ISSN: 2447-3359

REVISTA DE GEOCIÊNCIAS DO NORDESTE

Northeast Geosciences Journal

v. 10, nº 2 (2024)

https://doi.org/10.21680/2447-3359.2024v10n2ID35149

Direct Hydrocarbon Indicators Mapping via Joint Cluster Analysis: A Two-Step Approach over 3D Seismic Data

Mapeamento de Indicadores Diretos de Hidrocarbonetos por Análise Conjunta de Agrupamentos: Uma Abordagem em Duas Etapas sobre Dados Sísmicos 3D

Matheus R. S. Barbosa[1]; Vinicius Carneiro[2]; Alexsandro G. Cerqueira[3]

[1] Federal University of Bahia, Group for the Study and Application of Artificial Intelligence in Geophysics (GAIA-UFBA) Email: m.radames09@hotmail.com
ORCID: https://orcid.org/0000-0001-5656-3353

[2] Federal University of Bahia, UFBA Geophysics Graduate Program (PPGEOF-UFBA) Email: vinicius.geophysics@gmail.com
ORCID: https://orcid.org/0000-0001-8509-9254

[3] Federal University of Bahia, Group for the Study and Application of Artificial Intelligence in Geophysics (GAIA-UFBA) Email: alexsandrocerqueira@ufba.br
ORCID: https://orcid.org/0000-0003-3462-9336

Abstract: This paper presents a novel methodology developed in Python to map Direct Hydrocarbon Indicator (DHI) anomalies in 3D seismic data using the unsupervised machine learning algorithms K-Means and Gaussian Mixture Models. The joint cluster analysis consists of implementing the spatial density-based filtering after clustering analysis and investigates the groups interpreted as DHI aiming to distinguish sparsely dense samples and noisy information from samples that are, in fact, areas of interest for hydrocarbon exploration. The experiments were performed on the 3D seismic data F3 Block from Central Graben Basin, Dutch North Sea. The following seismic attributes were extracted to conduct the experiments: Spectral Decomposition of 25 and 45 Hz, Relative Acoustic Impedance, Coherence, Logarithm of Sweetness, and Reflection Strength. The working flowchart took advantage of good artificial intelligence practices to train the models, such as seismic attributes preconditioning, dimensionality reduction via Principal Component Analysis (PCA), and model validation through statistical tests. Despite the initial challenges faced in isolating DHI anomalies through the K-Means algorithm, the two-step approach ultimately succeeded in accurately mapping them.

Keywords: Joint Clustering Analysis; Direct Hydrocarbon Indicators; Spatial Filtering.

Resumo: Este trabalho apresenta uma metodologia original desenvolvida em Python para mapear anomalias de Indicadores Diretos de Hidrocarbonetos (DHI) em dados sísmicos 3D utilizando os algoritmos de aprendizado de máquina não-supervisionados K-médias e Modelo de Misturas Gaussianas. A análise conjunta de agrupamentos consiste em implementar o filtro espacial baseado em densidade após o agrupamento das amostras e investiga os grupos interpretados como DHI com o objetivo de distinguir grupos de amostras com densidade esparsa e informações ruidosas das amostras que são, de fato, áreas de interesse para a exploração de hidrocarbonetos. Os experimentos foram realizados no dado sísmico F3 Block, da Bacia do Graben Central, Mar do Norte holandês. Os seguintes atributos sísmicos foram obtidos: Decomposição Espectral de 25 e 45 Hz, Impedância Acústica Relativa, Coerência, Logaritmo do *Sweetness* e Amplitude Instantânea. O fluxograma de trabalho faz uso das boas práticas na inteligência artificial para treinar os modelos, como o pré-condicionamento dos atributos sísmicos, redução de dimensionalidade através da Análise de Componentes Principais (PCA) e a validação do modelo por meio de testes estatísticos. Apesar dos desafios iniciais encontrados ao tentar isolar as anomalias de DHI através do algoritmo K-Médias, a abordagem em duas etapas obteve sucesso ao mapeá-las com precisão.

Palavras-chave: Análise Conjunta de Agrupamentos; Indicadores Diretos de Hidrocarbonetos; Filtragem Espacial.

Received: 22/01/2024; Accepted: 12/08/2024; Published: 30/09/2024.

1. Introduction

Direct Hydrocarbon Indicator (DHI) anomalies are usually caused by changes in the elastic properties of rocks. They are commonly associated with the saturation of reservoirs in gas or oil (NANDA, 2012). Their interpretation significantly impacts exploratory risk assessment and well allocation for drilling, making it essential to identify economically viable reservoirs (FORREST *et al.*, 2010).

According to Hilterman (2001), advancements in methodologies for detecting and validating DHI anomalies have been notable since the 1970s, particularly through AVO (Amplitude vs Offset) analysis and improvements in seismic data acquisition and processing technologies. Another significant factor driving this progress is the application of seismic attributes to evaluate potential anomalous zones. Although effective for seismic interpretation, correlating a large number of seismic attributes simultaneously can be challenging. Regarding this scenario, in recent years, the application of machine learning techniques to this and other geophysical problems has seen steady growth (BARBOSA *et al.*, 2022; CERQUEIRA *et al.*, 2019; TROCCOLI et al., 2022; ZHAO *et al.*, 2016).

Machine learning algorithms are techniques used to extract information from datasets, automate different activities, and identify patterns of interest that may be imperceptible to human analysis (MITCHELL, 1997). These algorithms use different methodologies to learn iteratively from the dataset and adapt to produce reliable and reproducible results. Its popularity and efficiency quickly presented intelligent algorithms as a powerful tool for the study of various problems in the field of geophysics, such as seismic processing (MA & LUO, 2018; TSAI *et al.*, 2018), seismic interpretation (OLIVEIRA *et al*., 2023; BÖNKE *et al*., 2024), well logging (WANG *et al.*, 2023; CORDEIRO *et al.*, 2023), seismic imaging (HUANG & NOWACK, 2020), earthquake detection (YU & MA, 2021), and quantitative interpretation (MENG *et al.*, 2021; LI *et al.*, 2023).

Pattern recognition activity is an important step for interpreting seismic data related to structural characterization and understanding the tectonostratigraphic evolution of a depositional basin, or the aspects of a hydrocarbon reservoir. Its primary purpose is to segment seismic data according to some similarity. Barnes & Laughlin (2002) conducted a comparative study on K-Means, Hierarchical Agglomeration, and Self-organizing Maps (SOM) algorithms, attesting to their good performance in analyzing 3D seismic volumes in terms of accuracy, similarity between techniques, and label ordering. Roden & Chen (2017) incorporate a machine learning workflow where principal component analysis (PCA) and self-organizing maps (SOM) analyze combinations of seismic attributes for meaningful patterns that correspond to direct hydrocarbon indicators.

The Central Graben Basin (Figure 1a) is an area with a complex geological evolution (BOUROULLEC *et al*., 2018; MAUNDE & ALVES, 2022) and a long exploratory process. From the 60s, through the drilling of the pioneering well, exploration activities began in this region (LARMINIE, 1987). Since then, with the increase in the amount of data acquisition campaigns, the exploratory potential of the basin has been proven with noteworthy discoveries of large recoverable reserves of oil and, mainly, gas. One of those scenarios is formed when those fluids are trapped in Cenozoic sandstone reservoirs, generating well-marked seismic amplitude anomalies (DE BRUIN *et al*., 2022).

The primary objective of the paper is to present a methodology to identify anomalies indicative of DHI within 3D post-stack seismic data. Besides that, we established a comparison between two unsupervised shallow learning algorithms implemented in Python. The first among these techniques is the K-Means algorithm, which seeks to segment data under the cost of minimizing intracluster variance iteratively (JAMES *et al.*, 2013). Then, the Gaussian Mixture Model (GMM) was used. It is considered a simple linear superposition of Gaussian components that classifies probability density models more informatively than the classification made by a single normal distribution (BISHOP, 2006). A subsequent filtering step is introduced to this second technique, employing spatial density evaluation. This assesses the spatial density of cluster samples identified as DHI, aiming to differentiate relevant anomalies from smaller volumes labeled as such but deemed insignificant for potential hydrocarbon accumulation zones. A similar approach was implemented by Jiang (2017), in which the DBSCAN (Density-based Spatial Clustering in Applications with Noise) algorithm was applied to post-process supervised segmented seismic images to identify fluvial channels and faults through convolutional neural networks (CNNs). Its use reinforces the idea of continuity within the detected structures (JIANG, 2017). We conducted a multi-attribute clustering analysis of seismic attribute data in the domain of principal components, in which an ordinary marine seismic survey from offshore Netherlands was used in the experiments. They are part of the 3D open data set known as F3 Block, located in the Dutch North Sea, at the Central Graben Basin (Figure 1a). The incorporation of this two-step process, featuring an additional spatial density-based filter, is designed to enhance mapping results when compared to the conventional approach of single unsupervised clustering. By focusing on unsupervised learning techniques, we seek to offer a more efficient and potentially faster approach to extracting insights and aiding in the interpretation of hydrocarbon

indicators. This methodology provides a streamlined workflow, which could be particularly advantageous in scenarios where well data is scarce or the traditional labeling process is time-consuming.

Furthermore, our approach emphasizes the versatility and adaptability of artificial intelligence in geophysical studies. By introducing these unsupervised methods, we demonstrate that it is possible to achieve reliable results without heavily relying on conventional data preparation steps. This could open new avenues for exploration and analysis, encouraging further innovation and experimentation within the field.

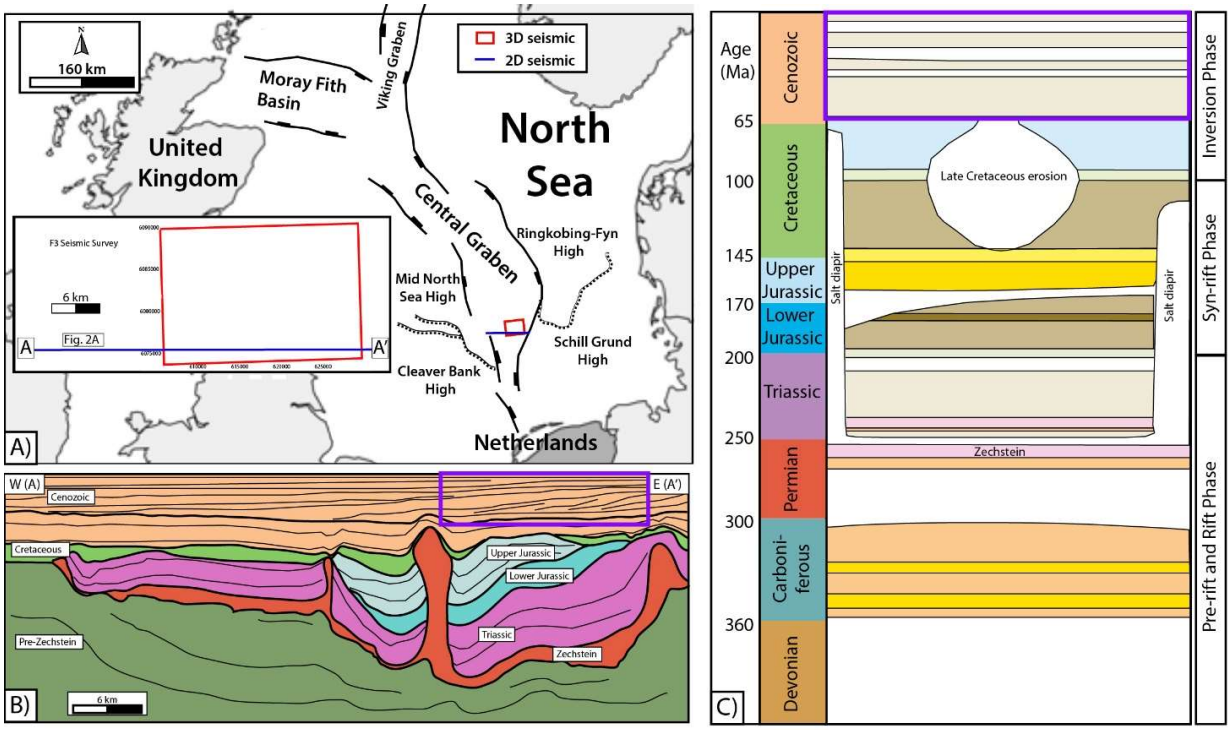

Figura 1 – (a) Location of the Central Graben Basin, structural highs, and dataset used in this study. (b) Regional geological section from the Dutch Central Graben based on 2D seismic after Rosendall et al. (2014). The purple polygon delimits the study area. (c) Simplified stratigraphic chart from the Dutch Central Graben (modified from Jakobsen et al., 2020). Note the sedimentary hiatus in the borders due to salt diapirs. The purple polygon indicates the temporal interval of the strata studies herein.
Source: Authors (2024).

2. Geological overview and dataset

The Central Graben Basin (Figure 1a) is located in the northeastern offshore portion of the Netherlands, has an area of approximately 25,000km², and is considered the southern member of the rifting system that contextualizes the North Sea geologically, reaching maximum depths of up to 9 km (WIJKER, 2014). It is bounded by several structural highs such as the Cleaver Bank High in southwest, Mid North Sea High in the west, Ringkobing-Fyn High in the east and the Schill Grund High in the southeast (ROSENDAAL *et al.*, 2014).

The tectonostratigraphic evolution of the Central Graben Basin (Figure 1b and 1c) is complex, and it has several phases including rifting, intense halokinesis, and tectonic inversion (BOUROULLEC *et al.*, 2018; MAUNDE & ALVES, 2022). It is dominated by rifting that occurred mainly in the Mesozoic with a Cenozoic post-rift phase.

Pre-Zechstein Group comprises sediments from the Carboniferous deposited under a lacustrine environment and terrestrial sandstone from the Permian. In the Late Permian cyclic marine evaporite deposition occurred forming the Zechstein group (MÜLLER *et al.*, 2022), which is responsible for the development of salt diapirs and the halokinesis during the Jurassic and Cretaceous (Figure 1b and 1c).

Initiated in the Triassic, the rifting system stabilized between the Jurassic and the Lower Cretaceous with the tectonic phases extensional Kimmeridgian, related to the opening of the Atlantic Ocean. From the Late Cretaceous to the present day, the rift phase was followed by the sag-type post-rift phase, mainly characterized by tectonic quiescence and subsidence of the basin, except for the presence of some tectonic pulses that occurred from the Late Cretaceous to the Cenozoic, which generated several faults associated to tectonic inversion (MAUNDE & ALVES, 2022). In the Cenozoic, prograding deltas and slope systems developed during periods of sea-level fall forming sand-prone successions that accumulated gas which is this paper's main object of study.

To study this interval, the seismic data F3 Block was used. It corresponds to a 3D streamer survey of approximately 386km² stacked time migrated data, containing 651 inlines and 951 crosslines. The survey follows the geometry: 25m inline spacing, 25m crossline spacing, and 4ms sample interval. The data illuminates up to 1.848 seconds (SILVA, 2019).

3. Seismic attributes

Estimating rocks' physical properties through the acquisition and processing of seismic data and analyzing their vertical and lateral variation in time, space, and frequency domains constitute the basis for seismic interpretation (NANDA, 2021). As Taner *et al*. (1979) stated, seismic attributes can be defined as any observations extracted from seismic data that directly or indirectly aid hydrocarbon exploration. Besides that, calculating seismic attributes can be seen as the application of filters that remove particular components of the seismic signal to highlight others (BARNES, 2016). These quantities support seismic interpretation by revealing hidden structural features, such as faults and fractures (HESTHAMMER & FOSSEN, 1997), and the basement's boundaries; identifying strata terminations; highlighting reflector's continuity; discretizing seismic facies (BAGHERI & RIAHI, 2015); detecting gas hydrate presence (CLAIRMONT *et al.*, 2021), as well as aiding in the mapping of DHI anomalies (RODEN & CHEN, 2017).

Seismic attributes have been frequently used as input data for different algorithms whose purpose is pattern recognition or cluster analysis. As per Barnes and Laughlin (2002), the accuracy of the pattern recognition results is primarily related to the choice of the set of seismic attributes to be used. Except for the increasing computation cost, there are no limitations to the number of seismic attributes used for clustering analysis. However, the exaggeration of this choice can be harmful to the result. It will reduce the interpretability of the method and favor redundancy in the input data (BARNES, 2016).

In order not to promote an exhaustive search for the ideal set of seismic attributes through successive tests using the combination of dozens of attributes developed to date, the decision regarding which to use in this research is initiated by the connection between the knowledge of the expected response of each attribute for the identification of DHI anomalies and the guidance found in specialized literature (RODEN & CHEN, 2017). A similar strategy can be found in Ismail *et al.* (2023), where the authors assess geometric and curvature attributes that provide detailed information on structural discontinuities to enhance fracture network interpretation.

Since the geometric, spectral, and amplitude settings stand out in characterizing DHI anomalies in seismic images, the set of attributes chosen as input to the clustering algorithms highlight characteristics of this nature, as recommended by Infante and Marfurt (2019). The gas anomalies giving rise to major accumulations in the Central Graben Basin are correlated with the intense fracturing of layers underlying the reservoirs, forming gas chimneys (DE BRUIN *et al.*, 2022). Coherence in the vicinity of a seismic trace can provide clues to such zones. The presence of this fluid in rocks and lithological variability within the petroleum system also suggests that relative acoustic impedance may contribute to the segmentation of DHI anomalies. Furthermore, the drastic reduction in spectral content in DHI zones reinforces the use of attributes associated with frequency response. Finally, these anomalies are often linked to high seismic amplitudes. Hence, it was determined that the cubes of attributes extracted from the amplitude data would be the Reflection Strength, Relative Acoustic Impedance, Similarity, Logarithm of Sweetness, and Spectral Decomposition (25Hz and 45Hz). Together, these features will support a multi-attribute analysis to identify class 3 DHI anomalies, since their properties allow us to verify anomaly consistency and conformance to downdip structure, phase change at downdip edge of anomaly, and flat spots (RODEN & CHEN, 2017). Class 2 anomalies require an appropriate AVO observation, however partial stacks were not available. Table 1 shows the category they are included in and their enhanced features, meanwhile Figure 2 shows the seismic attributes on FS8 seismic horizon.This seismic horizon is freely provided by dGB Earth Sciences via TerraNubis portal and marks the top of FS8 reservoir (Adeoti *et al.*, 2023). It is inserted in a context of plane-parallel reflectors with high amplitude.

Table 1 – Seismic attributes extracted to be the input data in the multi-attribute approach.

Seismic Attributes	Category*	Application
Relative Acoustic Impedance	Seismic inversion	Amplitude anomalies; DHI; channels; stratigraphy; lithology
Logarithm of Sweetness	Highlighting amplitudes	Amplitude anomalies; DHI; stratigraphy
Reflection Strength	Instantaneous	Amplitude anomalies; DHI; channels; shadow zones; stratigraphy
Coherence	Geometric	Continuity; faults and fractures; channels; stratigraphic variations
Spectral Decomposition – 25 Hz	Frequency	Channels; shadow zones; frequency content
Spectral Decomposition – 45 Hz	Frequency	Channels; shadow zones; frequency content

Fonte: Brown (1996).

4. Machine learning algorithms

Machine learning is a subarea of artificial intelligence whose focus remains on developing algorithms that automate tasks and improve their performance with experience (MITCHELL, 1997). According to James *et al.* (2013), most statistical learning problems can be classified into two categories: supervised and unsupervised. The dimensionality reduction and cluster analysis algorithms applied in this work are classified as unsupervised, in which there is no associated response for each observed sample. The objective of cluster analysis is to verify if certain samples have similar characteristics that are capable of differentiating them from others present in the dataset. In this section, we present the operating principles of the Principal Component Analysis (PCA), K-Means, Gaussian Mixture Model (GMM), and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithms.

4.1. Principal Component Analysis (PCA)

PCA is a data analysis tool often used to perform dimensionality reduction and data filtering (DEISENROTH *et al.,* 2020). It is a simple non-parametric method for extracting relevant information from a large dataset. In other words, this algorithm provides a way to mitigate the complexity of a problem by finding the most meaningful form to express a set of variables that make up a given system and describe one or more events (SHLENS, 2014).

PCA performs a linear transformation on the dataset to project it on a subspace of lower dimensionality (LEVER *et al.*, 2017). The algorithm's output yields an optimized representation of the original data by retaining as much information as possible through a reduced set of variables known as principal components.

Let $\mathbf{X}$ be a data matrix normalized by the mean and standard deviation, with dimension $N \times M$, in which N and M are the numbers of samples and variables, respectively. Its covariance matrix $\mathbf{S_x}$ is given by Equation 1:

$$\mathbf{S_X} = \frac{1}{N-1}\mathbf{X}\mathbf{X}^T. \tag{1}$$

It is possible to obtain a matrix $\mathbf{Y}$ capable of representing $\mathbf{X}$ in the principal component's domain, so that:

$$\mathbf{Y} = \mathbf{P}\mathbf{X}, \tag{2}$$

In which $\mathbf{Y} = [\boldsymbol{y_1}, \dots, \boldsymbol{y_M}]^T$ contains the principal components, and the matrix $\mathbf{P} = [\boldsymbol{p_1}, \dots, \boldsymbol{p_M}]^T$ represents the eigenvectors of the matrix $\mathbf{S_X}$. The matrix P is responsible for translating and rotating the original set of attributes to determine the orthogonal base that maximizes the variance of the elements of Y, and that better restates the datum in the domain of principal components (DEISENROTH *et al.*, 2020).

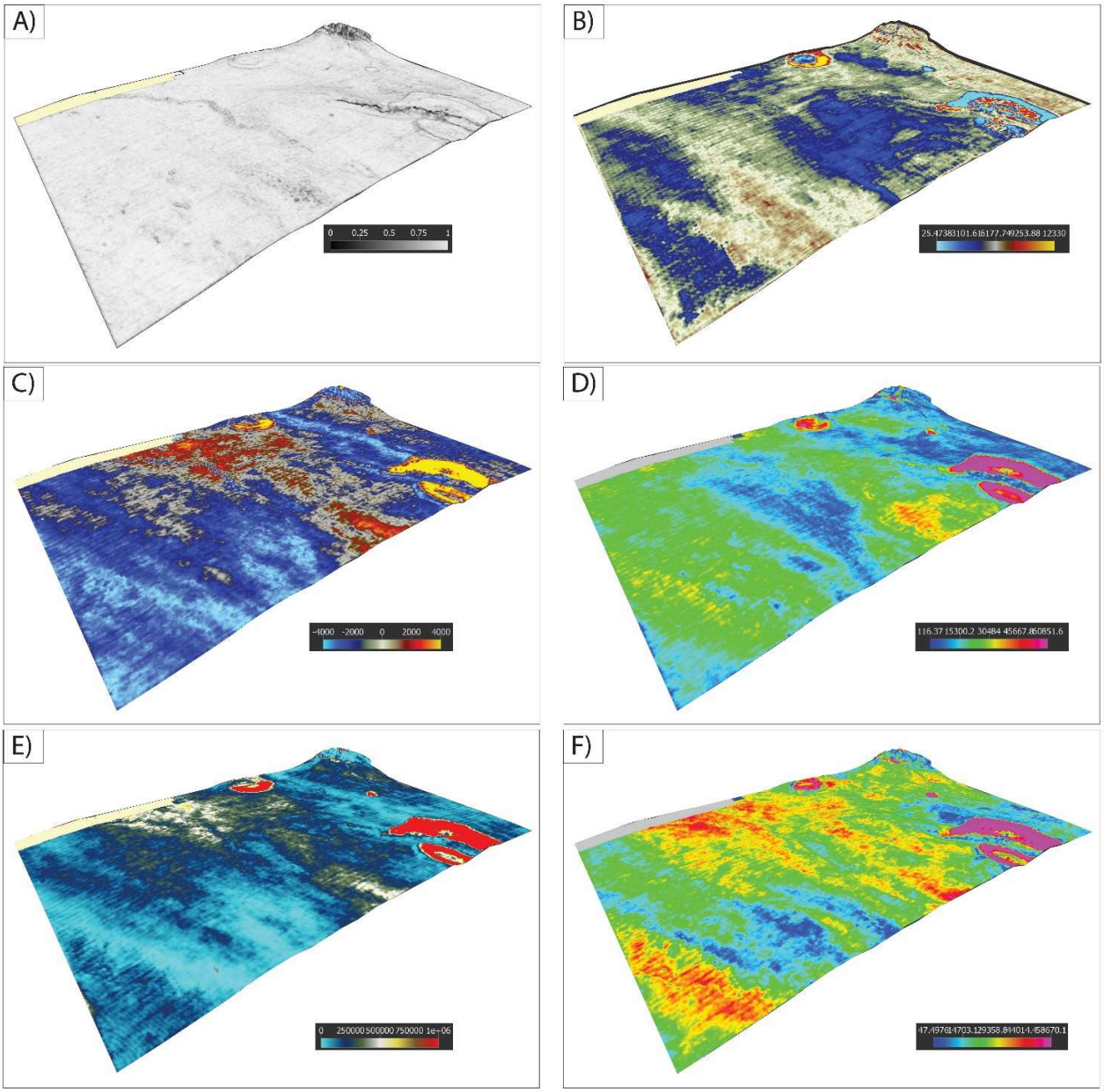

Figure 2 – Seismic attributes used as input of the clustering analysis: (a) Similarity, (b) Reflection Strength, (c) Relative Acoustic Impedance, (d) Spectral Decomposition (25 Hz), (e) Logarithm of Sweetness, and (f) Spectral Decomposition (45 Hz).
Source: Authors (2024).

4.2. K-Means

The K-Means algorithm is a clustering method that aims to segment a set of unlabeled data into a previously defined *K* number of groups. After this selection, it is verified that each point belongs to one and only one cluster in such a manner that similar samples are clustered and related to a centroid (JAMES *et al.*, 2013). The centroids are objects that store all the feature's means for each *K* cluster (TAN *et al.,* 2016). In other words, considering the matrix **X**, defined following the

same patterns as outlined in the preceding subsection, the K-Means algorithm aims to perform data partitioning in K distinct clusters that share characteristics similar to centroids $\boldsymbol{\mu}_K$. This objective is realized by ensuring that each centroid is correctly positioned within the observation space, aiming to minimize the intragroup variance across all clusters. The variance Var of a cluster C_K is the measure of how its observations differ (JAMES *et al.*, 2013). The mathematical description of the intragroup variance used in this work is related to the notion of Euclidean distance. As specified by Bishop (2006), it can be seen as the sum of the Euclidean distance between every sample $\boldsymbol{x}_n$ that belongs to the Kth group and its respective centroid C_K, divided by the total number of observations N. Equation 3 presents the intragroup variance as described below:

$$Var(C_K) = \frac{1}{N} \sum_{\substack{\forall \boldsymbol{x_n} \in K \\ n=1}}^{N_K} \sum_{k=1}^{K} \|\boldsymbol{x}_n - \boldsymbol{\mu}_k\|^2 \ . \tag{3}$$

Under these circumstances, the function to be minimized by the K-Means algorithm is defined as:

$$\underset{C_1, \dots, C_K}{\boldsymbol{minimize}} \left\{ Var(C_K) = \frac{1}{N} \sum_{\substack{\forall \boldsymbol{x_n} \in K \\ n=1}}^{N_K} \sum_{k=1}^{K} \|\boldsymbol{x}_n - \boldsymbol{\mu}_k\|^2 \right\}. \tag{4}$$

4.3. Gaussian Mixture Models (GMM)

The GMM is considered a simple linear superposition of Gaussian components that aims to establish a grouping of density models that is more informative than that generated by a single normal distribution (BISHOP, 2006). In the one-dimensional domain, a Gaussian probability distribution $\mathcal{N}(x|\mu, \sigma^2)$ of a variable x is given by:

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{2\pi\sigma^2} exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \tag{5}$$

In which μ and σ are the mean and variance, respectively. The generalization to the M-dimensional domain is based on the concept of covariance matrix $\boldsymbol{\Sigma}$ (BISHOP, 2006). Considering the matrix $\mathbf{X}$, its Gaussian distribution is defined as per Equation 6.

$$\mathcal{N}(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{M}{2}}} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} exp\left\{-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\right\} \tag{6}$$

Analogously to Equation 5, $\boldsymbol{\mu}$ is the mean of the data matrix, $\boldsymbol{\Sigma}$ and $|\boldsymbol{\Sigma}|$ are the covariance matrix and its module, respectively.

According to Deisenroth *et al.* (2020), the GMM algorithm provides a probability density model in which a finite number of K normal distributions are combined in such a way that the distribution equation of Gaussian mixtures $p(\mathbf{X})$ (Equation 7) be satisfied.

$$p(\mathbf{X}) = \sum_{k=1}^{K} \boldsymbol{\pi}_k \, \mathcal{N}(\mathbf{X}|\boldsymbol{\mu}_\mathrm{k}, \boldsymbol{\Sigma}_\mathrm{k}). \tag{7}$$

In this equation, π_k is the weight assigned to the occurrence of each normal distribution. The purpose of the GMM algorithm is to optimize the parameter set $\boldsymbol{\theta} \coloneqq \{\boldsymbol{\pi}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k : k = 1, \dots, K\}$ to maximize the likelihood function through an Expectation-Maximization method, described in Equation 8:

$$log\big(p(\mathbf{X}|\boldsymbol{\mu}_\mathrm{k}, \boldsymbol{\Sigma}_k)\big) = \sum_{n=1}^{N} log\Big(\sum_{k=1}^{K} \boldsymbol{\pi}_k \, \mathcal{N}(\mathbf{X}|\boldsymbol{\mu}_\mathrm{k}, \boldsymbol{\Sigma}_\mathrm{k})\Big). \tag{8}$$

It is possible to calculate the probability that a sample belongs to one of the normal distributions of the mixture model. For this purpose, the quantity

$$r_{nk} = \frac{\boldsymbol{\pi}_k \mathcal{N}(\mathbf{X}|\boldsymbol{\mu}_\mathrm{k}, \boldsymbol{\Sigma}_\mathrm{k})}{\sum_{k=1}^{K} \boldsymbol{\pi}_k \mathcal{N}(\mathbf{X}|\boldsymbol{\mu}_\mathrm{k}, \boldsymbol{\Sigma}_\mathrm{k})} \tag{9}$$

Is defined as the *responsibility* of the Kth component of the mixture by the nth instance in the distribution. The responsibility, as defined in the preceding molds, is directly proportional to the probability of the nth sample of the distribution belonging to the Kth component of the mixture. Therefore, one of the Gaussian distributions of the mix has high responsibility over a sample when, most likely, that sample belongs to this Gaussian component (BISHOP, 2006).

4.4. Density Based Spatial Clustering of Applications with Noise (DBSCAN)

First introduced by Ester *et al.* (1996), the DBSCAN algorithm assumes that a sample belongs to a cluster only if its neighborhood, defined by a radius ϵ, contains a minimum number of points (MinPts). In other words, the spatial density of points in its vicinity must exceed a threshold value. This principle is employed to recognize patterns in a dataset and segment it into informational subgroups with arbitrary shapes, considering the typical relationship between a sample cloud and the spatial density of noisy points (ESTER *et al.*, 1996). In the same work, the authors define concepts such as ϵ-neighborhood, reachable and directly reachable points, core points, among others. These concepts are essential for understanding how DBSCAN can distinguish genuine clusters from noisy samples.

5. Methodology

Figure 3 shows the workflow used for all experiments, including K-Means and Gaussian Mixture Models clustering analysis and the subsequent spatial density filtering. All workflow was implemented in Python, and the OpendTect software was used only to visualize the results. Key steps in this process will be briefly detailed below.

5.1. Study interval selection

Although it was possible to use all samples from the seismic cube, in this approach, cluster analysis was conducted within a delimited area of interest defined by two seismic horizons. Besides reducing computational costs, this focus is influenced by geological factors. For instance, Schroot and Schüttenhelm (2003) state that the primary hydrocarbon accumulations in the Dutch North Sea are associated with unconsolidated Miocene clastic sediments. Therefore, the upper limit of our study is the FS8 horizon, while the lower is the MFS4 horizon. We used 25 samples above the FS8 surface and the same amount below the MFS4 to delimit the area of interest, i. e. 100ms above and below the limits.

5.2. Data preprocessing

Seeking to eliminate null samples and spurious values, i.e., those lacking geophysical-geological significance. For each selected seismic attribute, we analyzed the interquartile amplitude of the distribution, aiming to assess the statistical dispersion of samples within a dataset around its central measure.

5.3. Principal Component Analysis (PCA)

Once the seismic attributes were preprocessed, the normalized by mean and variance dataset became the input for dimensionality reduction using the PCA algorithm. After assessing the explained variance ratio, it was determined that four principal components can explain 92.83% of the original dataset, as depicted in Figure 4.

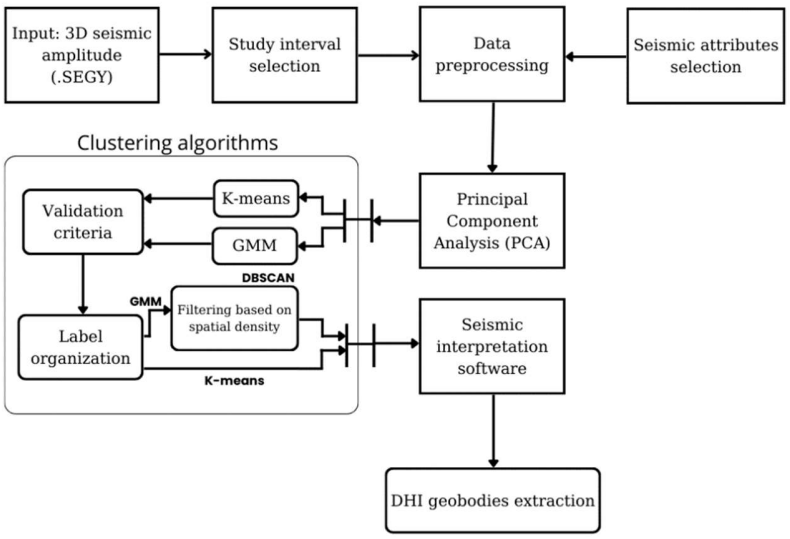

Figure 3 – Workflow applied to perform the seismic attributes clustering analysis over a 3D dataset.
Source: Authors (2024).

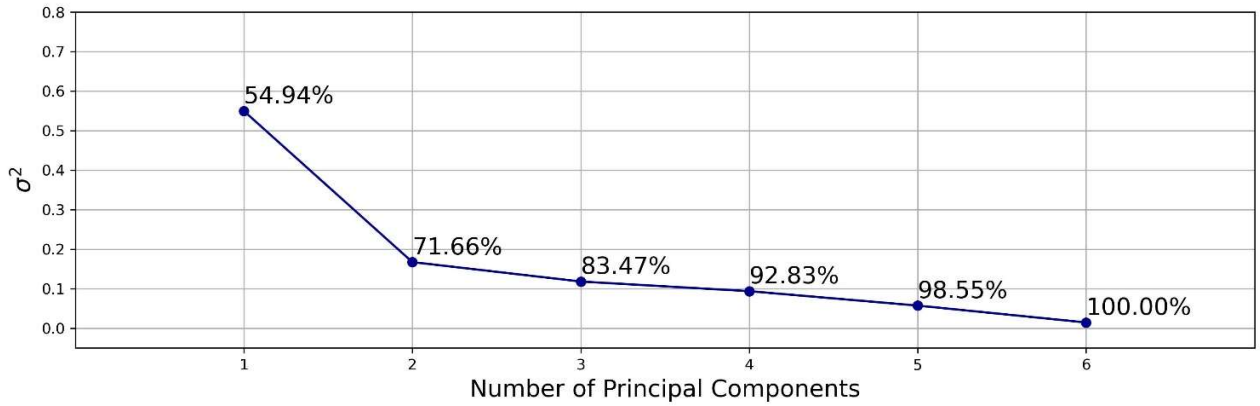

Figure 4 – Normalized percentage variance of each principal component and cumulative percentage variance applied to determine the total number of principal components.
Source: Authors (2024).

5.4. Clustering Algorithms

The multi-attribute cluster analysis is the first step of the combined approach used in this work. 60% of the dataset samples were randomly selected to compose the training set of the K-Means and GMM models. Next, the labels of the remaining samples were inferred. Finally, according to the methodology proposed by Troccoli *et al.* (2022), the obtained labels are organized concerning the origin to preserve some degree of similarity between the nearest centroids, adopting Euclidean distance as a comparison metric.

5.5. Validation Criteria

Using statistical tests to estimate the potential number of clusters that optimizes the dataset sample's segmentation. Once the elbow method could not present a satisfactory indication of this parameter due to the nature of the dataset, the Davies-Bouldin Index was also calculated (Figure 5). It suggested $K = 5$ as the optimal parameter for both algorithms.

Although the validation criterion is an essential stage for this methodological procedure, in this problem, we must consider the geological features, as well as the work's objective. The visualization of the clustering results pointed out that five groups were not able to isolate DHI anomalies. Setting $K = 7$ best met this objective: isolated DHI anomalies and showed a high correlation with the geomorphological features.

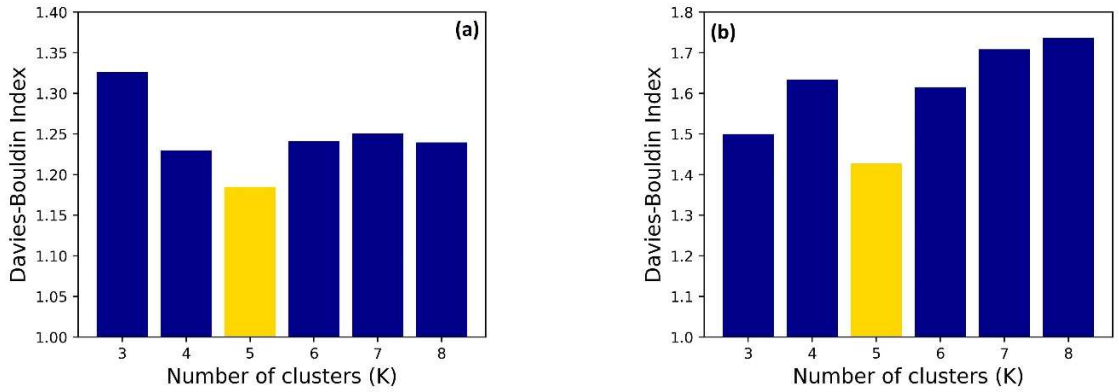

Figure 5 – Davies-Bouldin Index per number of clusters. $K = 5$ minimizes the metric for both (a) K-Means and (b) Gaussian Mixture Model algorithms, suggesting this is the optimal quantity of clusters.
Source: Authors (2024).

5.6. Filtering based on spatial density

In this second step, the DBSCAN algorithm aimed to comprehend the spatial distribution of DHI cluster samples, enabling the filtration of spurious points and low-spatial density zones. During this phase, the input data consists of normalized coordinates and time samples, ranging from zero to one, for instances interpreted as DHI. A series of experiments were conducted to fine-tune the model's hyperparameters, resulting in an ϵ-neighborhood radius of 0.01 and a minimum of 500 points for optimal performance. Under the specified initial conditions, the algorithm determines the number of subgroups, the sample count within each subgroup, and the identification of samples as noise. These subgroups are then organized in descending order based on the number of samples, with the top 12 being assigned the DHI label, while the remaining samples are categorized as noise. It is important to note that this numerical threshold may vary depending on the specific study region. Consequently, the expectation is that the DHI cluster will be entirely isolated within the anomalous zones.

6. Results and discussions

This section presents the clustering results derived from both K-Means and Gaussian Mixture Model algorithms. It also provides adequate discussions regarding the application of each technique in mapping direct hydrocarbon indicators anomalies and the association of multi-attribute cluster analysis with filtering based on spatial density. Once the preprocessing, dimensionality reduction, and validation steps were fulfilled, a visual analysis of the clustering results was made by varying the number of groups between three and eight. The optimized models have the following configuration (Table 2).

Table 2 – Parameters of the optimized models.

Algorithm	Seismic Attributes	Principal Component	Number of Clusters
K-Means	Logarithm of Sweetness, Coherence, Acoustic Impedance, Reflection Strength, Spectral Decomposition (25 and 45 Hz)	4	5
GMM	Logarithm of Sweetness, Coherence, Acoustic Impedance, Reflection Strength, Spectral Decomposition (25 and 45 Hz	4	7

Source: Authors (2024).

The acoustic response related to the presence of hydrocarbons characterizes a DHI when associated with a trapping configuration, identified, in this case, by low frequency shadow zones or velocity pull-down effects (RODEN & CHEN, 2017). Schroot and Schüttenhelm (2003) state that the Dutch North Sea presents a series of gas-related phenomena whose seismic expression resembles those mentioned above. Fault related amplitude anomalies, gas chimneys, buried gas-filled ice-scours, and mostly bright spots are among the seismic events interpreted in the area (SCHROOT & SCHÜTTENHELM (2003); SCHROOT *et al*., (2005); CONNOLLLY (2015)).

Looking at the clustering result over the FS8 horizon (Figure 6b), it is found that the K-Means algorithm, followed by the labels organization, assigned more than one label (0 and 2) to two zones with either bright spot and fault-related seismic anomaly, previously interpreted by many authors (SCHROOT & SCHÜTTENHELM (2003); SCHROOT *et al*., (2005); CONNOLLLY (2015); DE BRUIN *et al.*, (2022)). In addition, group 2 extends over much of the FS8 horizon, which suggests that this group was defined predominantly by the similarity of its samples concerning amplitude attributes, such as Reflection Strength and Logarithm of Sweetness, in detriment to the set of aspects that characterize DHIs. This behavior is repeated throughout the analyzed interval. Because it is a prototype-based clustering algorithm, all members of the cluster associated with a given centroid should be close to their corresponding prototype (PATEL & KUSHWAHA, 2020). For K-Means, the optimal distribution of samples in space consists of well-separated spherical groups. Therefore, the construction of the algorithm's minimization problem may suggest difficulties on clustering sets of samples that overlap in the four-dimensional space of the principal components used here. Thus, the algorithm proved ineffective in isolating DHI anomalies in the region.

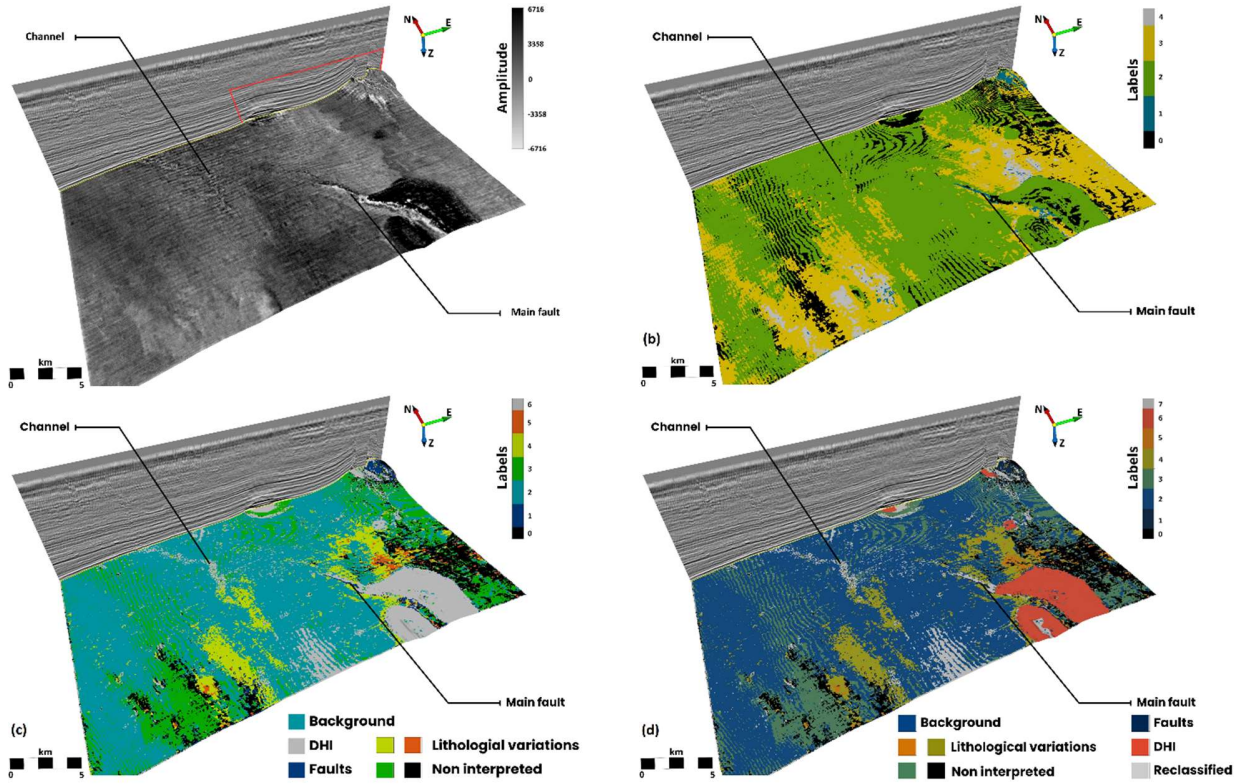

Figure 6 – (a) Original seismic amplitude; clustering results derived from (b) K-Means, (c) Gaussian Mixture Model, and (d) the joint application of GMM and filtering based on spatial density.
Source: Authors (2024).

The expectations created by the robustness of the Gaussian Mixture Model algorithm began to be achieved by observing the clustering result on the seismic horizon, as shown in Figure 6c. After organizing labels concerning the origin, to the DHI class of samples was assigned label 6, seen in gray. It is possible to notice that the GMM was way more effective in delimiting DHI anomalies zones than K-Means. Based on probability density estimations, each cluster is modelled as a Gaussian distribution with its particular mean and standard deviation, which guarantee that GMM will provide a better quantitative measure of fitness per number of cluster (PATEL & KUSHWAHA, 2020). As a reflection of this, there is the behavior of the class interpreted as DHI in Figure 6c. The most prominent anomalies are confined in this class, so that the set of attributes used to construct the hyperspace of principal components and that highlight structural, phase, frequency, and amplitude characteristics seems to have had a strong influence on this phenomena's segmentation.

Even though it was not the primary objective of this work, the GMM model made it possible to correlate clusters to interesting features such lithological variations, faults, and fractures. The latter may be related to class 1, seen in dark blue, as noticeable in the main fault and at the northeastern zone of the seismic horizon, a raised region with intense fault presence due to the halokinesis of the Zechstein Formation (MAUNDE & ALVES, 2022). It becomes clearer on Figure 7, where a vertical section view – Inline 668 - of the results is shown. It is a window, indicated by the red polygon, of the same inline that appears on Figure 6a. It can be seen that the presence of the cluster interpreted as DHI are often associated with the fault system, alongside with the class 1. It can indicate migration pathways from deeper sources to shallow gas-related DHI anomalies, in agreement with the hypothesis that establishes a relationship between these accumulations and deeper structures, as discussed by De Bruin *et al*. (2022). On the other hand, samples of the same class appear dispersed in other regions of the section that are not necessarily related to fractures or faults. This fact leads us to the possibility that low coherence zones are strongly influencing the conformation of this cluster, since the model was not optimized for this purpose.

As stated earlier, the organization of labels assigns a similar character to clusters represented by close colors in the color palette. This effect can be observed in groups 4 and 5, in yellow and orange (Figure 6c), possibly related to lithological variations.

Although the goal of isolating DHI anomalies had been reasonably achieved by the GMM model, it seems there are still regions whose distribution does not ratify their classification within the target cluster. The apparent paleochannel could be an example of this drawback. The DBSCAN's effect is clear in Figure 6d, where it is possible to observe that both samples associated with the paleochannel and the gas-filled ice-scours, which in this case presents a typical pattern of straight lineaments with N-S orientation (SCHROOT & SCHÜTTENHELM, 2003) were generically relabeled as "reclassified" and represented in gray (now, Group 7). Here, it is observed that the joint application of unsupervised algorithms was able to distinguish between anomalies characterized as bright spots and other gas-related seismic manifestations. It occurs due to the fact that the latter are regions containing low spatial density of samples. On Figure 7 a similar effect take place on the fractured zone. Besides that, the DBSCAN could also provide a noise removal outcome, since a few sparkling samples, found on TWTs superior to 800ms, were visibly relabeled. It will be a key point on 3D visualization, as follows.

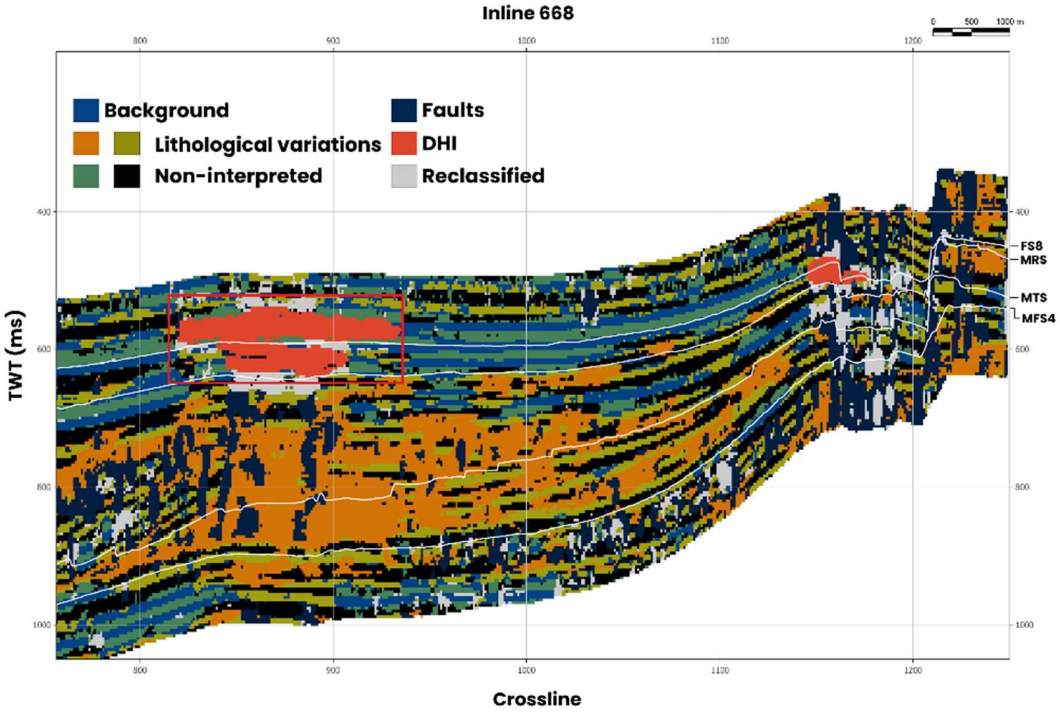

Figure 7 – Clustering results seen in a window from inline 668 obtained from the joint application of the filtering based on spatial density on samples labeled as DHI by the GMM algorithm.
Source: Authors (2024).

From a 3D perspective, the GMM-labeled DHI samples pollute the extent of the seismic survey entirely, impairing the three-dimensional visualization of this work's targets, as can be seen in Figure 8a. The improvement in the DHI anomalies' visualization in this perspective after the filtering based on spatial density is evident (Figure 8b). The geobodies composed by the labels interpreted as DHI present a much cleaner aspect, in comparison to the previous result.

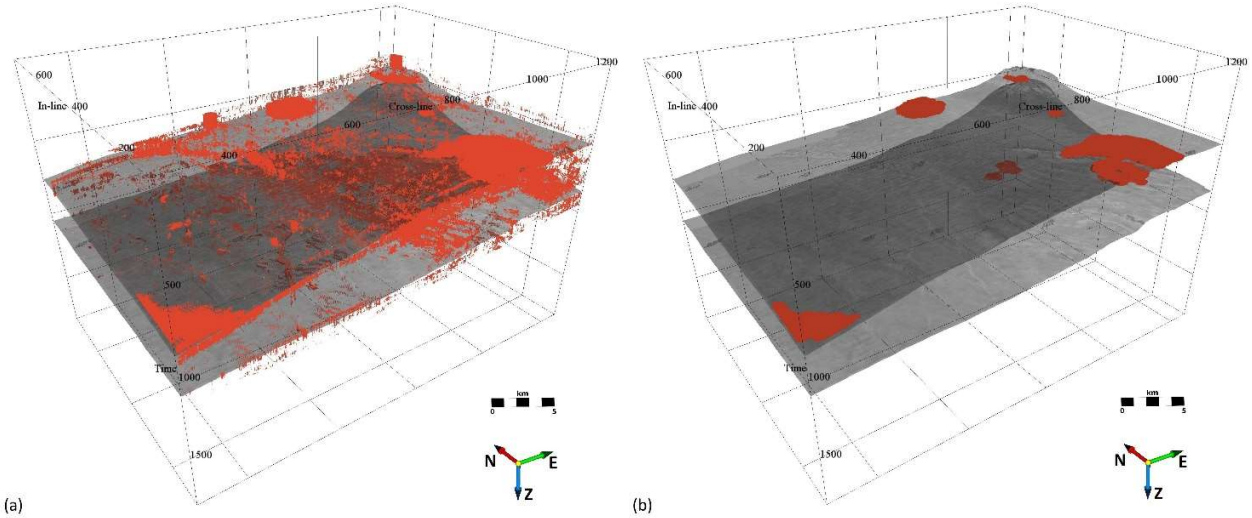

Figure 8 – The result of applying spatial density filtering through the DBSCAN algorithm on cluster 6's samples. The two seismic horizons were added to give the idea of the spatial location of each anomaly.
Source: Authors (2024).

7. Final remarks

This study introduces a novel methodology designed for clustering analysis on 3D seismic data to map direct hydrocarbon indicators (DHIs). By employing two unsupervised machine learning algorithms, K-Means and the Gaussian Mixture Model (GMM), the research examines the individual performances of each technique and explores the benefits of integrating the multi-attribute approach with spatial density-based filtering via the DBSCAN algorithm.

It was observed that using statistical tests, such as the elbow method or the Davies-Bouldin index, to determine the optimal number of clusters for the models was not efficient, at least when the study's goal is to segment a specific geological event or feature like DHI anomalies. Possibly, in an exploratory approach, optimizing this hyperparameter through statistical tests could be a good starting point.

The K-Means algorithm showed limitations in accurately delimiting the DHI anomalies. It assigned more than one class to regions of known DHIs. These facies are distributed across extensive regions of the seismic survey, suggesting that this group was defined only by the similarity of the amplitude content of their samples, to the detriment of the characteristics expressed by the DHIs. In contrast, the Gaussian Mixture Model achieved good results in identifying this work's targets. With a model composed of seven clusters, it could accurately delimit DHI anomalies. In addition, it efficiently highlighted other geological features, such as faults and fractures, paleochannels, and groups related to lithological variation.

The subsequent application of spatial density-based filtering to samples labeled by the GMM algorithm and interpreted as DHI - the heart of the joint clustering - identified subgroups of low spatial density. Thus, the DBSCAN algorithm offered the conditions to re-label these samples and increase the accuracy of the anomaly mapping. Based on these results, geobodies could be generated, and the perspective of three-dimensional observation became noise-free. This methodology is expected to be adapted and employed in other sedimentary basins to assist geophysicists in interpreting possible anomalies caused by oil and gas.

Acknowledgements

The authors thank Grupo de Estudo e Aplicação de Inteligência Artificial em Geofísica da Universidade Federal da Bahia (GAIA-UFBA) and the Institute of Geociensces from UFBA for the infrastructure provided. We also would like to thank dGB Earth Sciences for making the dataset available as an OpendTect project via their TerraNubis portal terranubis.com. Finally, the authors also acknowledge the Instituto Nacional de Ciência e Tecnologia de Geofísica do

Petróleo (INCT-GP) and National Council for Scientific and Technological Development (CNPQ) for Author 1 scholarship, Bahia State Research Support Foundation for the scholarship of Author 2, and (CNPq) for financing the project of number 409718/2022-0.

References

NANDA, N. C.. Seismic data interpretation and evaluation for hydrocarbon exploration and production. *Springer International Publishing*, 2021.

HILTERMAN, F. J.. *Seismic amplitude interpretation*. Society of Exploration Geophysicists and European Association of Geoscientists and Engineers, 2001.

FORREST, M.; RODEN, R.; HOLEYWELL, R.. Risking seismic amplitude anomaly prospects based on database trends. *The Leading Edge*, v. 29, n. 5, p. 570-574, 2010.

ZHAO, T.; ZHANG, J.; LI, F.; MAFURT, K. J.. Characterizing a turbidite system in Canterbury Basin, New Zealand, using seismic attributes and distance-preserving self-organizing maps. *Interpretation*, v. 4, n. 1, p. 79-89, 2016.

CERQUEIRA, A. G.; DE LIMA, O. A. L.; RIOS, R. A.. A nonparametric approach using clustering analysis to estimate shaliness in shaly-sand formations. *Journal of Applied Geophysics*. v. 164, p. 11-18, 2019.

TROCCOLI, E. B.; CERQUEIRA, A. G.; LEMOS, J. B.; HOLZ, M.. K-Means clustering using principal component analysis to automate label organization in multi-attribute seismic facies analysis. *Journal of Applied Geophysics*. v. 198, 2022.

BARBOSA, M. R. S.; CARNEIRO, V.; CERQUEIRA, A. G.. Seismic well tie using geophysical logs obtained from k-nearest neighbor regression algorithm. *Brazilian Journal of Geophysics*. v. 40, n. 1, 2022.

MITCHELL, T. M.. *Machine Learning*. McGraw-Hill Science/Engineering/Math, 1997.

TSAI, K. C.; HU, W.; WU, X.; CHEN, J.; HAN, Z.. First-break automatic picking with deep semisupervised learning neural network. *SEG Technical Program Expanded Abstracts 2018*. Society of Exploration Geophysicists. p. 2181-2185, 2018.

BRITO, L.; ALAEI, B.; TORABI, A.; LEOPOLDINO-OLIVEIRA, K.; VASCONCELOS, D.; BEZERRA, F.; NOGUEIRA, F.. Automatic 3D fault detection and characterization – A comparison between seismic attributes methods and deep learning. *Interpretation*. v. 11, T793-T808, 2023.

BÖNKE, W.; ALAEI, B.; TORABI, A.; OIKONOMOU, D.. Data augmentation for 3D seismic fault interpretation using deep learning. *Marine and Petroleum Geology*. v. 162, 106706, 2024.

MA, Y.; LUO, Y.. Automatic first-arrival picking with Reinforcement Learning. *International Geophysical Conference, Beijing, China, 24-27 April 2018*. Society of Exploration Geophysicists and Chinese Petroleum Society. p. 493-497, 2018.

MENG, J.; WANG, S.; CHENG, W.; WANG, Z.; YANG, L.. AVO Inversion Based on Transfer Learning and Low-Frequency Model. *IEEE Geoscience and Remote Sensing Letters*. pp. 1-1, 2021

LI, P.; LIU, M.; ALFARRAJ, M.; TAHMASEBI, P.; GRANA, D.. Probabilistic physics-informed neural network for seismic petrophysical inversion. *Geophysics*. v. 89, p. M17-M32, 2024

WANG, H.; ZHANG, M.; FAN, G.; XIAO, L.; ZUO, G.; YANG, L.; PANG, X.; WANG, C.; ZHANG, Y.. Prediction of lithology in lacustrine carbonates using well logs: The Cretaceous Barra Velha Formation in Santos Basin, offshore Brazil. *Geological Journal*. v. 58, n. 2, 2023.

CORDEIRO, F.; SOUZA, P.; CERQUEIRA, A.. Permeability Prediction in Geophysical Logs in the Barra Velha Formation of the Santos Basin. *18th International Congress of the Brazilian Geophysical Society & Expogef*, Brazilian Society of Geophysics, 2023.

HUANG, J.; NOWACK, R.. Machine Learning Using U-Net Convolutional Neural Networks for the Imaging of Sparse Seismic Data. *Pure and Applied Geophysics*. v. 177, n. 1, 2020.

YU, S.; MA, J.. Deep Learning for geophysics: Current and future trends. *Review of Geophysics*. v. 59, n. 3, 2021.

BARNES, A. E.; LAUGHLIN, K. J.. Investigation of methods for unsupervised classification of seismic data. *SEG Technical Program Expanded Abstracts 2002*. Society of Exploration Geophysicists. p. 2221-2224, 2002.

RODEN, R.; CHEN, C. W.. Interpretation of DHI characteristics with machine learning. *First Break*. v. 35, n. 5, 2017.

BOUROULLEC, R.; VERREUSSEL, R. M. C. H.; GEEL, C. R.; De BRUIN, G.; ZIJP, M. H. A. A.; KŐRÖSI, D., MUNSTERMAN, D. K.; JANSSEN, N. M. M.; KERSTHOLT-BOEGEHOLD, S. J.. Tectonostratigraphy of a rift basin affected by salt tectonics: synrift Middle Jurassic-Lower Cretaceous Dutch Central Graben, Terschelling Basin and neighbouring platforms, Dutch offshore. London: *The Geological Society of London*. v. 469, p. 269-303, 2018.

MAUNDE, F.; ALVES, T. M.. Effect of tectonic inversion on supra-salt fault geometry and reactivation histories in the Southern North Sea. *Marine and Petroleum Geology*. v. 135, 2022.

LARMINIE, F.. The history and future of the North Sea Oil and Gas: an environmental perspective. *Philosophical Transactions of the Royal Socitey of London*, Series B, Biological Sciences, v. 316, n. 1181, p. 487-493, 1987.

De BRUIN, G.; ten VEEN, J.; WILPSHAAR, M.; VERSTEIJLEN, N.; KEES, G.; VERWEIJ, H.; CARPENTIER, S.. Origin of shallow gas in the Dutch North Sea – Seismic versus geochemical evidence. *Interpretation*. v. 10, p. SB67, 2022.

ROSENDAAL, E.; KAYMAKCI, N.; WIJKER, D.; SCHROOT, B.. Structural development of the Dutch central graben – new ideas from recent 3D seismic.*76th EAGE Conference and Exhibition 2014*, European Association of Geoscientists & Engineers, p. 1-5, 2014.

JAKOBSEN, F. C.; BRITZE, P.; THÖLE, H.; JÄHNE-KLINGBERG, F.; DOORNENBAL, H.; VIS, G.. Harmonized stratigraphic chart for the North Sea area NL-DEDK. *3D Geomodeling for Europe project report*. 2020.

MÜLLER, S. M., JÄHNE-KLINGBERG, F.; THÖLE, H.; JAKOBSEN, F. C.; BENSE, F.; WINSEMANN, J.; GAEDICKE, C.. Jurassic to Lower Cretaceous tectonostratigraphy of the German Central Graben, southern North Sea. *Netherlands Journal of Geosciences*. v. 102, e4, 2023.

ISMAIL, A.; RADWAN, A.; MAHMOUD, L.; ABDELMAKSOUD, A.; ALI, M.. Unsupervised machine learning and multi-seismic attributes for fault and fracture network interpretation in the Kerry Field, Taranaki Basin, New Zealand. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*. v. 9, 2021.

INFANTE, L.; MARFURT, K.. Using machine learning as an aid to seismic geomorphology, which attributes are the best input?. *Interpretation*. v. 7, p 1-60, 2019.

ADEOTI, L.; BAKO, M.; ADEOGUN, O.; ANUKWU, G.; ADEGBITE, J.. Porosity prediction using 3D seismic genetic inversion at F3 Block, offshore Netherlands. *Ife Journal of Science*. v. 25, p. 159-174, 2023.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R.. An introduction to statistical learning. *Springer International Publishing*, 2013.

BISHOP, C. M.. *Patter recognition and Machine Learning* (Information Science and Statistics). Springer, 2006.

JIANG, Y. *Detecting geological structures in seismic volumes using deep convolutional neural networks*. Aachen, 2017. 76f. Thesis (Master if Engineering). Rheinisch-Westfälische Technische Hochschule Aachen, Fraunhofer-Gesellschaft, Aachen-Germany, 2017.

WIJKER, D.. *Fault mapping and reconstruction of the structural history of the dutch central graben*. Master thesis on Earth Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, 2014.

SCHROOT, B.; SCHÜTTENHELM, R.. Expressions of shallow gas in Netherlands North Sea. *Netherlands Journal of Geosciences*. v. 82, n. 1, p. 91-105, 2003.

SILVA, R.; BARONI, L.; FERREIRA, R.; CIVITARESE, D. S.; BRAZIL, E. V.. Netherlands dataset: a new public dataset for machine learning in seismic interpretation, 2019.

HESTHAMMER, J.; FOSSEN, H.. Seismic attribute mapping for structural interpretation of the Gullfaks Field, northern North Sea. *Petroleum Geoscience*. v. 3, p. 13-26, 1997.

BAGHERI, M.; RIAHI, M. A.. Modeling the facies of the reservoir using seismic data with missing attributes by dissimilarity-based classification. *Journal of Earth Sciences*. v. 28, n. 4, p. 703-708, 2017.

CLAIRMONT, R.; BEDLE, H.; MAFURT, K.; WANG, Y.. Seismic Attribute Analyses and Attenuation Applications for Detecting Gas Hydrate Presence. *Geosciences*. v. 11, p. 1-26, 2021.

TANER, M. T.; KOEHLER, F.; SHERIFF, R.. Complex seismic trace analysis. *Geophysics*. v. 44, n. 6, p. 1041-1063, 1979.

BARNES, A. E.. *Handbook of poststack seismic attributes*. Society of Exploration Geophysicists, 2016.

BROWN, A. R.. Seismic attributes and their classification. *The Leading Edge*. v. 15, n. 10, p. 1090-1090, 1996.

DEISENROTH, M. P.; FAISAL, A. A.; ONG, C. S.. *Mathematics for machine learning*. Cambridge University Press, 2020.

SHLENS, J.. *A tutorial on principal component analysis*, 2014.

LEVER, J.; KRZYWINSKI, M.; ALTMAN, N.; Points of significance: Principal component analysis. *Nature methods*. v. 14, n. 7, p. 641-643, 2017.

TAN, P. N.; STEINBACH, M.; KUMAR, V.. Introduction to data mining. *Pearson Education India*, 2016.

ESTER, M.; KRIEGEL, H.-P.; SANDER, J.; XU, X. A density-based algorithm for discovering clusters in large spatial databases with noise. *Second International Conference on Knowledge Discovery and Data Mining*, v. 96, p. 226-231, 1996.

SCHROOT, B.; KLAVER, G.; SCHÜTTENHELM, R.. Surface and subsurface expressions of gas seepage to the seabed: examples from the Southern North Sea. *Marine and Petroleum Geology*. v. 22, n. 4, p. 499-515, 2005.

CONNOLLY, D.. Visualization of vertical hydrocarbon migration in seismic data: Case studies from que Dutch North Sea. *Interpretation*. v. 3, p. 1A-T181, 2015.

PATEL, E.; KUSHWAHA, D.. Clustering Cloud Workloads: K-Means vs Gaussian Mixture Model. *Procedia Computer Science*. v. 171, p. 158-167, 2020.