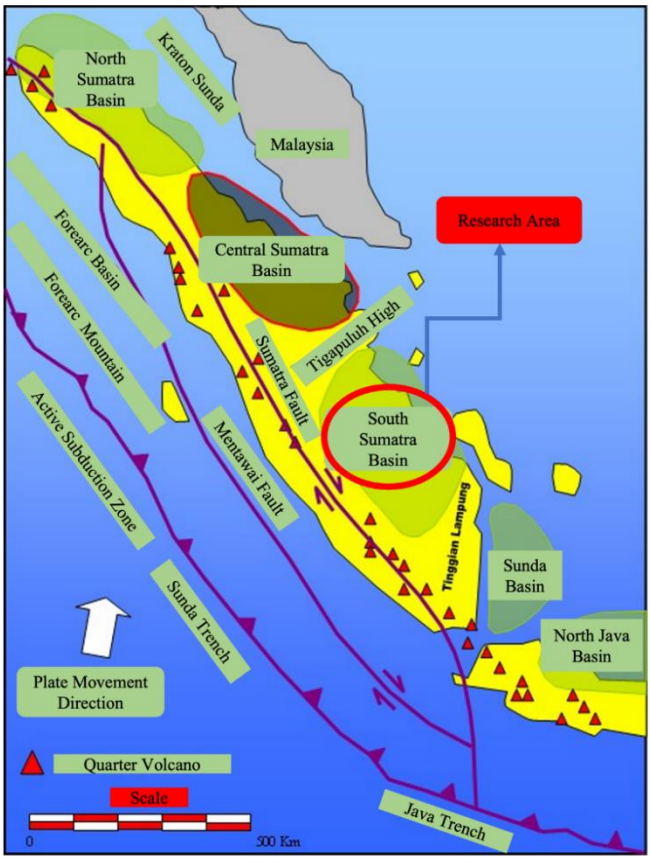

Fig. 1 South Sumatra Basin Physiography [18].

However, K-Means, BIRCH, and Gaussian Mixture algorithms using several seismic attributes data through principal components combined with Spectral Decomposition and conventional seismic attribute visualization, including RMS Amplitude has not been widely applied to Indonesian fields. This combination of several methods is aimed at improving the clustering of depositional facies in the "G" field, located in Jambi Sub-Basin, South Sumatra Basin. In addition, this combination is expected to distinguish the existent depositional facies in the field through the seismic attributes numerical data.

In physiography, the South Sumatra Basin is a northwest-southeast direction basin which is bounded by the Semangko and the Bukit Barisan faults in the southwest, the Sunda Shelf in the northeast, and the Lampung Highlands in the southeast (Figure 1).

2. RESEARCH SIGNIFICANCE

The significance of the study is how seismic data and several unsupervised machine learning methods work in resulting tin he mapping of depositional facies can be assessed. The absence of known label of the facies in the training data let the most suitable configuration of the algorithm work to give the best facies clustering map results from the data used while keep considering all characteristics of the data fairly. Using more than one algorithm to work in picturing the clusters of the data can explain what kind of algorithms fit best with the data characteristics used in the process.

3. METHODS

3.1 RMS Attribute

Amplitude attribute is commonly used to obtain the subsurface information, and the interpretative approach uses bright spot assumption on seismic, based on the amplitude's value, where this bright spot assumption may be related to the presence of hydrocarbon [16]. High hydrocarbon saturation condition, high porosity, and thick pay thickness is bound to produce high amplitude, meaning a more evident amplitude contrast implies a better prospect.

Furthermore, amplitude represents the amount of energy in the time domain on the seismic wave, and this attribute is generally used as the straight hydrocarbon indicator, as well as in the delineation of facies and thickness. The instance of primary amplitude is RMS Amplitude, the square root of the sum of amplitude square of seismic data, expressed in the following equation.

$$RMS\ Amplitude = \sqrt{\frac{1}{N}\sum_{i=1}^{N} A_i^2} \quad (1)$$

Where N is the magnitude of amplitude sample in the window analysis and A is the amplitude value.

3.2 Spectral Decomposition

The amplitude spectrum and phase to the specific wavelength are visible through this attribute. Furthermore, the thickness, density, and signal velocity parameters passing through a material provide a frequency characteristic description. The material layer usually comprises numerous rock layers, each with a unique frequency characteristic, and to obtain each layer's frequency, the thickness from the layer or layer volume must be included in the frequency interval until the desired maximum frequency is obtained.

3.3 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a technique of decreasing performed to orthogonally convert a big data set upon a smaller set while retaining the same information [4]. The principal components produced is a linear combination from the original data with maximum variance and is not inter-correlated [12]. This is defined in the following equations, using the X vector of SSA.

$$X = (X_1, X_2, \dots, X_m) \quad (2)$$

Xi is an SSA, therefore, the mean vector of X is stated by:

$$\boldsymbol{u_x = E\{X\}} \quad (3)$$

Other than the above concept, there is proof that eigenvectors are the principal components. In this study, in analyzing the attributes for principal components, we use the analysis of the eigenvalue and eigenvectors. To see if eigenvectors are the principal components, we can consider a linear combination that has maxim mum variance. The variance explained by any eigenvector is its eigenvalue since the eigenvector is selected by choosing an appropriate unit vector [17].

3.4 Machine Learning Clustering with Seismic Attributes (Unsupervised Machine Learning)

This step aims to use the seismic attributes data which have been transformed into two principal components through Principal Component Analysis to generate a facies clustering map. Both principal components (PCs) are used to create a facies model through several machine learning clustering algorithms. The algorithms used are K-Means Clustering, Balanced Iterative Reducing, and Clustering using Hierarchies (BIRCH), as well as Gaussian Mixture, and each algorithm uses a different method in deciding the effective number of clusters.

Continuous Wavelet Transform (CWT)-based Spectral Decomposition, is based on the presence of certain frequency characteristics response from the reflection seismic in a thin layer or below the seismic vertical resolution. Continuous Wavelet Transform separates signals into frequency distribution which is related to time with scaled translation and dilatation [11]. A detailed picture is obtainable by associating frequency value with the target area's thickness.

In the K-Means algorithm, the sum of squares of the data points' Euclidean distance for the nearest representative is used to measure the clustering objective function. Therefore, the following equation is obtained.

$$\boldsymbol{Dist(\overline{X_i},\overline{Y_j}) = \|\overline{X_i} - \overline{Y_j}\|_2^2} \quad (4)$$

Where, || ⋅ ||p represents Lp-norm. The expression of Dist (Xi, Yj) is evident as a squared error in data points with the nearest representative [14]. Generally, the purpose is to minimalize the sum of a square error on different data points.

The BIRCH algorithm works by two parameters: branching factor B, as well as threshold T, and obtains the input data set from N data points, represented as a real-valued vector, and the desired sum of clusters K [15]. Subsequently, the algorithm scans all leaf entry in cluster factor tree to build smaller cluster features while removing the outlier and clustering the full sub-cluster into a larger sub-cluster. The agglomerative hierarchical clustering algorithm is then immediately applied to the sub-cluster represented by the cluster feature vector. This also provides flexibility, enabling the desired amount of cluster or threshold diameter to be determined. With this algorithm, the parameter branching factor and threshold are used to obtain cluster prediction.

The Gaussian Mixture algorithm is a clustering method, where an ellipse form of the cluster is formed based on the density probability estimation using Maximization-Expectation. Each cluster is modeled as a Gaussian distribution, while the mean and covariance are compared only to mean in K-Means clustering, enabling the algorithm to provide a better quantitative measurement, compared to the compatibility per sum of the cluster [6]. Gaussian Mixture is represented as the linear combination from Gaussian probability

distribution and is expressed in the following equation.

$$\boldsymbol{p(X)} = \sum_{k=1}^{K} \boldsymbol{\pi_k} \mathrm{N(X|}\boldsymbol{\mu_k}, \sum_k) \quad (5)$$

Where K is the amount of component from the mixed model and πk is the mixing coefficient providing density estimation from each Gaussian component [6].

To determine the clusters' amount K, the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) are used. The concept of BIC is to select the model from a set of model candidates by maximizing the posterior probability [7]. In cases where a model depends on straight to the sum of the cluster, the BIC criterion for a set of X' is defined by the following relationship.

$$\boldsymbol{BIC\ (K) = -2L' * (K) + r \log(N')} \quad (6)$$

This is the criterion value in cases where there is no other information for the model with an independent parameter of r. The criterion also requires adjustments to consider side information. For the set of X, given the relation 1, the observation amount is N, while the possibility of the maximum log is L*(K), and this considers positive constraint measured differently.

The second criterion, AIC which was proposed in (Akaike, 1974) [8], selects the model, where the Kullback-Leibler distance between the models with the truth M_0, is minimalized. This is equivalent to choosing the model where the maximized value is the maximum value from the following equation.

$$\int_{-\infty}^{+\infty} P(X|M_0)P(X|M_i) \quad (7)$$

The AIC criterion without side information takes the following form.

$$\boldsymbol{AIC\ (K) = -2L' * (K) + 2r} \quad (8)$$

4. RESULT AND DISCUSSION

4.1 RMS Attribute

Figure 2 shows the form of distributary channel conceived by high RMS Amplitude value with the color of yellow to red at the southeast. Meanwhile, at the north, the form is interpreted as the fluvial floodplain correlated with the well-M, a well near the area interpreted as fluvial floodplain and distributary channel. At the area interpreted as a fluvial floodplain, the show of sand is found in the log data with a serrated form of gamma-ray log associated with aggrading deposition, the deposition with massive alternations of sandstone and mudstone. This also explains the area's high RMS Amplitude value, indicating a high-density sandstone distribution, and consequently, high amplitude anomaly.

4.2 Spectral Decomposition

In applying the spectral decomposition method, the seismic volume must first be extracted with three frequencies, from low, middle, and high. The frequencies used for the seismic cubes are 10 Hz, 20 Hz, and 45 Hz, based on the seismic data's spectrum analysis (Figure 3).

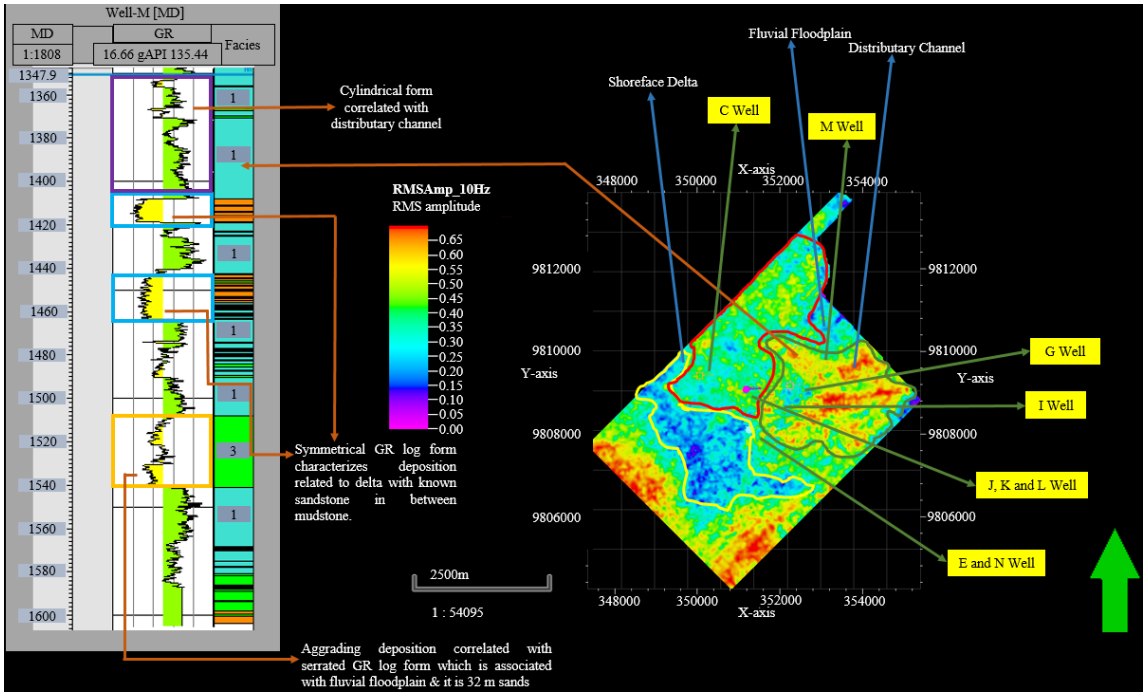

Fig. 2 RMS Amplitude visualization combined with gamma-ray (GR) log interpretation.

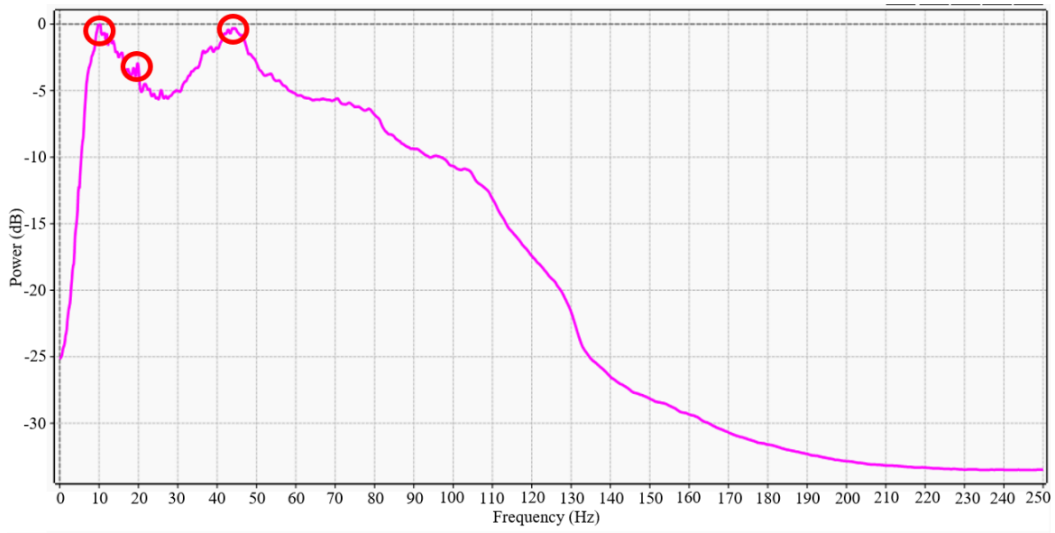

Fig. 3 Seismic Spectrum Analysis.

Figure 4 shows the north to northeast part, interpreted as the fluvial floodplain and close to the area interpreted as a channel (in the southeast part). This channel form is also found at the same area in the RMS Amplitude visualization (Figure 2). Meanwhile, the magenta to blue color at the northeast part interpreted as fluvial floodplain explains the show of high frequency associated with the thin sand layer. This is confirmed from the gamma-ray log near the area showing thin sand in between mudstone with serrated log form correlated with fluvial floodplain supporting this interpretation. Similar to Figure 3, the west direction's middle part is interpreted as shoreface delta form, indicating the form of the delta. This interpretation is supported by the deposition direction of NE – SW, and the deposition development begins from the fluvial floodplain to the distributary channel in the delta area (towards the marine).

4.3 Principal Component Analysis (PCA)

In this step, seismic attributes data is prepared for division into several principal components. The seismic attributes data used in this step are RMS Amplitude, Average Envelope, Average Instantaneous Frequency, Average Instantaneous Phase, Average Magnitude, Maximum Amplitude, and Standard Deviation of Amplitude. Data standardization was performed to ensure the data has mean and variance values of 0 and 1, respectively, and can produce the variance percentage of the components used [13].

Figure 5 shows only two features are used, and these are two principal components (PC). Therefore, the components used have variance above 0.1%. From Figure 5, feature 1 (PC 1) of PCA features axis (x-axis) is above 0.6%, while feature 2 (PC 2) is above 0.1%. Alternatively, the PCA feature is determined by creating explained variance plot from the components used in this step.

Based on Figure 6, the resultant variance (y-axis) depends on the quantity of the components used (x-axis). Practically, the PCA features used are the features covering about 70% of the variance. Therefore, two PCA features are used from these components.

Table 1 shows the variance ratio and Eigenvalue for PC 1 and PC 2, each with 65% of the data variance and 15% of the data variance, combined to obtain 80% of the data variance. PC 1and PC 2 also have Eigenvalues of 5.2 and 1.24, respectively. In addition, the Eigenvector is also explored to determine the PCs' significant features, as well as the percentage of each on a related PCA.

Table 2 shows the RMS Amplitude, Average Envelope, Average Magnitude, Maximum Amplitude, and Maximum Magnitude are the most important features for PC 1 because these features have significantly high Eigenvector. Meanwhile, the PC 2 counterparts are Average Instantaneous Frequency and Standard Deviation of Amplitude, because these features have significantly high Eigenvector.

4.4 Machine Learning Clustering with Seismic Attributes (Unsupervised Machine Learning)

The Elbow method approach in K-Means Clustering is carried out by observing the elbow on the inertia graph. In the graph, the form before the elbow is usually sharp enough and becomes smooth afterward. To obtain a clustering prediction with the K-Means algorithm, the parameters k and inertia are used.

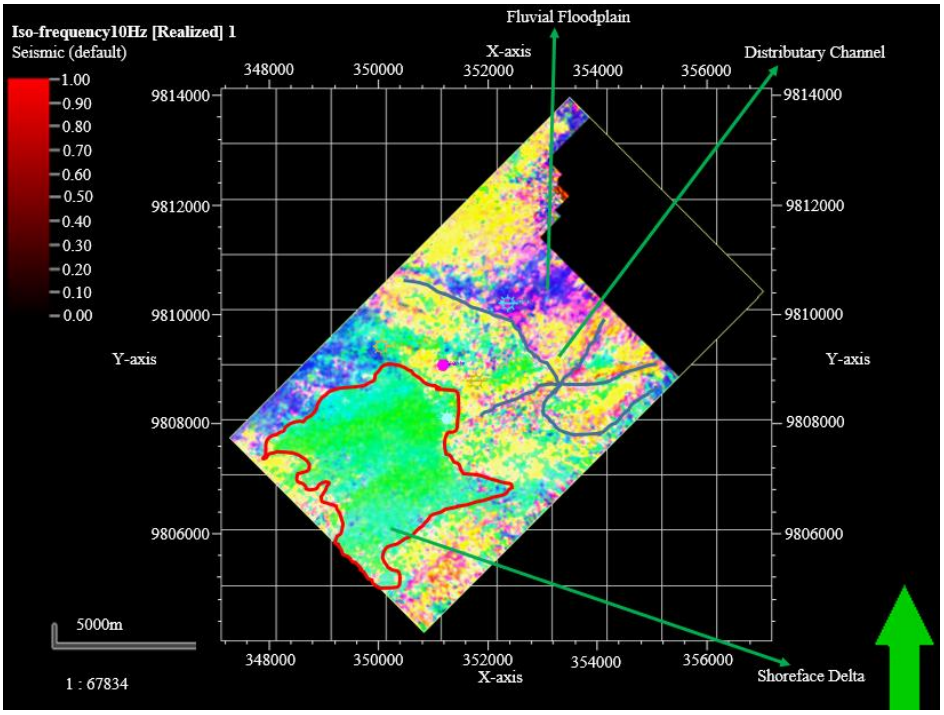

Fig. 4 Spectral Decomposition RGB Blending application on 10 Hz, 20 Hz, and 45 Hz.

The k used is within range 1 to 8, while the inertia is obtained from fitting results with the training model from the K-Means. Subsequently, the possibility of cluster amount is resulted by using the Elbow method (Figure 7).

Figure 7 shows the elbow form occurs at three clusters for the K-Means algorithm. Therefore, the clustering process is run using a k value of 3.

For the BIRCH algorithm, in the process, the branching factor used is within the range of 100 to 1000 with the interval of 50, while the threshold used is within the range of 0 to 1. This results in the possibility of an effective cluster (Figure 8). Figure 8 shows the effective sum of the cluster for the data used is three, therefore, the branching factor and threshold of 100 and 1, respectively, are used in the processing cluster.

To determine the cluster prediction, the criteria AIC and BIC, the parameter k with a range of 1 to 8, as well as the inertia are used. Figure 9 shows the effective sum of the cluster.

According to Figure 9, the effective sum of the cluster is three. In addition, AIC and BIC reflect each other quite well [9]. AIC and BIC are meant to be minimalized; therefore, the graph's least point is selected. Generally, the measurement of both criteria shows the same number on the sum of the cluster. However, in some cases, differences occur due to the BIC model's relative simplicity, and the AIC model's data complexity. Therefore, the sum of cluster k is applied.

Contrary to supervised learning, the clustering process with BIRCH, K-Means and Gaussian Mixture does not use facies labeling but maps the facies based on the parameters. These three machine learning clustering methods have been able to map the data into three facies where the cluster distribution of the cluster is qualitatively similar to the facies distribution in RMS Amplitude and RGB Blending of the Spectral Decomposition visualization. Based on the resultant facies distribution map from the three clustering methods (Figure 10), the BIRCH method was concluded to produce the closest result to RMS Amplitude and Spectral Decomposition RGB Blending visualization. This is possible because the BIRCH method performs hierarchical clustering over, particularly large dataset process [10]. Meanwhile, the K-Means and Gaussian Mixture method modified with the addition of expectation-maximization algorithm can produce results with similar accuracy as the BIRCH algorithm result. Therefore, at the east to south part (Figure 10), in each facies cluster map, the three methods are unable to show the channel form described from the picture with the red dominant color distribution. Figure 10 only shows the area as a single cluster, possibly due to the algorithms' characteristics, general data density, and data density at the area interpreted as the channel in the east of south on the facies map, or the density of one of some parameters in PC 1 or PC 2. Further improvement is available to this study where there is a need to try or generate an improved algorithms aware le to model the depositional facies in a more detailed form. Other attributes or newly formed

attributes that can be derived in the future that can picture the depositional facies well can also be explored along with experimenting with the most suited algorithms.

A previous study applied another machine learning clustering method in the central Ordos Basin, western China, and successfully determined the facies classification resulted from spectrally decomposed attribute with FSOM map. The result from clustering with FSOM map visualization showed similarity with the manual interpretation. Also, the elaborated spectral attributes are able in picturing the condition of subsurface deposition, however, misunderstandings tend to occur in the FSOM map on account of the good calibrations' absence on the unsupervised learning [5].

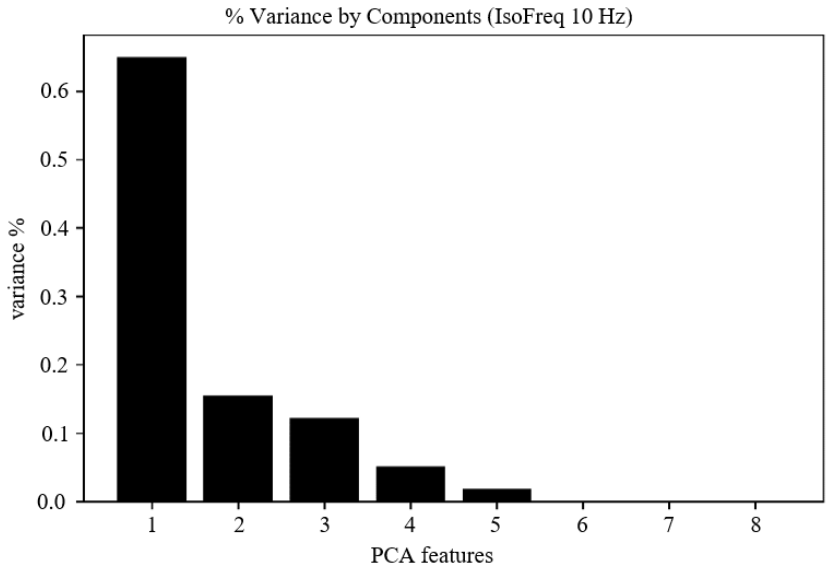

Fig. 5 The components' variance percentage.

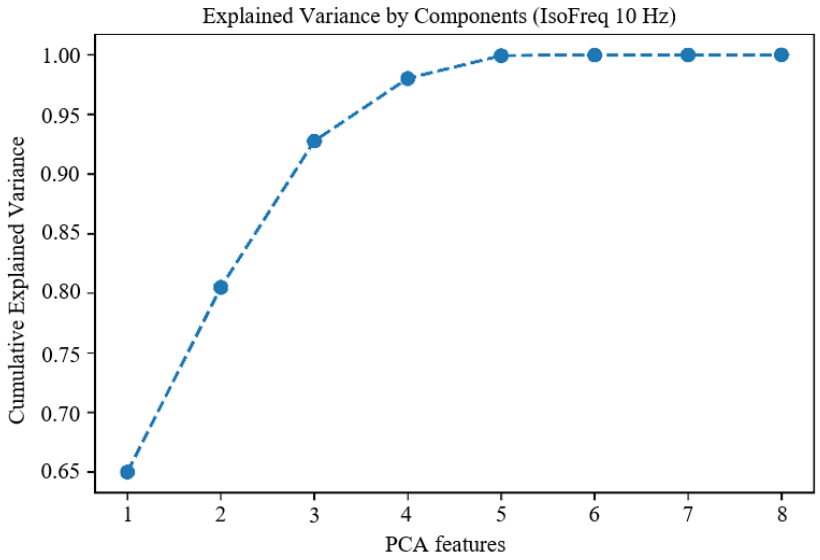

Fig. 6 Explained variance from the components.

Table 1 Variance Ratio and Eigen Value table for PCA 1 and PCA 2.

PC	Variance Ratio	Eigen Value
PC 1	0.65	5.2
PC 2	0.15	1.24

Table 2 Eigenvector table for PC 1 and PC 2.

Surface Attribute	Eigen Vector (PC 1)	Eigen Vector (PC 2)
RMS Amplitude	0.43	0.1
Average Envelope	0.42	0.16
Average Instantaneous Frequency	0.24	0.61
Average Instantaneous Phase	0.07	0.03
Average Magnitude	0.41	0.25
Maximum Amplitude	0.42	0.23
Maximum Magnitude	0.41	0.23
Standard Deviation of Amplitude	0.27	0.65

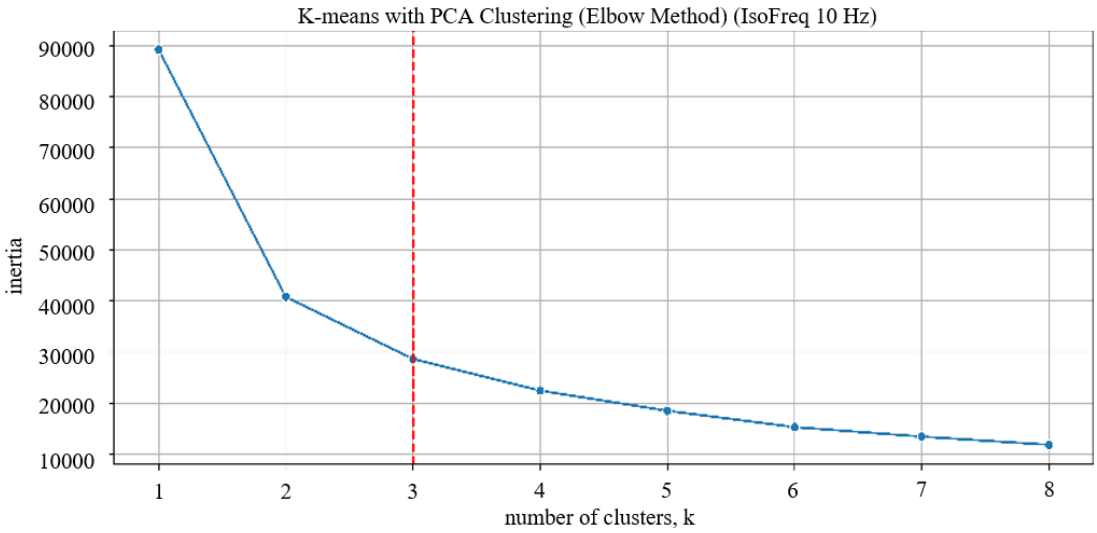

Fig. 7 Elbow Method Graph to determine the effective amount of cluster.

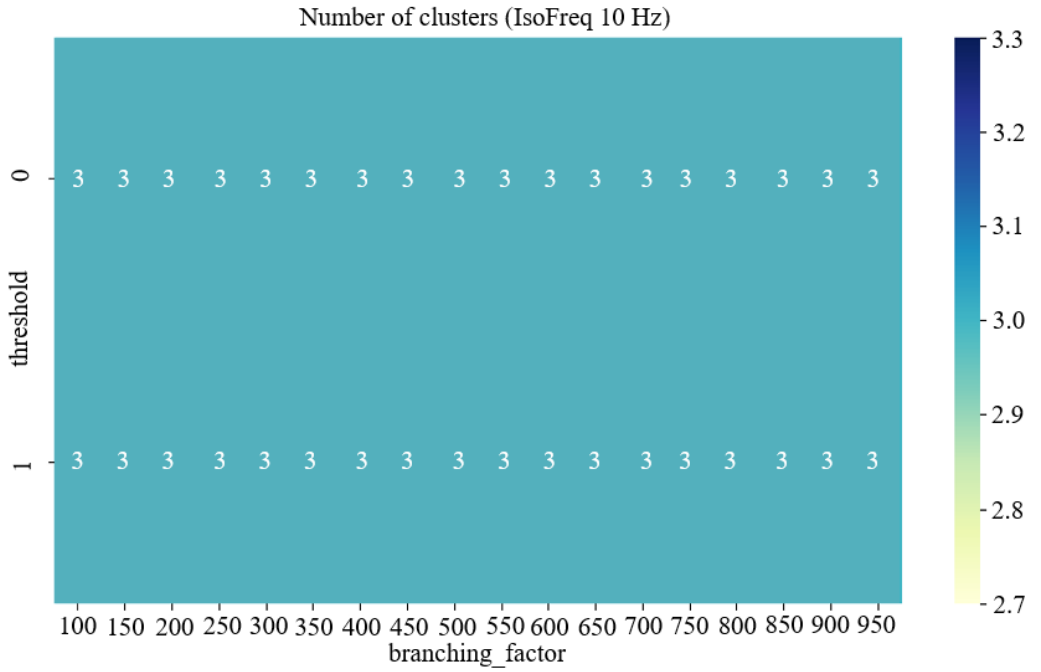

Fig. 8 Heatmap to determine the effective amount of cluster.

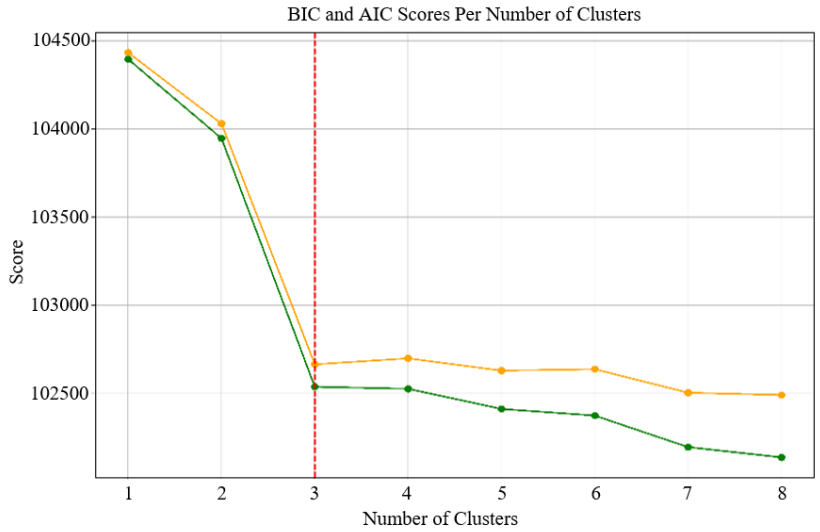

Fig. 9 AIC and BIC score graph in determining the effective sum of the cluster.

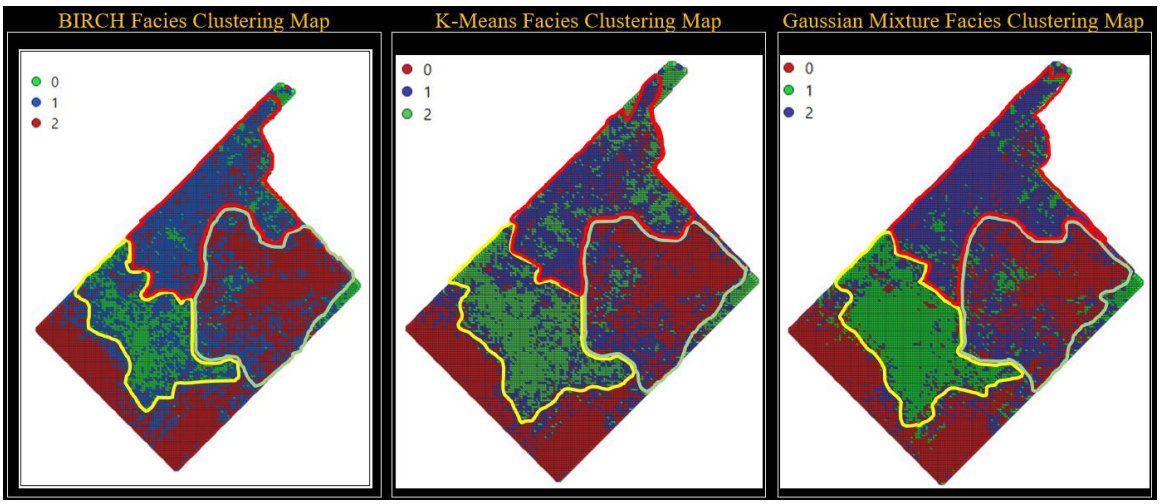

Fig. 10 Facies Cluster Maps with BIRCH (left), K-Means (middle), and Gaussian Mixture (right).

5. CONCLUSION

This study describes a new method combination of RMS Amplitude, Spectral Decomposition, and machine learning clustering using numerical seismic attributes values, and applied this method to the "G" Field of the South Sumatra Basin, to obtain facies clustering map from machine learning clustering algorithms. Subsequently, the map obtained is then compared to RMS Amplitude, Spectral Decomposition, and supported by the gamma-ray log interpretation. The result shows the facies found in the target area, and these are the fluvial floodplain, distributary channel, as well as shoreface delta. Furthermore, the results shown from the machine learning clustering have the same morphology and pattern as the RMS Amplitude and Spectral Decomposition visualization. Also, the interpretation from the machine learning clustering in the area interpreted as the channel does not show the form of the channel but shows the area as a single cluster, and this is possibly due to several factors, including the algorithms' characteristics, as well as the data density.

6. ACKNOWLEDGMENTS

The authors are grateful to the Universitas Indonesia PUTI scheme for providing financial support towards this study.

7. REFERENCES

[1] Grossmann, A. and Morlet, J. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. SIAM Journal on mathematical analysis, 15(4), 723-736.

[2] Chakraborty, A. and Okaya, D. (1995). Frequency-time decomposition of seismic data using wavelet-based methods. Geophysics, 60(6), 1906-1916.
[3] Sinha, S., Routh, P. S., Anno, P. D., and Castagna, J. P. (2005). Spectral decomposition of seismic data with continuous-wavelet transform. Geophysics, 70(6), P19-P25.
[4] Jolliffe, I. (2011). Principal component analysis (pp. 1094-1096). Springer Berlin Heidelberg. RESUME SELİN DEĞİRMECİ Marmara University, Goztepe Campus ProQuest Number: ProQuest). Copyright of the Dissertation is held by the Author. All Rights Reserved, 28243034, 28243034.
[5] Wang, Z., Gao, D., Lei, X., Wang, D., and Gao, J. (2020). Machine learning-based seismic spectral attribute analysis to delineate a tight-sand reservoir in the Sulige gas field of central Ordos Basin, western China. Marine and Petroleum Geology, 113, 104136.
[6] Patel, E. and Kushwaha, D. S. (2020). Clustering cloud workloads: k-means vs gaussian mixture model. Procedia Computer Science, 171, 158-167.
[7] Grall-Maes, E. and Dao, D. T. (2016, February). Assessing the Number of Clusters in a Mixture Model with Side-information. In ICPRAM (pp. 41-47).
[8] Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), 716-723.
[9] Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. Sociological methods and research, 33(2), 188-229.
[10] Krishna, D. P., Senguttuvan, A. and Latha, T. S. (2012). Clustering on Large Numeric Data Sets Using Hierarchical Approach: Birch. Global Journal of Computer Science and Technology.
[11] Haris, A., Morena, V., Riyanto, A., and Zulivandama, S. R. (2017, July). Reservoir hydrocarbon delineation using spectral decomposition: The application of S-Transform and empirical mode decomposition (EMD) method. In AIP Conference Proceedings (Vol. 1862, No. 1, p. 030184). AIP Publishing LLC.
[12] Sun, J., Passi, K., and Jain, C. (2016). Increasing Efficiency of Microarray Analysis by PCA and Machine Learning Methods. In Proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP) (p. 56). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
[13] Abosamra, G. and Faloudah, A. Machine Learning-Based Marks Prediction to Support Recommendation of Optimum Specialization and Study Track. International Journal of Computer Applications, 975, 8887.
[14] Telgarsky, M. and Vattani, A. (2010, March). Hartigan's method: k-means clustering without Voronoi. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (pp. 820-827). JMLR Workshop and Conference Proceedings.
[15] Lorbeer, B., Kosareva, A., Deva, B., Softić, D., Ruppel, P., and Küpper, A. (2018). Variations on the clustering algorithm BIRCH. Big data research, 11, 44-53.
[16] Farfour, M., Yoon, W. J., Ferahtia, J., and Djarfour, N. (2012). Seismic attributes combination to enhance detection of bright spots associated with hydrocarbons. Geosystem Engineering, 15(3), 143-150.
[17] Aschheim, M. A., Black, E. F., & Cuesta, I. (2002). Theory of principal components analysis and applications to multistory frame buildings responding to seismic excitation. *Engineering Structures*, *24*(8), 1091-1103.
[18] Heidrick, T. L., and Aulia, K. (1993). A structural and tectonic model of the coastal plains block, Central Sumatra Basin, Indonesia.