



ISSN: 2447-3359

REVISTA DE GEOCIÊNCIAS DO NORDESTE

Northeast Geosciences Journal

v. 10, n° 2 (2024)

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



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

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

Matheus R. S. Barbosa¹; Vinicius Carneiro²; Alessandro G. Cerqueira³

¹ Universidade Federal da Bahia, Grupo de Estudo e Aplicação de Inteligência Artificial em Geofísica (GAIA-UFBA) Email: m.radames09@hotmail.com

ORCID: <https://orcid.org/0000-0001-5656-3353>

² Universidade Federal da Bahia, Programa de Pós-Graduação em Geofísica da UFBA (PPGEOF-UFBA) Email: vinicius.geophysics@gmail.com

ORCID: <https://orcid.org/0000-0001-8509-9254>

³ Universidade Federal da Bahia, Grupo de Estudo e Aplicação de Inteligência Artificial em Geofísica (GAIA-UFBA) Email: alexsandrocerqueira@ufba.br

ORCID: <https://orcid.org/0000-0003-3462-9336>

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.

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.

Recebido: 22/01/2024; Aceito: 12/08/2024; Publicado: 30/39/2024.

1. Introdução

Anomalias de Indicadores Diretos de Hidrocarbonetos (DHI) são geralmente causadas por mudanças nas propriedades elásticas das rochas. Elas estão comumente associadas com a saturação de reservatórios em gás ou óleo (NANDA, 2012). A sua interpretação impacta significativamente na avaliação de riscos exploratórios e na alocação de poços para perfuração, o que a torna essencial para identificar reservatórios economicamente viáveis (FORREST *et al.*, 2010).

De acordo com Hilterman (2001), os avanços nas metodologias para detecção e validação de anomalias de DHI têm sido notáveis desde os anos 1970, particularmente através da análise de AVO (Amplitude vs Offset) e no avanço das tecnologias de aquisição e processamento de dados sísmicos. Outro fator significativo que orienta esse progresso é a aplicação de atributos sísmicos para avaliar potenciais zonas anômalas. Embora seja efetivo para a interpretação sísmica, correlacionar um grande número de atributos sísmicos simultaneamente pode ser desafiador. Diante desse cenário, nos últimos anos, a aplicação de técnicas de aprendizado de máquina a este e outros problemas da geofísica tem apresentado um crescimento considerável (BARBOSA *et al.*, 2022; CERQUEIRA *et al.*, 2019; TROCCOLI *et al.*, 2022; ZHAO *et al.*, 2016).

Algoritmos de aprendizado de máquina são técnicas utilizadas para extrair informações de conjuntos de dados, automatizar diferentes atividades e identificar padrões de interesse que possam ser imperceptíveis sob a análise humana (MITCHELL, 1997). Esses algoritmos usam diferentes metodologias para aprender iterativamente a partir dos dados e se adaptar para produzir resultados confiáveis e reproduzíveis. A sua popularidade e eficiência rapidamente os apresentaram como ferramentas poderosas para o estudo de vários problemas no campo da geofísica, como no processamento sísmico (MA & LUO, 2018; TSAI *et al.*, 2018), interpretação sísmica (OLIVEIRA *et al.*, 2023; BÖNKE *et al.*, 2024), dados de poços (WANG *et al.*, 2023; CORDEIRO *et al.*, 2023), imageamento sísmico (HUANG & NOWACK, 2020), detecção de terremotos (YU & MA, 2021) e interpretação quantitativa (MENG *et al.*, 2021; LI *et al.*, 2023).

A atividade de reconhecimento de padrões é uma etapa importante para a interpretação de dados sísmicos relacionados à caracterização estrutural, a compreensão da evolução tectonestratigráfica de uma bacia sedimentar e os aspectos de um reservatório de hidrocarbonetos. O seu propósito inicial é segmentar dados sísmicos de acordo com algum fator de similaridade. Barnes e Laughlin (2002) conduziram um estudo comparativo entre os algoritmos K-Médias, Hierárquico Aglomerativo e Mapas Auto-Organizáveis (SOM), atestando os seus bons desempenhos em analisar volumes sísmicos 3D em termos de acurácia, similaridade entre técnicas e ordenação de rótulos. Roden e Chen (2017) incorporaram um fluxo de trabalho de aprendizado de máquina em que a Análise de Componentes Principais (PCA) e os Mapas Auto-Organizáveis (SOM) analisam as combinações de atributos sísmicos para obter padrões significativos que correspondam a indicadores diretos de hidrocarbonetos.

A Bacia do Graben Central (Figura 1a) é uma área com uma complexa evolução geológica (BOUROULLEC *et al.*, 2018; MAUNDE & ALVES, 2022) e um longo histórico exploratório. Desde os anos 1960, através da perfuração do poço pioneiro, atividades exploratórias foram iniciadas nessa região (LARMINIE, 1987). Desde então, com o crescimento do número de campanhas de aquisição de dados, o potencial exploratório da bacia tem se provado por meio de notáveis descobertas de uma grande quantidade de reservas recuperáveis de óleo e, principalmente, gás. Um desses cenários é formado quando fluidos são traçados nos reservatórios clásticos do Cenozóico, gerando anomalias de amplitude sísmica bem marcadas (DE BRUIN *et al.*, 2022).

O objetivo principal desse trabalho é apresentar uma metodologia para identificar anomalias de DHI em dados sísmicos pós-empilhamento 3D. Além disso, nós estabelecemos uma comparação entre dois métodos não supervisionados de aprendizado de máquina raso implementados em Python. A primeira dentre essas técnicas é o algoritmo K-Médias, que visa segmentar os dados sob o custo de minimizar a variância intragrupos iterativamente (JAMES *et al.*, 2013). Em seguida, o algoritmo Modelo de Misturas Gaussianas (GMM) foi utilizado. É considerado uma superposição linear simples de componentes Gaussianas que classificam modelos de probabilidade de densidade de forma mais informativa do que a classificação feita por uma única distribuição normal (BISHOP, 2006). Uma etapa de filtragem subsequente é introduzida a essa segunda técnica, aplicando a avaliação da densidade espacial das amostras. Ela quantifica a densidade espacial do cluster interpretado como DHI, buscando diferenciar anomalias relevantes de menores volumes assim também rotulados, mas vistas com importância insignificante quanto ao potencial para armazenamento de hidrocarbonetos. Uma abordagem similar foi implementada por Jiang (2017), em que o algoritmo DBSCAN (Agrupamento Espacial Baseado em Densidade em Aplicações com Ruído) foi aplicado para efetuar o pós-processamento do resultado de uma segmentação supervisionada de imagens sísmicas para identificar canais fluviais e falhas por meio de Redes Neurais Convolucionais (CNNs). O seu uso reforça a ideia de continuidade nas estruturas detectadas (JIANG, 2017).

Nós conduzimos uma análise de agrupamentos multiatributos de dados sísmicos no domínio das principais componentes, na qual um dados de um levantamento sísmico marinho convencional foi utilizado nos experimentos. Eles são parte do conjunto de dados 3D aberto conhecido como F3 Block, localizado no Mar do Norte holandês, na Bacia do Graben Central (Figura 1a). A incorporação desse processo em duas etapas, realizando uma adicional filtragem espacial baseada em densidade, foi projetada para melhorar o resultados do mapeamento quando comparados à abordagem convencional de um único algoritmo não supervisionado de agrupamento. Focando nessa classe de algoritmos, buscamos oferecer uma abordagem mais eficiente e potencialmente mais rápida para extrair *insights* e dar suporte à interpretação de indicadores de hidrocarbonetos. Essa metodologia sugere um fluxo de trabalho simplificado, o que pode ser vantajoso em cenários onde dados de poços são escassos ou o processo tradicional de rotulagem é demorado.

Portanto, nossa abordagem enfatiza a versatilidade e adaptabilidade da inteligência artificial em problemas da geofísica. Com a implementação desses métodos não supervisionados, demonstramos que é possível obter resultados confiáveis sem depender fortemente de etapas convencionais de pré-condicionamento dos dados. Isto poderia abrir novos caminhos para a análise e exploração, incentivando mais inovação e experimentação nesse campo.

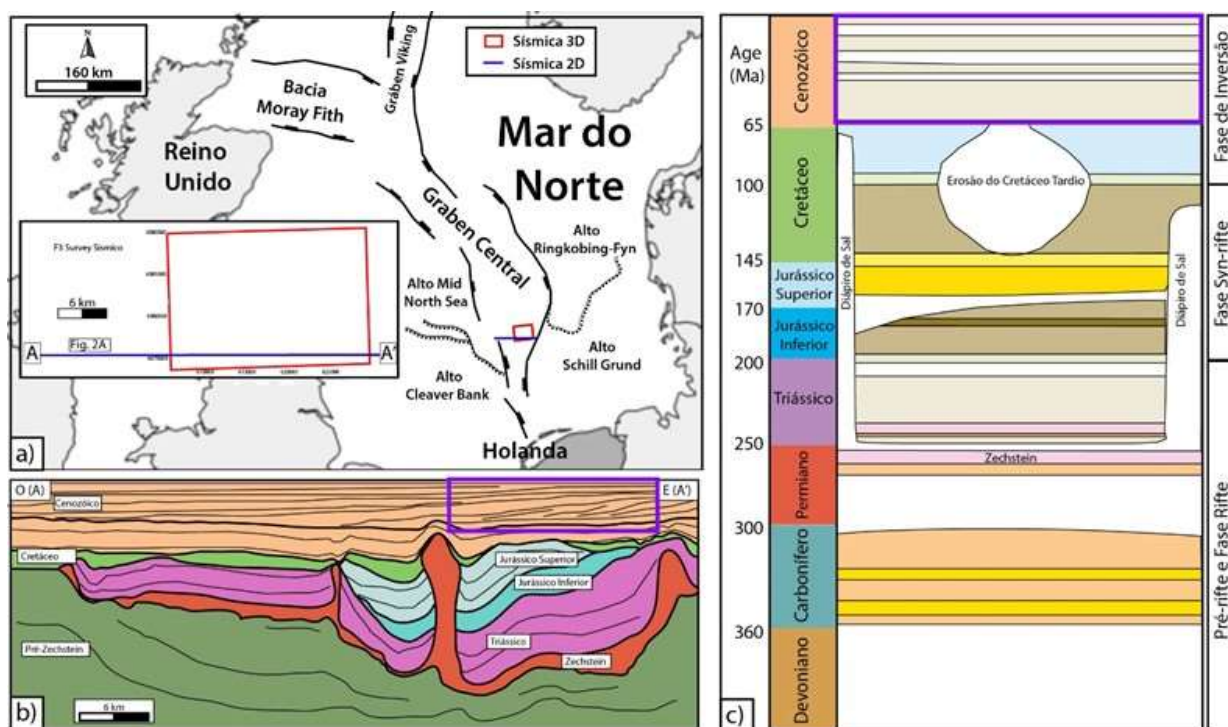


Figura 1 – (a) Localização da Bacia do Graben Central, altos estruturais e conjunto de dados utilizado nesse estudo. (b) Seção geológica regional da Graben Central holandês, baseada em dados sísmicos 2D, após Rosendall et al. (2014).

O polígono roxo delimita a área de estudo. (c) Carta estratigráfica simplificada do Graben Central holandês (modificado de Jakobsen et al., 2020). Note o hiato sedimentar nas bordas devido aos diápiros de sal. O polígono roxo indica o intervalo temporal dos estratos aqui estudados.

Fonte: Autores (2024).

2. Contexto geológico e conjunto de dados

A Bacia do Graben Central (Figura 1a) está situada na porção offshore a nordeste do território holandês, possui uma área de aproximadamente 25.000 km² e é considerada o membro mais ao sul do sistema rifte que contextualiza a região do Mar do Norte geologicamente, atingindo profundidades de até 9 km em seu depocentro (WIJKER, 2014). É circundada por uma série de altos estruturais como o Alto Cleaver Bank a sudoeste, Alto do Mar do Norte Central a oeste, Alto Ringkobing-Fyn a leste e o Alto Schill Grund a sudeste (ROSENDAAL et al., 2014).

A evolução tectonoestratigráfica da Bacia do Graben Central (Figura 1b e Figura 1c) é complexa e tem várias fases, incluindo rifteamento, intensa halocinese e inversões tectônicas (BOUROULLEC *et al.*, 2018; MAUNDE & ALVES, 2022). É dominada pelo rifteamento que ocorreu principalmente durante o período Mesozóico, com uma fase pós-rifte Cenozóica.

O Grupo Pré-Zechstein é composto por sedimentos do Carbonífero depositados em ambiente lacustrino e por arenitos continentais do Permiano. Durante o Permiano Superior ocorreu a deposição de cíclicos evaporitos marinhos que formam o grupo Zechstein (MÜLLER *et al.*, 2022), que é responsável pelo desenvolvimento de diápiros de sal e pela halocinese durante os períodos Jurássico e Cretáceo (Figure 1b e 1c).

Iniciado no período Triássico, o sistema rifte se estabilizou entre o Jurássico e o Cretáceo Inferior com o advento da fase extensional do Kimmeridgiano relacionada à abertura do Oceano Atlântico. A partir do Cretáceo Superior até os dias atuais, a fase rifte foi seguida por uma fase tipo sag, principalmente caracterizada pela quiescência tectônica e pela subsidência da bacia, com exceção da presença de alguns pulsos tectônicos que ocorreram entre o Cretáceo Superior e o período Cenozóico, que produziu uma série de falhas associadas a inversões tectônicas (MAUNDE & ALVES, 2022). Durante o Cenozóico, deltas progradacionais e sistemas de talude se desenvolveram em períodos de queda do nível do mar, formando sucessões ricas em areias que vieram a acumular gás, que são os principais objetos de estudo desse trabalho.

Para estudar esse intervalo, o dado sísmico F3 Block foi utilizado. Ele corresponde a um dado sísmico *streamer* PSTM 3D com aproximadamente 386 km², contendo 651 inlines e 951 crosslines. O levantamento possui a seguinte geometria: espaçamento entre inlines de 25 metros, espaçamento entre crosslines de 25 metros e amostragem a cada 4 ms. Possui 1,848 segundos de registro (SILVA, 2019).

3. Atributos sísmicos

A estimativa das propriedades físicas das rochas através da aquisição e processamento de dados sísmicos e a análise das suas variações verticais e laterais nos domínios do tempo, espaço e frequência constituem as bases para a interpretação sísmica (NANDA, 2021). Conforme dito por Taner *et al.* (1979), atributos sísmicos podem ser definidos como quaisquer observações extraídas dos dados sísmicos que direta ou indiretamente dão suporte a exploração de hidrocarbonetos. Além disso, o cálculo de atributos sísmicos podem ser vistos como a aplicação de filtros que removem componentes particulares do sinal sísmico para dar ênfase a outras (BARNES, 2016). Essas quantidades contribuem para a interpretação sísmica ao revelar feições estruturais previamente escondidas, como falhas e fraturas (HESTHAMMER & FOSSEN, 1997) e os limites do embasamento cristalino; ao indicar terminações de estratos; ao realçar a continuidade dos refletores sísmicos; ao discretizar fácies sísmicas (BAGHERI & RIAHI, 2015); detectar a presença de hidrato de gás (CLAIRMONT *et al.*, 2021), assim como a ajudar no mapeamento de anomalias de DHI (RODEN & CHEN, 2017).

Atributos sísmicos são frequentemente utilizados como dados de entrada para diferentes algoritmos cujo propósito é realizar o reconhecimento de padrões ou a análise de agrupamentos. De acordo com Barnes e Laughlin (2002), a acurácia dos resultados do reconhecimento de padrões está primariamente relacionado à seleção do conjunto de atributos sísmicos a serem empregados. Exceto pelo crescente custo computacional, não existem limitações para o número de atributos sísmicos para a análise de agrupamentos. No entanto, o exagero nessa escolha pode ser danoso ao resultado. Pode reduzir a interpretabilidade do método e favorecer a redundância no conjunto de dados (BARNES, 2016).

Para não promover uma busca exaustiva do conjunto ideal de atributos sísmicos através de testes sucessivos utilizando a combinação de dezenas de atributos desenvolvidos até hoje, a decisão sobre qual usar nesta pesquisa é iniciada pela conexão entre o conhecimento da resposta esperada de cada atributo para a identificação das anomalias de DHI e as orientações encontradas na literatura especializada. (RODEN & CHEN, 2017). Uma estratégia similar pode ser encontrada em Ismail *et al.* (2023), em que os autores avaliam atributos geométricos e de curvatura que fornecem informações detalhadas sobre descontinuidades estruturais para melhorar a interpretação de uma malha de fraturas.

Visto que os aspectos geométricos, espectrais e de amplitude se destacam na caracterização de anomalias de DHI em imagens sísmicas, o conjunto de atributos escolhidos como entrada para os algoritmos de análise de agrupamento buscam realçar características dessa natureza, como recomendado por Infante e Marfurt (2019). As anomalias de gás que dão origem às maiores acumulações na Bacia do Graben Central estão relacionadas com o intenso fraturamento das camadas subpostas aos reservatórios, formando chaminés de gás inclusive (DE BRUIN *et al.*, 2022). A coerência do sinal na vizinhança do sinal sísmico pode fornecer pistas para encontrar essas regiões. A presença desse fluido nos poros das rochas e a variabilidade litológica dentro de um sistema petrolífero também sugere que a impedância acústica relativa pode contribuir para a segmentação de anomalias de DHI. Além disso, a drástica redução no conteúdo espectral em zonas de DHI reforça o uso de atributos associados com a resposta em frequência. Por fim, as anomalias aqui tratadas estão

comumente associadas a altas amplitudes sísmicas. Portanto, foi determinado que os cubos de atributos extraídos dos dados sísmicos seriam o de Amplitude Instantânea, Impedância Acústica Relativa, Similaridade, Logaritmo do *Sweetness* e a Decomposição Espectral (25 Hz e 45 Hz). Juntas, essas feições sustentarão uma análise multiatributos para identificar anomalias de DHI classe 3, visto que as suas propriedades nos permitem verificar a consistência e a conformidade com o comportamento mergulhante das estruturas, inversão de polaridade no limite da anomalia e a presença de *flat spots* (RODEN & CHEN, 2017). Anomalias de classe 2 requerem uma análise de AVO apropriada, no entanto volumes de empilhamentos parciais não estavam disponíveis. A Tabela 1 apresenta a categoria em que cada atributo sísmico está incluído e as feições que enfatizam, enquanto a Figura 2 mostra os atributos sísmicos sobre o horizonte FS8. Esse horizonte sísmico é livremente distribuído pela dGB Earth Sciences através do portal TerraNubis e demarca o topo do reservatório FS8 (Adeoti *et al.*, 2023). Está inserido num contexto de refletores plano paralelos com alta amplitude sísmica.

Tabela 1 – Atributos sísmicos utilizados como entrada para a abordagem multiatributos.

Atributos Sísmicos	Categoria*	Aplicação
Impedância Acústica Relativa	Inversão sísmica	Anomalias de amplitude; DHI; canais; estratigrafia; litologia
Logaritmo do Sweetness	Realce de amplitudes	Anomalias de amplitude; DHI; estratigrafia
Amplitude Instantânea	Instantâneo	Anomalias de amplitude; DHI; canais; zonas de sombra; estratigrafia
Coerência	Geométrico	Continuidade; falhas e fraturas; canais; variações estratigráficas
Decomposição Espectral – 25 Hz	Frequência	Canais; zonas de sombra; conteúdo de frequência
Decomposição Espectral – 45 Hz	Frequência	Canais; zonas de sombra; conteúdo de frequência

Fonte: Segundo to Brown (1996).

4. Algoritmos de aprendizado de máquina

Aprendizado de máquina é uma subárea da inteligência artificial cujo foco está em desenvolver algoritmos que automatizem tarefas e melhorem o seu desempenho com a experiência (MITCHELL, 1997). Segundo *et al.* (2013), a maioria dos problemas de aprendizado estatísticos podem ser classificados em duas categorias: supervisionados e não supervisionados. Os algoritmos de redução de dimensionalidade aplicado e de análise de agrupamentos aplicado nesse trabalho são classificados como não supervisionados, de modo que não há associação entre uma resposta correta ou rótulo para cada amostra observada. O objetivo da análise de agrupamentos é verificar se uma as amostras de um conjunto de dados possuem características similares entre si que são capazes de distingui-las das demais. Nessa seção, apresentamos os princípios operacionais dos algoritmos de Análise de Componentes Principais (PCA), K-Médias, Modelo de Misturas Gaussianas (GMM) e Agrupamento Espacial Baseado em Densidade em Aplicações com Ruído (DBSCAN).

4.1. Principal Component Analysis (PCA)

PCA é uma ferramenta de análise de dados frequentemente utilizada para realizar redução de dimensionalidade e filtragem de dados (DEISENROTH *et al.*, 2020). É um método não paramétrico simples para extrair informações relevantes de grandes conjuntos de dados. Em outras palavras, fornece um caminho para minimizar a complexidade de um problema encontrando a forma mais significativa de expressar um conjunto de variáveis que constituem um dado sistema e descrevem um ou mais eventos. (SHLENS, 2014).

PCA realiza uma transformação linear no conjunto de dados para projetá-lo sobre um subespaço com menor dimensionalidade (LEVER *et al.*, 2017). A saída do algoritmo assegura uma representação otimizada do conjunto de dados original conservando tanta informação quanto possível numa quantidade reduzida de variáveis conhecidas como principais componentes.

Seja \mathbf{X} a matriz de dados normalizada pela média e desvio padrão, com dimensão $N \times M$, em que N e M são o número de amostras e variáveis, respectivamente. A sua matriz de covariância \mathbf{S}_x é dada pela Equação 1:

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

É possível obter a matriz \mathbf{Y} , capaz de representar \mathbf{X} no domínio das principais componentes, de modo que:

$$\mathbf{Y} = \mathbf{P}\mathbf{X}, \quad (2)$$

Em que $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_M]^T$ contém as principais componentes, e a matriz $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_M]^T$ contém os autovetores da matriz \mathbf{S}_X . A matriz \mathbf{P} é responsável por transladar e rotacionar o conjunto de atributos original para determinar a base ortogonal que maximiza a variância dos elementos de \mathbf{Y} e que melhor representa os dados no domínio das principais componentes (DEISENROTH *et al.*, 2020).

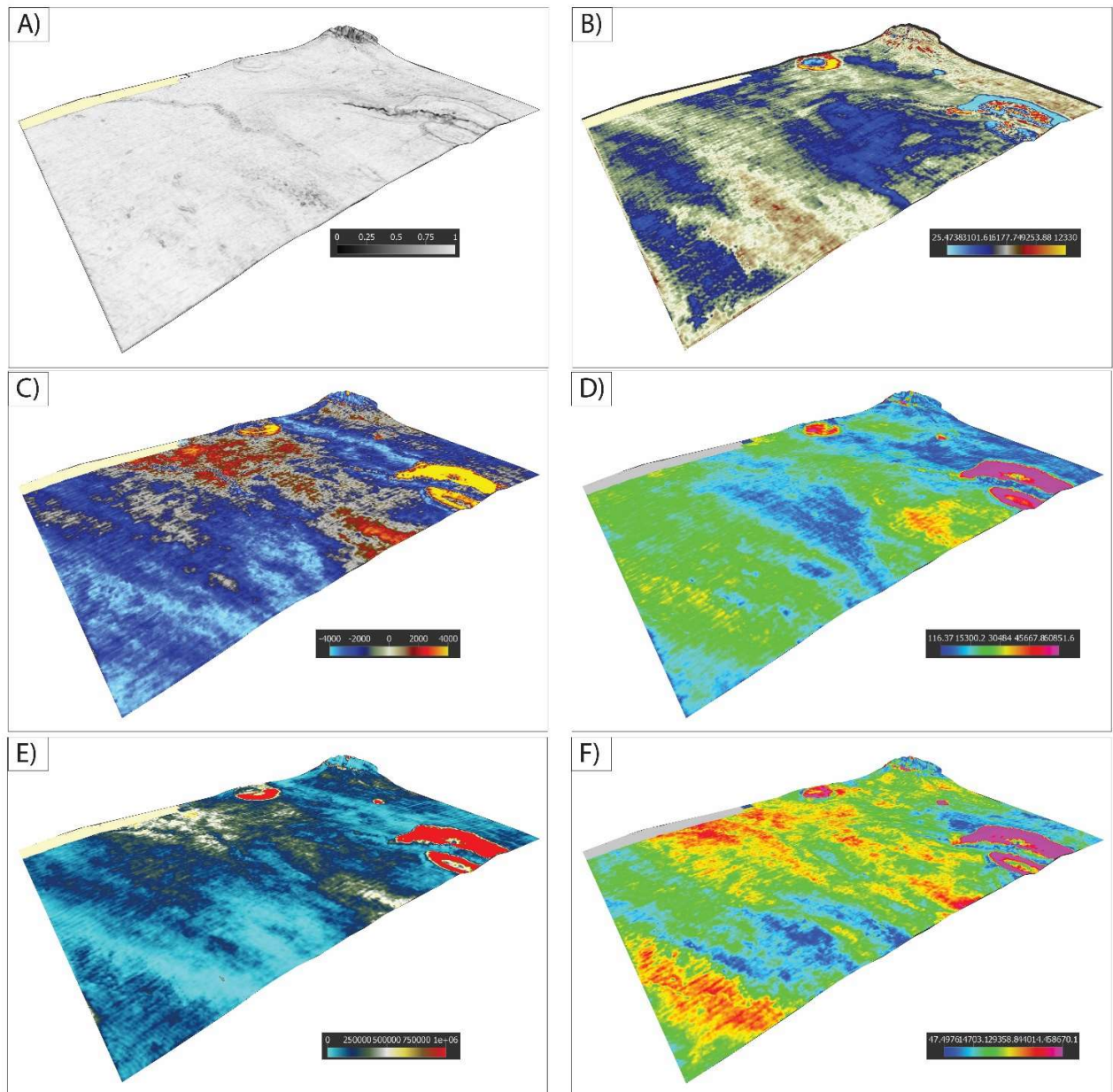


Figura 2 – Atributos sísmicos utilizados como entrada para os algoritmos de análise de agrupamento: (a) Similaridade, (b) Amplitude Instantânea, (c) Impedância Acústica Relativa, (d) Decomposição Espectral (25 Hz), (e) Logaritmo do Sweetness e (f) Decomposição Espectral (45 Hz).

Fonte: Autores (2024).

4.2. K-Médias

O algoritmo K-Médias é um método de análise de agrupamentos que visa segmentar um conjunto de dados não rotulados em um número K de grupos previamente definido. Após essa seleção, é verificado que cada ponto pertence a um e somente um *cluster* de modo que amostras similares são agrupadas e relacionadas a um centroide (JAMES *et al.*, 2013). Centroides são objetos que armazenam as medias das variáveis para cada um dos grupos (TAN *et al.*, 2016). Em outras

palavras, considerando a matriz \mathbf{X} , definida nos mesmos moldes da subseção anterior, o algoritmo K-Médias busca realizar a partição dos dados em K distintos clusters que compartilham características similares aos centroides $\boldsymbol{\mu}_K$. Esse objetivo é realizado garantindo que o centroide esteja corretamente posicionado no espaço de observações, visando minimizar a variância intragrupo dentre todos os clusters. A variância Var de um clusters C_K é a medida do quanto as suas amostras diferem entre si (JAMES *et al.*, 2013). A descrição matemática da variância intragrupo pode ser vista como a soma das distâncias Euclidianas entre cada amostra \mathbf{x}_n pertencente ao K -ésimo grupo e seu respectivo centroide C_K , dividido pelo número total de observações N . A Equação 3 apresenta a variância intragrupo conforme descrita:

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

Nessas circunstâncias, a função objetivo a ser minimizada pelo algoritmo K-Médias é definida como:

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

4.3. Modelo de Misturas Gaussianas (GMM)

O algoritmo GMM é considerado uma simples superposição linear de componentes Gaussianas que visa estabelecer um agrupamentos de modelos de densidade mais informativos do que aquele Gerado por uma única distribuição normal (BISHOP, 2006). No domínio unidimensional, uma distribuição Gaussiana de probabilidades $\mathcal{N}(x|\mu, \sigma^2)$ de uma dada variável x é dada por:

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

Em que μ e σ são a média e a variância da distribuição, respectivamente. A generalização para o domínio M-dimensional é baseado no conceito de matriz de covariância $\boldsymbol{\Sigma}$ (BISHOP, 2006). Considerando a matriz \mathbf{X} , a sua distribuição Gaussiana é definida como na Equação 6.

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

Analogamente à Equação 5, $\boldsymbol{\mu}$ é a media da matriz de dados, $\boldsymbol{\Sigma}$ e $|\boldsymbol{\Sigma}|$ são a matriz de covariância e seu módulo, respectivamente.

Segundo Deisenroth *et al.* (2020), o algoritmo GMM fornece um modelo de densidade de probabilidade cujo número finito K de distribuições normais são combinados de tal forma que a equação da distribuição de misturas Gaussianas $p(\mathbf{X})$ (Equation 7) seja satisfeita.

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

Nessa equação, π_k são os pesos designados para a ocorrência de cada distribuição normal. O propósito do algoritmo GMM é otimizar o conjunto de parâmetros $\theta := \{\pi_k, \mu_k, \Sigma_k: k = 1, \dots, K\}$ para maximizar a função de verossimilhança através do método de Maximização-Expectativa descrito na Equação 8:

$$\log(p(\mathbf{X}|\mu_k, \Sigma_k)) = \sum_{n=1}^N \log(\sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{X}|\mu_k, \Sigma_k)). \quad (8)$$

É possível calcular a probabilidade de uma amostra pertencer a uma das distribuições normais incluídas no modelo de misturas. Para tal, a quantidade

$$r_{nk} = \frac{\pi_k \mathcal{N}(\mathbf{X}|\mu_k, \Sigma_k)}{\sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{X}|\mu_k, \Sigma_k)} \quad (9)$$

É definida como a *responsabilidade* da K-ésima componente da mistura sobre a enésima instância na distribuição. A responsabilidade, definida como na Equação 9, é diretamente proporcional à probabilidade da enésima amostra da distribuição pertencer à K-ésima componente da mistura. Portanto, quanto maior for a responsabilidade de uma componente Gaussiana sobre uma amostra, mais provável será que aquela amostra pertença àquela componente do modelo de misturas Gaussianas (BISHOP, 2006).

4.4. Agrupamento Espacial Baseado em Densidade em Aplicações com Ruídos (DBSCAN)

Inicialmente introduzido por Ester *et al.* (1996), o algoritmo DBSCAN parte do princípio que uma amostra pertence a um agrupamento somente se a sua vizinhança, definida por um raio ϵ , contém um número mínimo de pontos (MinPts). Em outras palavras, a densidade espacial dos pontos na sua vizinhança deve exceder um limiar. Esse conceito é aplicado ao reconhecimento de padrões num conjunto de dados para segmentá-lo em subgrupos informacionais com formatos arbitrários, considerando a típica relação entre uma nuvem de pontos e a densidade espacial dos pontos ruidosos (ESTER *et al.*, 1996). No mesmo trabalho, os autores definiram conceitos como a ϵ -vizinhança, pontos alcançáveis e diretamente alcançáveis, pontos núcleo, entre outros. Esses conceitos são fundamentais para a compreensão de como o DBSCAN é capaz de distinguir agrupamentos genuínos de amostras ruidosas.

5. Metodologia

A Figura 3 apresenta o fluxo de trabalho utilizado para todos os experimentos, incluindo as análises de agrupamento feitas com o K-Médias e GMM, bem como a subsequente filtragem por densidade espacial. Todo o workflow foi implementado em Python e os software OpendTect foi utilizado somente para visualização dos resultados. Etapas chave desse processo será brevemente discutido abaixo.

5.1. Seleção do intervalo de estudo

Embora fosse possível utilizar todas as amostras do cubo sísmico original, nesta abordagem, a análise de agrupamentos foi conduzida dentro de uma área de interesse delimitada por dois horizontes sísmicos. Além de reduzir o custo computacional, esse foco é influenciado por fatores geológicos. Por ventura, Schroot e Schüttenhelm (2003) declararam que as principais acumulações de hidrocarbonetos no Mar do Norte estão associados com sedimentos clásticos do Mioceno. Portanto, o limite superior do nosso estudo foi o horizonte FS8, enquanto o limite inferior foi o horizonte MFS4. Utilizamos 25 amostras sobre o horizonte FS8 e a mesma quantidade abaixo do horizonte MFS4 para delimitar a área de interesse, ou seja, 100 ms abaixo e cima desses limites.

5.2. Pré-processamento dos dados

Visando eliminar amostras nulas e valores espúrios, isto é, amostras sem significado geofísico geológico. Para cada atributo sísmico selecionado, analisamos a amplitude interquartil da distribuição, buscando avaliar a dispersão estatística das amostras no conjunto de dados ao redor da mediana.

5.3. Análise de Componentes Principais (PCA)

Uma vez que os dados foram pré-processados, o conjunto de dados normalizado pela média e variância se tornou a entrada para a redução de dimensionalidade utilizando o algoritmo PCA. Após avaliar a razão de variância explicada, verificamos que quatro componentes principais podem representar 92,83% do conjunto de dados original, como apresentado na Figura 4.

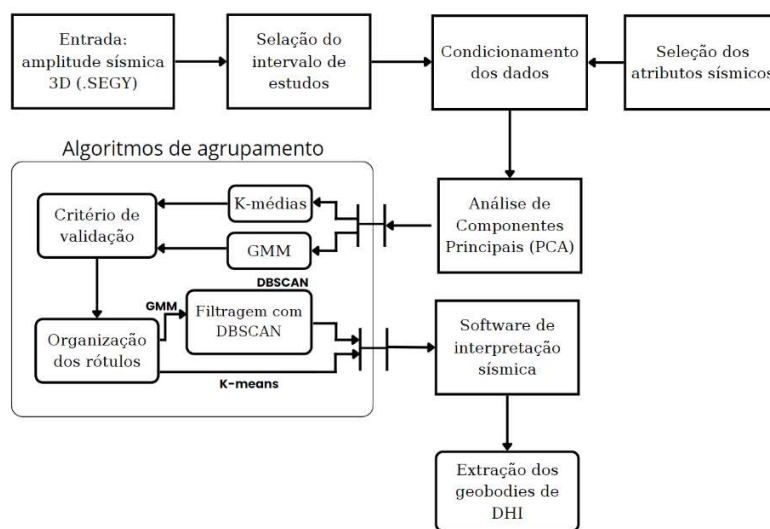


Figura 3 – Fluxo de trabalho aplicado para executar a análise de agrupamentos de atributos sísmicos num conjunto de dados 3D.

Fonte: Autores (2024).

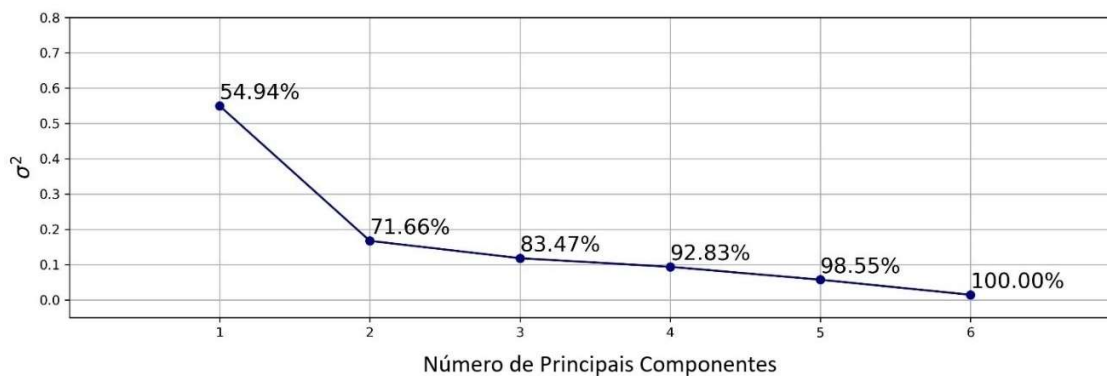


Figura 4 – Variância percentual normalizada de cada principal componente e a variância percentual acumulada aplicada para determinar o número total de principais componentes.

Fonte: Autores (2024).

5.4. Algoritmos de agrupamento

A análise multiatributos é o primeiro passo da abordagem combinada utilizada nesse trabalho. 60% das amostras do conjunto de dados foram aleatoriamente selecionadas para compor os dados de treinamento dos modelos K-Médias e GMM. Em seguida, os rótulos das amostras restantes foram inferidos. Finalmente, de acordo com a metodologia proposta

por Troccoli *et al.* (2022), os rótulos obtidos foram reorganizados com respeito à origem para preservar algum nível de similaridade entre os centroides mais próximos, adotando a distância Euclidiana como métrica de comparação.

5.5. Critérios de validação

Utilizar testes estatístico para estimar o potencial número de agrupamentos que otimiza a segmentação do conjunto de dados. Uma vez que o método do cotovelo não apresenta uma indicação satisfatória desse hiperparâmetro devido à natureza do conjunto de dados, o Índice de Davies-Bouldin foi também calculado (Figura 5). Ele sugeriu $K = 5$ como o hiperparâmetro ótimo para ambos os algoritmos. No entanto, embora o critério de validação seja uma etapa essencial para esse procedimento metodológico, devemos considerar os aspectos geológicos, assim como o objetivo de cada trabalho. A visualização do resultados dos agrupamentos indicaram que cinco grupos não foram capazes de isolar as anomalias de DHI. Utilizar $K = 7$ cumpriu melhor com o objetivo: isolar anomalias de DHI, além de apresentar alta correlação com as feições geomorfológicas da área.

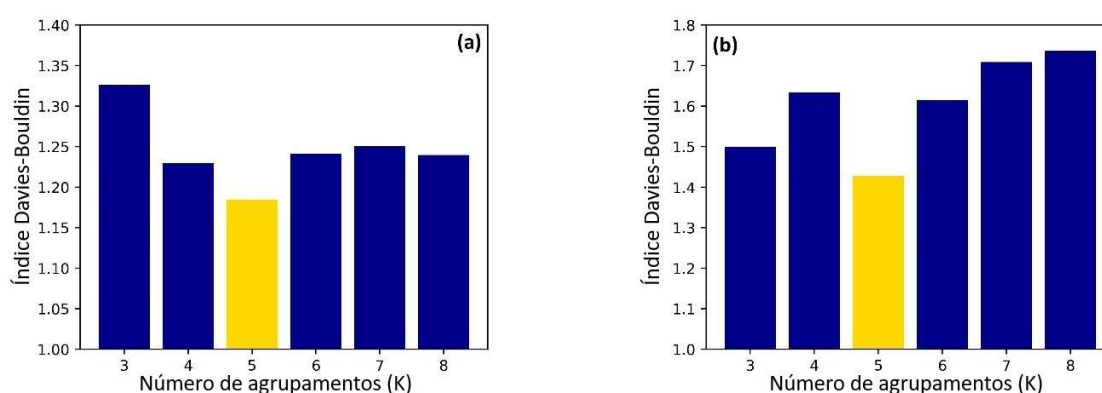


Figura 5 – Índice Davies-Bouldin por número de agrupamentos. $K = 5$ minimiza a métrica para ambos os algoritmos (a) K-Médias e (b) Modelo de Misturas Gaussianas, sugerindo que essa é a quantidade ótima de clusters.

Fonte: Autores (2024).

5.6. Filtragem Baseada em Densidade Espacial

Nessa segunda etapa, o algoritmo DBSCAN visou compreender a distribuição espacial das amostras do cluster interpretado como DHI, possibilitando a filtragem de pontos espúrios e zonas com baixa densidade espacial. Durante essa fase, os dados de entrada consistem das coordenadas espaciais e temporais normalizadas, variando entre 0 e 1, das amostras interpretadas como DHI. Uma série de experimentos foram conduzidos para refinar os hiperparâmetros do modelo, resultando num raio de ϵ -vizinhança de 0,01 e um número mínimo de 500 pontos para o desempenho ótimo. Sob essas condições, o algoritmo determinou o número de subgrupos, a contagem de amostras em cada subgrupo e a identificação das amostras espúrias como ruídos. Esses subgrupos foram organizados em ordem decrescente de número de amostras, no qual os 12 primeiros foram rotulados como DHI, enquanto as demais amostras foram categorizadas como ruído. É importante notar que esse limiar numérico pode variar a depender da região de estudo. Consequentemente, a expectativa é que o cluster de DHI será completamente isolado dentro das zonas anômalas.

6. Resultados e discussões

Esta seção apresenta os resultados advindos dos agrupamentos feitos tanto pelo algoritmo K-Médias quanto do GMM. Também fornece as discussões adequadas acerca das aplicações de cada técnica para mapear indicadores diretos de hidrocarbonetos e a associação entre análise de agrupamentos multiatributo com a filtragem baseada em densidade espacial. Uma vez que as etapas de pré-processamento, a redução de dimensionalidade e a validação foram completadas, uma análise visual dos resultados feita através da variação do número de grupos entre três e oito. Os modelos otimizados possuem a seguinte configuração (Tabela 2).

Tabela 2 – Parâmetros dos modelos otimizados.

Algoritmo	Atributos Sísmicos	Principais Componentes	Número de Agrupamentos
K-Médias	Logaritmo do <i>Sweetness</i> , Similaridade, Impedância Acústica Relativa, Amplitude Instantânea, Decomposição Espectral (25 e 45 Hz)	4	5
GMM	Logaritmo do <i>Sweetness</i> , Similaridade, Impedância Acústica Relativa, Amplitude Instantânea, Decomposição Espectral (25 e 45 Hz)	4	7

Fonte: Autores (2024).

A resposta acústica relacionada à presença de hidrocarbonetos caracteriza um DHI quando associado com uma configuração de aprisionamento, identificada, nesse caso, por zonas com sobras de baixas frequências ou efeitos de *pull-down* da velocidade (RODEN & CHEN, 2017). Schroot e Schüttenhelm (2003) afirmam que o Mar do Norte holandês apresenta uma série de fenômenos relacionados à acumulações de gás cuja expressão sísmica se assemelha àquelas supracitadas. Anomalias de amplitude relacionadas a fraturas e falhas, chaminés de gás, marcas de arraste de gelo preenchidas com gás e, principalmente, *bright spots* estão entre os eventos sísmicos interpretados na área (SCHROOT & SCHÜTTENHELM (2003); SCHROOT *et al.*, (2005); CONNOLLY (2015)).

Analisando o resultado do agrupamento sobre o horizonte FS8 (Figure 6b), foi observado que o algoritmo K-Médias, seguido pela organização de rótulos, designou mais de um rótulo (0 e 2) para duas zonas contendo *bright spots* e anomalias sísmicas relacionadas a falhamentos, conforme interpretado anteriormente por alguns autores (SCHROOT & SCHÜTTENHELM (2003); SCHROOT *et al.*, (2005); CONNOLLY (2015); DE BRUIN *et al.*, (2022)). Além disso, o grupo 2 se estende por grande parte do horizonte FS8, o que sugere que esse cluster foi definido predominantemente pela similaridade entre suas amostras com relação aos atributos de amplitude, como a Amplitude Instantânea e o Logaritmo do *Sweetness*, em detrimento do conjunto de aspectos que caracteriza DHIs. Esse comportamento se repete ao longo do intervalo analisado. Visto que se trata de um algoritmo de agrupamento baseado em protótipo, todos os membros do cluster associado com um dado centroide deve estar próximo ao seu protótipo correspondente (PATEL & KUSHWAHA, 2020). Para o K-Médias, a distribuição ótima das amostras no espaço consiste em grupos esféricos bem separados entre si. Portanto, a construção do problema de minimização do algoritmo sugere dificuldades no agrupamento de dados que se sobrepõem no espaço quadridimensional das principais componentes utilizadas nos experimentos. Assim, o algoritmo se mostrou ineficiente em isolar as anomalias de DHI na região.

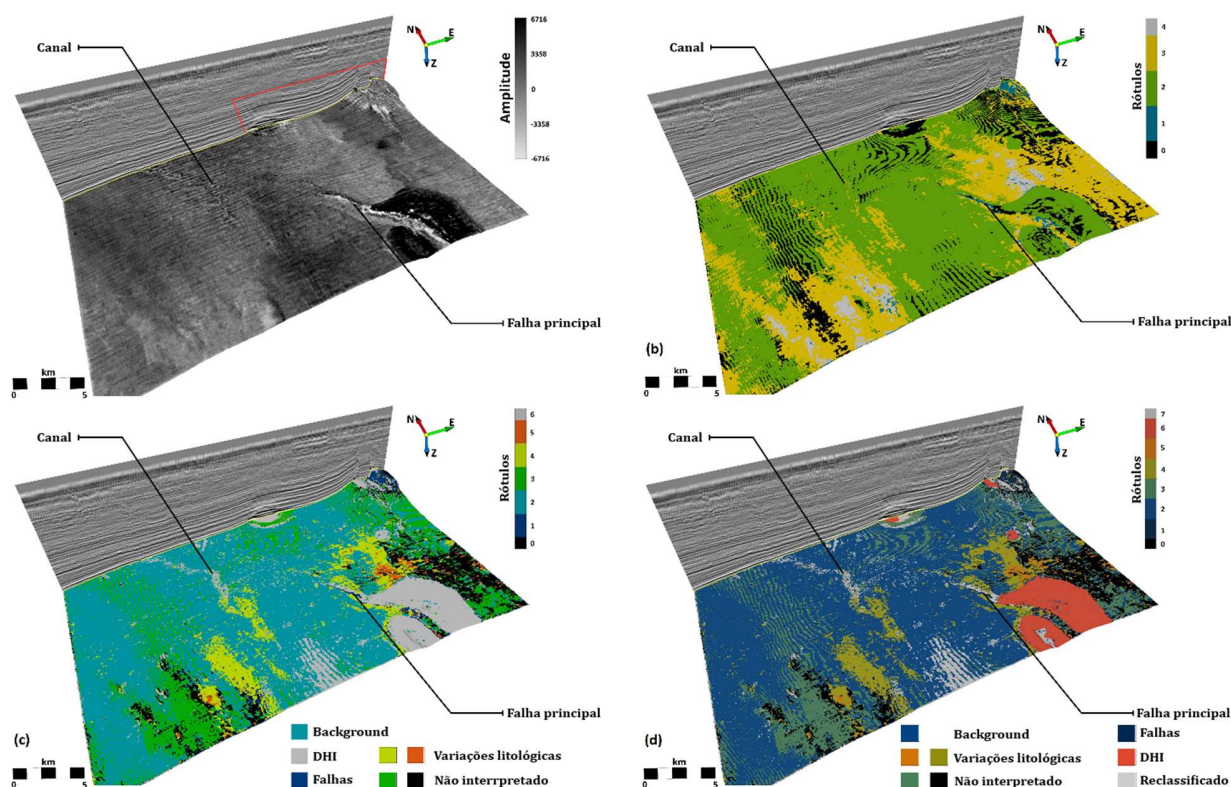


Figura 6 – (a) Amplitude sísmica original; resultado dos agrupamentos derivados do (b) K-Médias, (c) Modelo de Misturas Gaussianas e (d) a aplicação conjunta do GMM e da filtragem baseada em densidade espacial. Fonte: Autores (2024).

As expectativas criadas pela robustez do algoritmo Modelo de Misturas Gaussianas começaram a se concretizar através da observação do resultado do agrupamento sobre o horizonte sísmico, como pode ser visto na Figura 6c. Após a organização dos rótulos com relação à origem, às amostras da classe interpretada como DHI foram designadas o rótulo 6, visto em cinza. É possível notar que o GMM foi mais efetivo em delimitar as zonas com anomalias de DHI que o K-Médias. Baseado nas estimativas de probabilidade de densidade, cada agrupamento é modelado como uma distribuição gaussiana com suas particulares médias e desvios padrões, o que garante que o GMM fornecerá uma melhor medida quantitativa do ajuste por número de clusters (PATEL & KUSHWAHA, 2020). Como consequência disso está o comportamento da classe interpretada como DHI na Figura 6c. As anomalias mais proeminentes estão confinadas nessa classe, de modo que o conjunto de atributos utilizados para construir o hiperespaço das principais componentes e que realçam características estruturais, de fase, frequência e amplitude parecem ter tido forte influência na segmentação desse fenômeno.

Mesmo que esse não tenha sido o principal objetivo desse trabalho, o modelo GMM permitiu correlacionar clusters a feições interessantes como variações litológicas, falhas e fraturas. Essas últimas estão relacionadas à classe 1, vista em azul escuro, como pode-se notar na falha principal e na região no extremo nordeste do horizonte sísmico, uma zona elevada pelo intenso fraturamento devido à halocinése da Formação Zechstein. (MAUNDE & ALVES, 2022). Isso se torna ainda mais claro na Figura 7, onde uma seção vertical – Inline 668 – dos resultados é apresentada. É uma janela indicada pelo polígono vermelho na mesma inline que aparece na Figura 6a. Podemos observar que a presença do cluster interpretado como DHI está frequentemente associado ao sistema de falhas, juntamente com a classe 1. Isso pode indicar rotas de migração das fontes mais profundas para anomalias de DHI associadas com acumulações rasas de gás, em acordo com a hipótese que estabelece a relação entre essas acumulações e estruturas mais profundas, como discutido por De Bruin *et al.* (2022). Por outro lado, amostras com a mesma classe aparecem dispersas em regiões que não necessariamente estão relacionadas a falhas e fraturas. Isso nos leva a entender a possibilidade zonas de baixa similaridade influenciam fortemente a formação desse cluster, visto que o modelo não foi otimizado com esse propósito.

Como mencionado anteriormente, a organização dos rótulos designa uma característica similar a clusters representados por cores próximas na paleta de cores. Esse efeito pode ser observado nos grupos 4 e 5, em amarelo e laranja (Figure 6c), possivelmente relacionados a variações litológicas.

Embora o objetivo de isolar anomalias de DHI foi razoavelmente alcançado pelo modelo GMM, ainda parecem haver regiões cuja distribuição não ratifica a sua classificação dentro do cluster objetivo. O aparente paleocanal pode ser um exemplo desse inconveniente. O efeito do algoritmo DBSCAN fica claro na Figura 6d, onde é possível observar que tanto as amostras associadas com o paleocanal quanto às marcas de arraste de gelo preenchidas com gás, que nesse caso apresentam o típico padrão de lineamentos alongados com orientação N-S (SCHROOT & SCHÜTTENHELM, 2003), foram genericamente rotulados novamente como “reclassificado” e representados em cinza (agora, grupo 7). Aqui, pode ser observado que a aplicação conjunta de algoritmos não supervisionados foi capaz de distinguir entre anomalias caracterizadas como *bright spots* e outras manifestações sísmicas relacionadas à presença de gás. Isso ocorre devido ao fato de que essas manifestações são regiões contendo baixa densidade especial de amostras. Na Figura 7, um efeito similar ocorre na zona de fraturamentos. Além disso, o DBSCAN pode também fornecer um resultado com remoção de ruídos, visto que algumas amostras dispersas, vistas em tempos duplos superiores a 800 ms, foram visivelmente reclassificadas. Esse é um fator chave para a visualização tridimensional, como visto na Figura 8.

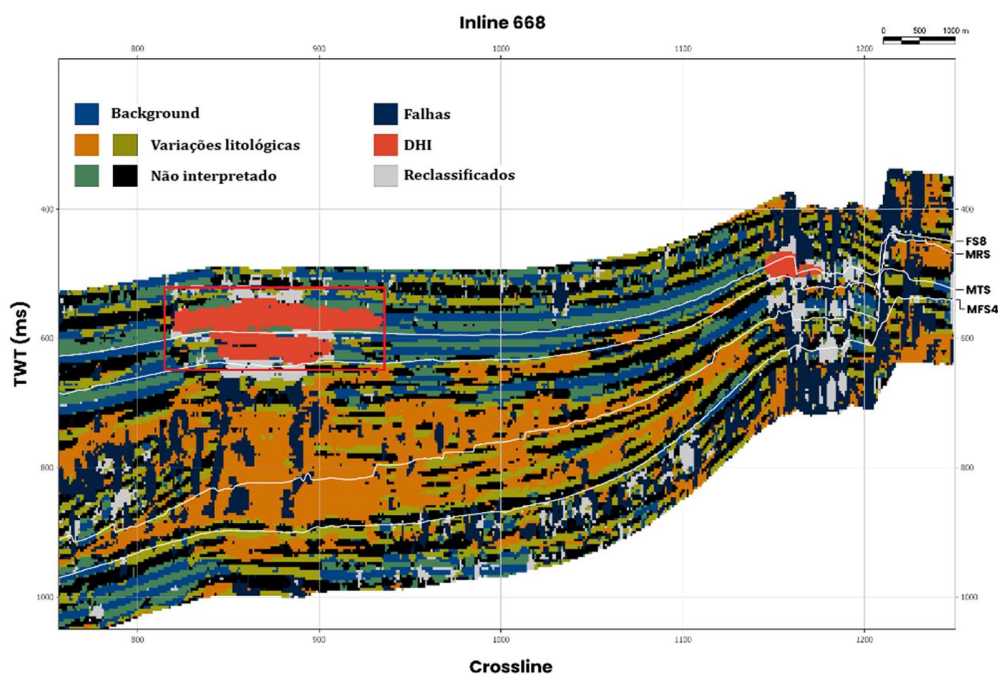


Figura 7 – Resultados do agrupamento visto numa janela da inline 668 obtido a partir da aplicação conjunta da filtragem baseada em densidade espacial em amostras rotuladas como DHI pelo algoritmo GMM.

Fonte: Autores (2024).

De uma perspectiva em três dimensões, as amostras rotuladas como DHI pelo modelo GMM poluem toda a extensão do dado sísmico, dificultando a visualização 3D do alvo desse trabalho, como pode ser visto na Figura 8a. A melhoria na visualização das anomalias de DHI nessa perspectiva após a filtragem baseada em densidade espacial é evidente (Figura 8b). Os geobodies compostos pelas amostras com o rótulo interpretado como DHI apresenta um aspecto muito mais claro, em comparação com o resultado anterior.

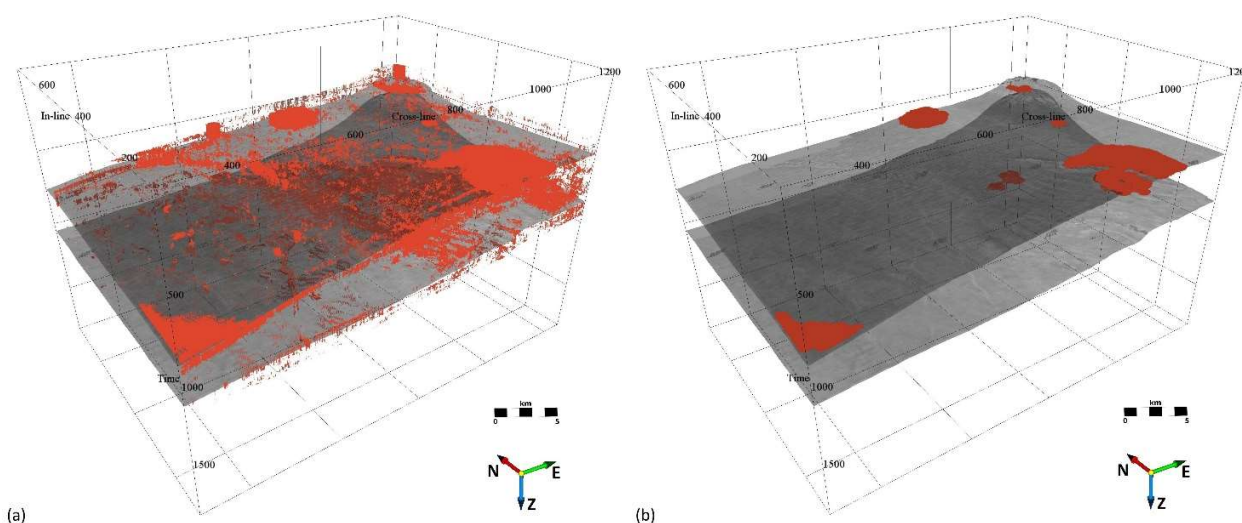


Figura 8 – O resultado da aplicação da filtragem baseada em densidade espacial através do algoritmo DBSCAN sobre as amostras do cluster 6. Os dois horizontes sísmicos foram adicionados para corroborar com a ideia da localização espacial de cada anomalia.

Fonte: Autores (2024).

Considerações finais

Esse trabalho introduz uma metodologia para a análise de agrupamento em dados sísmicos 3D para mapear indicadores diretos de hidrocarbonetos (DHIs). Pela aplicação de dois algoritmos não supervisionados de aprendizado de máquina, K-Médias e Modelo de Misturas Gaussianas (GMM), a pesquisa examina os desempenhos individuais de cada técnica e explora os benefícios de integrar a abordagem multiatributos com a filtragem baseada em densidade espacial através do algoritmo DBSCAN.

Observamos que utilizar testes estatísticos, tais quais o método do cotovelo ou o Índice Davies-Bouldin, para determinar o número ótimo de clusters para os modelos não foi eficiente, ao menos quando o objetivo do estudo é segmentar um evento geológico específico ou feições como anomalias de DHI. Possivelmente, em abordagens mais exploratórias, otimizar esse hiperparâmetro através de testes estatísticos poderia ser um bom ponto de partida.

O algoritmo K-Médias apresentou limitações em delimitar com acurácia as anomalias de DHI. Designou mais de uma classe a regiões com a presença de DHIs já conhecidos. Essas fácies estão distribuídas por grande extensão do dado sísmico, sugerindo que essas classes foram definidas somente pela similaridade do conteúdo de amplitude das suas amostras, em detrimento às características expressadas por DHIs. Em contraste, o Modelo de Misturas Gaussianas atingiu bons resultados em identificar o alvo desse trabalho. Com um modelo composto por sete clusters, foi capaz de delimitar com precisão as anomalias de DHI. Além disso, realçou com eficiência outras feições geológicas como falhas e fraturas, paleocanais e grupos relacionados à variações litológicas.

A subsequente aplicação da filtragem baseada em densidade espacial sobre as amostras classificadas pelo algoritmo GMM e interpretadas como DHI – o coração do agrupamento conjunto – identificou subgrupos com baixa densidade espacial de amostras. Assim, o algoritmo DBSCAN ofereceu condições para atualizar os rótulos dessas instâncias e aumentar a acurácia do mapeamento das anomalias. Baseado nesses resultados, geobodies puderam ser gerados, e a perspectiva tridimensional se tornou livre de ruídos para interpretar esse grupo. Essa metodologia pode ser adaptada e aplicada em outras bacias sedimentares para ajudar geofísicos a interpretar anomalias sísmicas causadas pela presença de possíveis anomalias de gás e óleo.

Agradecimentos

Os autores agradecem ao Grupo de Estudo e Aplicação de Inteligência Artificial em Geofísica da Universidade Federal da Bahia (GAIA-UFBA) e ao Instituto de Geociências da UFBA pela infraestrutura fornecida. Gostaríamos de agradecer

também à dGB Earth Sciences por tornar esse conjunto de dados disponível como um projeto do OpendTect através do portal TerraNubis terranubis.com. Finalmente, também agradecemos ao Instituto Nacional de Ciência e Tecnologia de Geofísica do Petróleo (INCT-GP) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) pela bolsa do autor 1, à Fundação de Amparo à Pesquisa do Estado da Bahia pela bolsa do autor 2, e ao CNPQ por financiar o projeto de número 409718/2022-0.

Referências

- 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.; KÖRÖSI, 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 Society 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.; ABDELMAKSOU, 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.. *Pattern 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.