Probability Estimation of Direct Hydrocarbon Indicators Using Gaussian Mixture Models

John B. Lemos
Universidade Federal da Bahia
Salvador, Brazil
jonh.brian@ufba.br

Matheus R. S. Barbosa
Universidade Federal da Bahia
Salvador, Brazil
matheus.radames@ufba.br

Edric B. Troccoli
Grupo de Estratigrafia Teórica e Aplicada
Universidade Federal da Bahia
Salvador, Brazil
edrictroccoli@ufba.br

Alexsandro G. Cerqueira
Grupo de Estratigrafia Teórica e Aplicada
Universidade Federal da Bahia
Salvador, Brazil
alexsandrocerqueira@ufba.br

***Abstract*—This work aims to delimit the Direct Hydrocarbon Indicators (DHI) zones using the K-means and mostly the Gaussian Mixture Models (GMM) algorithm, unsupervised machine learning methods, over the FS8 seismic horizon in the seismic data of the Dutch F3 Field. The dataset used to perform the clustering analysis was extracted from the 3D seismic dataset. It comprises the following seismic attributes: Sweetness, Spectral Decomposition, Acoustic Impedance, Coherence, and Instantaneous Amplitude. The Principal Component Analysis (PCA) algorithm was applied in the original dataset for dimensionality reduction and noise filtering, and we choose the first three principal components to be the input of the clustering algorithm. The cluster analysis using both K-means and Gaussian Mixture Models was performed by varying the number of groups from 2 to 20. The Elbow Method suggested a smaller number of groups than needed to isolate the DHI zones. Therefore, we observed that four is the optimal number of clusters to highlight this seismic feature. Furthermore, it was possible to interpret other clusters related to the lithology through geophysical well log data.**

***Index Terms*—Direct Hydrocarbon Indicators, Gaussian Mixture Models, Principal Component Analysis, Seismic Attributes, Cluster Analysis**

I. Introduction

The pattern recognition activity is essential for interpreting seismic data, whether related to studies of depositional environments or hydrocarbon reservoirs. Its major purpose is partitioning seismic data according to some similarities. The main challenge for interpreters is to recognize these patterns in a large seismic dataset that makes their quick interpretation, without intelligent algorithms, infeasible. These algorithms support determining rock physical parameters, data interpretation, noise elimination, and ambiguity analysis in geophysical models [1]–[3].

Barnes and Laughlin (2002) [4] carried out a comparative study between K-means, Hierarchical Clustering, and Self-Organizing Maps (SOM) and attested the good performance of these techniques in 3D seismic volumes. Zhao (2015) [5] conducted studies applying supervised and unsupervised methods in the Barnett Shale geological formation, known to be one of the main shale gas reservoirs in the United States.

The use of clustering analysis algorithms on 3D seismic data increases recurrently and has become even more necessary due to the abundance of available seismic data and the need for faster and more reliable interpretation. This paper presents an application of the Gaussian Mixture Model (GMM) algorithms over the FS8 seismic horizon, obtained from F3 block 3D seismic data, located in the North Sea - Netherlands. The K-Means algorithm was also implemented in order to compare with the technique mentioned above. We extracted some attributes that enhance the Direct Hydrocarbon Indicators (DHI) on the surface to perform the cluster analysis and generate a probability map of this anomaly from the seismic clustering.

II. Geophysical settings

A. Seismic Method

The seismic method is the most important among all geophysical techniques regarding investments, expenditures, and the number of professionals involved, mainly in the hydrocarbon industry. It is mostly employed due to its high accuracy, high resolution, versatility, and high penetration in the subsurface [6].

The main idea behind seismic methods consists of controlled generation of seismic waves by a known source to obtain an image of the subsurface by recording the seismic waves travel-time from depth [7]. Once the seismic waves start to propagate in all directions through the subsurface, they suffer reflection, refraction, or diffraction when the wavefront encounters elastic contrasts at the boundaries between rocks. The interaction amongst the seismic wave and high contrasts of elastic properties is presented as a high-amplitude reflector in seismograms.

The velocity of seismic waves is essential in seismic methods. According to Shuck and Lange (2007) [7], this parameter mainly depends on the rock elastic properties and varies with the mineral content, lithology, porosity, pore fluid saturation,

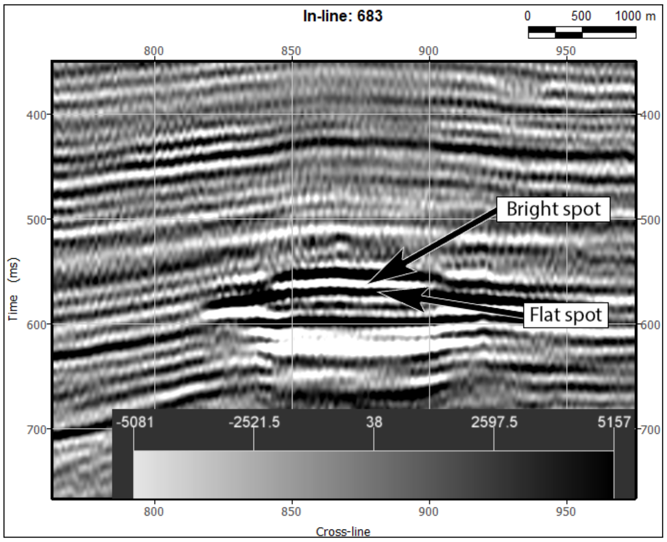

Fig. 1. Direct Hydrocarbon Indicators present in F3 seismic survey in the Netherlands, North Sea. A low amplitude anomaly represents the Bright spot, while a horizontal high amplitude event represents the Flat spot.

and degree of compaction. In ordinary geological situations, velocity also varies as a function of depth due to physical properties contrasts between layers. Horizontal variations are also common due to lithological changes within the individual layers [8].

Imaging complex structures as salt domes and major faults or estimating thickness and bed attitudes become possible tasks through studies of seismic waves arrival times and amplitude, frequency, phase, and wave shape variations [6].

B. Direct Hydrocarbon Indicator

Direct Hydrocarbon Indicators (DHI) are anomalous features in seismic amplitude typically caused by gas and oil in the reservoir. We can highlight two types of DHIs anomalies: Bright Spots and Flat Spots. Bright spots (Fig. 1) are caused by the high acoustic impedance contrast between the seal rock and the porous rock filled with hydrocarbon. Nevertheless, if the reservoir is thick enough, a high amplitude event can occur at the bottom of the bright spot - called a flat spot due to the oil-water contact [9], [10].

C. Seismic Horizon

According to Aykroyd and Hamed (2014) [11], seismic horizons can be defined as a three-dimensional surface originating from a strong reflecting interface and indicates a stratigraphic boundary between two regions with distinct elastic properties. As stated by Wu and Hale (2016) [12], the determination of the seismic horizons is essential obtained from seismic interpretation. Allied to mapped fault systems and unconformities, these surfaces provide subsurface structural maps and a chronostratigraphic framework of the survey region.

D. Seismic Attributes

The estimation of rock physical properties through the acquisition and processing of seismic data and its vertical and lateral variations in the time and space domain constitutes the basis for seismic interpretation [13]. In this sense, seismic attributes are quantities derived from seismic data based on recorded time, amplitude, frequency, and attenuation, that support seismic interpretation [14].

Among other objectives, calculating these attributes allows the interpreter to increase the Signal-Noise ratio, detect discontinuities, enhance the continuity of seismic reflectors and emphasize direct hydrocarbon indicators (DHI). In this sense, Machine Learning algorithms, especially unsupervised techniques, can be helpful to reveal important features related to oil and gas reservoirs [15]. Hence, seismic attributes are excellent tools for seismic data analysis that reduce ambiguity, supporting the interpreters to correct seismic events with real geological features.

III. Experimental Setup

A. K-means

K-means is a clustering algorithm that uses the distance between the clusters samples and centroids to partition data. It starts by choosing K representative points as the initial centroids. Each point is then assigned to the closest centroid $\mathbf{c_k}$ based on a particular proximity measure chosen [16]. Subsequently, the position of the centroids is updated to the center of samples on their domain.

Let $\mathbf{X}$ be a dataset matrix represented as a matrix (Eq. 1), in which D is the number of attributes, and N is the number of objects:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1D} \\ x_{21} & x_{22} & \cdots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{ND} \end{bmatrix}, \tag{1}$$

as stated by James (2013) [17], the idea behind K-means clustering is that a good clustering is one for which the within-cluster variation (Eq. 2) is as small as possible.

$$Var = \frac{1}{N} \sum_{\substack{n=1 \\ \forall x_n \in K}}^{N_k} \sum_{k=1}^{K} \|\mathbf{x}_n - \mathbf{c_k}\|^2. \tag{2}$$

B. Gaussian Mixture Models

According to Bishop (2006) [18] and Deisenroth (2020) [19], a Gaussian Mixture Model (GMM) is a weighted sum of K Gaussian densities components which can be written as a linear superposition of Gaussian distributions (Eq. 3), so that:

$$p(\mathbf{X}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{X}|\mu_\mathbf{k}, \mathbf{\Sigma_k}), \tag{3}$$

where $\mu_\mathbf{k}$ is the D-dimensional mean of the matrix $\mathbf{X}$, $\mathbf{\Sigma_k}$ is the co-variance matrix and $\pi_\mathbf{k}$ are the Gaussian weights. These parameters can be reached by an iterative process called Expectation-Maximization (EM), whose algorithm's goal is to maximize the following likelihood function (Eq. 4):

$$\mathcal{L} = \log p(\mathbf{X}|\pi, \mu, \mathbf{\Sigma}) = \sum_{n=1}^{N} \log \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x_n}|\mu_\mathbf{k}, \mathbf{\Sigma_k}), \quad (4)$$

and the best parameters can be computed using the gradient method, where every step in the EM algorithm increases the log-likelihood equation [20]. Thus, we can describe each step of the algorithm as follow:

1) The initial parameters π_k, $\mu_\mathbf{k}$, and $\mathbf{\Sigma_k}$ are obtained using the K-means algorithm:
2) (*Loop start*) Evaluate responsibilities r_{nk} of each dataset sample $\mathbf{x_n}$ using current parameters:

$$r_{nk} = \frac{\pi_k \mathcal{N}(\mathbf{x_n}|\mu_\mathbf{k}, \mathbf{\Sigma_k})}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x_n}|\mu_\mathbf{j}, \mathbf{\Sigma_j})} \quad (5)$$

3) Update the parameters: π_k, $\mu_\mathbf{k}$, and $\mathbf{\Sigma_k}$:

$$\pi_k = \frac{1}{N} \sum_{n=1}^{N} r_{nk} = \frac{N_k}{N}, \quad (6)$$

$$\mu_\mathbf{k} = \frac{\sum_{n=1}^{N} r_{nk} \cdot \mathbf{x_n}}{N_k}, \text{ and} \quad (7)$$

$$\mathbf{\Sigma_k} = \frac{1}{N_k} \sum_{n=1}^{N} r_{nk} (\mathbf{x_n} - \mu_\mathbf{k})(\mathbf{x_n} - \mu_\mathbf{k})^T. \quad (8)$$

The responsibility, $\mathbf{r_n} = [r_{n1}, r_{n2}, ..., r_{nK}]^T$, is a normalized probability vector, and it represents the probability that a sample has been generated by the kth mixture component [19].

C. Principal Component Analysis

Principal Component Analysis is an unsupervised algorithm, often used for dimensionality reduction, which performs a linear transformation to project the dataset onto a lower-dimensional subspace [21]. However, due to the number of features and the correlation between them, high-dimensional data is often overcomplete, and many dimensions are redundant [19]. Therefore, conforming to James (2013) [17], PCA provides an optimal representation of the original data whilst maintain as most of the information as possible using a smaller number of variables, the principal components, that collectively explain most of the variability in the original set.

Let us consider a dataset $\mathbf{X}$ (Eq. 1), with mean 0 that posses the data covariance matrix (Eq. 9):

$$\mathbf{S}_X = \frac{1}{n-1} \mathbf{X}\mathbf{X}^T. \quad (9)$$

We can obtain a matrix $\mathbf{P}$ able to represent the dataset $\mathbf{X}$ in the principal component domain:

$$\mathbf{PX} = \mathbf{Y}, \quad (10)$$

where $\mathbf{Y} = \begin{bmatrix}\mathbf{y}_1, & \dots, & \mathbf{y}_D\end{bmatrix}^T$ contain the principal components and the matrix $\mathbf{P} = \begin{bmatrix}\mathbf{p}_1, & \dots, & \mathbf{p}_D\end{bmatrix}^T$ represents the eigenvectors of the covariance matrix $\mathbf{S_x}$. The matrix $\mathbf{P}$ is responsible for rotating and stretching the original attributes. Furthermore, the variance of the data projected onto the principal components is equals to the eigenvalues associated with the basis vector [19].

D. Validation

The criterion adopted to choose the best number of clusters using a statistical validation was the Bayesian Information Criterion (BIC) combined with the Elbow Method. A brief discussion of them is presented as follows.

1) Intra-cluster Variance: The intra-cluster variance, also known as within-cluster variation, is the distance between cluster members rather than between two clusters. Applying this concept to all the centroids using the summation of each one, we can measure the clustering dispersion (Eq.2). Therefore, high values indicate that the samples are not close enough to the centroids and, consequently, the groups are not well defined.

2) Bayesian Information Criterion (BIC): The Bayesian Information Criterion is a probabilistic statistical measure to evaluate the best model concerning the number of mixture components. The BIC tends to define a small value for a model with a low-test error, and so generally we select the model that has the lowest BIC value [17]. According to Charu (2014) [16], the chosen kth value is the one that minimizes the BIC function:

$$BIC = \frac{1}{N} \times \log\left(\frac{N^k}{\mathcal{L}^2}\right) \quad (11)$$

where N is the number of samples, $\mathcal{L}$ is the likelihood function (Eq. 4), and k is the number of clusters.

3) Elbow Method: This technique is adopted to determine the optimal number of groups in clustering analysis. The elbow method plots the value of a cost function produced by different values of K, and the intra-cluster variance (Eq. 2) for K-means cluster analysis or BIC function (Eq. 11) to evaluate the cluster results obtained by the Gaussian Mixture Models algorithm. As the value of K rises, the distortion decreases because the clusters will divide their samples with the new ones, generating smaller and well-defined groups. However, there is a threshold where the improvement in distortion declines the most, and the data should not be divided into further clusters [21]. When the optimal value is achieved, the cost reaches a plateau, and the cost slowly decreases. Therefore, indicating that adding more clusters will not improve the model.

E. Dataset

The dataset used to extract the seismic attributes in the present work, which will be the input of the cluster analysis, is the F3 Block located at the Offshore region of the North Sea in the Netherlands [22] (Fig. 2). This dataset consists of 386 km^2 of 3D time migrated seismic, with 651 inlines, and 951 crosslines sections in a time range of 1,848 ms with a sampling rate of 4 ms. Furthermore, four well logs are available with petrophysical information - Gamma Ray, Vagarosity, Density, and Resistivity - and they were used to interpret the cluster analysis. The FS8 seismic horizon alongside a cross-line, an in-line, and the four wells are shown in Fig. 3.

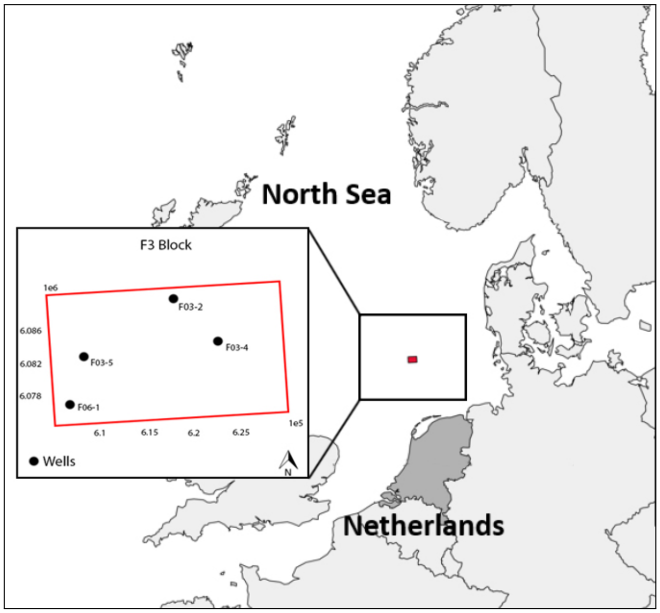

Fig. 2. F3 Seismic Survey delimited by the red rectangle, and the well logs location used to interpret the cluster analysis results.

According to Schroot and Schüttenhelm (2003) [9], in the Netherlands North Sea, the geological formations containing the gas are mostly unconsolidated clastic sediments of Miocene. The FS8 seismic surface chosen for the analysis is in a region characterized by plane-parallel reflectors of high amplitude in the context of marine transgression deposited during the Cenozoic.

Concerning the Seismic Attributes, the following features were extracted from the 3D seismic dataset using the OpendTect software from dGB Earth Science to emphasize the DHI anomaly in the cluster analysis. Conforming to Barnes (2016) [23], the attributes employed in the analysis can be described in the following way:

Sweetness: The attribute is meant to identify places that are oil and gas prone. Sweetness is defined as the amplitude divided by the square root of frequency.

Spectral Decomposition: This attribute is generated through filter banks characterized by a center frequency. These attributes are especially useful to distinguish reflections and thin beds on the basis of their tuning response.

Acoustic Impedance: A rock property that quantifies the resistance offered to propagate compressional seismic waves. This attribute is usually associated with lithologies and DHI anomalies.

Coherence: The attribute is often used to reveal breaks caused by faults, diapirs, channels, and other structures. It is relevant to highlight Faults that can compartmentalize pressures and fluids in producing reservoirs [24].

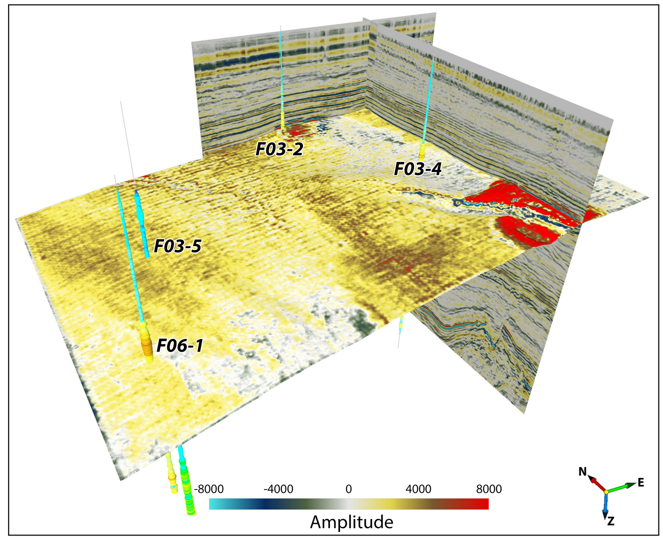

Fig. 3. FS8 Seismic horizon amplitude, in-line section amplitude number 722, cross-line section amplitude number 1065, and the well logs in the original dataset. The DHI anomalies are related to the high amplitudes in this horizon (red color).

Instantaneous Amplitude: Also known as reflection strength, it refers to the magnitude of the trace envelope obtained from the Hilbert transform. Its polarity and phase influence the brightness of a seismic reflection, but Instantaneous Amplitude removes the difference, rendering hydrocarbon indicators more visible.

F. Workflow

The code used to perform the experiments was implemented in Python Language, and the Scikit-learn library [25] was adopted to run the cluster analysis. The workflow employed is illustrated by Fig. 4, and each step is scrutinized as follow:

1) **Seismic Horizon Pick**: The chosen seismic surface was the FS8, from the F3 survey, due to its location concerning the appearance of the direct hydrocarbon indicator.
2) **Seismic Attributes Generation**: the seismic attributes used in the analysis ought to stand out the gas anomaly in this distinct region. The attributes mathematically derived from the 3D dataset were: Acoustic Impedance, Spectral Decomposition 10Hz, Coherence, Instantaneous Amplitude, and Sweetness.
3) **Preprocessing**: the null values and outliers present in each seismic attribute were removed in this step.
4) **Principal Component Analysis**: as mentioned in sec.III-C, the PCA is chiefly applied for dimensionality reduction, but this algorithm also provides a noise filtering effect depending on the number of principal components selected. This technique was applied, and the first three principal components were the input data of the clustering algorithm.
5) **Clustering Analysis**: the K-means and Gaussian Mixture Model algorithms were performed, varying the

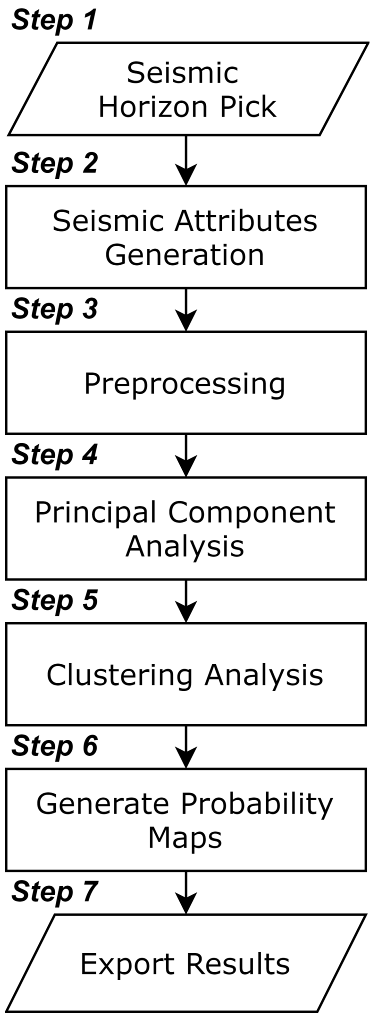

Fig. 4. Workflow used to perform the seismic clustering analysis in the FS8 seismic surface.

number of centroids and mixture components, respectively, from 2 to 20.

6) **Probability Map Generation**: according to sec.III-B, the GMM algorithm generates a normalized probability vector for each kth mixture component vector. Thus, the results of the seismic maps, generated using the GMM labels, were scrutinized, and the nth label corresponding to the gas anomaly was recognized. Hereafter, the probability map was produced using the values from the responsibility of the cluster correspondent to the DHI anomaly.
7) **Export Results:** finally, the results were displayed on maps to be analyzed using the OpendTect software.

IV. Results

As described in the workflow (subsection III-F), we performed several experiments using different values of the parameter K of the K-means and Gaussian Mixture Models technique to analyze the effects of the number of clusters' increases in the direct hydrocarbon indicator anomaly isolation. Although the validation criterion is a helpful technique to select the value of K, geological characteristics should also be considered. The Elbow Method suggested four as the optimal cluster number for the K-means and three for the GMM algorithm. Therefore, our analysis demonstrated that the number of groups equal to four was able to isolate the DHI anomaly and showed a high correlation to the geomorphological setup, which will be the clustering model adopted to present the analysis about the FS8 seismic horizon (Fig. 5 and 6). The K-means results where the label designated for the DHI anomaly represents an extensive area on the seismic surface that is not necessarily related to the DHI (Fig. 5). However, the GMM algorithm was the most accurate on delimiting the interest zone through clustering analysis (Fig.6), according to the geophysical interpreter.

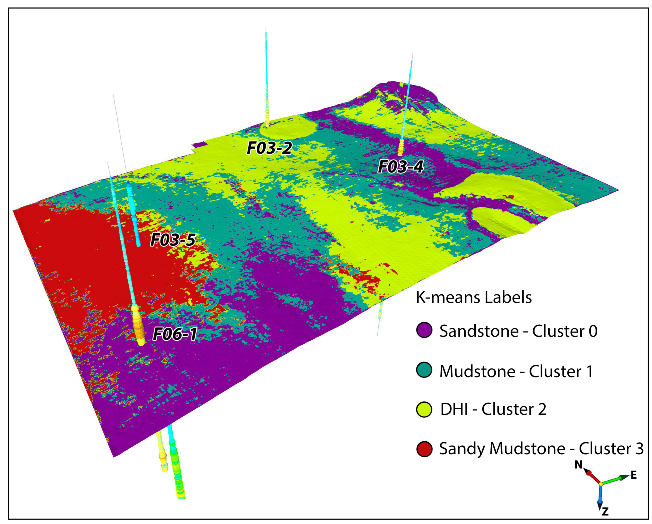

Fig. 5. Cluster analysis map obtained from the K-means algorithm, with $K = 4$, including the interpretation of the groups over the FS8 seismic horizon.

Geophysical well log data support the interpretation of the cluster's distribution unrelated to direct hydrocarbon indicators. The physical properties measurements guided the analysis of the lithology that characterizes each of the other groups. The FS8 seismic horizon crosses the well log F06-1 at a depth of 772 meters, and, in this region, the lithologies present low proportions of clay content. The composite well log indicates that the labeled cluster 0 (purple) is characterized by sandstones. Similarly, the composite well log F03-5 suggests that the lithologies in group 2 (displayed in yellow) over the studied seismic horizon are sandy mudstone. Lastly, the well log F03-4 points out that the dominant lithology present in cluster 1 (exhibited in blue) is composed of mudstones.

Endorsed by the previous analysis [26], [27], the direct hydrocarbon indicators anomalies samples were assigned to cluster 3 (shown in red). Consequently, the values for the cluster 3 responsibility, generated from the GMM density function, can be used to produce a normalized probability map displayed in Fig.7. This map shows the likelihood that each sample has to be a member of group 3, revealing the

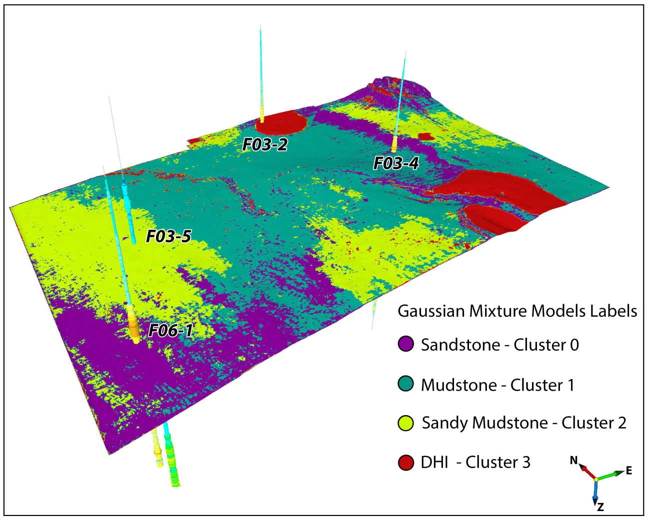

Fig. 6. Cluster analysis map obtained from the Gaussian Mixture Models algorithm, with $K = 4$, including the interpretation of the groups over the FS8 seismic horizon.

regions that were classified already as anomalies linked with hydrocarbons and the domain with an intermediary to low probability to be a DHI anomaly domain. In addition, the algorithm also highlighted some areas with high probabilities of DHI occurrence close to the faulted areas.

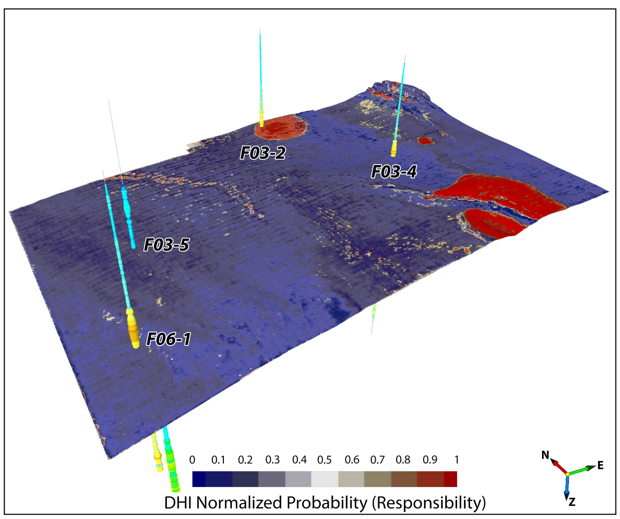

Fig. 7. Direct Hydrocarbon Indicator normalized probability - cluster 3 (red) - obtained from the interpreted Gaussian Mixture Models clusters of the FS8 seismic horizon. Note that the warmer colors are associated with a higher probability of the cluster to be a DHI anomaly.

V. Final Remarks

Nowadays, clustering algorithms, such as K-means, have been an extremely useful tool to generate interpretable maps for seismic analysis and hydrocarbon exploration. Nevertheless, unsupervised machine learning algorithms based on the Bayesian approach can provide a probabilistic approach. In this sense, the Gaussian Mixture Models can be employed to determine the uncertainty about clusters related to Direct Hydrocarbon Indicators anomalies in seismic data. The algorithm validation process and the delivered cluster maps indicate the effectiveness of the method. Furthermore, the specialist's knowledge combined with petrophysical measurements and lithological observation derived from well logs contributed to interpreting the Hydrocarbon Indicator anomaly and clusters unrelated to the DHI anomaly. Thus, the methodology proposed achieved the goal of this research, and similar studies can be performed in other sedimentary basins with occurrences of oil and gas.

Acknowledgment

The authors thank the Instituto Nacional de Ciência e Tecnologia de Geofísica do Petróleo (INCT-GP) for the scholarship of John Lemos, and Matheus Barbosa. Alexsandro Cerqueira also thank the FINEP Project - Apoio à Rede de P & P&D em Gás Não Convencional do Brasil (GASBRAS) for supporting this research and for the his scholarship. This study was also financed in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenanção de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and the Fundação de Apoio à Pesquisa do Estado da Bahia (FAPESB), project PIE00005/2016 Edital de Infraestrurura FAPESB 003/2015. Edric Troccoli acknowledges Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for his doctoral scholarship process #141516/2018-6. The authors also thank Grupo de Estratigrafia Teórica e Aplicada da Universidade Federal da Bahia (GETA-UFBA), dGB Earth Science for the seismic dataset and OpendTect software license, and GeoPark E&P Brazil for supporting the research project.

References

[1] T. Mukerji, P. Avseth, G. Mavko, I. Takahashi, and E. F. González, "Statistical rock physics: Combining rock physics, information theory, and geostatistics to reduce uncertainty in seismic reservoir characterization," *The Leading Edge*, vol. 20, no. 3, pp. 313–319, 2001.

[2] M. K. Dubois, G. C. Bohling, and S. Chakrabarti, "Comparison of four approaches to a rock facies classification problem," *Computers & Geosciences*, vol. 33, no. 5, pp. 599–617, 2007.

[3] A. G. Cerqueira, O. A. L. de Lima, and R. A. Rios, "A nonparametric approach using clustering analysis to estimate shaliness in shaly-sand formations," *Journal of Applied Geophysics*, vol. 164, pp. 11–18, 2019.

[4] A. E. Barnes and K. J. Laughlin, "Investigation of methods for unsupervised classification of seismic data," in *SEG Technical Program Expanded Abstracts 2002*. Society of Exploration Geophysicists, 2002, pp. 2221–2224.

[5] T. Zhao, V. Jayaram, A. Roy, and K. J. Marfurt, "A comparison of classification techniques for seismic facies recognition," *Interpretation*, vol. 3, no. 4, pp. SAE29–SAE58, Nov. 2015. [Online]. Available: https://doi.org/10.1190/int-2015-0044.1

[6] W. M. Telford, L. P. Geldart, and R. E. Sheriff, *Applied geophysics*. Cambridge university press, 1990.

[7] A. Schuck and G. Lange, *Environmental Geology*. Springer, 2007, vol. 1.

[8] P. Kearey, M. Brooks, and I. Hill, *An introduction to geophysical exploration*. John Wiley & Sons, 2002, vol. 4.

[9] B. Schroot and R. Schüttenhelm, "Expressions of shallow gas in the netherlands north sea," *Netherlands Journal of Geosciences - Geologie en Mijnbouw*, vol. 82, no. 1, p. 91–105, 2003.

[10] A. L. R. Rosa, *Análise do Sinal Sísmico*. Sociedade Brasileira de Geofísica, 2010, vol. 1.

[11] R. G. Aykroyd and F. M. O. Hamed, "Horizon detection in seismic data: An application of linked feature detection from multiple time series," *Advances in Statistics*, vol. 2014, 2014.
[12] X. Wu and D. Hale, "Automatically interpreting all faults, unconformities, and horizons from 3d seismic images," *Interpretation*, vol. 4, no. 2, pp. T227–T237, 2016.
[13] N. C. Nanda, *Seismic data interpretation and evaluation for hydrocarbon exploration and production: A practitioner's guide*. Springer, 2016.
[14] R. E. Sheriff, *Encyclopedic dictionary of applied geophysics*. Society of exploration geophysicists, 2002.
[15] L. Azevedo, L. M. Pinheiro, A. Kaschaka, and H. Abbassi, "Caracterização de reservatórios de hidrocarbonetos utilizando atributos sísmicos," 2010.
[16] C. Aggarwal, *Data clustering : algorithms and applications*. Boca Raton, FL: CRC Press, 2014.
[17] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013, vol. 112.
[18] C. M. Bishop, "Pattern recognition and machine learning (information science and statistics)," 2007.
[19] M. P. Deisenroth, A. A. Faisal, and C. S. Ong, *Mathematics for machine learning*. Cambridge University Press, 2020.
[20] R. M. Neal and R. S. Zemel, *Unsupervised Learning: Foundations of Neural Computation*. Massachusetts Institute of Technology (MIT), 1999.
[21] P. Dangeti, *Statistics for machine learning*. Packt Publishing Ltd, 2017.
[22] dGB Earth Sciences, "The netherlands offshore, the north sea, f3 block-complete," 1987.
[23] A. E. Barnes, *Handbook of poststack seismic attributes*. Society of Exploration Geophysicists, 2016.
[24] S. Jolley, H. Dijk, J. Lamens, Q. Fisher, T. Manzocchi, H. Eikmans, and Y. Huang, "Faulting and fault sealing in production simulation models: Brent province, northern north sea," *Petroleum Geoscience - PETROL GEOSCI*, vol. 13, pp. 321–340, 11 2007.
[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
[26] A. S. Bahri, P. R. Aripin, A. Banuboro, and A. A. Robi, "Analysis of seismic attributes and band-limited inversion for re-determining the hydrocarbon prospect zone in data f3 netherland," *IPTEK Journal of Proceedings Series*, vol. 3, no. 6, 2017.
[27] Q. Qiang Guo, N. Nayyer Islam, and W. D. Pennington, "Tuning, avo, and flatspot effects in north sea block f3," in *2014 SEG Annual Meeting*. OnePetro, 2014.
[28] R. M. Silva, L. Baroni, R. S. Ferreira, D. Civitarese, D. Szwarcman, and E. V. Brazil, "Netherlands dataset: A new public dataset for machine learning in seismic interpretation," 2019.