Cluster Analysis to Detect Reflectors associates with Salt Dome Structures

André Luiz Souza da Silva[1,2], Vinicius Carneiro[1,2], and Alexsandro G. Cerqueira[1,2].
[1]Universidade Federal da Bahia (UFBA) / [2]Grupo de Estudo e Aplicação de Inteligência Artificial em Geofísica (GAIA·UFBA)

This paper was prepared for presentation during the 18th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 16-19 October 2023.

ABSTRACT

This study introduces a methodology developed to automatically detect and delimit reflectors associated with salt bodies in 3D seismic data using the unsupervised machine learning algorithms K-means and DBSCAN (Density-based spatial clustering of applications with noise). The data used in this study were 3D seismic data from the F3 block in the Central Graben Basin of the North Sea. The seismic attributes used to perform the experiments were Variance, Entropy, Instantaneous Amplitude, Dissimilarity, Homogeneity, and Contrast. The K-means algorithm was able to classify most reflectors close to the salt body, distinguishing it from other geological structures. However, it produced groups that clearly could not be associated with the salt. Therefore, using DBSCAN as a density filter algorithm was crucial to differentiate salt from locations categorized as chaotic seismic facies and noisy data.

INTRODUCTION

Salt domes are crucial geological structures with implications for robust characterization and modeling of subsurface hydrocarbon reservoirs. For example, a belt of salt domes is found in the Gulf of Mexico Sedimentary Basin, which allows the formation of large stratigraphic traps /that make the region one of the most abundant in hydrocarbon reservoirs (Di et al., 2018a). In 3D seismic, the presence of a salt body is easily recognized due to its weak and chaotic reflection interior patterns as well as in most cases a well-maker reflection in the top due to strong impedance contrast with the other rocks. However, the detection and interpretation of saline bodies is still challenging, especially in exploration areas where several of those structures developed at different times.

In this research, an unsupervised machine learning technique was used to perform cluster analysis of seismic attribute data in the domain of principal components, obtained through Principal Component Analysis (PCA), which the main objective is to identify layers associated with the Salt body to delimit them in a 3D seismic data. The technique used was the K-means algorithm that tries to group the data under the cost of minimizing the intra-group variance iteratively (James et al., 2013). First, we ran the PCA, and the last three principal components were used to perform the cluster analysis using the K-means algorithm – a practice that is not common, but it proved efficient in this work. Next, the DBSCAN algorithm was employed as a density filtering step, which evaluates the spatial density of the samples in the group, interpreted as possible reflectors associated with salt, to differentiate the salt dome from structures of no interest with characteristics similar to it.

GEOLOGICAL SETTINGS AND DATASET

The research area is the Central Graben Basin in the northeast portion of the Dutch Deep Sea. It has an area of approximately 25,000 km² and is considered the southern member of the rifting system that geologically contextualizes the North Sea, reaching depths of up to 9 km (Wijker, 2014). It is limited to the southwest by the top

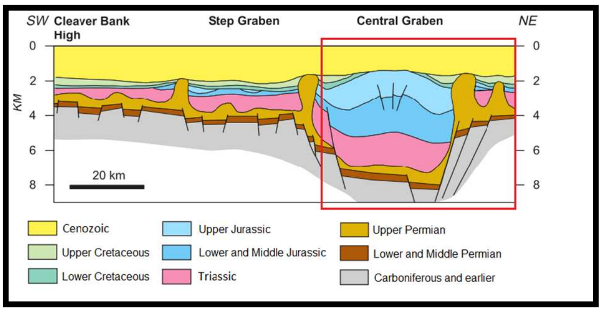

Figure 1: Regional geological model featuring the Central Graben basin highlighted (red polygon). Modified from de Jager (2007).

of Cleaver Bank and to the northeast by the top of Schill Grund (Rosendaal et al., 2014), as seen in Figure 1.

This basin is characterized by rifting mainly occurring in the Mesozoic with post-rift Cenozoic phase. Initiated in the Triassic, the rifting system stabilized between the Jurassic and the Lower Cretaceous, with the Kimeridgian extensional tectonic phases related to the opening of the Atlantic Ocean. From the Upper Cretaceous to the present day, the rift phase was followed by the post-rift phase of the sag type, mainly characterized by tectonic quiescence and basin subsidence, except for the presence of some tectonic pulses that occurred from the Upper Cretaceous to the Tertiary. In this depositional stage, thick layers accumulated that composed depositional mega-sequences (Schroot e Schüttenhelm, 2003).

The data used in this work are part of a public 3D dataset called the Netherlands Offshore F3 Block, located in the

North Sea, in the deep waters of the Dutch territory (Figure 2). This seismic cube corresponds to an area of approximately 386km² of data stacked and migrated in time, containing 651 inlines and 951 crosslines (Silva et al., 2019). In time, the data spans up to 1848 milliseconds and is sampled in 4 ms - 462 samples.

Although it is possible to use the entire seismic cube together with its respective seismic attributes, in this work, clusters were analyzed only in the region of interest that could encompass the most Salt bodies in the area, as shown in Figure **3**.

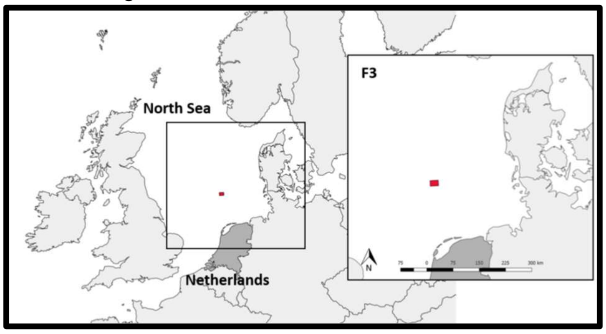

Figure 2: Regional geological model featuring the Central Graben basin highlighted (red polygon). Modified from Jager (2007).

Seismic Attribute Selection

Seismic attributes are used as input data for different cluster analysis algorithms. According to Barnes and Laughlin (2002), the accuracy of specific pattern recognition results is closely related to the choice of the set of attributes to be used. Therefore, if computational cost is disregarded, there are no limits to the number of seismic attributes that can be used for clustering analysis. However, exaggerating the choice of this criterion can be detrimental to the result, reducing the interpretability of the method since some attributes present very similar results, as Barnes (2016) discussed.

The selection of attributes is based on the association between the prior knowledge of the expected response of each seismic attribute in the enhancement of salt bodies and the recommendation of specialized literature. The following seismic attributes were used for the experiments: seismic amplitude, instantaneous amplitude, contrast, dissimilarity, entropy, homogeneity, variance, and coherence, which are presented in the Figure 4. It is crucial to highlight that these attributes were chosen based on the

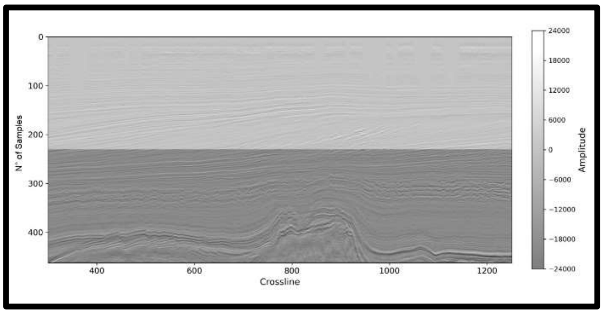

Figure 3: Seismic section showing the region chosen for the clustering analysis.

literature survey conducted by Barnes (2016) and Di et al. (2018b).

METHOD

The Principal Components Analysis algorithm was executed first to evaluate the components most related to the salt bodies -- the components used in the cluster analysis. Once the principal components were defined, the K-means algorithm was executed, and the label related to the salt bodies was selected to be spatial filtered through the DBSCAN to enhance the reflector associated with the salt.

Principal Component Analysis

Let us consider the dataset, where each attribute has a mean equal to zero. The data covariance matrix can be calculated through Eq. 1:

$$\mathbf{S}_x = \frac{1}{n-1}\mathbf{X}\mathbf{X}^{\mathrm{T}}. \qquad \textit{Eq. 1}$$

We can obtain a matrix $\mathbf{P}$ capable of representing X in the principal component domain:

$$\mathbf{PX} = \mathbf{Y}. \qquad \textit{Eq. 2}$$

where $\mathbf{Y} = [\boldsymbol{y_1},\ldots,\boldsymbol{y_D}]^{\boldsymbol{T}}$ contains the principal components, and the matrix $\mathbf{P} = [\boldsymbol{p_1},\ldots,\boldsymbol{p_D}]^{\boldsymbol{T}}$ represents the eigenvectors of the covariance matrix $\mathbf{S_x}$. The matrix $\mathbf{P}$ is responsible for rotating and translating the original attribute set. In addition, the variation of the data projected onto the principal components is equal to the eigenvalues associated with the base vector (James et al., 2013). To obtain this result, it is possible to decompose the covariance matrix into a set of matrices in which at least one is diagonal. This procedure is performed through Singular Value Decomposition (SVD).

We can evaluate the principal components to perform a dimensionality reduction, which means removing information contained in the less important principal components, keeping only the components with the highest variances. The fraction of variance removed is a good estimate of how much information is discarded by the PCA process (Lemos et al., 2021 and Troccoli et al, 2022).

K-means

The K-means algorithm is usually used to split a dataset into groups based on its approximation with a centroid. Let us consider, again, the dataset $\mathbf{X}$. As stated by James et al (2013), the K-means is a clustering algorithm based on the idea that good clusters are that the intra-cluster variation (Eq. 3) is as small as possible:

$$Var = \frac{1}{N}\sum_{\substack{n=1\\ \forall x_n \in K}}^{N_k}\sum_{k=1}^{K}\|\mathbf{x_n} - \mathbf{c_k}\|. \qquad \textit{3}$$

In this work were preformed cluster analysis with K equals to 2 and 3.

Density-based spatial clustering of applications with noise (DBSCAN)

The objective of density filtering was to investigate and delimit the spatial distribution of the group representing salt using the DBSCAN algorithm. This was done to verify the presence of noisy points and to separate data groups unrelated to the salt body. This approach is based on the idea that true salt anomalies are characterized by a dense cloud of points, i.e., many samples close to each other.

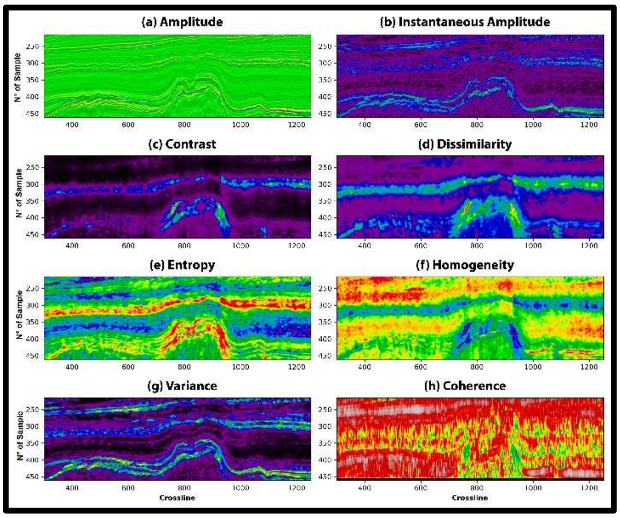

Figure 4: Visual analysis was performed on the seismic attributes selected on inline 400, namely (a) amplitude, (b) seismic amplitude, (c) contrast, (d) dissimilarity, (e) entropy, (f) homogeneity, (g) variance, and (h) coherence.

For this filtering, the DBSCAN algorithm used the samples labeled as possible salt bodies by the K-means algorithm as input, and the selected variables were the inline number, crossline number, and normalized sample number. A series of tests were conducted to optimize the fundamental parameters of DBSCAN, such as the minimum number of points ($MinPts$) and the neighborhood radius (ϵ). The tests showed that using $\epsilon = 0.018$ and $MinPts = 150$ for the dataset would be ideal.

As DBSCAN is a density-based algorithm, it automatically determines the number of existing subgroups, the number of samples each has, and how many samples were defined as noise. The identification of the salt-related group in the F3 Block was made by organizing the samples in descending order of number and observing that the second most populous cluster was related to the salt dome. It allowed the separation of the group related to the salt body from other seismic facies labeled by the K-means algorithm, similar to the analysis carried out by Barbosa (2022), to detect direct indicators of hydrocarbons.

Workflow

As shown in Figure 5, this work was divided into several stages. Starting the workflow, the .SEGY files containing seismic attributes were loaded. Next, the data interval of interest was selected to optimize the processing. Then, the most relevant seismic attributes were chosen for this work's objective, and the principal components were analyzed and determined. After this step, the K-means and DBSCAN algorithms were executed to obtain the geobodies, which constitute the final result of the analysis.

RESULTS

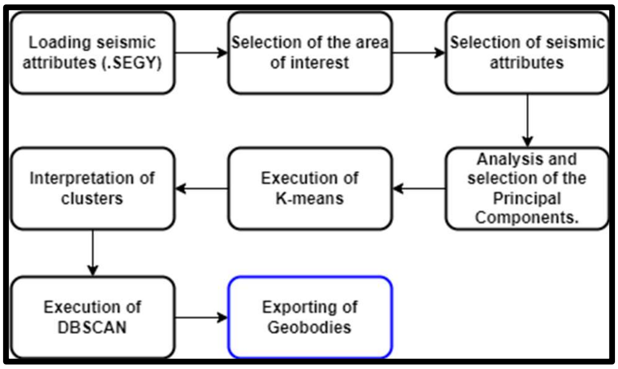

Figure 5: Workflow diagram used to perform the clustering analysis.

Principals Components Analysis

After the attribute selection, the Principal Component Analysis (PCA) algorithm was used to analyze the dataset on a new basis (through graphical investigation) and check if it was necessary to reduce the dimensionality of the problem before performing the clustering analysis with the K-means algorithm. Figure 6 illustrates the accumulated variance of the principal components, and Figures 7 and 8 show the five main components obtained by the PCA algorithm for inline 400 and crossline 760 of block F3. Note that in Figure 6, the first two principal components carry about 90.97% of the dataset's information.

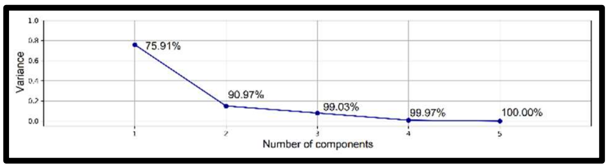

Figure 6: Normalized variance of each principal component and cumulative percent variance.

K-Means

After using the Principal Component Analysis algorithm to select relevant attributes for delimiting the salt body and reducing the dimensionality of the problem, a visual analysis of the clustering results was performed using the K-means algorithm.

As can be seen in Figures 9 and 10, it is notable that, due to the principal components chosen as input for the K-means algorithm, it was able to consistently label a large portion of the interface between the salt body and carbonate, assigning a single class to this region, represented by the color red and labeled 1. However, due to the significant similarity in seismic responses between the salt body, located at depths indicated on average by the sample number equal to 400, and the chaotic seismic facies, found at depths indicated on average by the sample number equal to 300, the K-means algorithm was not able to distinguish these two structures and assigned the same label to both regions.

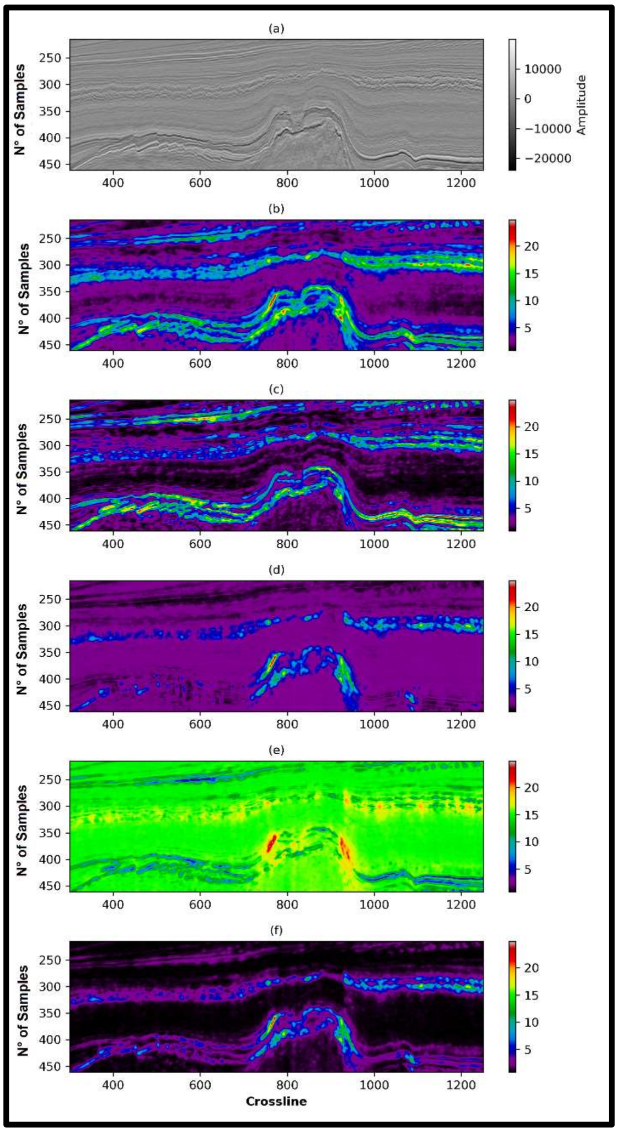

Figure 7: Figure (a) shows the seismic amplitude section of inline 400, and Figures (b), (c), (d), (e), and (f) show the 1st, 2nd, 3rd, 4th, and 5th principal components, respectively, obtained from the PCA algorithm using seismic attributes as input data: contrast, dissimilarity, entropy, homogeneity, and variance.

Performing tests with the number of clusters $K = 3$ (see Figure 9), it is noticeable that even with the increase in the number of groups, the K-means algorithm cannot satisfactorily isolate the chaotic seismic facies from the salt body. Furthermore, when observing the salt dome, which was consistently grouped when the number of clusters $K = 2$, it is possible to perceive a thinning of the lateral continuity of the group, as well as a drastic reduction in the number of samples belonging to the salt group, which is reflected spatially as a group of lower density.

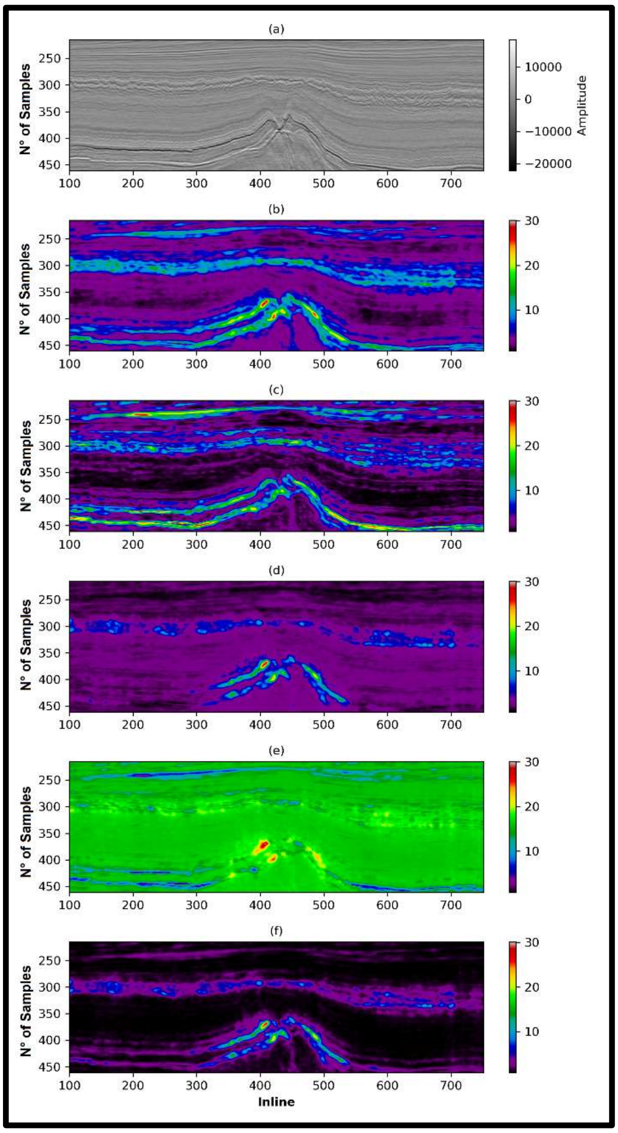

Figure 8: Figure (a) shows the seismic amplitude section of crossline 760, and Figures (b), (c), (d), (e), and (f) show the 1st, 2nd, 3rd, 4th, and 5th principal components, respectively, obtained from the PCA algorithm using seismic attributes as input data: contrast, dissimilarity, entropy, homogeneity, and variance.

Therefore, it can be stated that, with the seismic attributes used, the K-means algorithm was not able to distinguish the salt body from other structures that were not of interest. However, the results generated by the K-means algorithm can be used as input for the DBSCAN algorithm, as with the number of clusters equal to 2 ($K = 2$), K-means densely labeled the region related to the salt dome, presenting a spatial separation between the dome and the chaotic pattern reflectors. Thus, the DBSCAN algorithm was employed to filter the samples related to the salt to achieve the objective.

Density Filtering

The visualization of the samples grouped with the label "salt body" showed that the K-means algorithm presented undesired spatial areas with densities similar to what was assumed to be the salt body, as seen in Figure 10. Therefore, it was decided to use Density-Based Spatial Clustering of Applications with Noise (DBSCAN) as a spatial noise filter, where the input data were the samples with K-means label equal to 1, parameterized with the number of groups $K = 2$, and the input data were the numbers of normalized inlines, crosslines, and timeslices.

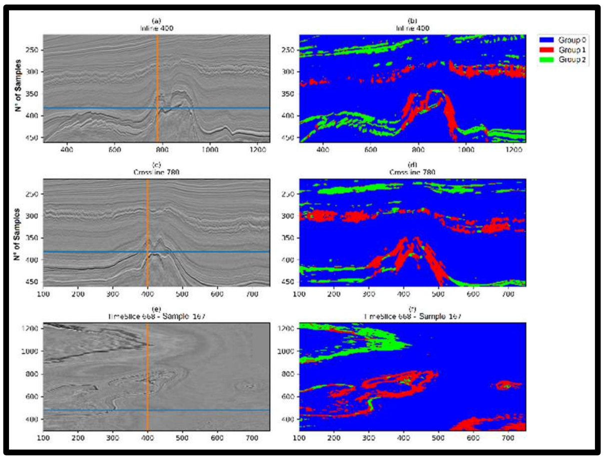

Figure 9: Cluster analysis was performed using principal components 3, 4, and 5 from the PCA algorithm as input data with the K-means algorithm and K=3 clusters. Figures (a), (c), and (e) show seismic amplitude sections of inline 400, crossline 760, and timeslice 668, respectively. Figures (b), (d), and (f) show the results of the clustering analysis in the aforementioned inline, crossline, and timeslice.

Due to the large number of samples related to the salt body contour and the existing computational limitation, it was not possible to use all available data from the clustering analysis obtained by the K-means algorithm to perform this task using the DBSCAN algorithm. Therefore, the input dataset for the algorithm was reduced to include the maximum of the salt body in the region, composed of 400 inlines (range of 100-500) and 470 crosslines (range of 650-1120). The selected inline range was divided into three intervals (100-300, 301-400, and 401-500) so that the DBSCAN algorithm could be run separately in each dataset slice. This division was made to avoid machine stops due to the limited memory available in the graphics processing unit.

Several experiments were performed by varying the parameters that define the neighborhood radius (ε) and the minimum number of points (MinPts) of the algorithm, with the objective of obtaining a result in which the salt body is completely isolated from chaotic seismic facies without loss of interface continuity. The best selected model was parameterized with the maximum distance of each sample $\varepsilon = 0.018$ and the minimum number of samples $MinPts = 150$, which can be seen in inline, crossline, and timeslice sections in Figures 11 and 12.

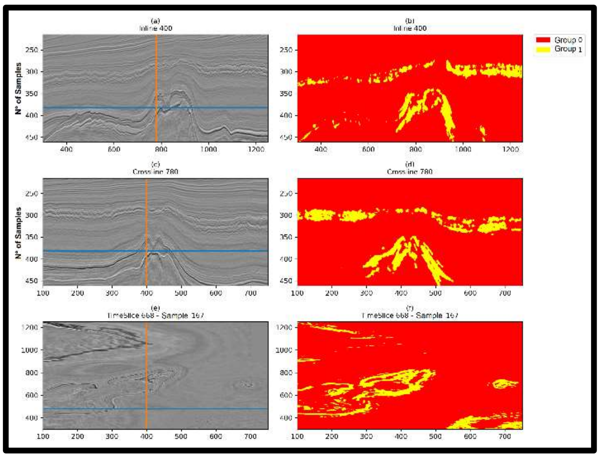

Figure 10: Cluster analysis was performed using principal components 3, 4, and 5 from the PCA algorithm as input data with the K-means algorithm and K=2 clusters. Figures (a), (c), and (e) show seismic amplitude sections of inline 400, crossline 760, and timeslice 668, respectively. Figures (b), (d), and (f) show the results of the clustering analysis in the aforementioned inline, crossline, and timeslice.

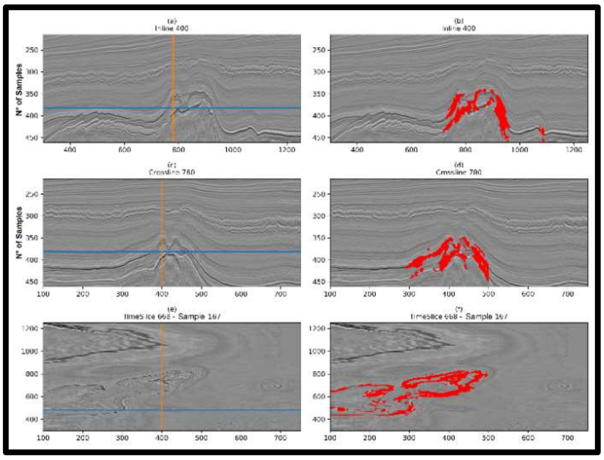

Figure 11: Result of density filtering using class 1 obtained by the K-means algorithm with the number of clusters K=2 from the DBSCAN algorithm with maximum distance between samples ϵ=0.018 and minimum number of samples MinPts=150 as input data. Figures (a), (c), and (e) show the seismic amplitude sections of inline 400, crossline 760, and timeslice 668, respectively, and Figures (b), (d), and (f) show the results of the density filtering applied to the mentioned inline, crossline, and timeslice, respectively.

Interestingly, in Figures 11 and 12, the DBSCAN algorithm was able to isolate the salt body as a different cluster from the portion indicated as chaotic seismic facies, as well as not carrying information related to the carbonate ~~platform~~.

Conclusions

In this work, we employed a methodology to analyze 3D seismic data to identify and delimit salt bodies from associated geological structures. To achieve this, we used the unsupervised machine learning algorithm K-means. After using this algorithm, we applied the Density-Based

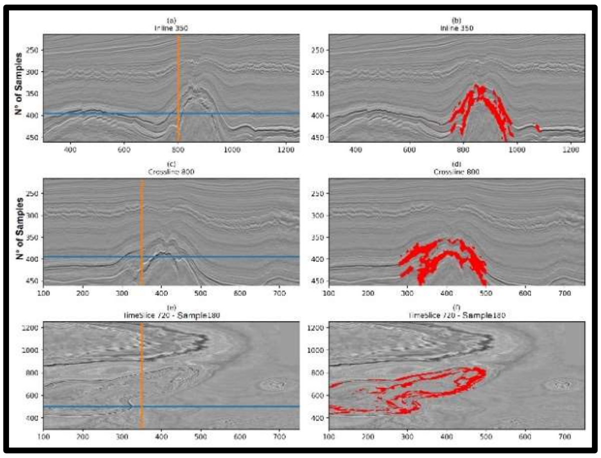

Figure 12: Result of density filtering applied using class 1 obtained from K-means algorithm with the number of clusters K=2, from DBSCAN algorithm with maximum distance between samples ϵ=0.018 and minimum number of samples MinPts=150. Figures (a), (c), and (e) show seismic amplitude sections of inline 350, crossline 800, and timeslice 720, and figures (b), (d), and (f) show the results of density filtering applied to the inline, crossline, and timeslice mentioned.

Spatial Clustering of Applications with Noise (DBSCAN) algorithm to the label related to the salt dome to exclude possible noise and unwanted structures.

For K-means, we performed a series of experiments varying the number of clusters from 2 to 3. We observed that the model that could be optimally combined with DBSCAN had two clusters. However, upon analyzing the results of the seismic cube through group overlaps in inlines, crosslines, and timeslices, we found that the K-means algorithm was not effective in grouping only the salt body, as it assigned another structure (scruff group carbonate layer) to the salt body label and classified several noises with the same label.

Despite this, K-means played an important role in generating a suitable model for using density filtering through DBSCAN. Density filtering separated structures that did not belong to the salt body and labeled dissimilar samples as noise when applied to the group samples interpreted as salt. Thus, the DBSCAN algorithm, with parameters $\epsilon = 0.018$ and $MinPts = 150$, satisfactorily delimited the salt body.

Acknowledgments

The authors would like to thank the INCT-GP for the scientific initiation scholarship provided to the first author during graduation. We also would like to thank the Geoscience Instituto of the UFBA and the Grupo de Estudo de Aplicação de Inteligência Artificial em Geofísica (GAIA-UFBA) for providing the necessary equipment and infrastructure. V.C acknowledges the Fundação de Apoio a Pesquisa do Estado da Bahia (FAPESB) for the PhD scholarship.

References

Barbosa, M. R. S. (2022) Análise de agrupamentos de dados sísmicos 3d para detecção de indicadores diretos de hidrocarbonetos., Trabalho de Graduação em Geofísica, Universidade Federal da Bahia, Salvador, Brasil.

Barnes, A. E. (2016) Handbook of poststack seismic attributes, Society of Exploration Geophysicists.

Barnes, A. E. e Laughlin, K. J. (2002) Investigation of methods for unsupervised classification of seismic data, In: SEG Technical Program Expanded Abstracts 2002, pp. 2221–2224, Society of Exploration Geophysicists.

Di, H.; Shafiq, M. e AlRegib, G. (2018a) Multi-attribute k-means clustering for salt-boundary delineation from three-dimensional seismic data, Geophysical Journal Inter-52 53 national, 215(3):1999–2007.

Di, H.; Shafiq, M. e Alregib, G. (2018b) Multi-attribute k-means clustering for salt-boundary delineation from three-dimensional seismic data, Geophysical Journal International, 215:1999–2007.

de Jager, J. (2007) Geological development, Geology of the Netherlands, pp. 5–26.

James, G.; Witten, D.; Hastie, T. e Tibshirani, R. (2013) An introduction to statistical learning, vol. 112, Springer.

Lemos, J. B., Barbosa, M. R., Troccoli, E. B., and Cerqueira, A. G. Probability Estimation of Direct Hydrocarbon Indicators Using Gaussian Mixture Models. XV Brazilian Congress on Computational Intelligence, 10.21528/CBIC2021-131.

Schroot, B. e Schüttenhelm, R. (2003) Expressions of shallow gas in the netherlands northsea, Netherlands Journal of Geosciences, 82(1):91–105.

Rosendaal, E.; Kaymakci, N.; Wijker, D. e Schroot, B. (2014) Structural development of the dutch central graben-new ideas from recent 3d seismic, In: 76th EAGE Conference and Exhibition 2014, vol. 2014, pp. 1–5, European Association of Geoscientists & Engineers.

Troccoli, E. B., Cerqueira, G. C. Lemos, J. B. and, Holz, M. K-means clustering using principal component analysis to automate label organization in multi-attribute seismic facies analysis. Journal of Applied Geophysics, 10.1016/j.jappgeo.2022.104555.

Wijker, D. (2014) Fault Mapping and Reconstruction of the Structural History of the Dutch. Central Graben, Tese de Doutorado, MSc thesis (EBN), Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.