



Universidade do Estado do Rio de Janeiro

Centro de Tecnologia e Ciências

Faculdade de Geologia

Victor Lebre Fiaux Rodrigues

**Análise multiatributos sísmicos e aprendizado de máquinas aplicados à
detecção de rochas ígneas no pré-sal da Bacia de Santos**

Rio de Janeiro

2023

Victor Lebre Fiaux Rodrigues

Análise multiatributos sísmicos e aprendizado de máquinas aplicados à detecção de rochas ígneas no pré-sal da Bacia de Santos

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Geociências, da Universidade do Estado do Rio de Janeiro. Área de concentração: Geofísica Aplicada.

Orientador: Prof. Dr. Webster Ueipass Mohriak

Coorientador: Dr. Cesar José Calderón Filho

Rio de Janeiro

2023

CATALOGAÇÃO NA FONTE
UERJ / REDE SIRIUS / BIBLIOTECA CTC/C

R696 Rodrigues, Victor Lebre Fiaux.
Análise multiatributos sísmicos e aprendizado de máquinas aplicados à detecção de rochas ígneas no pré-sal da Bacia de Santos. / Victor Lebre Fiaux Rodrigues. – 2023.
157 f. : il.

Orientador: Webster Ueipass Mohriak.
Coorientador: Cesar José Calderón Filho.
Dissertação (Mestrado) – Universidade do Estado do Rio de Janeiro, Faculdade de Geologia.

1. Geologia estratigráfica – Teses. 2. Rochas ígneas – Santos, Bacia de – Teses. 3. Reservatórios de hidrocarbonetos – Teses. 4. Pré-sal - Santos, Bacia de - Teses. I. Mohriak, Webster Ueipass. II. Calderón Filho, Cesar José. III. Universidade do Estado do Rio de Janeiro. Faculdade de Geologia. IV. Título.

CDU: 551.7(815.6)

Bibliotecária Responsável: Priscila Freitas Araujo/ CRB-7: 6390

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada a fonte.

Assinatura

Data

Victor Lebre Fiaux Rodrigues

Análise multiatributos sísmicos e aprendizado de máquinas aplicados à detecção de rochas ígneas no pré-sal da Bacia de Santos

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Geociências, da Universidade do Estado do Rio de Janeiro. Área de concentração: Geofísica Aplicada

Aprovada em 27 de julho de 2023.

Orientador: Prof. Dr. Webster Ueipass Mohriak
Faculdade de Geologia - UERJ

Coorientador: Prof. Dr. Cesar José Calderón Filho
Petróleo Brasileiro S.A. - Petrobras

Banca examinadora: _____

Prof. Dr. André Luis Albuquerque dos Reis (Titular)
Faculdade de Geologia - UERJ

Dr^a. Ana Krueger (Titular)
Bluware Inc.

Prof. Dr. Leonardo Costa de Oliveira (Titular)
Petróleo Brasileiro S.A. - Petrobras

Rio de Janeiro

2023

AGRADECIMENTOS

Aos meus pais e à minha prima Kívia;

Ao meu Orientador e Co-Orientadores;

À Banca examinadora;

Ao PPGG e especialmente, à Marianni, pelo apoio.

À Petrobras e aos meus gerentes e coordenadores que me deram essa oportunidade,
Marília, Diogo e Marcão.

À Petrobras/CIMANTEC/Ogbon pela infraestrutura.

RESUMO

RODRIGUES, Victor Lebre Fiaux. *Análise multiatributos e aprendizado de máquinas aplicados à detecção de rochas ígneas no pré-Sal da Bacia de Santos*. 2023.157 f. Dissertação (Mestrado em Geociências) – Faculdade de Geologia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2023.

Os reservatórios carbonáticos da seção pré-sal estão entre as descobertas recentes mais importantes das últimas décadas, transformando a Bacia de Santos num dos principais focos exploratórios de hidrocarbonetos no Brasil. Estes reservatórios são marcados por intercalações com intrusões vulcânicas, de diversas idades. A presença dessas rochas diminui o potencial das reservas e pode representar riscos à etapa de perfuração. A alta profundidade da seção de interesse, e conseqüente reduzida resolução sísmica, associada às características petrofísicas das intrusões ígneas, tornam sua identificação desafiadora. Técnicas de classificação automática, na forma de algoritmos de agrupamento, permitem a análise simultânea de dados de alta dimensionalidade, agrupando-os em classes que compartilham características ao longo dessas dimensões. No contexto dos dados de sísmica de reflexão, a alta dimensionalidade é alcançada a partir da extração de diversos atributos sísmicos com o objetivo de identificar diferentes fácies sísmicas, com o potencial de identificar feições não visíveis no dado de amplitude. O potencial de três algoritmos, K-Means, Mapas Auto-organizáveis de Kohonen e Modelos de Mistura Gaussiana, é avaliado na identificação de fácies correspondentes à presença das intrusões ígneas em dados sísmicos da região do Campo de Mero, com o objetivo de desenvolver uma rotina para aplicação em outras áreas de contexto geológico similar, em frente exploratória, quando não há disposição de processamentos sísmicos avançados, na forma de inversões sísmicas elásticas ou outros processamentos de alto custo computacional, que possa ser realizado pelo geocientista de exploração, em sua própria estação de trabalho.

Palavras-chave: aprendizado de Máquinas; algoritmos de agrupamento; Bacia de Santos.

ABSTRACT

RODRIGUES, Victor Lebre Fiaux. *Seismic multi-attribute analysis and machine learning applied to the detection of igneous rocks in the pre-salt Santos Basin*. 2023. 157 f. Dissertação (Mestrado em Geociências) – Faculdade de Geologia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2023.

The carbonate reservoirs of the pre-salt section are among the most important recent discoveries in recent decades, transforming the Santos Basin into one of the main exploration centers for hydrocarbons in Brazil. These reservoirs are marked by intercalations with volcanic intrusions of different ages. The presence of these rocks reduces the potential of the reserves and may pose risks to the drilling stage. The high depth of the section of interest, and consequent reduced seismic resolution, associated with the petrophysical characteristics of igneous intrusions, make its identification challenging. Automatic classification techniques, in the form of clustering algorithms, allow the simultaneous analysis of high-dimensional data, grouping them into classes that share characteristics along these dimensions. In the context of reflection seismic data, high dimensionality is achieved from the extraction of several seismic attributes to classify different seismic facies, with the potential to identify features not visible in the amplitude data. The potential of three algorithms, K-Means, Self-Organizing Kohonen Maps and Gaussian Mixture Models, is evaluated in the identification of facies corresponding to the presence of igneous intrusions in seismic data from the Mero Field area, aiming to develop a routine for application in other areas of similar geological context, on the exploratory front, while there is no availability of advanced seismic processes, such as elastic seismic inversions or other high computational cost products, which can be carried out by the exploration geoscientist, in your own workstation.

Keywords: machine learning; clustering algorithms; Santos Basin.

LISTA DE FIGURAS

Figura 1 - Área de Estudo.....	15
Figura 2 - Geometrias de intrusões.....	20
Figura 3 - Principais elementos ígneos.....	21
Figura 4 - Dedos de magma na soleira de Golden Valley	22
Figura 5 - Visão geral de afloramento.....	24
Figura 6 - Representação esquemática das fácies de soleiras.....	25
Figura 7 - Linha sísmica 3D ilustrando problemas comuns na geração de imagens de complexos de soleiras	28
Figura 8 - Vista detalhada para comparação das seções sísmicas sintéticas	29
Figura 9 - Dados de afloramento e sismogramas sintéticos	30
Figura 10 - Resposta petrofísica característica de uma intrusão máfica	32
Figura 11 - Painel de correlação de poços exploratórios públicos	33
Figura 12 - Carta estratigráfica da Bacia de Santos	35
Figura 13 - Perfil com principais elementos tectônicos e sedimentares na Bacia de Santos.....	37
Figura 14 - Cenário de Desenvolvimento estrutural estratigráfico proposto por KARNER et. al., 2021	38
Figura 15 - Mapa de espessura crustal das Bacias de Campos e Santos e principais estruturas geológicas.....	39
Figura 16 - Mapa esquemático mostrando os quatro compartimentos do bloco de Libra.....	40
Figura 17 - Classificações das ocorrências magmáticas em Mero/Bloco de Libra.....	41
Figura 18 - Representação esquemática das três dimensões espaciais e o espaço N- dimensional.....	51
Figura 19 - Facetas redundantes e irrelevantes:	53
Figura 20 - Ilustração de agrupamento	59
Figura 21 - Comparação dos algoritmos em um espaço tridimensional de atributos.....	60
Figura 22 - K-Means aplicado a um mesmo conjunto de dados	62
Figura 23 - Mapa estrutural da base do sal e localização dos poços cedidos pela ANP	67
Figura 24 - Linha arbitrária S-N interceptando poços e o Campo de Mero	69
Figura 25 - Linha arbitrária W-E interceptando poços e o Campo de Mero.....	70

Figura 26 - Mapas estruturais dos horizontes embasamento e DPA e mapa de isópacas entre os dois horizontes.....	72
Figura 27 - Mapa estrutural da base do sal e isópacas entre base do sal e DPA.....	73
Figura 28 - Mapa do topo do sal (a) e isópacas entre topo do sal e base do sal (b).	74
Figura 29 - Mapa do topo do Cretáceo (a) e isópacas entre Cretáceo e topo do sal (b).....	76
Figura 30 - Amplitudes RMS entre base do sal e DPA (a) e Cretáceo e topo do sal (b).....	77
Figura 31 - Trecho de linha sísmica no entorno do poço 4-BRSA-1346.....	78
Figura 32 - Mapa estrutural do fundo do mar e atributo “Edge Detection” aplicado	80
Figura 33 - Curvas e litologia do Poço 1-SHEL-5-RJS	83
Figura 34 - Poços sem a presença de rochas ígneas	84
Figura 35 - Poços com intrusões ígneas de espessura relevante próximas à base do sal.....	85
Figura 36 - Poços com intrusões ígneas de espessura menor na base do sal.....	86
Figura 37 - Extração de espectro de frequência do dado sísmico	88
Figura 38 - Modelagem de cunha sedimentar adaptada para intrusão ígnea.....	89
Figura 39 - Traço sísmico da convolução com a refletividade (esquerda), Integração (direita)	92
Figura 40 - Fluxograma para condicionamento dos atributos de entrada dos algoritmos de agrupamento.....	93
Figura 41 - Disposição das soleiras na área de interesse.....	95
Figura 42 - Atributos sísmicos gerados (continua).....	97
Figura 43 - Exemplos de alguns atributos com resolução vertical reduzida.....	102
Figura 44 - Representação ilustrativa do algoritmo K-means	103
Figura 45 - Inércia plotada em função do número de grupos e o ponto de inflexão chamado “cotovelo”.....	104
Figura 46 - Representações de três funções gaussianas com diferentes parâmetros.....	105
Figura 47 - Um modelo de mapa de Kohonen, em amarelo o neurônio vencedor.....	107
Figura 48 - Correlação Spearman entre todos os atributos.....	111
Figura 49 - Composição dos componentes principais a partir dos atributos de entrada	114
Figura 50 - Variância dos componentes principais, gráfico "Scree"	115
Figura 51- Fluxograma do processo de agrupamento	116
Figura 52 - Testes do cotovelo para ambos os grupos.....	117
Figura 53 - Mapa e Linhas do volume de amplitudes	120

Figura 54 - Resultado K-Means K=7, Acima: Grupo 01. Abaixo: Grupo 02	121
Figura 55 - Resultado K-Means K=8, Acima: Grupo 01. Abaixo: Grupo 02	122
Figura 56 - Resultado K-Means K=9, Acima: Grupo 01. Abaixo: Grupo 02	123
Figura 57 - Resultado K-Means K=10. Acima: Grupo 01. Abaixo: Grupo 02	124
Figura 58 - Pontuação BIC para os diferentes tipos de covariância.....	126
Figura 59 - Resultados para GMM 12 componentes: Covariância Completa (acima); Covariância Diagonal (abaixo).....	127
Figura 60 - Resultado GMM 12 componentes Covariância Completa (Acima) Covariância Diagonal (Abaixo).....	129
Figura 61 - Erros de Quantização e Topográficos para o Grupo 01 (a) e Grupo 02 (b).....	132
Figura 62 - Resultados do algoritmo SOM do Grupo 01 para os parâmetros selecionados a partir dos testes de EQ e ET	134
Figura 63 - Resultado do algoritmo SOM do Grupo 01 para os parâmetros não oriundos dos testes de EQ e ET	135
Figura 64 - Resultados do algoritmo SOM do Grupo 02 para os parâmetros selecionados a partir dos testes de EQ e ET	136
Figura 65 - Resultados do algoritmo SOM do Grupo 02 para parâmetros não oriundos dos testes de EQ e ET	137

LISTA DE TABELAS

Tabela 1: - Poços fornecidos pela ANP e marcadores litoestratigráficos..	82
--	----

LISTA DE QUADROS

Quadro 1 - Diferentes tipos de atributos sísmicos	43
Quadro 2 - Valores utilizados para modelagem	87
Quadro 3 - Lista dos atributos selecionados.....	94
Quadro 4 - Atributos selecionados por escolha dirigida.....	96
Quadro 5 - Atributos com alta correlação	112
Quadro 6 - Atributos selecionados após seleção de facetas	113
Quadro 7 - Parametrização SOM para cada grupo de atributos	131

LISTA DE ABREVIATURAS E SIGLAS

ANP	Agência Nacional do Petróleo
API	<i>American Petroleum Institute</i>
BBOE	bilhões de barris de óleo equivalente
FWI	<i>Full Waveform Inversion</i>
FMI	Fullbore Formation Microimager
Ma	Milhões de anos
AVO	<i>Amplitude versus offset</i>
RTM	Reverse Time Migration
SOM	Self-Organizing Map
TTI	<i>Tilted Transversely Isotropic</i>
DHI	<i>Direct Hydrocarbon Indicator</i>
Lidar	Light Detection and Ranging
ZTL	Zona de Transferência de Libra
3D	Tridimensional
RCSB	Rifte Continental do Sudeste Brasileiro
LC	Lineamento de Curitiba
GLCM	<i>Gray-Level Co-occurrence Matrices</i>
GPU	Unidades de Processamento Gráfico
GTM	<i>Generative Topographic Map</i>
PCA	Análise de Componente Principal
GMM	Modelo de Mistura Gaussiana
ICA	Análise de componente Independente
2D	Bidimensional
DCEC	<i>Deep Convolutional Embedded Clustering</i>
DPA	Discordância Pré-Alagoas
RJS	Rio de Janeiro Submarino
MRO	Mero
BRSA	Petróleo Brasileiro S.A. (sigla ANP)
EQ	Erro de Quantização
ET	Erro Topográfico

BMU	<i>Best Matching Unit</i>
BIC	<i>Bayesian information criterion</i>

LISTA DE SÍMBOLOS

\approx	Aproximadamente
NNW	Norte-Noroeste
SSE	Sul-Sudeste
NE	Nordeste
ENE	Leste-Nordeste
SE	Sudeste
WSW	Oeste-Sudoeste
ESE	Leste-Sudeste
ms	milissegundos
\pm	mais ou menos
L	Leste
W	Oeste
N	Norte
\in	pertencente
μ	média
Σ	covariância

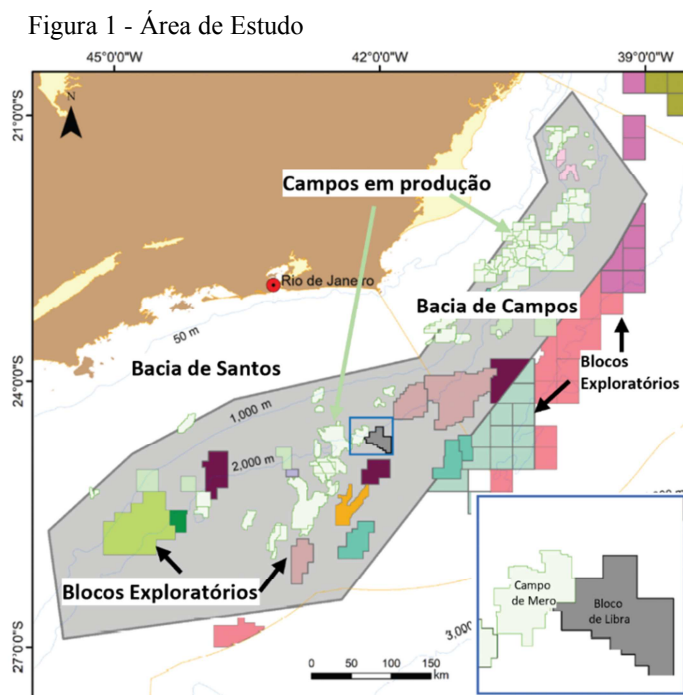
SUMÁRIO

	INTRODUÇÃO	15
1	LEVANTAMENTO BIBLIOGRÁFICO	19
1.1	Rochas Ígneas em Bacias Sedimentares	19
1.1.1	Afloramentos	22
1.1.2	<u>Sísmica</u>	24
1.1.2.1	Desafios do imageamento sísmico de intrusões	26
1.1.3	<u>Poços</u>	32
1.2	<u>Evolução Sedimentar e Tectonomagmática da Bacia de Santos</u>	35
1.3	Atributos Sísmicos Pós-Empilhamento	42
1.4	Aprendizado de Máquinas	47
1.4.1	<u>Técnicas de Redução de Dimensionalidade</u>	50
1.4.1.1	Seleção de Facetas.....	54
1.4.1.2	Extração de Facetas.....	55
1.4.2	<u>Técnicas de Agrupamento</u>	58
2	MATERIAL E MÉTODOS	66
2.1	Interpretação Sísmica	66
2.2	Análise de Poços	82
2.3	Atributos Sísmicos	92
2.4	Técnicas de agrupamento	103
2.4.1	<u>Algoritmo K-Means</u>	103
2.4.2	<u>Algoritmo Modelos de Mistura Gaussiana</u>	105
2.4.3	<u>Algoritmo Mapas Auto-organizáveis de Kohonen</u>	107
3	RESULTADOS	110
3.1	Seleção de facetas	110
3.2	Extração de Facetas	114
3.3	Técnicas De Agrupamento	116
3.3.1	<u>Algoritmo K-Means</u>	118
3.3.2	<u>Modelos de Mistura Gaussiana</u>	126
3.3.3	<u>Mapas Auto-organizáveis (SOMs)</u>	131
	CONSIDERAÇÕES FINAIS	139

REFERÊNCIAS	143
APÊNDICE A - Rotinas dos algoritmos	154

INTRODUÇÃO

A Bacia de Santos é localizada no sudeste da margem continental passiva brasileira e se estende por, aproximadamente, 350 mil km², com lâmina d'água que chega a ultrapassar 3000 metros. É limitada a Norte pelo Alto de Cabo Frio e a Sul pelo Alto de Florianópolis (MOREIRA et al., 2007). Sua formação está ligada à separação do supercontinente Gondwana, cujo processo é até hoje fruto de intenso debate na comunidade científica (MOHRIAK; LEROY, 2013; SALOMON; PASSCHIER; KOEHN, 2017; REUBER; MANN, 2019; PEACE et al., 2020; ALVES et al., 2020). Atualmente, ela é um dos principais focos de exploração de hidrocarbonetos do Brasil e seus reservatórios carbonáticos da seção pré-sal estão entre as descobertas mais importantes das últimas décadas (BELTRAO et al., 2009; KARNER et al., 2021). A área deste estudo é o Bloco de Libra (**Erro! Fonte de referência não encontrada.**), onde se localiza o Campo de Mero, descoberto pelo poço exploratório 2-ANP-2A-RJS, o qual possui reservas estimadas em 11.94 bilhões de barris de óleo equivalente (BBOE) considerado de boa qualidade, a 29° API (PETERSOHN et al., 2021), sendo o setor NW do Bloco de Libra, ofertado na 1ª Rodada de Partilha de Produção Pré-Sal, em 2013 (ANP, 2013).



Fonte: Modificado de PETERSOHN et al., 2021.

Até o momento, cinco grandes fases de eventos magmáticos foram identificadas na bacia. Duas fases durante a fase rifte, uma durante o pós-rifte (~120 Ma) e duas durante a fase drifte (~80 Ma e ~50Ma). Na região do Bloco de Libra, apenas três eventos ígneos são descritos: um evento inicial intrusivo no Aptiano (126 Ma), grandes corpos extrusivos entre 118-120 Ma e intrusões durante o Santoniano-Campaniano (80 Ma). As rochas ígneas do Aptiano são compostas de basaltos toleíticos, enquanto as mais recentes são compostas de diabásio alcalino e gabro (MOREIRA et al., 2007; FORNERO et al., 2019; MAGEE et al., 2021; PETERSOHN et al., 2021).

A presença de rochas ígneas na seção pré-sal da bacia foi identificada em diversos poços perfurados na bacia, em alguns dos quais estas rochas possuem características de reservatório e colunas de hidrocarboneto significativas (DE LUCA et al., 2015), entretanto, essas rochas também podem significar risco exploratório, a partir do seu potencial de sobrepressão (SCHOFIELD et al., 2020) ou devido à dificuldade de diferenciação, a partir de sua impedância acústica, de rochas carbonáticas e ígneas, um complexo ígneo pode ser confundido com uma estrutura de recife (e.g. WATSON et al., 2019; MOHRIAK; GORDON; MELLO, 2021).

Em seções sedimentares siliciclásticas com intercalações de folhelhos e arenitos, em geral, de baixa impedância acústica, como é caso da seção pós-sal das bacias do sudeste brasileiro, o contraste de impedância provocado por intrusões ígneas resulta em refletores com alta amplitude e morfologia específica (MAGEE et al., 2015; CORREIA et al., 2019), o que permite sua rápida identificação, mesmo por intérpretes com pouca experiência. Em contraste, nas seções sedimentares caracterizadas por sedimentos de alta impedância acústica, i.e. carbonatos e evaporitos, a identificação de intrusões é dificultada pelo reduzido contraste de impedância entre as camadas e a identificação dessas intrusões é desafiadora mesmo para intérpretes experientes (EIDE et al., 2018; RABELL et al., 2018).

A identificação das soleiras nos limites entre as Formações Ariri e Barra Velha é desafiadora (DE OLIVEIRA; RANCAN; OLIVEIRA, 2019; KANGXU et al., 2019; MOHRIAK; GORDON; MELLO, 2021; GORDON et al., 2022), entretanto, processamentos de alta complexidade, utilizando Inversão Completa da Forma da Onda (FWI) e anisotropia TTI (“*Tilted Transversely Isotropic*”, isotropia transversa ao plano inclinado), associados a inversões elásticas, já mostraram bons resultados (PENNA et al., 2019). Todavia, esse tipo de processamento demanda altíssima capacidade técnica, requer supercomputadores dedicados e, no caso das inversões elásticas, dados de poços já perfurados.

Em novas áreas, onde a disponibilidade de dados de poços é extremamente limitada, os dados sísmicos costumam possuir processamento acelerado, com modelos de velocidades feitos de maneira expedita, especialmente em projetos multiclientes, apesar dos recentes avanços nessa esfera, como o uso de algoritmos avançados de migração, em geral, Migração Reversa no Tempo (Reverse-Time Migration, RTM) e uso de anisotropia TTI na maioria dos projetos de processamento. Dessa forma, dados como os apresentados por PENNA et al. (2019) demoram meses a anos para serem disponibilizados para intérpretes.

Estima-se que esse contexto geológico esteja distribuído ao longo das outras bacias do sudeste que compartilham o chamado “polígono do pré-sal” e, no contexto de frente exploratória, no qual os intérpretes recebem apenas o resultado final do processamento acelerado, na forma do dado sísmico empilhado, em profundidade, muitas vezes sem os dados parciais utilizados para inversões, faz-se necessária uma alternativa rápida e de baixa complexidade para a identificação das intrusões ígneas, de modo a diminuir as incertezas em relação à presença de reservatórios e zonas de sobrepressão, para novos poços exploratórios.

Os atributos sísmicos são diferentes tipos de medições quantitativas ou filtros aplicados ao dado sísmico de amplitude que permitem destacar ou observar características desejadas (BARNES, 2016). Desde os anos 1930, quando os geofísicos mediram pela primeira vez tempos de percurso em reflexões coerentes de registros sísmicos de campo, passando pela identificação dos “bright spots” nos anos 1960 e pelas primeiras estações de interpretação dos anos 1980, com suas telas coloridas e a possibilidade de usar diferentes barras de cores nos dados sísmicos, até as atuais dezenas (se não centenas) de atributos, estes foram integrais aos sucessos de exploração (CHOPRA; MARFURT, 2005).

Há aproximadamente três décadas, os atributos sísmicos são utilizados associados a técnicas de classificação automática de fácies sísmicas (COLÉOU; POUPON; AZBEL, 2003), com sucessos limitados muitas vezes pelo poder de processamento de computadores antigos. Algoritmos como K-Means (MACQUEEN, 1967) e Mapas Auto-Organizáveis de Kohonen (SOMs, KOHONEN, 1982) utilizam grupos naturais (“natural clusters”), utilizam um ou mais atributos sísmicos para dividir o dado em grupos que apresentam alta similaridade, destacando de forma automática as diferentes fácies sísmicas em um dado. Em geral, as diferentes fácies, indicam diferentes litologias.

Este trabalho, portanto, visa avaliar o potencial da classificação automática de fácies sísmicas aplicada à identificação das intrusões ígneas no limite entre as Formações Ariri e Barra Velha, a partir de um fluxo de trabalho que pode ser realizado individualmente, na estação do intérprete, utilizando apenas os atributos extraídos do dado sísmico empilhado,

sem o uso de inversões elásticas, dados pré-empilhamento e sem a necessidade de reprocessamentos avançados.

Além das técnicas já citadas, K-Means e SOMs, será também avaliada a técnica de Modelos de Misturas Gaussianas (“*Gaussian Mixture Models*”, GMMs DUDA; HART, 1974). Inicialmente, apresenta-se uma revisão acerca da presença de rochas ígneas nas bacias sedimentares ao redor do mundo e das diferentes expressões geométricas dos complexos ígneos, com exemplos de afloramentos, dados sísmicos (junto com os desafios de seus imageamento) e poços. Em seguida, trata-se da evolução sedimentar e tectonomagnética da Bacia de Santos, seguida de atributos sísmicos e conceitos de aprendizados de máquinas. Por fim, são apresentados os resultados da interpretação sísmica e dos poços (fornecidos pela Agência Nacional do Petróleo para este estudo), bem como os resultados das classificações de fácies sísmicas, utilizando duas abordagens distintas (com menos ou mais atributos de entrada) e várias parametrizações das técnicas. O sucesso na aplicação dessas técnicas teria o potencial de redução de incertezas em poços de frente exploratória e de aumento na segurança da perfuração, com utilização reduzidas de dados, o que pode representar economia financeira significativa ao processo exploratório.

1 LEVANTAMENTO BIBLIOGRÁFICO

1.1 Rochas ígneas em bacias sedimentares

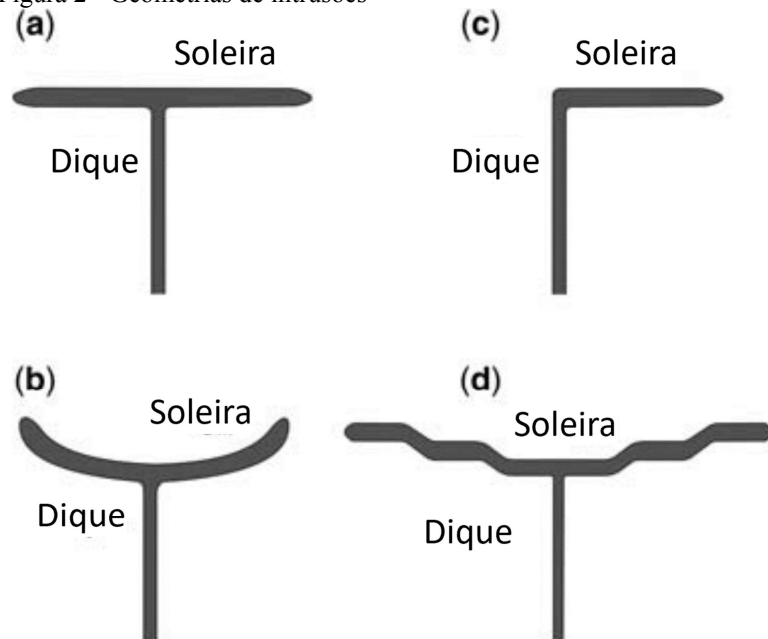
Pelo menos três quartos das margens rifteadas do Atlântico são estimadas como vulcânicas e talvez até 90% das margens ao redor do mundo (NELSON et al., 2009). Nestes locais, fluxos subaéreos de lava foram extravasados formando as províncias de derrames basálticos e, em alguns desses lugares, esses derrames cobrem bacias com alto potencial para a exploração de hidrocarbonetos, por exemplo, na Província Ígnea do Atlântico Norte e na província de Paraná-Etendeka, atualmente distribuída entre o sudeste do Brasil e oeste da África. Essas grandes bacias vulcânicas estão presentes ao longo de margens oceânicas, nas bacias intracontinentais (Amazonas/Solimões) e em margens compressivas (Bacia de Neuquén, Argentina) (PLANKE et al., 2018).

A maior parte do derretimento crustal gerado por atividade ígnea não consegue chegar à superfície da terra e, em bacias sedimentares, esse magma costuma formar extensas redes de canalizações que se espalham lateralmente por dezenas de quilômetros antes de entrarem em erupção (MAGEE et al., 2018). O movimento do magma através das camadas mais rasas da crosta continental é controlado pela interação com rochas encaixantes heterogêneas e falhas, dessa forma, em bacias sedimentares ocorre a prevalência de intrusões laminares uma vez que o magma tende a se propagar através dos planos de fraqueza dos estratos encaixados e suas falhas (BISCHOFF et al., 2021).

Pesquisas recentes mostram que, nesses locais, a presença de grandes volumes de intrusões ígneas, formando complexos intrusivos, com suas soleiras vulcânicas e lacólitos, podem exercer um grande impacto na dinâmica das bacias sedimentares e seus sistemas petrolíferos relacionados, entender seus processos controladores é essencial na exploração de hidrocarbonetos (PLANKE et al., 2005; INFANTE-PAEZ; MARFURT, 2017; SENGER et al., 2017; EIDE et al., 2018). Entre os possíveis efeitos estão maturação da rocha geradora, formação de trapas a partir da deformação da rocha encaixante e sua sobrecarga, criação de barreiras ou caminhos para fluxo de fluidos ou, nos casos em que as intrusões são fraturadas, as mesmas podem formar reservatórios atípicos (RABELL et al., 2018).

Soleiras são, por definição, intrusões ígneas tabulares predominantemente paralelas à rocha encaixante e sub-horizontais. Diques são lâminas de magma que cortam transversalmente os estratos, geralmente subverticais (SENGER et al., 2017). As soleiras apresentam uma grande variação tanto na sua geometria, como no seu tamanho. Podem possuir, em espessura, desde poucos metros até centenas de metros. Em relação à sua geometria, na maioria das vezes são concordantes às camadas encaixantes e essa característica costuma ser usada para distingui-las de outras intrusões laminares como diques e lâminas inclinadas. As diferenças geométricas nas soleiras costumam ser relacionadas à resposta da rocha encaixante à intrusão laminar e como essa resposta afeta o campo de stress local (GUDMUNDSSON; LØTVEIT, 2014). Dessa forma, a anisotropia da rocha encaixante e a profundidade da soleira em relação ao seu diâmetro, são fatores importantes que afetam os detalhes da propagação das soleiras e sua eventual geometria (Figura 2).

Figura 2 - Geometrias de intrusões

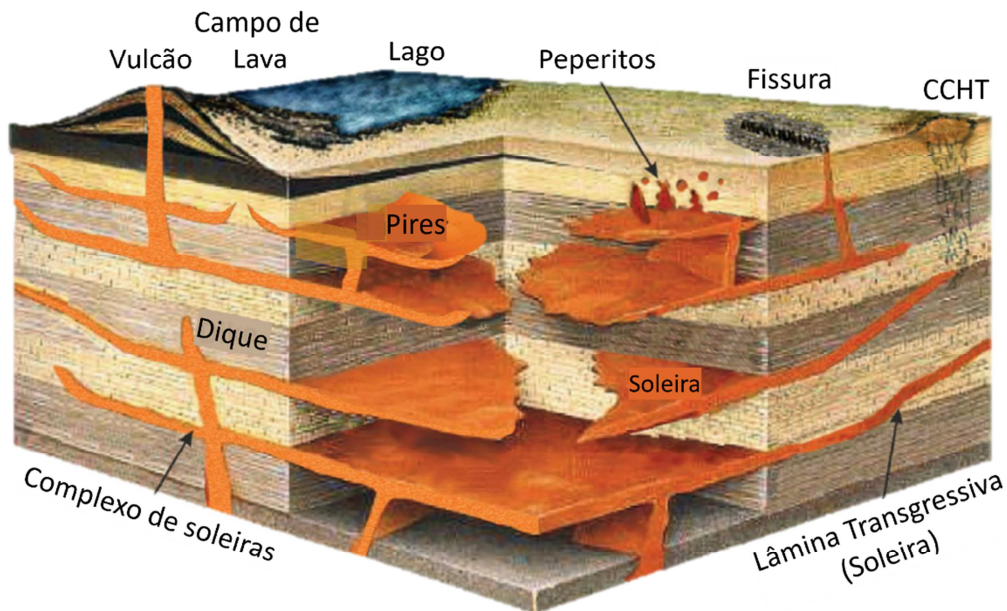


Legenda: (a) e (c) Intrusões profundas; (b) e (d) intrusões rasas.
 Fonte: GUDMUNDSSON; LØTVEIT, 2014.

PLANKE et al. (2018) defendem a ocorrência de soleiras paralelas ao acamamento em partes mais profundas da bacia (acima de 5 km) e as geometrias de pires em profundidades intermediárias. Soleiras sub-horizontais seriam mais comuns no quilômetro mais próximo da superfície. Em algumas bacias ocorrem ainda complexos de chaminés hidrotermais, representando conexões entre as intrusões e a paleosuperfície, formados a partir da geração de gás e aumento de pressão em auréolas metamórficas no entorno de intrusões em processo de

resfriamento, são compostas por sedimentos fraturados e remobilizados. A Figura 3 representa os principais elementos ígneos em uma bacia vulcânica.

Figura 3 - Principais elementos ígneos



Legenda: Elementos ígneos típicos, CCHT: Complexo de Chaminés Hidrotermais.
 Fonte: Modificado de PLANKE et al., 2018

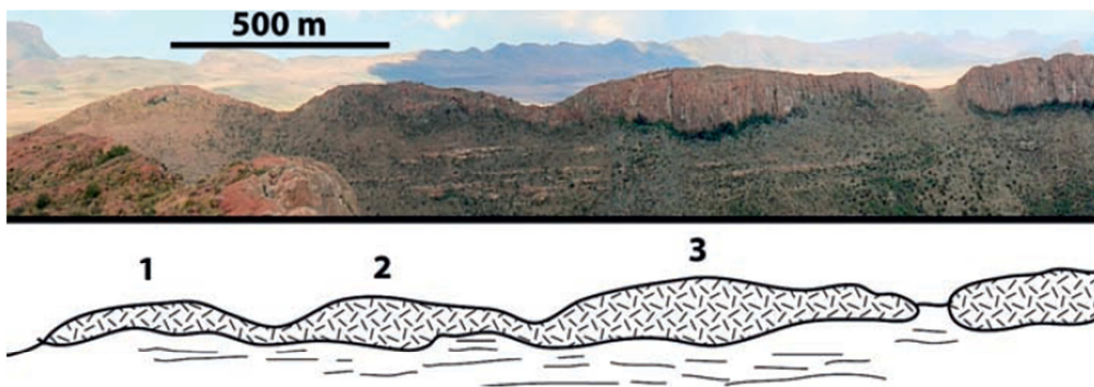
Os dois grupos de PLANKE et al. (2005) são subdivididos a partir de sua suavidade, tamanho, continuidade e profundidade de paleosuperfície. Além dos dois grupos principais, duas outras unidades foram definidas: transgressiva planar e de bloco falhado. Esses diagramas esquemáticos são uma representação das diversas formas em que intrusões ocorrem nas bacias de Møre e Vøring e podem ser aplicados às diferentes bacias sedimentares ao redor do mundo.

As próximas seções buscam ilustrar e caracterizar as ocorrências de intrusões ígneas em afloramentos, dados sísmicos e poços, em diferentes bacias sedimentares, com o objetivo de expandir o conhecimento a respeito de sua disposição geométrica e das fácies sísmicas características das intrusões e suas rochas encaixantes, além de destacar desafios para a imagem sísmica e os principais perfis de poços utilizados na caracterização de intrusões ígneas.

1.1.1 Afloramentos

As intrusões de soleira em formato de pires ocorrem em menores profundidades, por isso, possuem mais estudos de afloramento detalhados e, portanto, são mais bem compreendidas (EIDE et al., 2017). SCHOFIELD; STEVENSON; RESTON (2010) descrevem a soleira de Golden Valley, na Bacia de Karoo, África do Sul, na qual é possível observar “dedos” de magma com dezenas a centenas de metros, irradiando a partir de um eixo central e formando a borda mais externa (Figura 4). A formação dos dedos de magma se deu ao longo de uma circunferência de, aproximadamente, 50 quilômetros, escala até então inédita. Os autores defendem que a formação desses dedos em grande escala ocorre a partir da fluidização causada pela ebulição rápida dos fluidos dos poros, o que possibilitou a quebra da heterogeneidade mecânica que controlava a propagação concordante da soleira, permitindo a transgressão da intrusão.

Figura 4 - Dedos de magma na soleira de Golden Valley



Legenda: Fotografia de dedos de magma na Bacia de Karoo, África do Sul (acima). Representação esquemática (abaixo).

Fonte: SCHOFIELD et al., 2012.

HANSEN et al. (2011) descrevem soleiras em formato de pires referentes à atividade ígnea da Província Ígnea do Atlântico Norte, responsável por gerar extensos complexos de soleira nas bacias sedimentares nas margens noroestes europeias e pelas pilhas de lava basáltica contemporâneas das Ilhas Faroese, objeto do estudo. As grandes soleiras dessas ilhas são intrudidas na pilha de lava principal e, para os autores, representam informações novas e importantes sobre a natureza de sistemas de canalização ígnea ativos durante os estágios finais da atividade em províncias de derrames basálticos.

A partir da observação das soleiras, os autores também sugerem que a liberação rápida de gases, abrindo sistemas hidráulicos previamente fechados, associada à quebra de forças de estrangulamento da rocha encaixante, resultaram na reorientação dos eixos de stress principal, agindo nas margens das soleiras.

Descrições detalhadas de intrusões ocorridas em partes mais profundas das bacias são mais raras. SCHOFIELD et al. (2012) descrevem intrusões de doleritos do Oligoceno-Mioceno em camadas de carvão do Cretáceo, na bacia de Raton, no Colorado (EUA). É estimado que as intrusões ocorreram pelo menos 30 milhões de anos após a deposição das camadas de carvão, que já estariam a uma profundidade de 4,5-6 km.

Ao longo da bacia, seções transversais das soleiras de dolerito intrudindo as camadas de carvão estão expostas. As soleiras possuem formato elíptico, similar a dedos, e são rodeadas por zonas caóticas e retorcidas de carvão coqueificado com aproximadamente 40 cm de espessura. Esses afloramentos corroboram com a observação de que dedos de magma se formam onde a rocha encaixante se comporta como fluido.

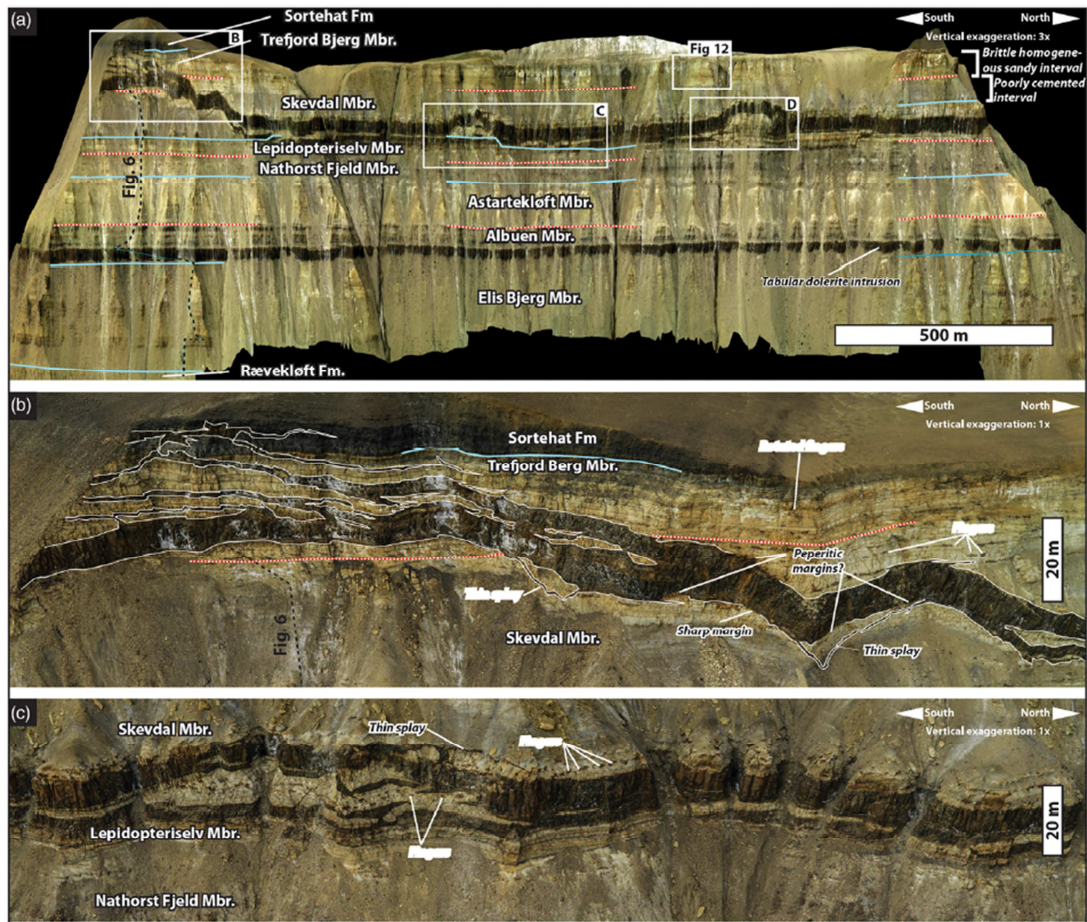
EIDE et al. (2017) utilizaram dados de Lidar (“Light Detection and Ranging”) e afloramentos de uma seção com 22 km de comprimento e 250 metros de espessura de rochas sedimentares do Jurássico, na bacia de Jameson Land, leste da Groenlândia, para descrever as intrusões ocorridas a uma profundidade aproximada de 3 km à época do posicionamento do magma.

A Figura 5 mostra diferenças entre as geometrias de intrusão em arenitos homogêneos rúpteis e arenitos inconsolidados do membro Skevdal, além de características tabulares das soleiras dos lamitos do Membro Albuen. Para os autores, as observações corroboram o controle crítico das propriedades litológicas no posicionamento das intrusões, descendo a preferência, pelas soleiras, de unidades ricas em lamitos, as quais são intrudidas ao longo de vários quilômetros, apresentando bordas retas com grande estabilidade vertical. Em contraste, nos arenitos inconsolidados, as soleiras apresentam texturas globulares, caóticas e peperíticas.

O complexo de soleiras das Montanhas Theron, na Antártica, são descritos por HUTTON (2009) e considerados pelo autor como talvez as melhores exposições desse tipo de complexo no mundo. Essas soleiras ocorrem encaixadas em sedimentos siliciclásticos do Permiano, com a intrusão ocorrendo no Jurássico. As soleiras possuem uma abundância de estruturas de ponte, observadas em uma grande variedade de escalas, incidam que pequenas protossoleiras se propagam à frente do corpo principal, explorando horizontes mais frágeis como camadas de carvão e folhelho. Para o autor, evidências em grande e pequena escala

indicam que o mecanismo principal de posicionamento é soerguimento forçado do teto da camada.

Figura 5 - Visão geral de afloramento



Legenda: Visão geral do afloramento (a) com destaque para geometrias e características de intrusão em arenitos rúpteis (b) e inconsolidados (c).

Fonte: Modificado de EIDE et al., 2017.

1.1.2 Sísmica

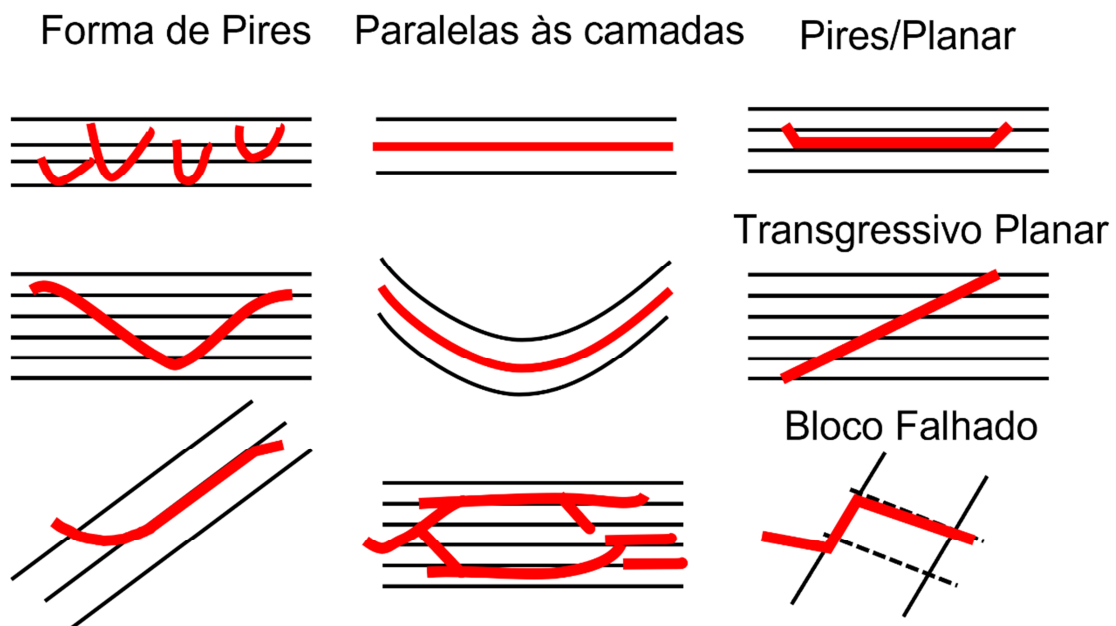
Os dados de sísmica de reflexão possibilitam uma oportunidade única para visualizar e analisar os sistemas de canalização magmática que influenciam os processos de intrusão ígnea (MAGEE et al., 2015). Os dados tridimensionais representaram um avanço significativo nos modelos e na compreensão dos sistemas magmáticos nas bacias sedimentares (EIDE et al., 2018). A sísmica de reflexão fornece a imagem mais completa e detalhada de intrusões individuais e de complexos intrusivos, particularmente nas bacias sedimentares, devido ao

alto contraste de impedância entre as rochas ígneas e os estratos encaixantes (SMALLWOOD; MARESH, 2002).

Estudos sobre as arquiteturas em grande escala das intrusões ígneas costumam ser baseados em dados sísmicos uma vez que estudos semelhantes em afloramentos de intrusões ígneas que possuem uma escala comparável à reflexão sísmica são raros (HUTTON, 2009; MUIRHEAD et al., 2012; WALKER et al., 2017; EIDE et al., 2017). O resultado é um conhecimento escasso a respeito das relações entre as estruturas observadas em sísmica e as observações detalhadas possibilitadas por estudos de afloramentos, tornando difícil saber o quanto as características dos complexos de soleiras, observadas na reflexão sísmica, se relacionam com geometrias reais desses complexos na subsuperfície (EIDE et al., 2018).

Os dados de reflexão sísmica possibilitam observar imagens de corpos intrusivos inteiros, permitindo a descrição de seus aspectos geométricos, suas dimensões lateral e vertical, além de sua interconectividade em detalhe (BISCHOFF et al., 2021). PLANKE et al. (2005) descreveram nove unidades de fácies de soleiras (Figura 6), divididas em dois grupos principais: paralelos às camadas e em formato de pires (“saucer-shaped”). Estas possuem uma característica discordante à atitude das rochas encaixantes e permitiram o aperfeiçoamento da compreensão a respeito da função de intrusões interconectadas na migração de magma, demonstrando o papel crítico de soleiras na transferência de magma de camadas mais profundas para mais rasas.

Figura 6 - Representação esquemática das fácies de soleiras



Fonte: Modificado de PLANKE et al., 2005.

Poucos trabalhos descrevem as intrusões ígneas nas bacias do sudeste brasileiro e seu impacto na evolução das bacias. MAGEE et al. (2021) realizaram uma análise de 38 intrusões ígneas na bacia de Santos, na região do Gráben de Merluza, onde sugerem que o deslocamento das soleiras intrudidas em sedimentos pós-sal em relação às intrassal ocorre devido a diferenças na litologia dos evaporitos. Os autores sugerem, ainda, que as intrusões nos evaporitos sobre esse gráben aumentaram o movimento do sal durante o Albiano-Santoniano e a cristalização dessas rochas restringiu o diapirismo pós-Santoniano.

CORREIA et al. (2019) fazem caracterização geométrica de soleiras em dado sísmico adquirido em águas rasas próximo ao Alto de Cabo Frio, limite sul da Bacia de Campos. Os autores caracterizaram cinco tipos de soleiras: em formato de pires, levemente em formato de pires, formato de pires ascendente, transgressiva e paralela ao acamamento. As intrusões possuem área variando entre 0,4 e 14,3 km², mergulho máximo de 15° e estão em uma profundidade até 3 km.

Os diversos trabalhos indicam certa homogeneidade na ocorrência das soleiras. As intrudidas em sedimentos relativamente rasos costumam possuir capacidade de intrusão e transgressão das camadas sedimentares, culminando com o típico formado de pires (“saucer-shape”), enquanto as intrudidas em sedimentos mais profundos possuem a capacidade de transgressão dos sedimentos muito limitadas, devido às altas pressões exercidas pelas camadas sobrepostas.

Em geral, a característica que permite maior distinção desses elementos com as camadas a seu entorno é a alta amplitude gerada pelo elevado coeficiente de reflexão. Entretanto, o sinal sísmico pode sofrer perdas e as rochas encaixantes podem possuir características petrofísicas similares às das intrusões ígneas, diminuindo a capacidade de identificação das soleiras. A próxima seção visa detalhar esses problemas.

1.1.2.1 Desafios do imageamento sísmico de intrusões

Uma das principais razões para o avanço no mapeamento sísmico de intrusões é a proeminência característica de reflexões de alta amplitude, provocadas pela alta impedância das soleiras em relação à rocha encaixante (PLANKE et al., 2018). Entretanto, existem diversos problemas relacionados ao imageamento sísmico das soleiras. Imageamento é a apresentação das formas de onda adquiridas por um método geofísico numa forma em que elas simulam uma imagem da estrutura na subsuperfície (KEAREY; BROOKS; HILL, 2002).

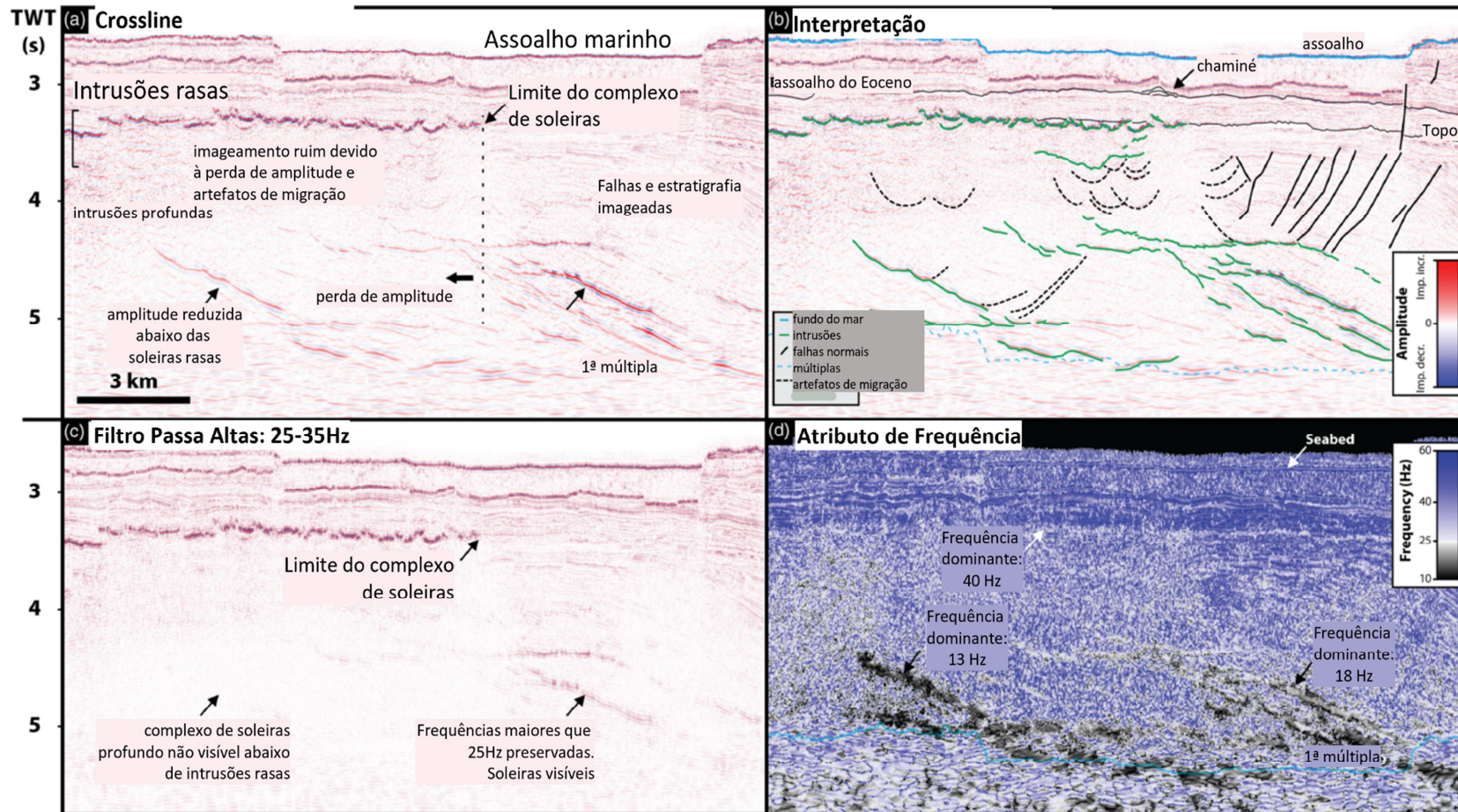
Em uma análise de poços perfurados na região do Complexo de Soleiras Faroe-Shetland, SCHOFIELD et al. (2017) estimam que até 88% das soleiras do complexo possuem espessura inferior a 40 metros e, portanto, estão abaixo do limite de resolução vertical do dado sísmico nas profundidades em que elas ocorrem.

EIDE et al., (2018) destacam três limitações inerentes ao método sísmico 3D: (1) diminuição da qualidade e resolução sísmica devido à absorção de altas frequências e da energia sísmica e ao aumento da velocidade sísmica com a profundidade; (2) efeitos das rochas sobrepostas, onde o sinal sísmico é afetado por litologia complexa, especialmente em bacias com rochas ígneas (Figura 7); e (3) a incapacidade do método em visualizar interfaces de alto mergulho ou verticais. Somado a essas limitações, os autores apontam que os dados sísmicos de regiões com complexos de soleiras normalmente não são processados de maneira a preservar detalhes desses sistemas, raramente utilizando algoritmos de migração mais avançados, tal como a Migração Reversa no Tempo (Reverse Time Migration).

Para entender melhor a problemática relacionada ao imageamento das intrusões ígneas é importante compreender resolução sísmica e detectabilidade sísmica. Devido à variação limitada de frequências disponíveis em um levantamento sísmico, a espessura de camadas a serem observadas tem um limite mínimo que, em geral, corresponde a um quarto do comprimento de onda dominante (produto entre a velocidade e a frequência), esse limite é denominado resolução vertical (EIDE et al., 2018). A detectabilidade sísmica se refere à capacidade da reflexão de uma determinada interface ser reconhecida ou não no dado sísmico. PLANKE et al. (2005) sugerem que, apesar da dificuldade em quantificá-lo, o limite de detecção pode ser aproximado a um décimo do comprimento de onda dominante. Entretanto, EIDE et al. (2018) argumentam que, devido às muitas variáveis envolvidas nesse limite, não é possível determinar uma regra de base, como no caso da resolução vertical. Para os autores, é importante que a reflexão tenha, pelo menos, amplitude maior que a média de amplitude apresentada pelo ruído sísmico do levantamento, uma vez que processamentos realizados para remover ruído são capazes de eliminar reflexões.

O emprego de modelagem sísmica a partir de análogos de campo é uma ferramenta importante na interpretação sísmica de complexos intrusivos (RABELL et al., 2018), assim como na redução de incertezas em relação à detectabilidade e resolução sísmica. Infelizmente, há poucos estudos na literatura com modelagem de soleiras. MAGEE et al., (2015) apresentam modelagens com mudanças de litologia da rocha encaixante, espessura de intrusão e frequência dominante do dado. Entretanto, os autores limitam as rochas encaixantes a folhelhos e arenitos e não tratam da profundidade das intrusões.

Figura 7 - Linha sísmica 3D ilustrando problemas comuns na geração de imagens de complexos de soleiras



Legenda: (a) Linha sísmica não interpretada. (b) Versão interpretada de (a). (c) Versão filtrada passa-altas de (a) com um filtro inclinado de 25 a 35 Hz.

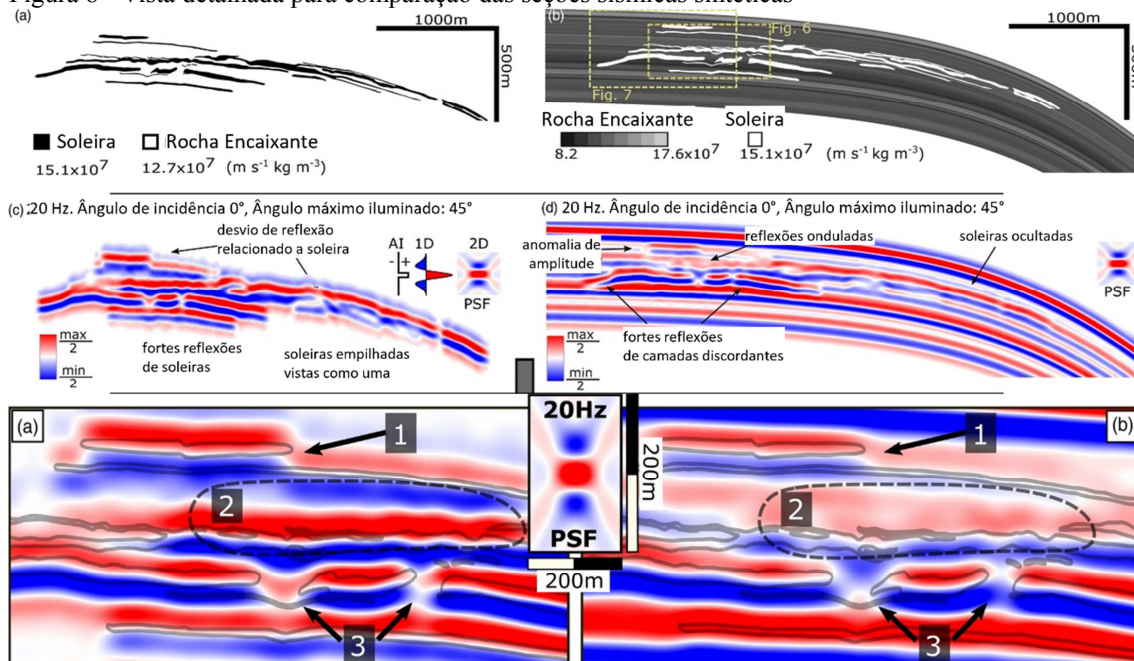
Observe a perda de altas frequências abaixo das intrusões rasas. (d) Atributo de frequência de (a).

Fonte: Modificado de EIDE et. al. (2018).

Buscando modelos mais realísticos, RABEL et al., (2018) combinam representações virtuais de afloramento, associados a dados de poços e filtragem tridimensional em modelos com diferentes propriedades elásticas. De acordo com observações dos autores, interferências sísmicas dependentes da frequência e fortes variações de amplitude parecem ser motivadas por variações das propriedades elásticas tanto das rochas encaixantes quanto das soleiras.

A Figura 8 mostra os resultados da modelagem realizada pelos autores, apenas a modelagem para frequência da wavelet de 20 Hz é exibida por ser mais próxima ao objetivo deste trabalho. O modelo 1 é um modelo litológico binário simples, onde tanto soleiras quanto as rochas encaixantes são homogêneas com velocidade compressiva (V_p) 5500 m/s e 4200 m/s, respectivamente. A velocidade da onda cisalhante (V_s) é de 2890 e 2470, enquanto a densidade é 2800 e 2600 kg/m^3 . O modelo 2 apresenta camadas de rocha encaixante variáveis obtidas a partir de dados de poço, cujas V_p variam de 3350 m/s a 5950 m/s, V_s variando entre 1970 e 3380 m/s e densidade entre 2480 e 2950 kg/m^3 , as soleiras possuem duas velocidades compressivas de 4700 e 5500 m/s, V_s de 2470 e 2890 m/s e densidade de 2800 kg/m^3 .

Figura 8 - Vista detalhada para comparação das seções sísmicas sintéticas



Legenda: (a) Modelo 1; (b) Modelo 2; (c) resultados do Modelo 1; (d) resultados do modelo 2; (e) destaque de (c); (f) destaque de (d).
Fonte: RABEL et al., 2018.

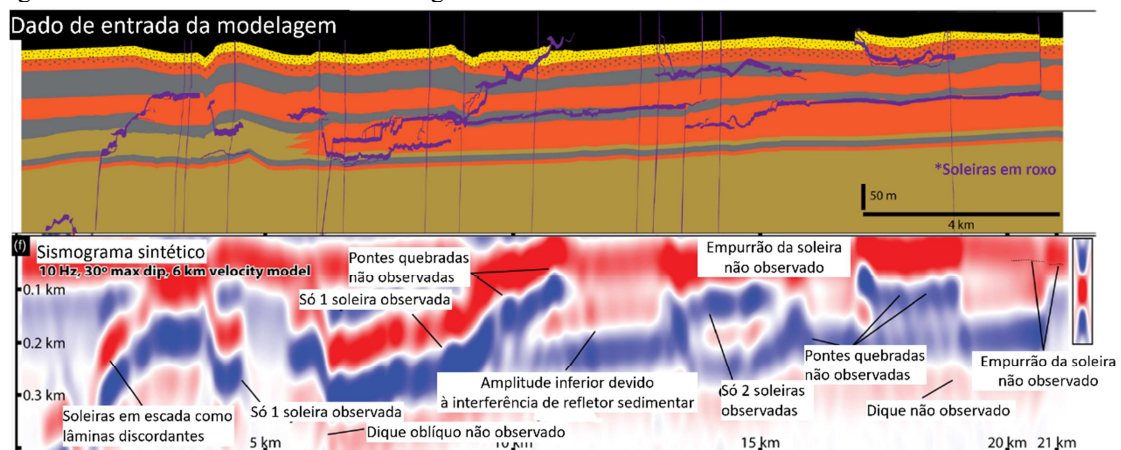
As modelagens mostram que, com uma frequência dominante de 20 Hz, no modelo 1, nenhum dos segmentos individuais consegue ser observado, os empilhamentos das soleiras se aglomeram em uma única reflexão contínua de topo e uma reflexão de base levemente irregular com algumas discontinuidades, além disso, alguns pontos de término da soleira

provocam um deslocamento aparente da reflexão criando uma falsa impressão de falha. No modelo mais realístico, no entanto, se torna muito mais difícil identificar as intrusões, em especial nos pontos onde elas são paralelas ao acamamento, locais onde apenas leves modificações das reflexões da rocha encaixante são observadas. Onde as soleiras apresentam geometria ondulante e se dividem em camadas mais delgadas, a interferência entre suas reflexões e as da rocha encaixante causam um padrão de formas de onda trançadas. Estas interferências podem ter um impacto positivo na estimação de atributos texturais, contribuindo na identificação das soleiras.

Em uma abordagem similar, EIDE et al. (2018) utilizam dados de afloramento da Bacia de Jameson, leste da Groenlândia, onde soleiras paralelas ao acamamento são intercaladas a lamitos, arenitos homogêneos e arenitos inconsolidado, divididos em sete associações de fácies. Um dos modelos apresentados tem o diferencial de analisar o alvo a 6 km da superfície, possuindo, portanto, uma frequência de 10 Hz e capaz de observar mergulhos de no máximo 30° e com espessura de afinação de aproximadamente 100 metros.

Na Figura 9, é possível observar soleiras entre 7-16 metros de comprimento à esquerda, porém onde várias soleiras ocorrem juntas, à direita, os refletores se somam e o complexo inteiro aparece como um único refletor. Soleiras mergulhantes mais proeminentes, em 2 e 9 km do perfil, são vistas como lâminas oblíquas, entretanto as com mais de 30° de mergulho (em 7 km) não são observadas. Para os autores, os modelos indicam que onde há empilhamento das soleiras, camadas com 1/30 do comprimento da onda podem chegar a ser detectadas.

Figura 9 - Dados de afloramento e sismogramas sintéticos



Legenda: Acima: Modelo de entrada para modelagem sísmica; abaixo: sismograma sintético correspondente ao modelo de 6 km de profundidade.

Fonte: Adaptado de EIDE et al., 2018.

A detecção e visualização das intrusões ígneas em grandes profundidades é, portanto, um desafio ao intérprete e vários fatores devem ser levados em consideração: a seção sedimentar sobreposta, as rochas encaixantes, a profundidade total, o conteúdo de frequências do dado sísmico, as técnicas de construção do modelo de velocidade e os algoritmos de migração utilizados. Entretanto, as modelagens apresentadas por RABEL et al., (2018) e EIDE et al. (2018) mostram que camadas delgadas podem causar interferência sísmica suficiente para indicar a presença de intrusões, ainda que uma reflexão específica não seja observada.

1.1.3 Poços

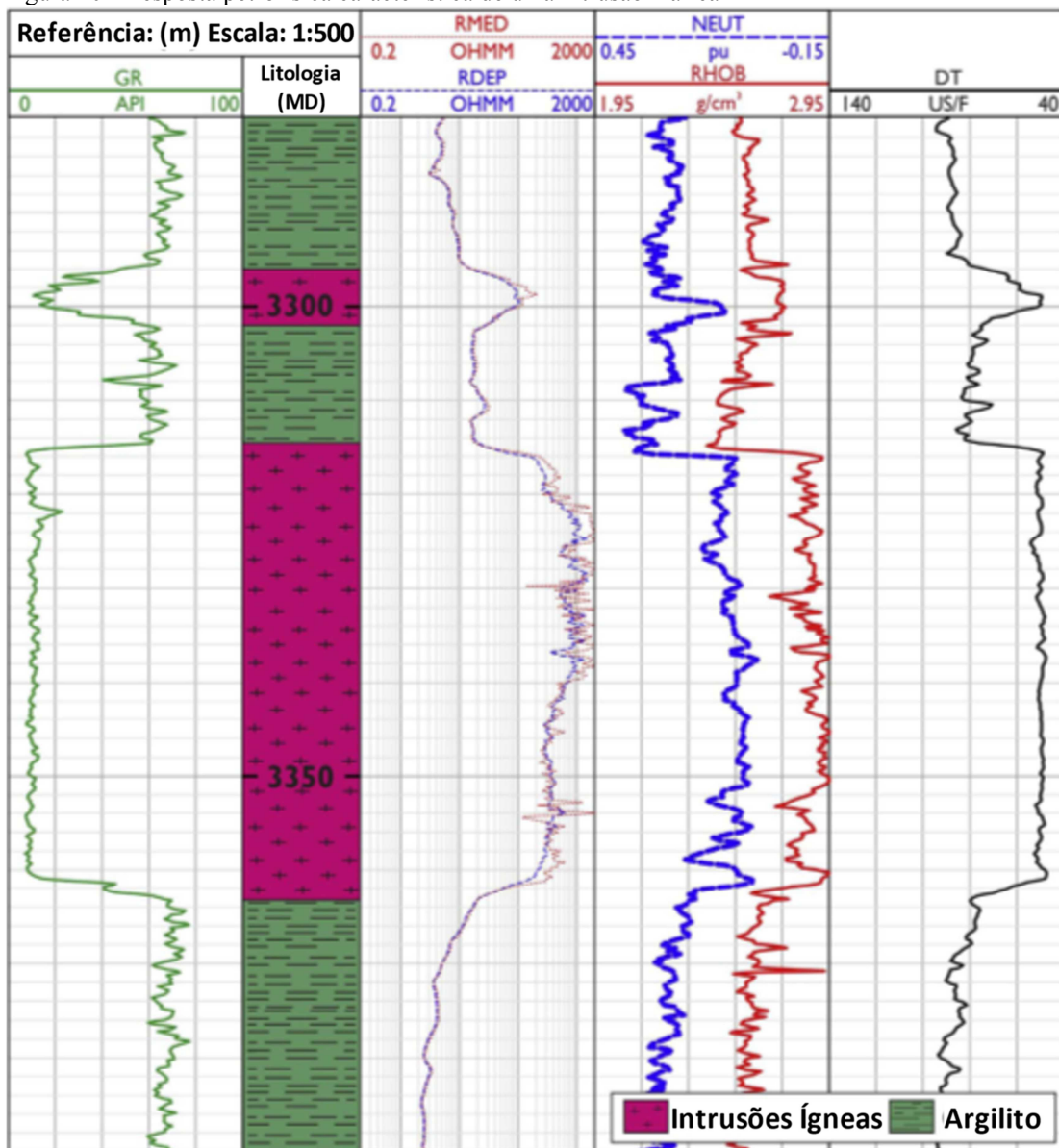
O uso de dados sísmicos aumentou consideravelmente o conhecimento a respeito das geometrias dos complexos intrusivos nos últimos anos, entretanto, como foi destacado na seção anterior, esses dados possuem limites de resolução vertical. O uso de dados de poços é uma importante ferramenta na complementação deste conhecimento. MARK et al. (2018) descrevem a resposta característica de intrusões máficas que as torna discerníveis em relação aos sedimentos encaixantes na escala de poço.

O magma máfico é rico em minerais como olivina e piroxênio, os quais possuem velocidades compressivas são 8420 m/s e 7200 m/s, respectivamente, o que leva às altas velocidades das rochas intrusivas. Além disso, velocidades cisalhantes entre 2400-3400 m/s também são superiores às dos sedimentos encaixantes. Devido à distribuição relativamente uniforme dos minerais nessas rochas, a resposta do perfil sônico geralmente é na forma de bloco, com poucas variações em uma camada (Figura 10).

Além das altas velocidades sísmicas, os minerais máficos também possuem alta densidade (olivina: 3,31 g/cm³; piroxênio: 3,3 g/cm³), o que leva a rocha a possuir entre 2.8 e 3 g/cm³ e perfil petrofísico de densidade também com forma de bloco. O perfil de densidade de nêutrons para intrusões ígneas costuma apresentar valores inferiores aos dos sedimentos encaixantes, uma vez que esse equipamento mede o conteúdo de hidrogênio em uma camada, o qual costuma ser baixo em rochas ígneas cristalinas, cuja porosidade reduzida limita sua capacidade de armazenar água. A relativa ausência de minerais radioativos, tais como Potássio, Tório e Urânio, faz com que a resposta à perfilagem de raios gama permaneça no intervalo entre 10-30 API.

A resistividade elétrica das intrusões máficas é alta, devido à baixa porosidade e permeabilidade e por conter relativamente pouca água, levando a leitura do instrumento a extrapolar o intervalo típico do perfil de resistividade. Em algumas intrusões o perfil de resistividade pode ter caráter caótico e flutuar significativamente em uma curta distância, o que indica a presença de fraturas. Devido à natureza mecanicamente resistente das intrusões, o perfil de cáliber costuma ser uniforme, desvios podem ser causados por camadas muito delgadas ou pela presença de fraturas.

Figura 10 - Resposta petrofísica característica de uma intrusão máfica

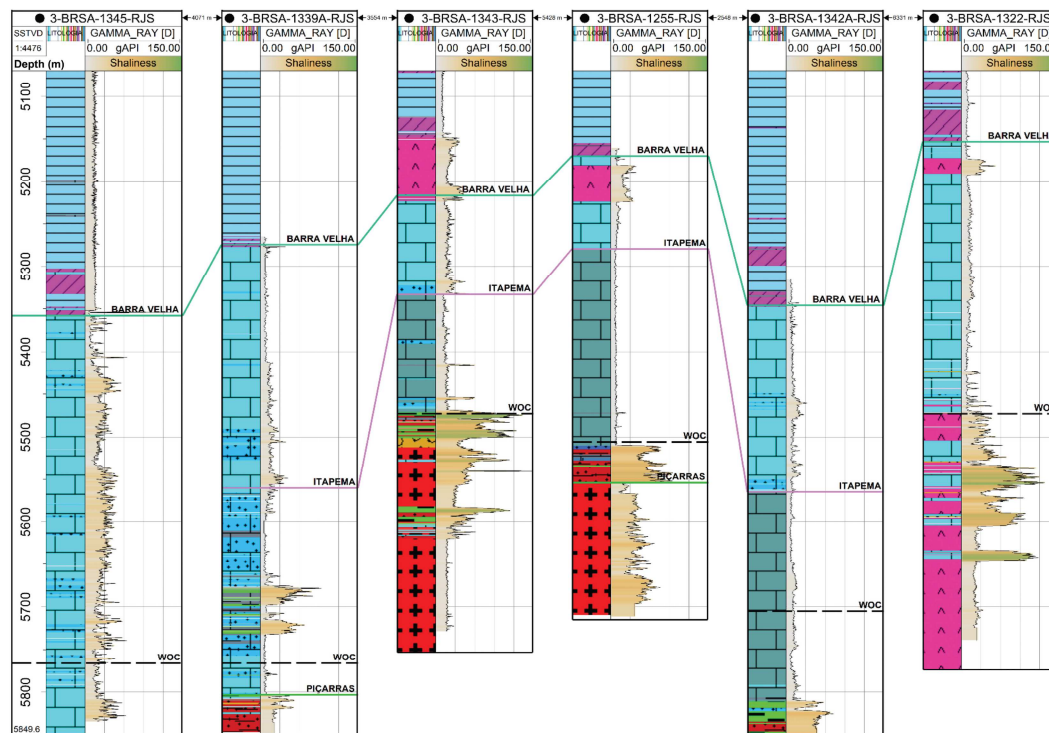


Fonte: Adaptado de MARK et al. (2018).

Na região do Campo de Mero, PETERSOHN et al. (2021) exibem os perfis de raios-gama dos poços públicos, os quais, aparentemente, mostram uma variação dos valores API discordante às estabelecidas por MARK et al. (2018). Nesses poços, os valores variam desde 30° API até 120° API em alguns pontos (

Figura 11), entretanto, é importante frisar que os maiores valores ocorrem nas rochas ígneas basais dos poços, as quais correspondem a derrames basálticos, enquanto as que ocorrem próximo ao topo da Formação Barra Velha e possuem valor API médio próximo a 30 são efetivamente intrusões.

Figura 11 - Painel de correlação de poços exploratórios públicos



Fonte: PETERSOHN et al. (2021).

Recentemente alguns estudos utilizaram uma ferramenta de perfilagem mais avançada, denominada Fullbore Formation Microimager (FMI), para imagear e descrever rochas ígneas perfuradas por poços. Esse tipo de perfil de imagem fornece grande apoio os estudos de exploração de petróleo quando ocorre a presença de rochas ígneas pois permitem a caracterização de fácies que ajudam a descrever a história deposicional dessas rochas e, no interesse de localizar zonas potenciais de reservatórios, auxiliam na identificação de zonas porosas e permeáveis, o que não é possível em perfis elétricos.

FAROOQUI et al. (2009) utilizaram a informação do FMI associada à ressonância magnética e outros perfis petrofísicos para identificar 11 tipos de rocha ígneas presentes no maior campo de petróleo da China, Daqing. Na Bacia de Santos, FORNERO et al., (2019) utilizaram dados de FMI associados a petrografia e análises geoquímicas para descrever quatro fácies vulcânicas em um poço (**Erro! Fonte de referência não encontrada.**), tendo sido a primeira vez que lavas pahoehoe foram descritas em uma bacia “*offshore*” brasileira.

A análise petrofísica de poços permite descrever intrusões e, recentemente, o advento dos perfis de imagem possibilita, inclusive, a caracterização de derrames basálticos e o potencial dessas camadas como reservatórios. Perfis clássicos, como o cáliper, auxiliam na estimativa de fraturas, enquanto perfis sônico e de densidade contribuem para modelagens sísmicas mais realísticas e, por consequência, de resultados mais confiáveis.

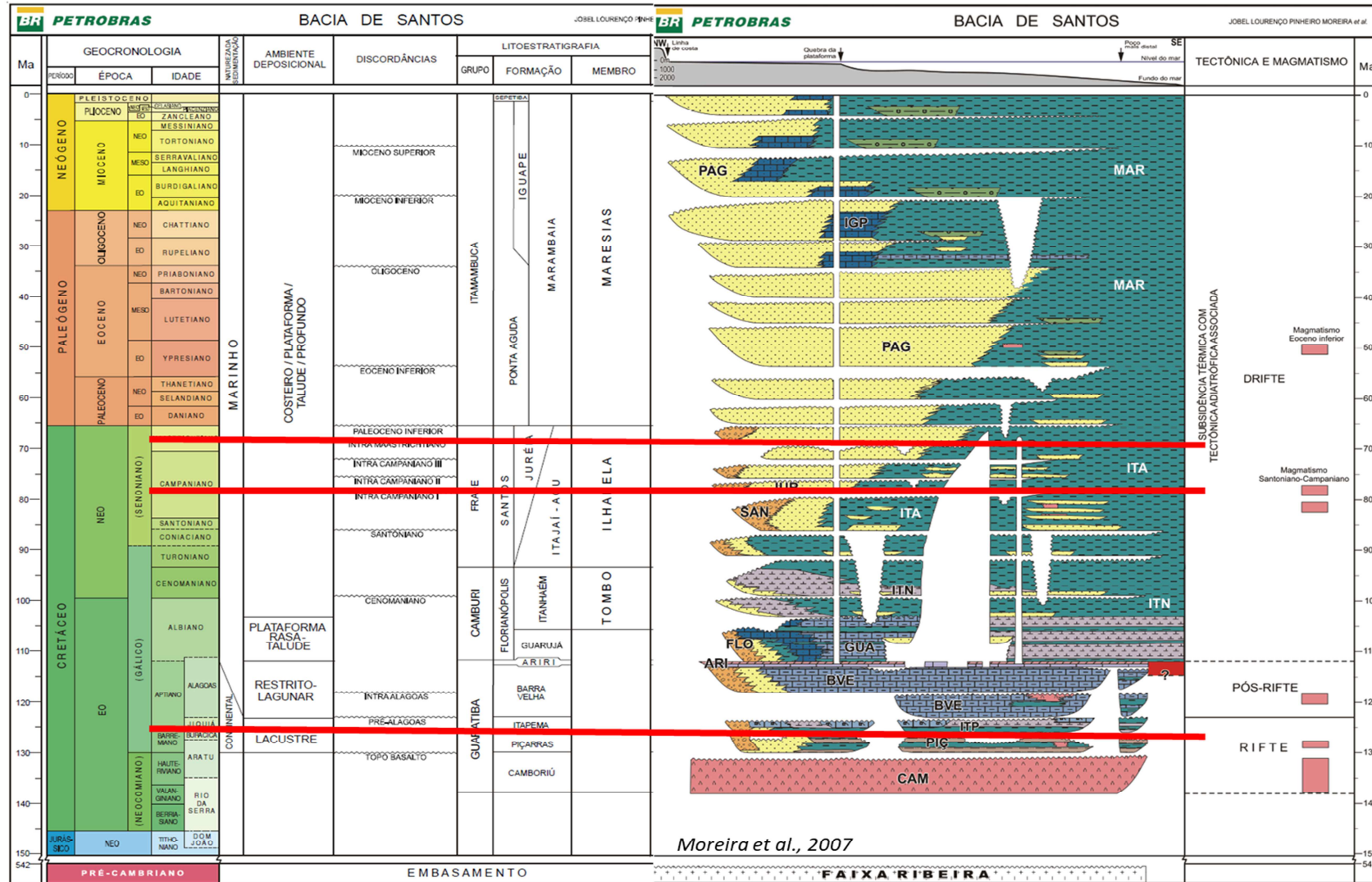
1.1 Evolução Sedimentar e Tectonomagmática da Bacia de Santos

A Bacia de Santos é, atualmente, um dos principais focos de exploração de hidrocarbonetos do Brasil e seus reservatórios carbonáticos da seção pré-sal estão entre as descobertas recentes mais importantes das últimas décadas (BELTRAO et al., 2009; KARNER et al., 2021). O Campo de Mero, descoberto pelo poço exploratório 2-ANP-2A-RJS, possui reservas estimadas em 11.94 bilhões de barris de óleo equivalente (bboe) considerado de boa qualidade, a 29° API (PETERSOHN et al., 2021), sendo o setor NW do Bloco de Libra, ofertado na 1ª Rodada de Partilha de Produção Pré-Sal, em 2013 (ANP, 2013).

Localizada no sudeste da margem continental passiva brasileira, a Bacia de Santos se estende por, aproximadamente, 350 mil km², com lâmina d'água que chega a ultrapassar 3000 metros. É limitada a Norte pelo Alto de Cabo Frio e a Sul pelo Alto de Florianópolis, situada no contexto da Faixa Ribeira, a bacia apresenta embasamento cristalino composto por granitos e gnaisses do pré-cambriano (MOREIRA et al., 2007). Sua formação está ligada à separação do supercontinente Gondwana durante o Cretáceo, cujo processo é, até hoje, fruto de intenso debate na comunidade científica (MOHRIAK; LEROY, 2013; SALOMON; PASSCHIER; KOEHN, 2017; REUBER; MANN, 2019; PEACE et al., 2020; ALVES et al., 2020).

O evento de separação continental foi marcado por extenso e volumoso magmatismo que afetou tanto as bacias intracratônicas (Província Magmática Paraná-Etendeka) quanto a crosta rifteada (PEACE et al., 2020). Este magmatismo, definido por basaltos toleíticos localmente agrupados na Formação Camboriú e datados em 130-121 Ma, compõe o embasamento econômico da bacia (MOREIRA et al., 2007), o qual é observado em vários poços do Campo de Mero, na base da seção sedimentar (PETERSOHN et al., 2021).

Figura 12 - Carta estratigráfica da Bacia de Santos



Legenda: Linhas vermelhas indicam datações na área do Bloco de Libra, executadas por RANCAN et al. (2018).

Fonte: Adaptado de Moreira et. al. (2007).

O registro sedimentar pode ser dividido em três grandes supersequências: rifte, pós-rifte e drifte (Figura 12). Os sedimentos da sequência rifte foram depositados durante o Barremiano (130-125 Ma) e são depósitos lacustrinos, leques aluviais de conglomerados e arenitos, siltitos e folhelhos, litoestratigraficamente denominados Formação Piçarras. Do Neobarremiano ao Eoaptiano (125-120 Ma), ocorrem calcirruditos e grainstones de alta energia constituídos de fragmentos de bivalves intercalados a folhelhos escuros, ricos em matéria orgânica em sua porção distal, e correspondem à Formação Itapema (MOREIRA et al., 2007).

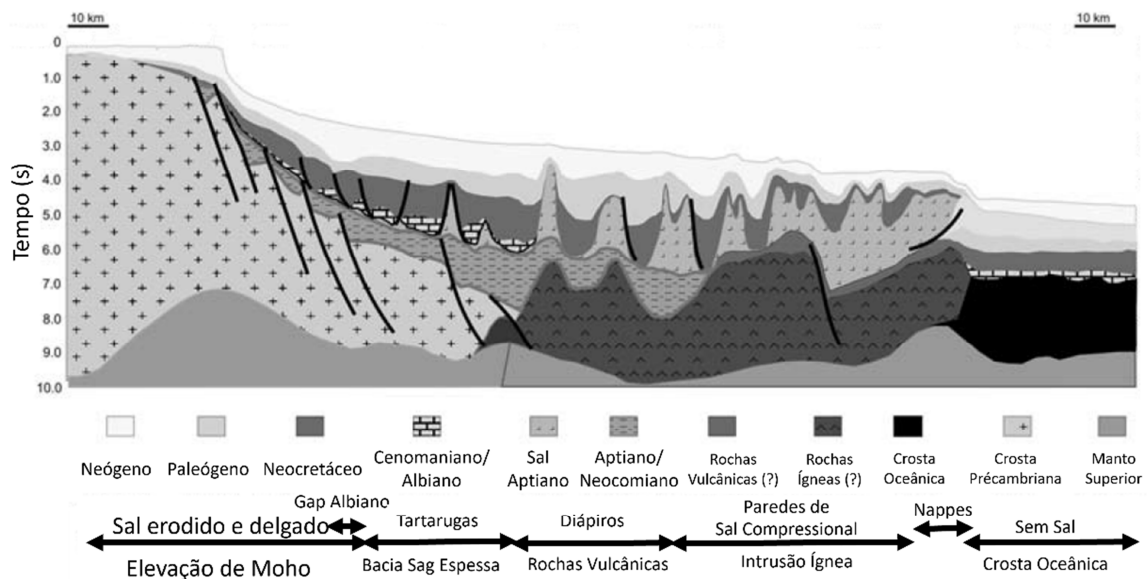
Na sequência pós-rifte, caracterizada pela transição de um ambiente continental para o marinho raso (intervalo sag), a porção inferior da Formação Barra Velha foi depositada, durante o Albiano (113-100 Ma), consistindo em calcários microbiais, estromatólitos e laminitos nas porções proximais, folhelhos nas porções distais. Sua porção superior é caracterizada por carbonatos “in situ”, como estromatólitos e laminitos. Após o estabelecimento de um ambiente de mar aberto, ao fim do Aptiano, uma espessa sequência evaporítica foi depositada, em um tempo estimado de apenas um milhão de anos e denominada Formação Ariri. Composta principalmente de halita e anidrita, com outros sais exóticos associados. A Formação Ariri constitui o selo efetivo do sistema petrolífero do pré-sal. A sequência drifte é composta pelos sedimentos depositados do Albiano ao recente e é caracterizada pelos sedimentos marinhos dos grupos Camburi, predominantemente carbonático, Frade e Itamambuca, predominantemente siliciclásticos (MOREIRA et al., 2007; CARLOTTO et al., 2017).

Até o momento, cinco grandes fases de eventos magmáticos foram identificadas na bacia. Duas fases durante a fase rifte, uma durante o pós-rifte (≈ 120 Ma) e duas durante a fase drifte (≈ 80 Ma e ≈ 50 Ma). Na região do Bloco de Libra, apenas três eventos ígneos são descritos: um evento inicial intrusivo no Aptiano (126 Ma), grandes corpos extrusivos entre 118-120 Ma e intrusões durante o Santoniano-Campaniano (80 Ma). As rochas ígneas do Aptiano são compostas de basaltos toleíticos, enquanto as mais recentes são compostas de diabásio alcalino e gabro (MOREIRA et al., 2007; RANCAN et al., 2018).

A abertura do Oceano Atlântico formou vários pares conjugados de bacias rifte, caracterizados por bacias com evolução tectono-sedimentar similar, porém distintas entre si, desde bacias muito largas e com espessas camadas de sal Aptiano, a bacias estreitas sem a presença de sal (MOHRIAK; NEMČOK; ENCISO, 2008; BIARI et al., 2021). Tais diferenças podem estar associadas às fases complexas de rifteamento e consequente quebra bem como à provável influência do arcabouço estrutural (PEACE et al., 2020).

O modelo de evolução das bacias do sudeste de MOHRIAK; NEMČOK; ENCISO (2008) descreve a ocorrência de cinco fases tectônicas (Figura 13). A primeira fase representa o início da extensão litosférica, com afinamento regional da crosta e depocentros locais, na segunda fase, ocorre aumento do estiramento litosférico e rifteamento intracontinental associado aos basaltos toleíticos do Eocretáceo, falhamento normal ao longo da crosta, formando hemigrábens a leste da Faixa Ribeira. A terceira fase é representada pela aquiescência tectônica e erosão dos blocos rifte e início da deposição de sedimentos de ambiente transicional a marinho raso, deposição das Formações Barra Velha e Ariri. A quarta fase é caracterizada pelo estabelecimento do centro de espalhamento do assoalho oceânico e, por fim, na quinta fase, ocorre aumento da batimetria, início da tectônica de sal além de eventos tectono-magmáticos na forma de intrusões ígneas em Santos e Campos, do fim do Cretáceo ao início do Paleógeno, e a extrusão, no Eoceno, do Complexo de Abrolhos, na Bacia do Espírito Santo.

Figura 13 - Perfil com principais elementos tectônicos e sedimentares na Bacia de Santos



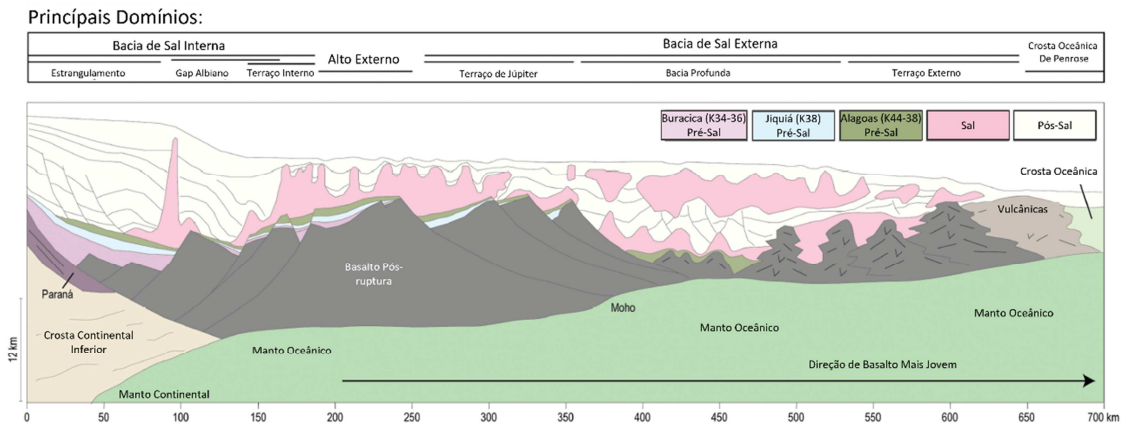
Fonte: Modificado de MOHRIAK; NEMČOK; ENCISO (2008)

KARNER et al., (2021) sugerem um novo modelo (

Figura 14), no qual uma espessa crosta magmática ocorre a leste de um ponto de exumação, progressivamente se torna mais jovem nessa direção e sucessivamente sobreposta por falhamento rúptil, os sistemas deposicionais correspondentes também são transgressivos no tempo, determinando o ritmo da formação da crosta subjacente, com um espalhamento oceânico inicial e quebra continental no Valanginiano tardio, início do Hauteriviano formando a bacia interna e o alto externo, o que permitiu a deposição da Formação Piçarras.

Numa segunda etapa ocorre um salto da crista de separação durante o Barremiano-Aptiano formaria o Terraço de Júpiter e deposição dos sedimentos da Formação Itapema. Um terceiro salto no fim do Aptiano, teria formado os terrenos distais de exumação do manto associados à bacia externa e deposição dos sedimentos da Formação Barra Velha.

Figura 14 - Cenário de Desenvolvimento estrutural-estratigráfico proposto por KARNER et. al., 2021



MATOS et al. (2021, 2022) descrevem quatro estágios tectônicos atuantes durante o desenvolvimento da Bacia de Santos e seu par conjugado, a Bacia da Namíbia. No primeiro estágio, entre 140 e 126 Ma, ocorre o magmatismo das Províncias Borborema e Paraná-Etendeka e propagação do rifte de Norte para Sul. No estágio seguinte, entre 126 e 123 Ma, inicia a hiperextensão do segmento Santos-Namíbia, seguido pelo início do movimento divergente entre África e América do Sul, entre 123 e 117 Ma e, por fim, entre 117 e 113 Ma, ocorre o estabelecimento das Bacias de Sal Centrais.

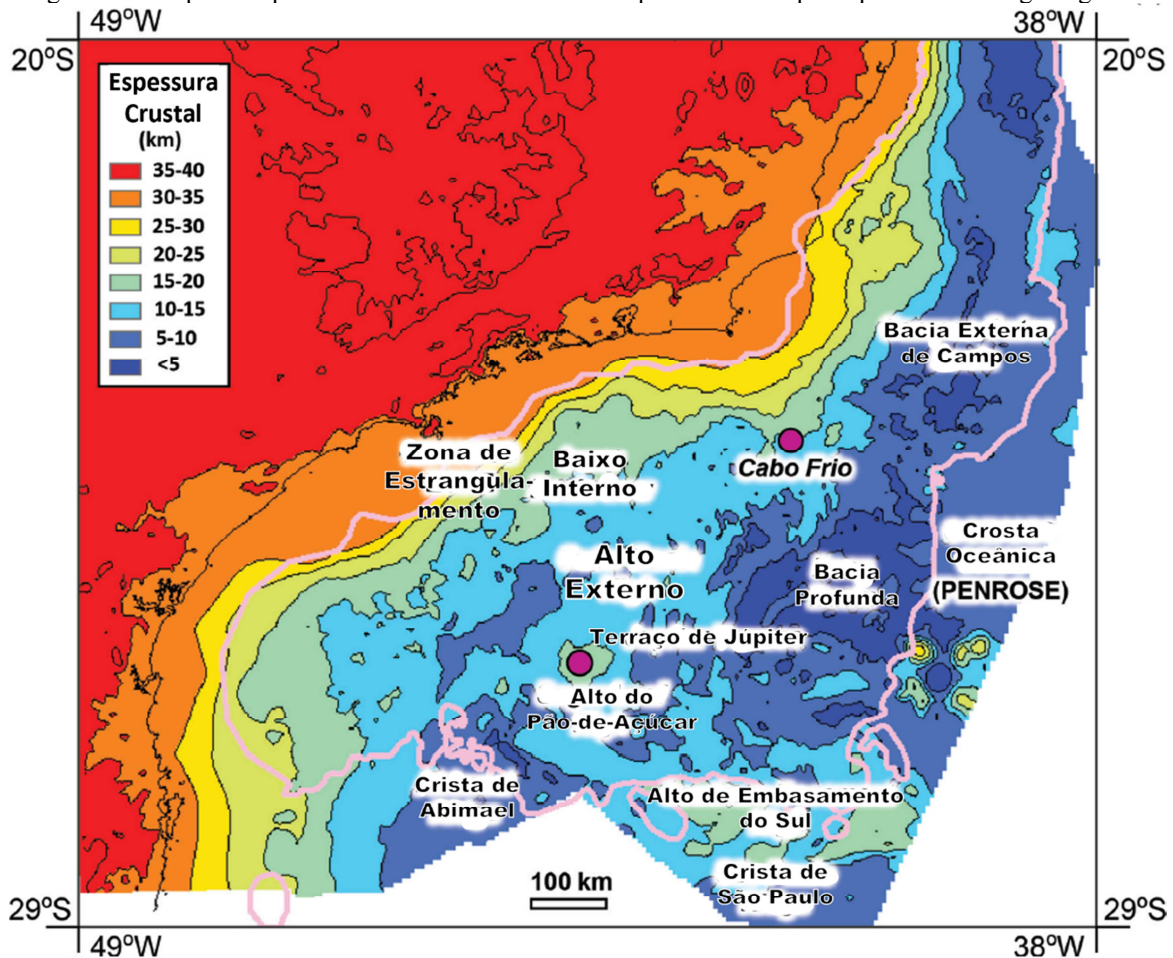
Ao longo da fase de estiramento pré-rifte, diversos compartimentos rifte intracontinentais se formaram ao longo de um caminho sinuoso de mais de 900 km de comprimento, junto a serras litorâneas com 2000 m de altitude, nas bacias de Santos, Campos e Espírito Santo, e são descritos como o Rifte Continental do Sudeste Brasileiro (RCSB). Estes compartimentos teriam se desenvolvido a partir de um campo de extensão com direção NNW-SSE, reativando lineamentos de direção NE-ENE do Proterozóico (RICCOMINI; SANT'ANNA; FERRARI, 2004).

A maior parte da porção distal da Bacia de Santos ocorre sobre uma estrutura gigante denominada Platô de São Paulo, a qual muitos autores acreditam representar uma grande área de extensão crustal anômala (referências em CARLOTTO et al. (2017), p. 170). Outra feição estrutural de destaque é o Alto Externo (Figura 15), na direção do qual ocorre adelgaçamento das sequências pré-sal. Supõe-se que esse alto estrutural represente uma parte da crosta menos

estirada, com 340 km de comprimento, por 130 km de largura, o manteria a região mais elevada que seu entorno.

A estrutura do Bloco de Libra se situa no extremo nordeste do Alto Externo, próximo ao Alto de Cabo Frio. Devido a um posicionamento elevado desde o estágio de rifte e por sua distância de fontes siliciclásticas significativas, o local era ideal para a deposição de carbonatos de alta energia em águas rasas, tais como coquinas, grainstones e packstones (CARLOTTO et al., 2017).

Figura 15 - Mapa de espessura crustal das Bacias de Campos e Santos e principais estruturas geológicas

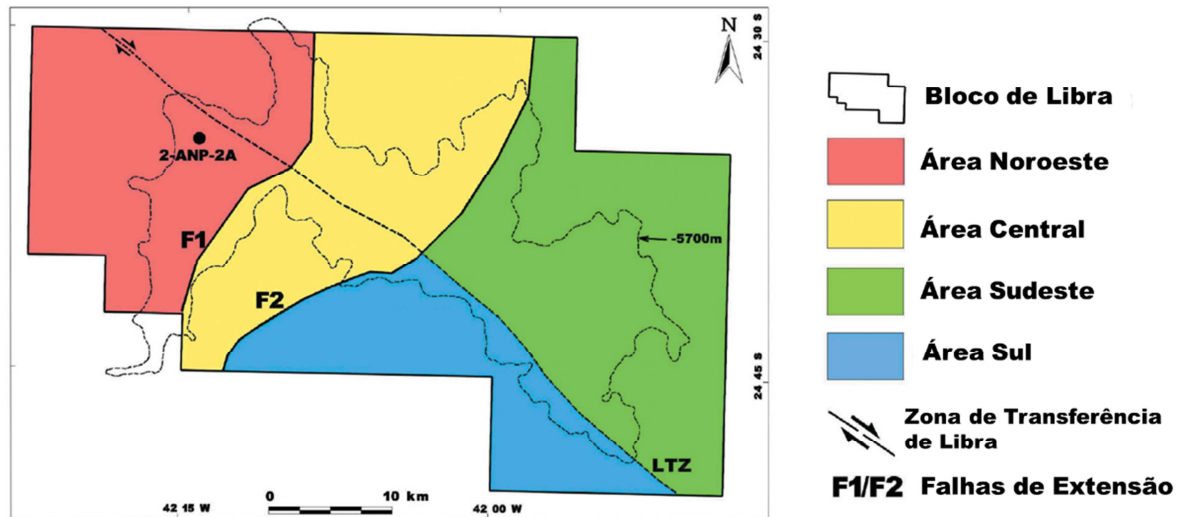


Fonte: Adaptado de KARNER et. al., 2021.

Quatro distintos compartimentos estruturais (Figura 16), noroeste, central, sudeste e sul, associados à presença de duas feições lineares importantes, a Zona de Transferência de Libra (ZTL, direção NW-SE) e o Lineamento de Curitiba (LC, direção WSW-ESE) caracterizam o Bloco de Libra. Além da heterogeneidade típica de reservatórios carbonáticos, a ampla ocorrência de rochas ígneas impacta negativamente a modelagem de reservatórios e a previsão de produção. A compartimentação tectônica foi responsável pelo condicionamento e

distribuição de magmatismo ao longo do bloco: apenas na seção pré-sal em Mero, no setor noroeste, nas seções pré e intrassal no setor central e distribuído em todas as seções (pré, intra e pós-sal) do setor sudeste. (CARLOTTO et al., 2017; RANCAN et al., 2018; ZHAO et al., 2019)

Figura 16 - Mapa esquemático mostrando os quatro compartimentos do bloco de Libra

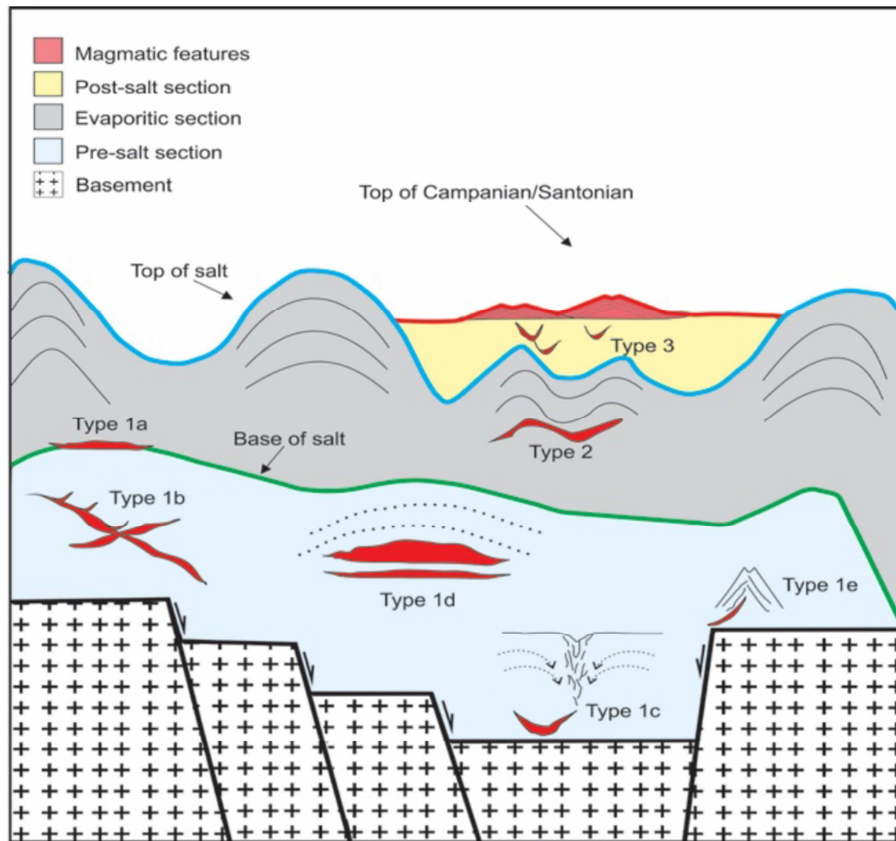


Fonte: Adaptado de CARLOTTO et al., 2017

O magmatismo do Eoaptiano a Campaniano, datado por RANCAN et al. (2018), utilizando datação $^{40}\text{Ar}/^{39}\text{Ar}$, afetaram a região do campo de Mero e do Bloco de Libra, onde sísmica e poços exibem rochas magmáticas nas seções pré-sal e sal (DE OLIVEIRA; RANCAN; OLIVEIRA, 2019; KANGXU et al., 2019; GORDON et al., 2022), ambos de alta impedância acústica e que representam maior desafio à interpretação sísmica.

As camadas de rocha intrusiva no pré-sal de Libra são geralmente da ordem de dezenas de metros e, embora geralmente associadas ao metamorfismo de contato, que forma uma auréola termal de cozimento paralela à intrusão, o impacto na qualidade do reservatório de rochas hospedeiras carbonáticas é previsto para ser relativamente baixo. (KANGXU et al., 2019; MOHRIAK; GORDON; MELLO, 2021). As ocorrências de soleiras foram caracterizadas Área de Libra (DE OLIVEIRA; RANCAN; OLIVEIRA, 2019) e foram classificadas em três tipos (Figura 17), de acordo com a geometria e os níveis estratigráficos, onde as soleiras em forma de molho próximas à base de sal foram chamadas de Tipo 1a e descritas como tendo uma geometria plana apresentando alta dificuldade de mapeamento devido à mistura de resposta do sinal sísmico entre anidrita basal, rocha carbonática e a própria soleira. Essas soleiras são o foco principal deste estudo.

Figura 17 - Classificações das ocorrências magmáticas em Mero/Bloco de Libra



Fonte: DE OLIVEIRA; RANCAN; OLIVEIRA, 2019.

1.2 Atributos Sísmicos Pós-Empilhamento

A sísmica de reflexão consiste, basicamente, em gerar ondas sísmicas artificiais e medir o tempo necessário para que as ondas se desloquem da fonte até uma série de geofones (ou hidrofones) dispostos ao longo de uma ou várias linhas. A partir dos tempos de viagem necessários até a localização de cada grupo de geofones e da velocidade das ondas, suas trajetórias são reconstruídas. A predominância do método sísmico sobre outros métodos geofísicos se deve a vários fatores, sendo os mais importantes a alta precisão, alta resolução e grande penetração de que o método é capaz. As chegadas de energia nos geofones representam os eventos de reflexão (SHERIFF; GELDART, 2012). As reflexões são provenientes de interfaces onde as propriedades acústicas das rochas mudam, dessa forma, a impedância acústica de uma camada de rocha é o produto da densidade e da velocidade dessa

camada e a reflexão é gerada pelo contraste de impedância acústica entre a camada superior e a inferior. É um fato, portanto, que a impedância e litologia seguem uma à outra e seus limites normalmente coincidem (BROWN, 2011).

A velocidade com que as ondas sísmicas se propagam é determinada pelas propriedades elásticas das rochas: densidade, módulos de Young, de cisalhamento e bulk, velocidades cisalhante e compressiva. No caso de rochas vulcânicas, a partir de medidas de laboratório, CARROLL (1969) demonstra que essas características, com exceção da densidade, podem ser estimadas com confiança a partir das velocidades das ondas compressivas. Em geral, no entanto, as rochas possuem várias composições, texturas, porosidades e até diferentes presenças de fluidos, dessa forma, um mesmo tipo de rocha pode exibir um grande intervalo de velocidades e densidades (KEAREY; BROOKS; HILL, 2002).

Os dados oriundos da aquisição sísmica são visualizados em sismogramas de reflexão, que podem ser exibidos de diferentes formas durante o processamento sísmico (e.g. tiro, ponto médio comum) e no volume sísmico final entregue ao intérprete, na forma de visualização de linhas, traços ou mesmo linhas arbitrárias. Esses sismogramas são agrupamentos de traços sísmicos. O traço sísmico representa a resposta combinada das camadas da subsuperfície em relação ao pulso sísmico, em outras palavras, é a convolução do pulso com uma série temporal denominada função refletividade, composta de uma série de picos. Cada pico possui uma amplitude que é relacionada ao coeficiente de reflexão entre camadas e um tempo de trajeto que é equivalente ao tempo de ida e volta (two-way) da propagação da onda sísmica. Dessa forma, um dos objetivos finais do processamento sísmico é estimar as impedâncias acústicas das rochas. (KEAREY; BROOKS; HILL, 2002)

Como diferentes rochas podem possuir propriedades acústicas e elásticas semelhantes, a caracterização das reflexões sísmicas não permite estimar, sem ambiguidades, o tipo de rocha à qual aquela impedância pertence. Devido à sua capacidade de revelar características ocultas a partir da identificação de padrões e quantificação de propriedades específicas, os atributos sísmicos costumam ser empregados com o objetivo de redução da ambiguidade. São capazes de identificar anomalias, extrair características e até prever litologias. Como os dados sísmicos podem ser descritos de inúmeras maneiras, é impossível contabilizar todas as possibilidades, existem centenas e esse número cresce continuamente, levando à necessidade de análise de múltiplos atributos simultaneamente. Tarefa hercúlea para um humano, porém básica para computadores superpotentes (BARNES, 2016).

Atributos sísmicos possuem uma ampla definição, TANER et al. (1994) os definem como toda e qualquer observação extraída do dado sísmico, para BROWN (1996) são

derivados de uma medidas sísmica básica, CHEN; SIDNEY (1997) os consideram medidas específicas de características geométricas, cinemáticas, dinâmicas ou estatísticas derivadas do dado sísmico, CHOPRA; MARFURT (2005) os resumem a medidas quantitativas de uma característica sísmica de interesse, entretanto, em trabalho posterior, o segundo autor refina a definição para uma variedade de medidas simples da forma da onda sísmica até estimativas altamente sofisticadas de anisotropia azimutal, ruptibilidade, atenuação e intensidade de difração (MARFURT, 2018). Em um trabalho mais recente DEWETT; PIGOTT; MARFURT (2021), os atributos englobam um número infinito de permutações matemáticas, algoritmos ou características observáveis que podem dividir, filtrar, classificar ou descrever a forma da onda sísmica, um subconjunto dela, por si só ou no contexto de formas de onda vizinhas, contíguas ou diacrônicas.

Dessa forma, os atributos podem ser considerados desde simples observações até estimativas sofisticadas, em sua última classificação, inclusive resultados de algoritmos de classificação automática de fácies podem ser considerados atributos sísmicos. Neste trabalho, consideraremos a definição básica de CHOPRA; MARFURT (2005), atributos sísmicos são apenas aqueles extraídos diretamente do dado sísmico, ou seja, saídas de extração de facetas, como a PCA, ou de técnicas de agrupamento não serão consideradas atributos. Essa escolha visa evitar confusões com os resultados a serem discutidos nas técnicas de redução de dimensionalidade.

BARNES, (2016) divide os atributos sísmicos em geológicos, geofísicos ou matemáticos. Os geológicos registram propriedades estruturais, estratigráficas ou litológicas. Os geofísicos são referentes a propriedades das ondas sísmicas ou de suas wavelet, como amplitude, fase, frequência, entre outros. Dificilmente mostram informação geológica direta. Os matemáticos são os mais numerosos por sua facilidade em serem criados, entretanto, carecem de significância geológica. Os atributos são gerados a partir de uma grande variedade de métodos, que incluem: estatística, análise de traço complexo, correlação, filtragem, decomposição espectral etc.

Quadro 1 - Diferentes tipos de atributos sísmicos (continua)

Atributo	Representação Matemática	Aplicação
Amplitude Instantânea (envelope do traço, força da reflexão)	Medida da amplitude, independente de polaridade ou fase do dado sísmico.	Destaca anomalias de amplitude em geral (por exemplo, “Bright spots” e “dim spots”)
Amplitude RMS (“Root Mean Squared”)	Raiz media quadrática da amplitude instantânea	Destaca anomalias de amplitude, utilizada como função de ponderação para

		atributos de média
Fase Instantânea	Medida angular, em graus, da posição relativa de uma forma de onda sinusoidal, matematicamente descontínuo	Destaca reflexões sísmicas.
Cosseno da fase	Semelhante à fase instantânea, porém matematicamente contínuo	Reflexões sísmicas destacadas, ganho automático definitivo
Frequência Instantânea	Derivada no tempo da fase instantânea	Amplia altas frequências e suprime baixas frequências
Mudança de Amplitude Relativa	Contrapartida em amplitude da frequência instantânea	Revela detalhes escondidos e destaca zonas de interferência de reflexão
Sweetness	Resposta da amplitude dividida pela raiz quadrada da resposta da frequência	Capaz de identificar “Sweet spots” em sedimentos siliciclásticos jovens
Frequência Média	Obtida a partir da ponderação média da frequência instantânea em uma janela	Representa estimativa variante no tempo da frequência espectral média de Fourier
Azimute	Quantifica a direção do mergulho da reflexão.	Utilizado como entrada para atributos 3D
Mergulho	Quantifica o mergulho em relação ao plano horizontal	Utilizado como entrada para atributos 3D
Gradiente aparente	Combina mergulho e azimuth	Aparente topografia iluminada com exibição monocromática
Relevo Sombreado (“shaded relief”)	Combina mergulho, azimuth e técnicas de relevo	Produz imagens interpretáveis que revelam anticlinais, sinclinais, dobras, bacias, falhas e canais
Curvatura	Quantifica mudanças no mergulho e azimuth ao longo de reflexões	Revela detalhes estruturais, como falhas e bordas de canais
Descontinuidades	Família de atributos que detecta quebras na continuidade dos refletores	Destaca falhas, canais, diápiros, “pinch-outs”, artefatos e ruídos. Ideal para mapeamento de falhas e canais.
Decomposição Espectral	Aplica análise de Fourier para dividir o dado sísmico em volumes de frequências	Conseguem revelar características ausentes em outros atributos, muito utilizados na detecção de canais e camadas finas.
Impedância Acústica Relativa	Destaca a impedância acústica das camadas a	Mostra as camadas na subsuperfície

	partir de inversão recursiva	
Texturas GLCM	Medidas estatísticas de texturas a partir de matriz quadrada que revela a ocorrência dos níveis de cinza dos pixels.	Cada atributo textural representa propriedades específicas da imagem sísmica, como homogeneidade, contraste e complexidade textural. Muito aplicados em classificação de fácies sísmicas

Fonte: BARNES, 2016

Texturas são definidas como “um termo cotidiano relacionado ao toque que inclui conceitos como áspero, sedoso e irregular. Quando uma textura é áspera ao toque, a superfície exhibe diferenças acentuadas de elevação no espaço da ponta do dedo. Em contraste, as texturas sedosas exibem diferenças muito pequenas na elevação” por HALL-BEYER, (2017). DE MATOS et al. (2011) argumentam que texturas sísmicas funcionam de maneira análoga, onde a elevação é substituída pela amplitude, e o toque dos dedos por uma janela de análise orientada ao longo do mergulho e azimute locais da estrutura.

WEST et al. (2002) introduziram o conceito de análise textural 3D em volumes sísmicos, a partir da combinação da análise de imagem textural com uma classificação de rede neural para mapear quantitativamente as fácies sísmicas em dados tridimensionais. O objetivo da análise textural é descrever matematicamente a distribuição dos valores de pixels em uma sub-região do dado, quantificando de maneira efetiva a organização espacial das reflexões sísmicas. A técnica mais usada para quantificar textura é denominada “matrizes de co-ocorrência de níveis de cinza” (“*Gray-Level Co-occurrence Matrices*”, GLCM).

A Erro! Fonte de referência não encontrada. representa a quantização de um traço sísmico, de polaridade europeia (vales negativos, preenchidos com preto, picos positivos preenchidos com branco) usando nove níveis de cinza. Os níveis 1-4 representam vales, o nível 5 representa o cruzamento em zero, e os níveis 6-9 representam o pico. Para calcular as matrizes de co-ocorrência sensíveis a mudanças laterais na refletividade, DE MATOS et al. (2011) apresentam um novo método de cálculo com quatro parâmetros principais: o nível de quantização da imagem (a quantidade de níveis de cinza), o tamanho da janela deslizante, a direção e distância dos pares de voxels e as estatísticas usadas como atributos de textura. Por fim, vários atributos que descrevem a ocorrência dos níveis de cinza são gerados: Contraste, Correlação, Dissimilaridade, Energia, Entropia, Homogeneidade, Média e Variância.

Os atributos sísmicos são capazes de destacar diversas características “escondidas” em um dado sísmico. Cada atributo sísmico pode ser considerado uma nova dimensão do dado sísmico e, devido à limitação humana e dos programas comerciais de interpretação sísmica, é aparentemente impossível analisar muitas dimensões simultaneamente. Entretanto, há décadas matemáticos desenvolvem algoritmos capazes de analisar simultaneamente dados de alta dimensionalidade e, graças ao recente avanço do poder computacional, esses algoritmos podem ser implementados com relativa facilidade, como poderá ser visto na próxima seção.

1.3 Aprendizado de Máquinas

Em 1958, o conceito de perceptron, o primeiro “neurônio artificial”, foi introduzido por ROSENBLATT (1958), SAMUEL (1959) aplicou aprendizado de máquina ao jogo de damas e previu as muitas outras possibilidades da técnica computacional. Infelizmente, o amplo uso dessas ferramentas ainda estaria décadas distante devido à limitadíssima capacidade dos processadores à época. Em 2003, um dos primeiros artigos a reunir e comparar diferentes técnicas de classificação de fácies sísmicas com aprendizado de máquinas e análise multiatributos foi publicado na revista *The Leading Edge* (COLÉOU; POUPON; AZBEL, 2003). Ainda assim, capacidade de processamento – gráfico e tradicional – e memória de acesso rápido limitados impediram a adoção de técnicas promissoras nos crescentes cubos sísmicos tridimensionais. A plataforma de pesquisa de artigos científicos Google Acadêmico estima 75 citações do artigo mencionado após os primeiros dez anos de publicação (ou seja, até 2012), na década seguinte, no entanto, o mesmo artigo foi citado mais de 230 vezes, um crescimento superior a 300%, indicando uma clara explosão da aplicação da técnica, auxiliada pelo crescimento do poder computacional atrelado à criação de placas de processamento gráfico (GPU) dedicadas. Outro exemplo de revisão a ser citado é a de BARNES; LAUGHLIN (2002), o qual possui uma quantidade de citações mais modesta, em torno de 80.

As técnicas de aprendizado de máquina são algoritmos matemáticos relativamente simples que podem aprender e gerar clusters/classes com base em padrões (ou inter-relações entre) os dados. Dependendo da disponibilidade de dados, podemos usar algoritmos supervisionados ou não supervisionados (INFANTE-PAEZ; MARFURT, 2019). A classificação não supervisionada atribui automaticamente cores diferentes a diferentes grupos de dados, enquanto a supervisionada exige que o intérprete humano rotule explicitamente um

subconjunto de dados como pertencente a uma determinada classe ou fácies sísmica, estes dados rotulados servem como “treinamento” e são usados para estabelecer a relação entre entrada e saída (HA et al., 2021). Uma vez treinado, o algoritmo é então aplicado a outro subconjunto de “dados de validação” (fácies interpretadas não usadas na etapa de treinamento) para determinar se o algoritmo é suficientemente preciso. Se a validação for bem-sucedida, o algoritmo é então aplicado a todo o volume de dados sísmicos.

O objetivo principal no aprendizado supervisionado é aprender a partir de dados de treinamento rotulados que permitem fazer previsões a respeito de dados não vistos ou futuros. Uma tarefa de aprendizado supervisionado com classes de rótulos discretas é chamada tarefa de classificação. Outra subcategoria é a regressão, onde a saída é um sinal de valor contínuo (RASCHKA; MIRJALILI, 2019). Nos casos onde apenas alguns rótulos predefinidos serão preditos, é utilizado um classificador, se o número a ser previsto é possui um grande intervalo, é utilizada uma regressão (AMR, 2020).

As classificações supervisionadas consistem em dois tipos principais: classificações semissupervisionadas, nas quais um algoritmo de agrupamento é treinado (ou rotulado) por um intérprete para associar um cluster a uma classe específica ou fácies sísmica; e classificações de rede neural puramente supervisionadas, nas quais uma amostra de entrada é alimentada através de muitas camadas de neurônios ocultas para determinar a qual classe, ou “rótulo”, ela pertence. Algoritmos não supervisionados podem ser adaptados para uma classificação semissupervisionada, QI et al. (2016) modificam um mapa topográfico generativo (GTM) onde amostras de treinamento definidas como pertencentes a N classes são mapeadas e associadas às funções de densidade gaussiana correspondentes, permitindo à classificação a geração de uma função de densidade de probabilidade para cada fácies.

As classificações não supervisionadas podem ser divididas em duas categorias: técnicas de projeção, nas quais N atributos de entrada são projetados em um espaço latente de menor dimensionalidade (geralmente duas a três dimensões permitindo mapeamento contra uma matriz de cores contínuas 2D ou 3D); e técnicas de agrupamento, nas quais diferentes pontos de dados são organizados em diferentes agrupamentos, com cada agrupamento atribuído a uma cor distinta (HA et al., 2021).

A técnica de projeção mais conhecida é a análise de componentes principais (Principal Component Analysis - PCA), na qual os atributos de entrada são projetados em dois ou mais autovetores de sua matriz de covariância para capturar a variação máxima dos dados de entrada. A PCA reduz um grande conjunto de atributos sísmicos para indicar variações nos

dados, as quais muitas vezes se relacionam com características geológicas de interesse (GUO; LEWIS; MARFURT, 2008; CHOPRA; MARFURT, 2014a).

A PCA reduz a dimensionalidade e redundância dos dados multivariados de entrada, mas pode omitir características geológicas associadas à menor refletividade (GUO; LEWIS; MARFURT, 2008). A PCA é baseada na suposição de que os dados exibem estatísticas gaussianas, permitindo o uso de estatísticas de segunda ordem para decompor os dados em componentes ortogonais classificados com base em sua variabilidade. Os componentes principais são classificados de acordo com a energia dos dados de entrada que representam, dessa forma, uma fraqueza do método é a possibilidade de mistura de diferentes fácies sísmicas (RODEN; SMITH; SACREY, 2015). Essa questão é abordada por HONÓRIO et al. (2014), que introduz a análise de componentes independente (ICA), baseada em estatísticas de ordem superior e separa um sinal multivariado em subcomponentes independentes, mas não necessariamente ortogonais, encontrando uma representação linear de dados não gaussianos. O conceito de “independência” fornece um meio de capturar informações mais interessantes dos dados multivariados.

Entre os métodos de projeção mais recentes está uma técnica de projeção estocástica e não linear descrita por WALLETT; HA (2021), que usa uma rede neural de aprendizado profundo de autoencoder para “codificar” N atributos de entrada em um espaço latente 3D a ser exibido via mistura das cores vermelho, verde e azul. Auxiliando essa etapa e levando em consideração a importância da normalização dos dados para compensar a grande variação de intervalos dos dados e suas médias, HA et al. (2021) propõem a transformação logarítmica dos dados como uma alternativa mais sofisticada ao tratamento de dados com estatística gaussiana usualmente empregada.

Algoritmos de agrupamento têm como objetivo dividir as amostras dos dados em “clusters”, ou grupos, de modo a minimizar as distâncias internas nos grupos (distâncias intragrupo) e maximizar a distância entre os grupos (intergrupo). Em outras palavras, os dados de um mesmo grupo são o mais similar possível, amostras de diferentes grupos devem ser mais diferentes possíveis. A solução trivial é tratar cada amostra como seu próprio grupo, dessa forma as distâncias intragrupo são todas zeradas e as distâncias intergrupos são maximizadas. Como esse não é o objetivo do agrupamento e para evitar essa situação, uma restrição deve ser implementada, esta pode ser um número de grupos ou tamanho do grupo previamente definidos. Uma maneira de medir as distâncias intragrupo é calcular as distâncias entre cada ponto do cluster e o centroide do grupo. O centroide é basicamente a média de todas as amostras nos clusters. A distância euclidiana média entre algumas amostras e sua

média é o desvio padrão. A mesma medida de distância pode ser usada para medir a dissimilaridade entre os centroides dos grupos (AMR, 2020).

O método de agrupamento mais simples e rápido é o k-means, no qual o dado de entrada para treinamento é “particionado” em um número de grupos definido pelo usuário baseado na distância de um ponto no dado até o centro do grupo. Em geral, o método não leva em consideração o tamanho e a forma de cada grupo durante a classificação. O Modelo de Mistura Gaussiana (Gaussian Mixture Model – GMM) é uma extensão do k-means, neste cada grupo é associado a suas próprias funções de densidade gaussianas multidimensionais. O tamanho, a forma e a posição de cada função de densidade gaussiana são modificados em um esquema iterativo de maximização de expectativa. Uma desvantagem comum dos algoritmos k-means e GMM é que o resultado da classificação não mostra graficamente a proximidade de um cluster a outro. Em contrapartida, os mapas auto-organizáveis (Self-Organizing Maps – SOMs) mapeiam diretamente os clusters para uma variedade deformada, de modo que a posição relativa de cada cluster seja definida (WALLET; HARDISTY, 2019; HA et al., 2021).

Os algoritmos originais de SOM realizam o agrupamento ao encontrar o vetor protótipo (ou neurônio) mais próximo de uma amostra de dados e não fornecem uma medida da confiança do processo de agrupamento (CHOPRA; MARFURT, 2014a; ZHAO et al., 2015). Inovações mais recentes constroem uma distribuição gaussiana após o agrupamento, que fornece uma medida da confiança de cada ponto de dados agrupado. Por fim, o mapeamento topográfico generativo fornece uma medida direta da probabilidade de que um ponto de dados caia dentro de uma determinada classe, estimando as contribuições de todos os pontos latentes da grade espacial para uma amostra de dados usando uma mistura de funções de densidade gaussianas (ROY et al., 2014).

1.3.1 Técnicas de Redução de Dimensionalidade

O geocientista de interpretação possui uma enorme quantidade de dados à sua disposição, desde dados sísmicos tridimensionais com inúmeras versões de processamento e uma grande família de poços e seus dados associados até dúzias de atributos sísmicos que, atualmente, facilmente alcançam terabytes de informação. É humanamente impossível para o intérprete processar tanta informação de maneira eficiente, RODEN; SMITH; SACREY

(2015)apontam a possibilidade e necessidade de extrair informação geológica a partir dos inúmeros atributos de uma só vez. Em outras palavras, através da execução da análise multiatributos (CHOPRA; MARFURT, 2014b).

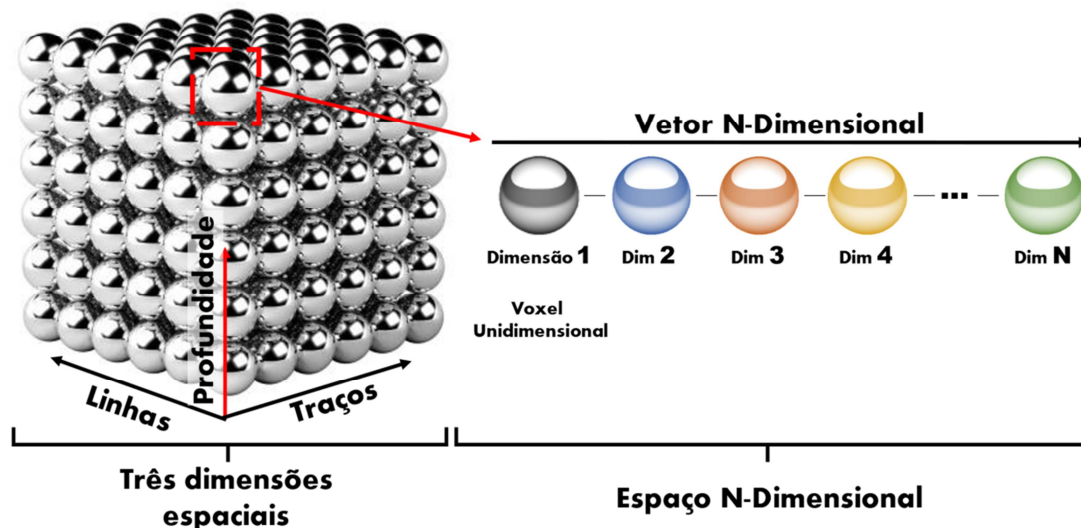
Imagens digitalizadas costumam ser tratadas como um conjunto de pixels, em duas dimensões, ou voxels (pixels volumétricos), em três dimensões (PROUSSEVITCH; SAHAGIAN, 2001). Nos levantamentos de sensoriamento remoto realizados por satélites, em áreas como estudos ambientais, monitoramento agrícola, planejamento urbano, previsão do tempo, entre outros (KUMAR et al., 2020), sensores hiperespectrais observam todo o espectro eletromagnético para fornecer imagens com centenas de bandas espectrais curtas e adjacentes entre si. Entretanto, ao contrário de imagens comuns, cujos pixels costumam consistir em três canais (vermelho, verde e azul), nas imagens hiperespectrais, cada pixel pode possuir centenas de valores (MANTRIPRAGADA et al., 2022; MOHAN; SAPIRO; BOSCH, 2007; DEEPA; THILAGAVATHI, 2015).

Nos dados sísmicos convencionais, cada voxel possui um valor de amplitude que, por si só, tem pouco significado e precisa ser avaliado em contexto a seus vizinhos. Os atributos sísmicos podem ser considerados qualquer medida que quantifica essas relações entre os voxels (MARFURT, 2018). Num paralelo entre as imagens hiperespectrais do sensoriamento remoto e a análise multiatributos, cada voxel do dado sísmico receberá, na forma de um vetor, todos os atributos selecionados para a análise, portanto um cubo sísmico de análise multiatributos pode ser considerado um cubo hiperespectral e, devido à maior quantidade de trabalhos sobre imagens hiperespectrais e às semelhanças destas com a análise multiatributos, este capítulo tratará os dois formatos de modo intercambiável.

Em geral, se imagina que mais informação sempre será melhor que menos, entretanto, na era do “Big Data”, a enorme quantidade de variáveis em uma única amostra pode ser problemática e esse excesso de riqueza de informações é denominado “a maldição da dimensionalidade” (ALTMAN; KRZYWINSKI, 2018; HUGHES, 1968). Cada variável representa uma nova dimensão, que significa uma banda espectral, no caso das imagens hiperespectrais, ou um atributo, no caso da análise multiatributos. A Figura 18 mostra uma representação esquemática do cubo sísmico, onde cada esfera está situada em um ponto de dimensões espaciais (linha, traço e profundidade, ou x, y e z) e ao mesmo tempo cada esfera pode representar um atributo, ou dimensão, diferente, representados pelas esferas de cores diferentes, representando o espaço N-dimensional e, por consequência as dimensões a serem tratadas. Cada vetor N-dimensional representa uma das esferas e possui coordenadas no espaço tridimensional, o conjunto de vetores forma o espaço N-dimensional. O

desenvolvimento de técnicas efetivas e consistentes para lidar com a extensa quantidade de dados representa um desafio significativo, uma vez que a alta dimensionalidade afeta a acurácia de classificação dos métodos de aprendizado de máquina (ISLAM; AHMED; HOSSAIN, 2019).

Figura 18 - Representação esquemática das três dimensões espaciais e o espaço N-dimensional



Fonte: O Autor, 2023.

Dados sísmicos são altamente contínuos, muito redundantes e significativamente ruidosos. A continuidade da reflexão sísmica é uma expressão da continuidade lateral geológica, sendo a base para o processamento de dados sísmicos e possibilitando a interpretação estrutural e estratigráfica. A redundância desse tipo de dado ocorre vertical e lateralmente, é controlada na vertical por uma largura de banda usualmente muito inferior à frequência de Nyquist e lateralmente pela continuidade dos refletores. Dessa forma, amostras consecutivas ao longo dos traços sísmicos são altamente correlacionadas e, quando visualizados em crossplots n-dimensionais, dados sísmicos são expressos em nuvens alongadas, que podem ser infladas pela presença de ruídos, dificultando a tarefa das técnicas de agrupamento (COLÉOU; POUPON; AZBEL, 2003). Um cubo sísmico hiperespectral, portanto, aumenta significativamente a redundância e quantidade de ruído do dado a ser analisado.

Devido à presença de dimensões redundantes, ruidosas e irrelevantes, os algoritmos de aprendizados podem se tornar muito lentos e ter a performance de suas tarefas de aprendizado reduzidas, além de dificultar a interpretação dos resultados (MIAO; NIU, 2016), tornando a seleção de atributos que melhor quantificam as características procuradas num determinado

projeto, seja ele na detecção de padrões geológicos, ambientes deposicionais ou mesmo diferentes litologias, como é o caso deste trabalho, uma etapa crucial para reduzir o custo computacional associado e aumentar a qualidade dos resultados dos algoritmos de classificação (KIM; HARDISTY; MARFURT, 2019; QI et al., 2020).

As bandas espectrais também são denominadas facetas espectrais, em inglês, “spectral features”. Neste estudo, a tradução “faceta” é utilizada em detrimento à tradução comum, “característica”, para aumentar a clareza e evitar confusões como na tradução de “feature characteristics” para “características da característica”.

A alta redundância e correlação de facetas espectrais adjacentes para propósitos de classificação associadas à possibilidade de que tais bandas não sejam as mais efetivas na separação de objetos de interesse levou ao desenvolvimento de técnicas de mineração de facetas, de modo que um grupo efetivo delas possa ser identificado antes da etapa de classificação (BIOUCAS-DIAS et al., 2013).

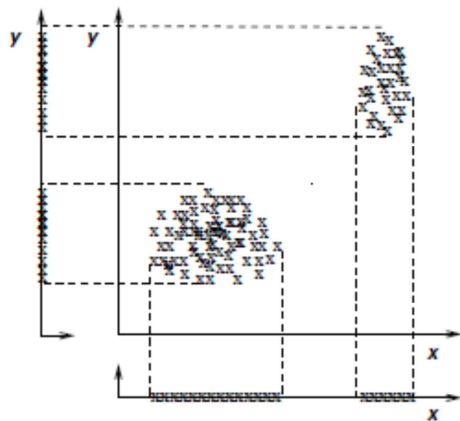
Para ilustrar o problema de redundância entre facetas, (DY; BRODLEY, 2004) mostram que, quando um dado pode ser agrupado da mesma forma utilizando duas facetas diferentes, estas são consideradas redundantes. Se uma faceta não contribui na discriminação de grupos ela é considerada irrelevante (Figura 19). As facetas irrelevantes podem atrapalhar agrupamentos, especialmente quando há mais facetas irrelevantes que relevantes.

As técnicas de mineração de facetas também podem ser denominadas técnicas de redução de dimensionalidade e seus benefícios incluem: diminuição da complexidade computacional e melhora no condicionamento estatístico, a partir do descarte de facetas redundantes (WANG et al., 2020); redução de viés observacional (SOLORIO-FERNÁNDEZ; CARRASCO-OCHOA; MARTÍNEZ-TRINIDAD, 2020); modelos mais simples e ágeis (SÁNCHEZ-MAROÑO; ALONSO-BETANZOS; TOMBILLA-SANROMÁN, 2007); aumento da acurácia de aprendizado (MIAO; NIU, 2016).

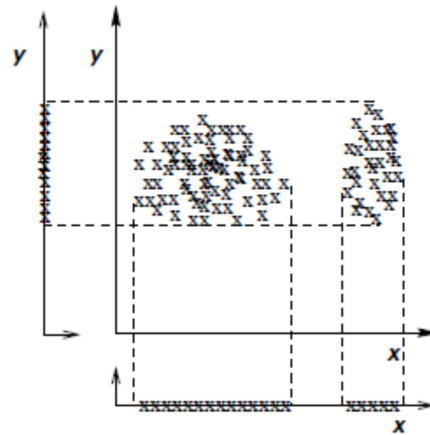
A redução de dimensionalidade é uma técnica bem conhecida de pré-condicionamento de dados, sendo um campo da matemática que lida com as complexidades de conjuntos grandes de dados (HANCER; XUE; ZHANG, 2020). Os diferentes métodos almejam reduzir a dimensionalidade, removendo facetas irrelevantes e redundantes, enquanto preservam características importantes do dado (DEEPA; THILAGAVATHI, 2015), uma vez que nem todas as facetas são necessárias para processos de segmentação ou classificação (MANTRIPRAGADA et al., 2022).

Figura 19 - Facetas redundantes e irrelevantes:

(a)



(b)



Legenda: (a) Facetas x e y são redundantes, x fornece mesma informação que y; (b) Faceta y é irrelevante pois se x for omitida, y descreve apenas um grupo.

Fonte: DY; BRODLEY, 2004.

1.3.1.1 Seleção de Facetas

As técnicas de redução de dimensionalidade podem ser divididas em duas categorias: extração de facetas e seleção de facetas (BIOUCAS-DIAS et al., 2013; ZHENG; YUAN; LU, 2017; HANCER; XUE; ZHANG, 2020; KUMAR et al., 2020; WANG et al., 2020). A extração de facetas transforma todas as facetas originais para construir espaços de facetas com menos dimensões, a dimensionalidade é reduzida a partir da criação de novas facetas oriundas das originais, com isso exploram todas as medidas espectrais disponíveis para extrair características relevantes do dado. Na seleção de facetas, um grupo menor de atributos disponíveis é selecionado a partir dos originais, utilizando um critério de avaliação pré-definido ou uma estratégia de busca, descartando facetas consideradas irrelevantes ou redundantes. É considerado boa prática executar, primeiramente, a seleção de facetas e, em seguida, a extração de facetas. Nas técnicas não-supervisionadas de seleção de facetas, a ausência de rótulos, comumente utilizados para guiar a busca por facetas discriminadoras, aumenta consideravelmente a dificuldade do problema (DY; BRODLEY, 2004).

Os critérios de avaliação medem a distância estatística entre as facetas para avaliar a capacidade de discriminação de um dado subgrupo de facetas, entretanto, são muito custosos

computacionalmente e determinar o número de facetas ótimo costuma ser desafiador, exemplos incluem: distância espectral, variância, divergência espectral, entre outros (KUMAR et al., 2020). Nas estratégias de busca, uma abordagem de busca otimizada identifica o melhor subgrupo de facetas de acordo com um critério de seleção (DEEPA; THILAGAVATHI, 2015), exemplos mais famosos são busca progressiva sequencial (“sequential forward search”, LEE; PARK; LEE, 2017) e eliminação regressiva sequencial (“sequential backward elimination”, VERGARA; ESTÉVEZ, 2014). Ambos são processos iterativos que acrescentam, na busca progressiva, ou removem, na eliminação regressiva, atributos até que não haja melhora do resultado de classificação. As performances de redução desses métodos, como acurácia e custo computacional, são altamente variáveis (WANG et al., 2020).

Outra abordagem popular para distinguir as técnicas de seleção de facetas de acordo com a estratégia de seleção é a divisão em três métodos: filtragem, invólucro (“wrapper”) e híbrido (SOLORIO-FERNÁNDEZ; CARRASCO-OCHOA; MARTÍNEZ-TRINIDAD, 2020). Os métodos de filtragem selecionam facetas relevantes a partir do próprio dado, ou seja, as facetas são avaliadas baseadas em propriedades intrínsecas do dado, independente dos classificadores utilizados na sequência, em contrapartida, métodos de invólucro avaliam subgrupos de facetas usando resultados de algoritmos de agrupamento, por exemplo, probabilidade máxima, vizinhos próximos, regressão logística e máquinas de vetores de suporte (JIA; KUO; CRAWFORD, 2013). Os métodos híbridos tentam explorar as qualidades de ambas as abordagens.

1.3.1.2 Extração de Facetas

Na extração de facetas, uma representação com número de dimensões inferior ao dado original é gerada através da incorporação dos dados de alta dimensionalidade em um subespaço de menor dimensionalidade. De maneira ideal, a representação dimensional reduzida deve preservar as propriedades intrínsecas do dado de alta dimensionalidade, seu resultado, portanto, facilita a classificação subsequente (ZHENG; YUAN; LU, 2017). Os exemplos principais de métodos de extração de facetas são: análise de componente principal (“Principal Component Analysis”, PCA), análise linear discriminante (“Linear Discriminant Analysis”, LDA), decomposição em valores singulares (“Singular Value Decomposition”,

SVD), incorporação linear local (“Local Linear Embedding”, LLE), mapas de autovalores laplacianos (“Laplacian Eigenmaps”), entre outros (HANCER; XUE; ZHANG, 2020).

Em outras palavras, a extração de facetas produz novas facetas ao combinar os atributos originais, preservando a maior parte da informação importante (KUMAR et al., 2020). Existem, no entanto, limitações nos métodos de redução de dimensionalidade. Os baseados na forma da onda consideram apenas a informação espectral, descartando a informação espacial, diminuindo a performance de classificação (SHI et al., 2021). As técnicas de extração de facetas geralmente alcançam maior precisão de classificação, mas seus resultados podem ser difíceis de interpretar, enquanto as técnicas de seleção preservam informação original relevantes mas têm melhora de precisão de classificação muito limitada (ZHENG; YUAN; LU, 2017). Além disso, técnicas que utilizam algoritmos de agrupamento para seleção de facetas automaticamente apresentam dois problemas, a necessidade de descobrir o número de grupos e a necessidade de normalizar o viés do critério de seleção em respeito à dimensão (DY; BRODLEY, 2004).

As primeiras pesquisas em redução de dimensionalidade e classificação de imagens hiperespectrais focaram em métodos de projeção linear (MOHAN; SAPIRO; BOSCH, 2007), como a análise de componente principal (“Principal Component Analysis”, PCA), o que a tornou a técnica de extração de facetas mais comumente utilizada, apesar de seu alto custo computacional, requisição de alta capacidade de memória e baixa eficácia com dados de alta dimensionalidade (DEEPA; THILAGAVATHI, 2015). A PCA emprega propriedades estatísticas nas diferentes bandas hiperespectrais para examinar a dependência e correlação entre elas, baseada no princípio matemático conhecido como decomposição de autovalores (“eigenvalue decomposition”) da matriz de covariância das bandas hiperespectrais (ou atributos) a serem analisados (RODAMEL; SHAN, 2002).

Considerada uma boa abordagem para identificar a combinação dos atributos sísmicos mais significativos (RODEN; SMITH; SACREY, 2015), a PCA se baseia apenas na variância e covariância do dado (HONÓRIO et al., 2014), para organizar dados multivariados em componentes linearmente não correlacionados (GUO; LEWIS; MARFURT, 2008). O primeiro componente principal é o que melhor representa a variabilidade do dado e cada componente subsequente (ortogonal a cada precedente) responde pelo máximo da variabilidade restante. A partir de um grupo de atributos sísmicos, gerados a partir do mesmo volume sísmico original, a PCA é capaz de identificar os atributos que produzem a maior variabilidade no dado, sugerindo que essas combinações aumentam a capacidade de identificação de características geológicas específicas.

Uma limitação da PCA é assumir uma distribuição gaussiana dos dados de entrada (LUBO-ROBLES; MARFURT, 2019). WALDEN (1985) aponta que dados sísmicos de reflexão possuem distribuição não-gaussiana de curtose positiva, enquanto alguns atributos sísmicos podem apresentar distribuição de Poisson. Portanto, maximizar a “não-gaussianidade” captura informação valiosa que pode ser usada para recuperar fontes geológicas independentes ocultas. Esse conceito de “independência” tem a capacidade de fornecer um jeito de capturar informações mais interessantes dos dados multivariados, na forma da análise de componentes independentes, “Independent Component Analysis” ICA (HONÓRIO et al., 2014).

Baseada em estatística de alta ordem, a ICA separa o sinal multivariado em subcomponentes independentes, mas não necessariamente ortogonais, encontrando uma representação linear de dados não-gaussianos (HYVÄRINEN; OJA, 2000). Os componentes independentes não são apenas não-ortogonais, sua ordem também não é definida, isto é, ao contrário dos componentes principais, classificados em ordem de redução de variabilidade, os independentes não podem ser classificados em alguma ordem específica.

Apesar da maior disponibilidade de trabalhos com técnicas de redução de dimensionalidade em imagens hiperespectrais, existem muitos exemplos bem-sucedidos de sua aplicação nas geociências. COLÉOU; POUPON; AZBEL (2003) revisam e comparam métodos de classificação não-supervisionados usados na análise de fácies sísmicas e destacam seu sucesso em trapas estratigráficas, onde apenas a amplitude é incapaz de caracterizar as propriedades do reservatório.

SINGH (2007) discute a aplicação de PCA em atributos derivados de AVO para discriminação de litofácies e detecção de fluidos, neste estudo, o segundo componente principal se mostrou um atributo discriminador mais robusto em comparação a atributos clássicos como a impedância acústica e impedância cisalhante. GUO; LEWIS; MARFURT (2008) aplicaram PCA em 86 componentes espectrais de um dado sísmico, variando entre 5 e 90 Hz, e observaram que os três primeiros componentes respondiam por 83% da variância espectral observada ao longo do horizonte de interesse, reduzindo a dimensionalidade de 86 para 3.

CHOPRA; MARFURT (2014) utilizaram uma mistura de vários atributos, tais como descontinuidades, “sweetness”, texturais e decomposições espectrais, para aumentar a definição e exibir variações internas em canais de areias de um campo na província de Alberta, Canadá. ROY et al. (2014) aplicaram mapeamento topográfico generativo (Generative Topographic Mapping, GTM) na classificação multiatributos de fácies sísmicas

de um campo de petróleo em reservatório conglomerado carbonático na Bacia de Veracruz, no sul do México que, associado a cálculos de probabilidade da ocorrência das fácies, permitiu um aumento na confiança da distinção de fácies que correspondiam a boa ou má produção.

RODEN; SMITH; SACREY, (2015) apresentam três estudos de caso. Foram aplicados PCA seguida de Self-Organizing Maps (SOMs), em uma região offshore do Golfo do México para identificar a relação entre DHIs e efeitos de atenuação para reduzir riscos exploratórios, e definir falhas e padrões de fraturas de forma mais precisa nos folhelhos de Eagle Ford, no Texas. Nos dados de Shallow Yegua, na costa do Texas, a análise multiatributos associada a SOMs foi realizada para entender melhor a distribuição dos reservatórios que produziam mais do que o modelado a partir da anomalia de amplitude.

LUBO-ROBLES; MARFURT, (2019) aplicam ICA em dados de decomposição espectral para aumentar resolução e separar características geológicas de ruído Além disso, foram capazes de analisar geologicamente as diferentes fácies sísmicas resultantes e relacioná-las a fácies propensas a terra ou lama.

1.3.2 Técnicas de Agrupamento

Em aprendizado não-supervisionado, as técnicas de agrupamento são consideradas as tarefas mais simples, típicas e poderosas (XU; WUNSCH II, 2005; AHMAD, 2020; HANCER; XUE; ZHANG, 2020). Seu objetivo é dividir dados em um determinado número de grupos de forma a minimizar distâncias internas e maximizar as distâncias entre os grupos, ou seja, de modo que as amostras de um mesmo agrupamento sejam o mais similares possível, enquanto amostras de diferentes agrupamentos sejam o mais diferente possível (AMR, 2020).

As técnicas são usadas em uma grande variedade de aplicações como segmentação de clientes, análise de dados, detecção de anomalias, motores de busca, segmentação de imagens, entre outros (GERON, 2022). Essas técnicas buscam conjuntos considerados “naturais” nos dados (BARNES; LAUGHLIN, 2002; AHMAD, 2020). XU; WUNSCH II (2005) apontam que não há uma definição universalmente aceita para os agrupamentos e a maioria dos autores descrevem um considerando sua homogeneidade interna e separação

externa, isto é, padrões de um mesmo grupo devem ser similares uns aos outros e dissimilares aos de outros grupos.

Quando aplicadas a dados sísmicos, as técnicas não-supervisionadas têm como objetivo dividir o dado em fácies sísmicas sem o uso de informação prévia para guiar a classificação do traço sísmico, dessa forma, a divisão em agrupamentos fornece uma estrutura de agrupamento natural (XU; WUNSCH II, 2005; MARROQUÍN; BRAULT; HART, 2009). A falta de dados de poços, nos primeiros estágios de exploração, torna a classificação automática de fácies sísmicas dependente das técnicas não-supervisionadas combinadas com a análise multiatributos (LIU et al., 2019).

Não há uma definição universal para o que um agrupamento é e diferentes algoritmos apresentam diferentes tipos (XU; WUNSCH II, 2005; GERON, 2022). Alguns algoritmos buscam instâncias (amostras) centradas em torno de um ponto particular, chamado centroide, outros procuram regiões contínuas de instâncias densamente compactadas e seus grupos podem ter qualquer forma.

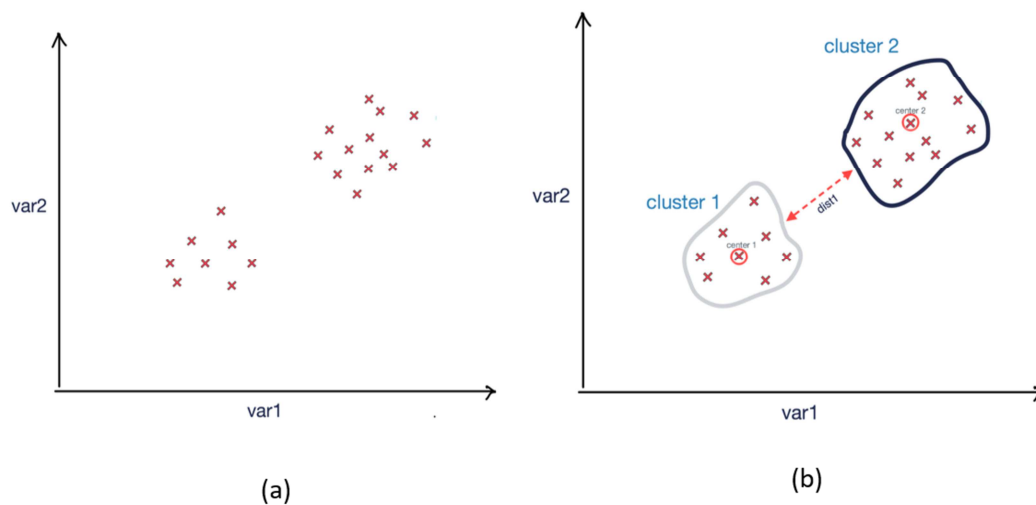
Como já foi apontado, as técnicas procuram minimizar distâncias intragrupos, uma das formas de se medir essas distâncias é calcular as distâncias entre cada ponto do agrupamento e do seu centroide. Existem diversas formas de se calcular essa distância, por exemplo, Euclidiana, Manhattan, Cossenoide (AHMAD, 2020). A distância euclidiana é a distância entre as amostras e sua média, em outras palavras, o desvio padrão, que também pode ser empregada para medir a dissimilaridade entre os centroides dos agrupamentos (Figura 20) (AMR, 2020).

O número de algoritmos relatado ou usado em análise de agrupamentos é muito alto e caracterizar todos vai além do escopo desse trabalho. Trabalhos mais antigos (XU; WUNSCH II, 2005; MARROQUÍN; BRAULT; HART, 2009) dividem os algoritmos em quatro categorias de acordo com o critério adotado para definir os clusters: particionais, probabilísticos, hierárquicos e competitivos. Entretanto, publicações mais recentes reduzem as categorias para apenas duas: particionais ou hierárquicos (DE AMORIM, 2016; AHMAD, 2020; AMR, 2020)

Entre as técnicas de agrupamentos mais populares de classificação está também a SOM (Mapas Auto-Organizáveis de Kohonen), descrita em KOHONEN (1982) como uma técnica de projeção múltipla (ZHAO et al., 2015), enquanto COLÉOU; POUPON; AZBEL (2003) a descrevem como um tipo de rede neural. A técnica SOM utiliza aprendizado competitivo no lugar de abordagens de correção de erro, como no caso do K-Means, para preservar as propriedades de espaço de entrada. Dessa forma, é muito útil na tarefa de

visualização, especialmente porque agrupamentos similares ficam próximos uns aos outros no espaço latente (NAPOLI et al., 2020). BARNES (2007) aponta que a técnica descobre os mesmos padrões observados em agrupamentos por K-Means sendo diferenciados pela forma com que atribuem números de classe, o autor afirma que, para reconhecimento de padrões, é de pouca importância qual método é utilizado.

Figura 20 - Ilustração de agrupamento



Legenda: (a) Nuvens de pontos antes do agrupamento. (b) Grupos após algoritmo K-means e seus centroides.

Fonte: AHMAD, 2020.

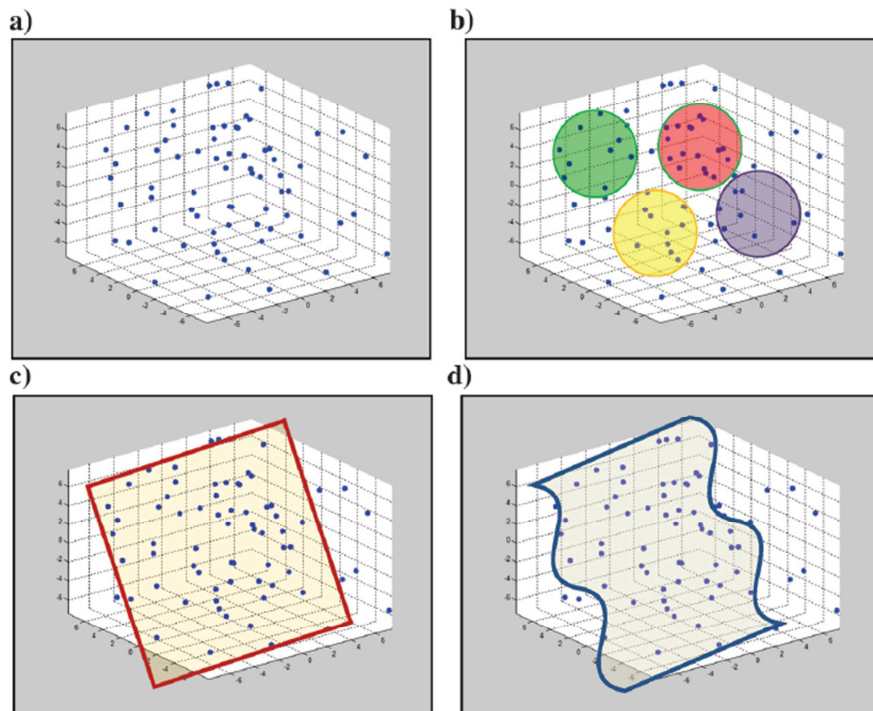
Nas técnicas hierárquicas, também denominadas aglomerativas, inicialmente, cada amostra pertence ao próprio agrupamento, ou seja, existem tantos agrupamentos quanto amostras no dado (AHMAD, 2020; AMR, 2020). A partir daí, a cada iteração, as duas amostras mais próximas são agregadas em um agrupamento. A cada iteração o número de agrupamentos diminui, dessa forma, o número de iterações é proporcional ao número de amostras, o que torna esse tipo de algoritmo computacionalmente custoso, algoritmos particionais tendem a convergir em menos tempo, dessa forma, apesar de conceitualmente simples, implementar esse algoritmo em dados com alto número de amostras é desafiador (BARNES; LAUGHLIN, 2002; DE AMORIM, 2016).

No contexto da interpretação sísmica, os SOMs são ferramentas valiosas para gerar mapas de fácies sísmicas, preservando as conexões topológicas entre os aglomerados (RODEN; SMITH; SACREY, 2015). SOMs fornecem um meio para converter relações estatísticas entre dados multidimensionais em relações geométricas simples, permitindo a

visualização de diferentes características de dados e estruturação de grupos (TANER et al., 2001). O algoritmo SOM organiza os atributos sísmicos de entrada em um mapa 2D com vetores protótipos interconectados, que se adaptam à nuvem de dados de entrada (DE MATOS; OSORIO; JOHANN, 2007).

A Figura 21 exibe a distribuição de pontos em um espaço 3D de atributos (a), o algoritmo *K-Means* agrupa os dados em um número de distribuições definido pelo usuário (b), o plano que encaixa melhor os dados (c), definido pelos dois primeiros autovalores da matriz de covariância calculada, a projeção dos dados 3D nesse plano fornecem os primeiros componentes principais do dado e em (d) a variedade geométrica 2D deformada que melhor encaixa os dados. Cada ponto dessa variedade é mapeado em um espaço latente retangular 2D, o qual define os grupos finais do algoritmo SOM (ZHAO et al., 2015).

Figura 21 - Comparação dos algoritmos em um espaço tridimensional de atributos



Fonte: ZHAO et al. 2015.

Apesar da extensa literatura sobre SOMs, ainda há necessidade de orientações sobre sua aplicação para agrupamento e visualização, bem como entender a relação entre esses propósitos (FLEXER, 2001). ZHAO et al. (2015) consideram limitações do algoritmo a subjetividade na escolha da função de vizinhança para cada iteração com diferentes escolhas

resultando em diferentes soluções e a ausência de medida quantitativa da confiança na classificação de fácies.

No entanto, os SOMs foram aplicados com sucesso em vários estudos de caso. Por exemplo, eles têm sido usados para caracterizar sistemas turbidíticos, analisar campos offshore no Golfo do México, estudar formações de xisto e caracterizar sistemas fluviais (ZHAO et al., 2016; RODEN; SMITH; SACREY, 2015; BABIKIR et al., 2020). Nesses estudos de caso, os SOMs se mostraram eficazes na análise de atributos sísmicos, na identificação de padrões de fácies e na melhoria dos esforços de caracterização de reservatórios.

A técnica de Modelos de Mistura Gaussiana (Gaussian Mixture Models, GMMs, desenvolvidos por DUDA; HART, 1974) também merece destaque entre as técnicas de agrupamento, apresenta alta similaridade ao K-Means e implementação similar, com exemplos de aplicação em agrupamento de amostras de poço (SINGH; OJHA, 2022) e detecção de fácies sísmicas (WALLET; HARDISTY, 2019; ZHU et al., 2022).

A técnica GMM é um poderoso modelo probabilístico usado para analisar dados que assume que as instâncias foram geradas a partir de uma mistura de distribuições gaussianas com parâmetros desconhecidos. Cada distribuição gaussiana representa um cluster e as instâncias geradas a partir de um único gaussiano formam clusters que normalmente se assemelham a elipsoides. O GMM utiliza o algoritmo Expectation-Maximization (EM), semelhante ao algoritmo K-Means, para estimar os parâmetros do cluster iterativamente (BISHOP, 2006; GERON, 2022).

Existem, no entanto, desafios ao usar o GMM, especialmente ao lidar com dados de alta dimensionalidade, um grande número de clusters ou um pequeno número de instâncias (Bishop, 2006; Geron, 2022). Nesses casos, a convergência para a solução ótima pode ser difícil de alcançar. Para resolver isso, pode ser necessário reduzir a complexidade da tarefa limitando o número de parâmetros que o algoritmo precisa aprender. Uma abordagem é impor restrições às matrizes de covariância, como assumir formas esféricas ou diagonais ou amarrar todos os clusters para ter a mesma forma e orientação (BISHOP, 2006; GERON, 2022).

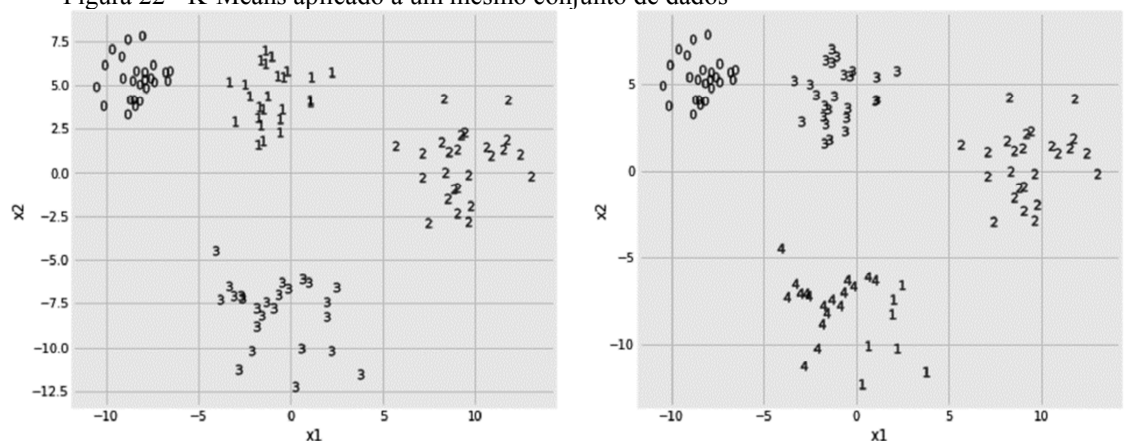
Na análise de fácies sísmicas, WALLET; HARDISTY (2019) aplicaram os modelos de mistura gaussiana com sucesso para modelar distribuições multimodais e identificar agrupamentos. Os autores argumentam que, ao incorporar o GMM na análise, uma avaliação mais quantitativa dos clusters pode ser alcançada devido à sua natureza probabilística. Além disso, a combinação de GMM com mapas auto-organizados (SOM) foi proposta como uma nova abordagem para análise de fácies sísmicas. O SOM atua como um passo inicial benéfico

para reduzir a dimensionalidade dos dados e formar clusters naturais, enquanto o GMM fornece avaliação objetiva do número de clusters e classificação com medidas de incerteza.

Apesar de promissoras, as técnicas não-supervisionadas de agrupamento podem carecer interpretabilidade (LUBO-ROBLES et al., 2022), além de outras limitações. AHMAD, (2020) destaca a simplicidade intencional no projeto do algoritmo K-Means, resultando na necessidade de predeterminação do número de agrupamentos. Como a atribuição inicial do centro do grupo é aleatória, cada vez que o algoritmo é executado, os agrupamentos podem ser sutilmente diferentes. Para AMR (2020), não há alternativa para determinar o valor de K a não ser executá-lo várias vezes, com diferentes números de agrupamentos, e escolher o melhor resultado. Além disso, a escolha do centroide inicial também afeta a sua precisão e uma escolha inicial ruim pode levar o algoritmo a convergir em um mínimo local indesejado. DE AMORIM (2016) aponta o tratamento igualitário do K-Means a cada faceta como uma fraqueza. Para o autor, diferentes facetas podem ter diferentes níveis de relevância e abordagens do K-Means com suporte à ponderação de facetas são necessárias.

A Figura 22 representa o problema do desconhecimento prévio da quantidade de grupos disposta no conjunto de dados a ser analisados. Neste caso, a análise visual indica a presença de quatro grupos, entretanto, se o K-Means é executado com parâmetro $K=5$, a divisão em cinco grupos é forçada, introduzindo erro indesejado provocado pela garantia de convergência do algoritmo, isto é, o algoritmo executa do modo como é parametrizado, sem capacidade de discernimento do erro.

Figura 22 - K-Means aplicado a um mesmo conjunto de dados



(a)

(b)

Legenda: (a) parâmetro $K=4$ (real). (b) $K=5$

Fonte: AMR, 2020.

Quando comparada ao K-Means, BARNES; LAUGHLIN (2002) afirmam que a técnica SOM produz bons resultados, porém muito próximos, sendo a maior diferença o ordenamento natural das classes por similaridade. Por essa característica, os autores consideram a técnica a mais robusta, no entanto, o agrupamento por K-Means produz, consistentemente, resultados levemente mais ajustados. MARROQUÍN; BRAULT; HART, (2009) apontam uma troca, realizada pelo algoritmo de SOM, entre a precisão do agrupamento e a suavidade do mapeamento topológico, como responsável por um possível desempenho prejudicado do agrupamento. FLEXER (2001) mostra que, além da performance comprometida, nos casos em que a técnica é usada tanto para agrupamento quanto para visualização, a SOM é equivalente ao K-Means, entretanto destaca que essas observações se dão a respeito da formulação clássica do algoritmo.

A incerteza em relação ao número de clusters e à precisão do K-Means foi a motivação para a criação de um algoritmo modificado de agrupamento acamado convolucional profundo (“Deep Convolutional Embedded Clustering”, DCEC) com GMM adaptativo proposto em ZHU et al. (2022). Essa abordagem combina a incorporação e o agrupamento de formas de onda sísmicas simultaneamente, melhorando o desempenho do agrupamento e abordando questões relacionadas ao K-Means e à incerteza do agrupamento.

O dado sísmico e seus atributos representam outro desafio às técnicas de agrupamento, uma vez que, na realidade, não existe nenhuma estrutura de grupos bem definida e geologicamente significativa em um espaço de classificação de facetas devido à variabilidade complexa de fácies sísmicas no espaço tridimensional (GAO, 2007). Dessa forma, diferentes fácies geológicas podem ser agrupadas em um mesmo grupo (subclassificadas), ou a mesma fácies geológica pode ser dividida em mais de um grupo (superclassificada). É importante, portanto, avaliar visualmente e validar os resultados de classificação utilizando o conhecimento geológico de modo a evitar erros nos resultados.

O ruído, como componente significativo do dado sísmico, também contribui negativamente nos resultados. Em uma plotagem cruzada N-dimensional, o ruído tende a inflar a nuvem de grupos e conectar subgrupos, removendo o isolamento entre eles. Como as técnicas particionais posicionam os centroides o mais distante entre si possível, elas são mais sensíveis aos ruídos, prevenindo uma ordenação significativa e levando a resultados instáveis que são fortemente impactados pelo número escolhido de grupos (COLÉOU; POUPON; AZBEL, 2003).

Como apontado anteriormente, diversos autores consideram o número predeterminado de agrupamentos a maior limitação do algoritmo K-Means (AHMAD, 2020; AMR, 2020;

BARNES; LAUGHLIN, 2002; GERON, 2022; HANCER; XUE; ZHANG, 2020; XU; WUNSCH II, 2005). Uma garantia do algoritmo é sua convergência, entretanto, esta pode não ser a solução correta, o centroide pode convergir para um ponto ótimo local em detrimento ao ponto global.

No aprendizado não-supervisionado, a falta de rótulos impede o uso das técnicas de avaliação realizadas no aprendizado supervisionado, sendo, portanto, necessário utilizar métricas intrínsecas e comparar os resultados de diferentes execuções do algoritmo escolhido (RASCHKA; MIRJALILI, 2019). Outros métodos foram desenvolvidos para métricas de avaliação interna de agrupamentos, tais como: a compactação, separação, silhueta, R-quadrado, índice de Dunn, entre outros (PANDIT; PIMPALE; DUBEY, 2020). No entanto, para este trabalho, optou-se por usar o método do cotovelo por sua simplicidade e pronta aplicação.

2 MATERIAIS E MÉTODOS

2.1 Interpretação Sísmica

O dado sísmico foi disponibilizado pela ANP, tendo sido adquirido em 2012, com processamento da região do bloco de Libra ocorrido entre 2014 e 2016. O processamento sísmico teve como objetivo gerar volumes migrados em profundidade com as mais modernas técnicas de atenuação de múltiplas, construção de modelos de velocidade utilizando Full Waveform Inversion e volumes anisotrópicos, e migração final utilizando os algoritmos Kirchhoff e Reverse Time Migration (RTM).

Na aquisição, de orientação Norte-Sul, foram utilizados 8 a 10 cabos com 7950 m de comprimento e 636 canais por cabo, com multiplicidade final de 80 canais. O intervalo de amostragem da aquisição foi 2 ms e 10 segundos de registro por tiro, com filtro corta alta em 200 Hz. O intervalo de linhas de receptores foi de 100 metros e o intervalo de pontos de tiro foi de 25 metros.

O fluxo de processamento seguiu um fluxograma padrão, com etapas iniciais de atenuação de ruídos, fantasmas e deconvolução da assinatura da fonte. Para a construção do modelo de velocidades foram utilizadas técnicas de tomografia, anisotropia e FWI. Esta técnica é capaz de gerar volumes de velocidade sísmica com altíssimo detalhe, sendo ideal para a complexidade litológica presente na área de estudo.

Com o modelo de velocidades finalizado, foram realizadas migrações anisotrópicas em profundidade utilizando o algoritmo Kirchhoff, além da técnica de migração reversa no tempo (Reverse Time Migration, RTM), introduzida por BAYSAL; KOSLOFF; SHERWOOD (1983), e tem como característica ser capaz de resolver a migração de estruturas de alto mergulho e com alto contraste de velocidade. A malha de processamento final apresentou celas de 25x25m com azimute das linhas em 358°. A polaridade do dado segue a convenção adotada pela Petrobras, onde um aumento na impedância da rocha constitui um pico positivo, além de ter sido aplicada deconvolução da assinatura da fonte em fase zero, o que significa que o pico da “wavelet” corresponde à superfície de reflexão.

O uso de ferramentas modernas, com o objetivo de imagear adequadamente as reflexões do pré-sal, não pode ser minimizado, uma vez que diversos autores argumentam que

os dados sísmicos de regiões com complexos de soleiras normalmente não são processados de maneira a preservar detalhes desses sistemas (e.g., EIDE et al., 2018; PLANKE et al., 2018; RABELL et al., 2018). É importante frisar aqui que o processamento desse dado sísmico não teve como foco os derrames basálticos e intrusões ígneas presentes na região, entretanto, a aplicação de FWI associada à migração RTM implica em modelos de velocidade não suavizados e maior resolução vertical e lateral mesmo em regiões de grande profundidade, aumentando, dessa forma a capacidade de visualização de reflexões mais sutis.

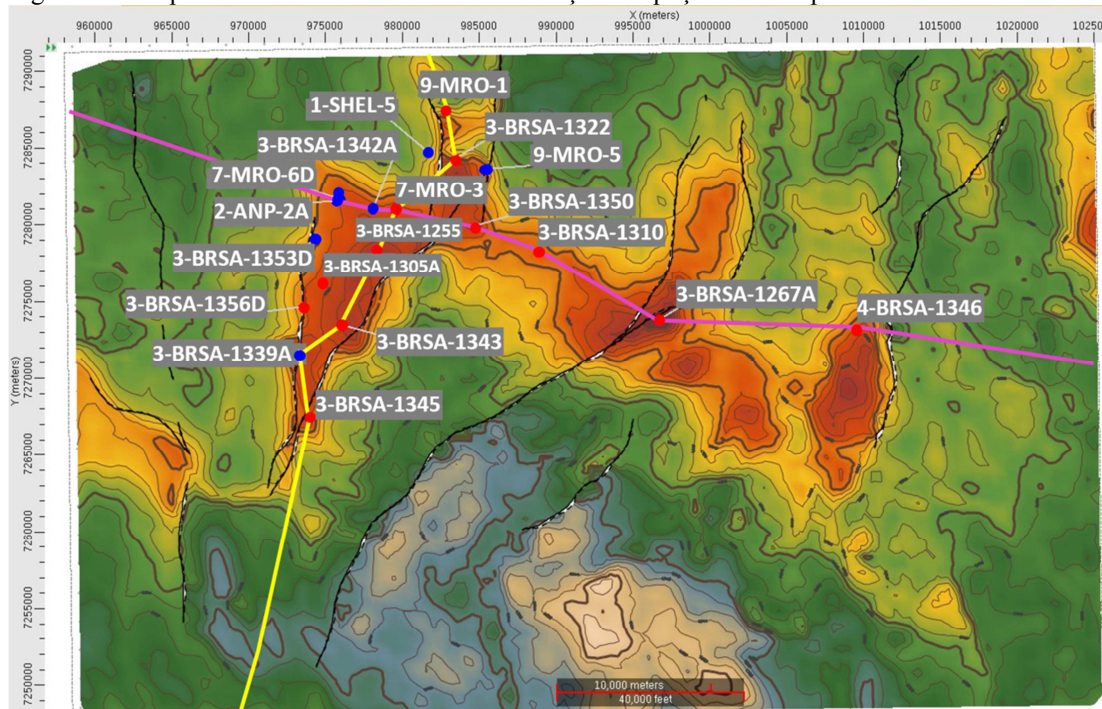
O dado possui aproximadamente 3000 km² e foram escolhidos os principais horizontes com o intuito de descrever a geologia local, sendo eles: Embasamento, Discordância Pré-Alagoas (DPA), Base do Sal, Topo do Sal, Topo do Cretáceo e Fundo-do-Mar. Os horizontes foram interpretados com uma malha relativamente grande, de 1 km x 1 km e interpolados utilizando refinamento por gridagem. A escala é compatível com o objetivo do estudo. Além dos horizontes, foram interpretadas diversas falhas referentes à fase rifte, das quais 14 foram selecionadas para a confecção dos mapas estruturais apresentados a seguir. Não foi realizada interpretação estrutural para a porção pós-sal do dado. A descrição das fácies sísmicas segue a nomenclatura apresentada em MITCHUM JR; VAIL; SANGREE (1977) e adaptada por XU; HAQ (2022).

A Figura 23 mostra o mapa estrutural da base do sal, o caminho percorrido pelas linhas arbitrárias e os poços deste trabalho: em vermelho, aqueles que amostraram rochas ígneas em qualquer porção sedimentar; em azul, os que não amostraram esse tipo de rochas. Foram selecionadas duas linhas arbitrárias, interceptando alguns poços fornecidos (Figura 23), representativas do volume sísmico, as Figura 24 e Figura 25 exibem a linha de orientação aproximada S-N e W-E, respectivamente, junto com o modelo de velocidades intervalares usado para a migração sobrepondo a sísmica, apenas a sísmica, migrada em profundidade com algoritmo RTM, e a sísmica com os seis horizontes e falhas interpretados e poços interceptados.

Conforme mencionado anteriormente, PENNA et al. (2019) utilizaram tomografia multicamadas associada ao FWI, o que permitiu um modelo de velocidades compatível com a geologia esperada, acrescentando ainda a incorporação da estratificação dentro dos evaporitos, o que pode ser observado nas seções superiores das Figura 24 e Figura 25, as quais mostram variações de velocidade na camada de sal estratificado, observadas na porção central, sul e leste das linhas. A tabela de cores, referente às velocidades sísmicas, foi ajustada para exibir de forma mais clara as variações na região de interesse, portanto, os sedimentos pós-sal aparentam ter sido migrados com uma velocidade constante, o

que não foi o caso. Essa porção dos sedimentos possui velocidade intervalar variando entre 1500 e 2600 m/s, e ultrapassando esses valores nas regiões de minibacias encaixadas entre os paredões de sal. A lâmina d'água possui velocidade constante de 1500 m/s.

Figura 23 - Mapa estrutural da base do sal e localização dos poços cedidos pela ANP

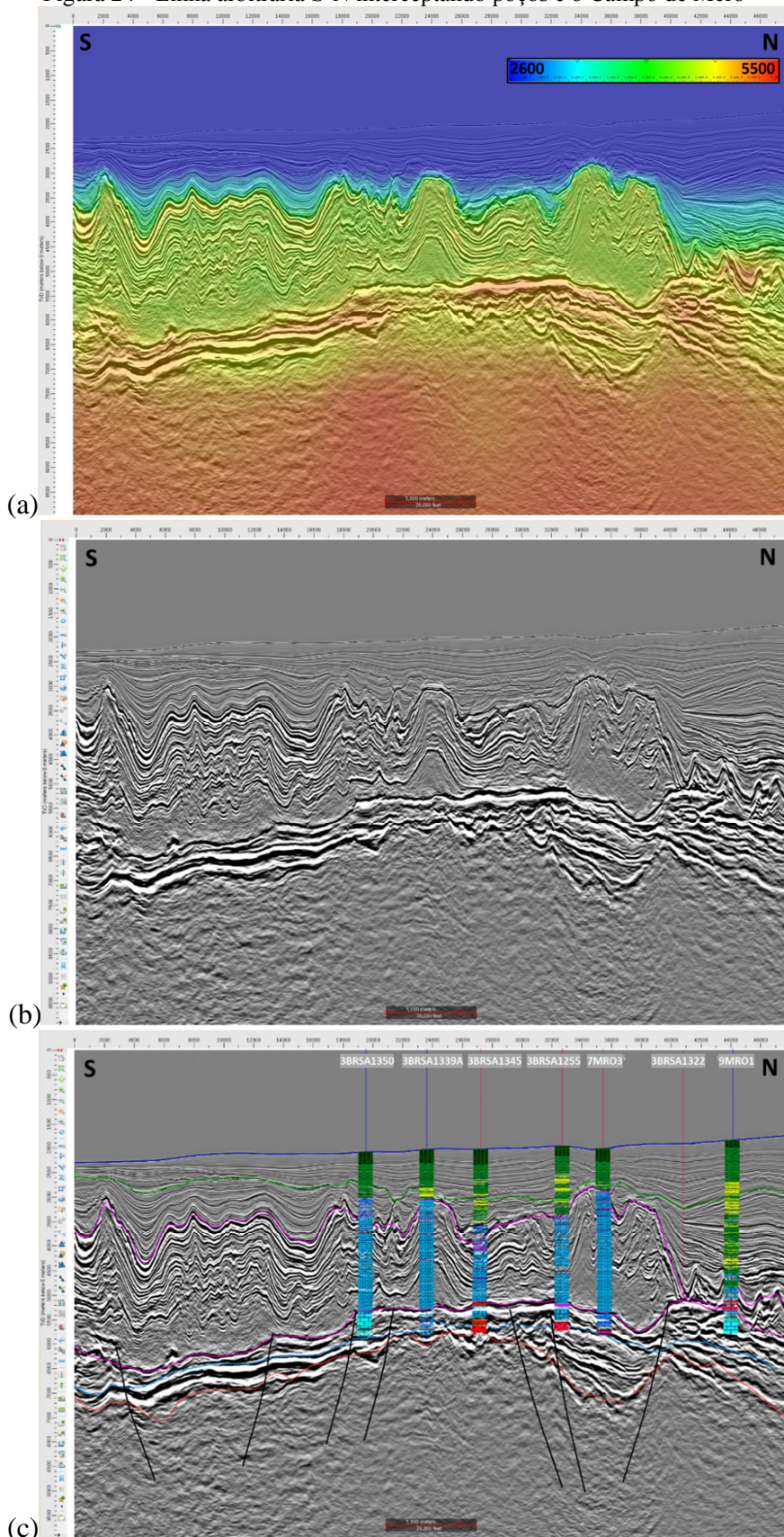


Legenda: Em rosa a linha arbitrária W-E e em amarelo a linha S-N, exibidas nas figuras a seguir.
Fonte: O Autor, 2023

A variação de velocidades na seção pré-sal também merece destaque, exibindo uma velocidade superior, ultrapassando 5000 m/s, na região do Campo de Mero, além de camadas de velocidade similarmente altas a sul e norte. Esse aumento da velocidade ocorre devido à presença de rochas ígneas e o nível de detalhe do modelo é consequência do uso do FWI. Na Figura 25, um complexo intrusivo, localizado entre os paredões de sal, na porção central do dado, possui velocidades ainda superiores às da região de Mero, e intrusões ígneas intrassal de ângulo moderado apresentam velocidade superior à do sal encaixante. A inversão de velocidades, observada nas camadas inferiores, é um advento relativamente recente do processamento sísmico, possibilitada pela construção do modelo de velocidades com as técnicas de tomografia multicamadas e FWI atreladas à migração RTM. Algoritmos de migração clássicos, como o Kirchhoff, não são capazes de lidar adequadamente com inversões de velocidade e seus modelos precisam ser mais suavizados.

Nas seções sísmicas é possível observar a estruturação da região, a porção pré-sal exibe o comportamento típico de rifteamento, grábens e hemi-grabens limitam os altos estruturais, nos quais se concentram os poços perfurados. A seção basal pré-sal exibe em diversos pontos refletores contínuos bem focalizados, cortados por falhas e que representam os derrames basálticos associados ao pós-rifte. O topo dessa sequência exibe uma base do sal anormalmente estruturada, cortada por intrusões ígneas, particularmente na porção central e norte da Figura 24 e na porção leste da Figura 25. A camada de sal exibe fácies sísmicas livre de reflexões em sua base, em direção a seu topo as reflexões referentes ao sal observadas com refletores paralelos seguindo a estruturação do topo do sal. Na região leste da Figura 25 a camada de sal é intensamente intrudida por soleiras próximo a um complexo intrusivo sobreposto à camada e intercalado às rochas do pós-sal. Na região, a camada pós-sal possui espessura maior apenas nas regiões de microbacias encaixadas entre os paredões de sal, e estruturação controlada pela halocinese persistente da região.

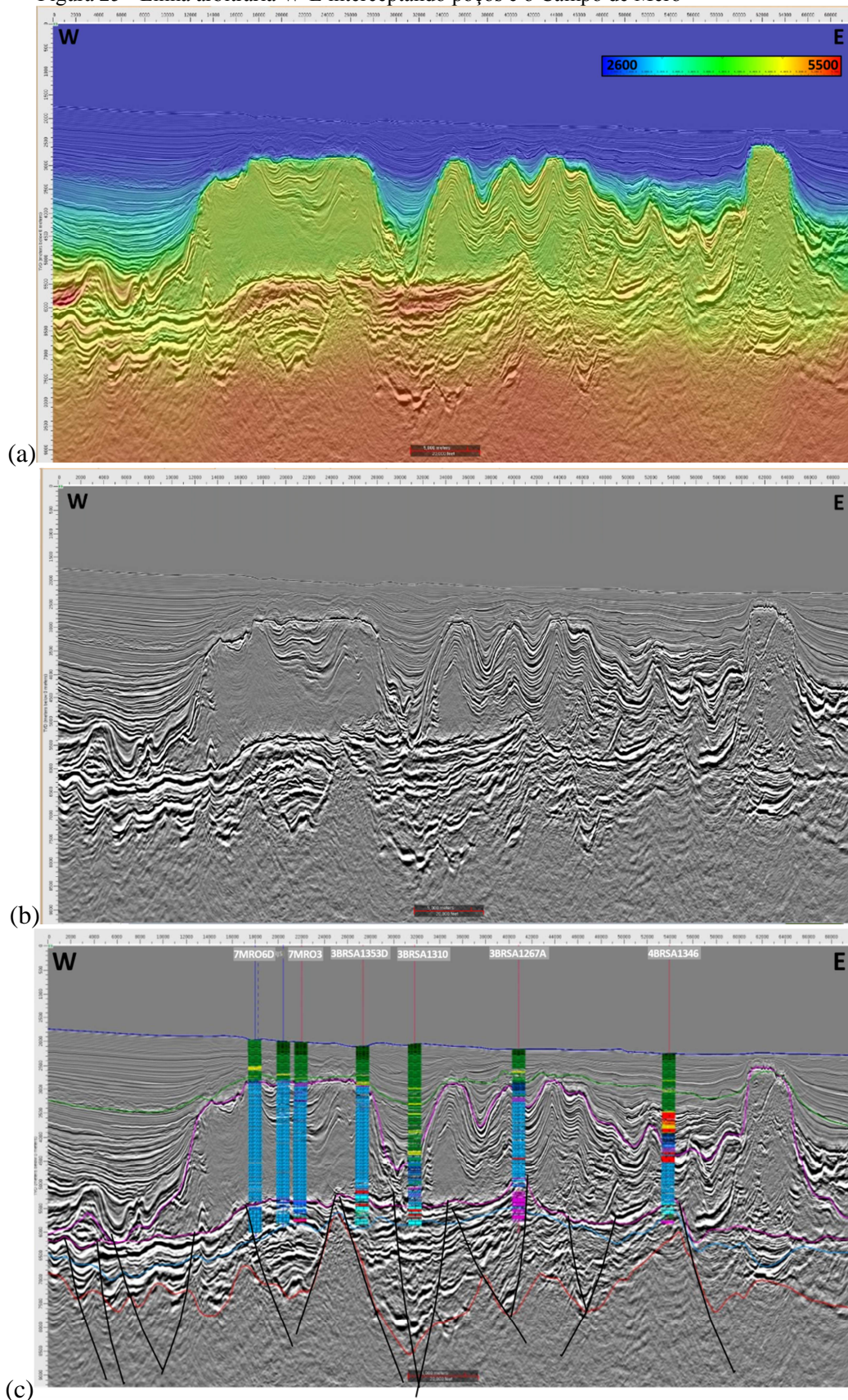
Figura 24 - Linha arbitrária S-N interceptando poços e o Campo de Mero



Legenda: (a) Modelo de velocidade sobreposto à sísmica; (b) sísmica não interpretada; (c) sísmica interpretada e poços.

Fonte: O Autor, 2023.

Figura 25 - Linha arbitrária W-E interceptando poços e o Campo de Mero



Legenda: (a) Modelo de velocidade sobreposto à sísmica; (b) sísmica não interpretada; (c) sísmica interpretada e poços.

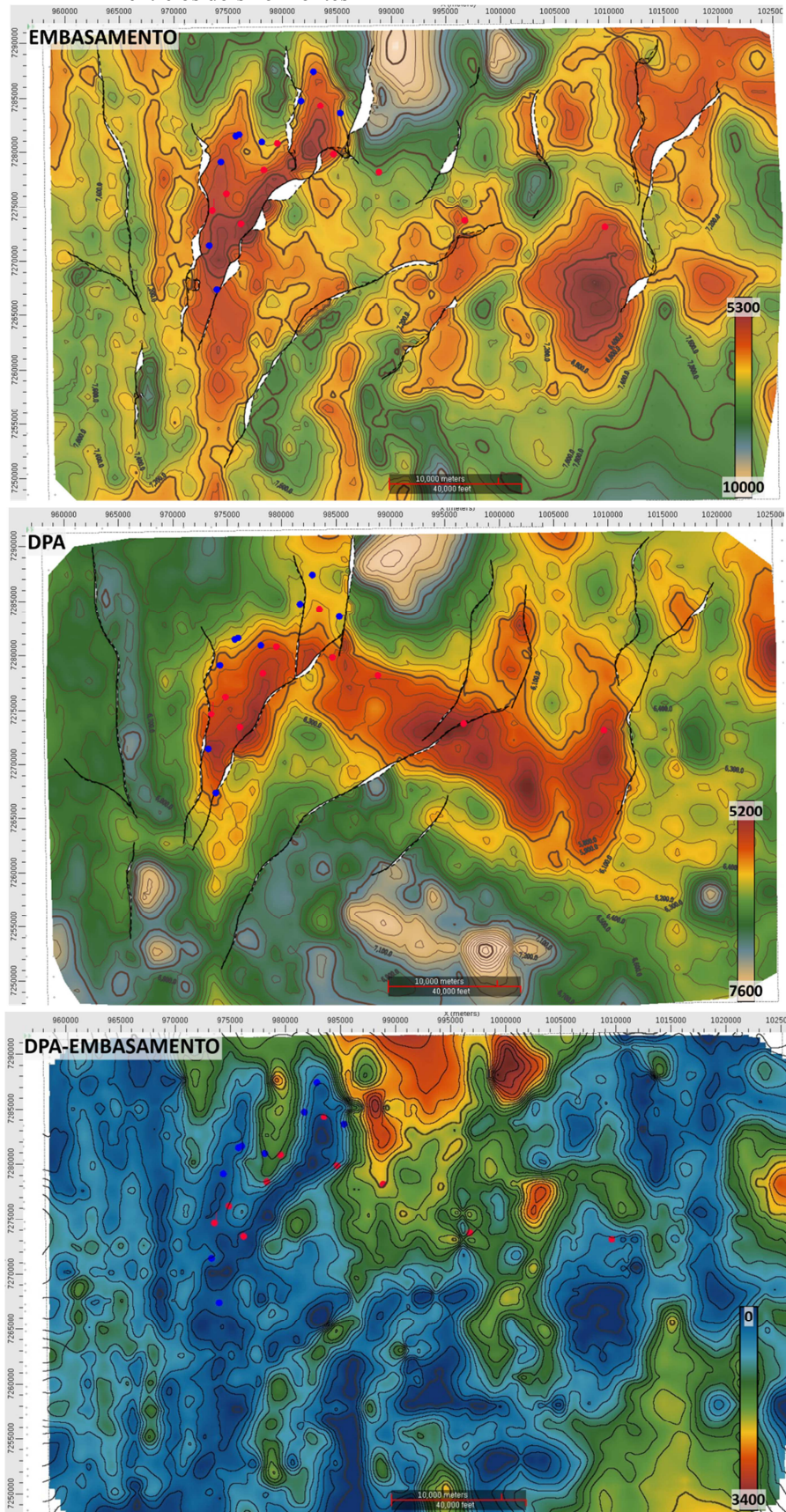
Fonte: O Autor, 2023.

O embasamento (Figura 26) é caracterizado por dois altos principais, limitados por falhas predominantemente antitéticas e orientação NNE e mergulho para sudeste. A noroeste ocorre o alto concomitante ao campo de Mero, onde se concentra a maior parte dos poços já perfurados na região e, por consequência, disponibilizados para esse trabalho pela ANP. A leste, ocorre um segundo alto, entretanto, um intenso vulcanismo, intercalado aos sedimentos depositados durante o Cretáceo, reduz drasticamente a iluminação dos refletores da porção pré-sal.

A Discordância Pré-Alagoas (Figura 26) representa o topo da Sequência K38, descrita por MOREIRA et al. (2007) que corresponde à Formação Itapema do Grupo Guaratiba, seus sedimentos foram depositados do Neobarremiano ao Eoaptiano. O mapa de isópacas entre a superfície e o embasamento (Figura 26) mostra espessuras variando entre 100 e 600 metros na porção oeste do dado, com depocentros ao norte e sudeste. A espessura dessa camada supera 2000m no depocentro a norte.

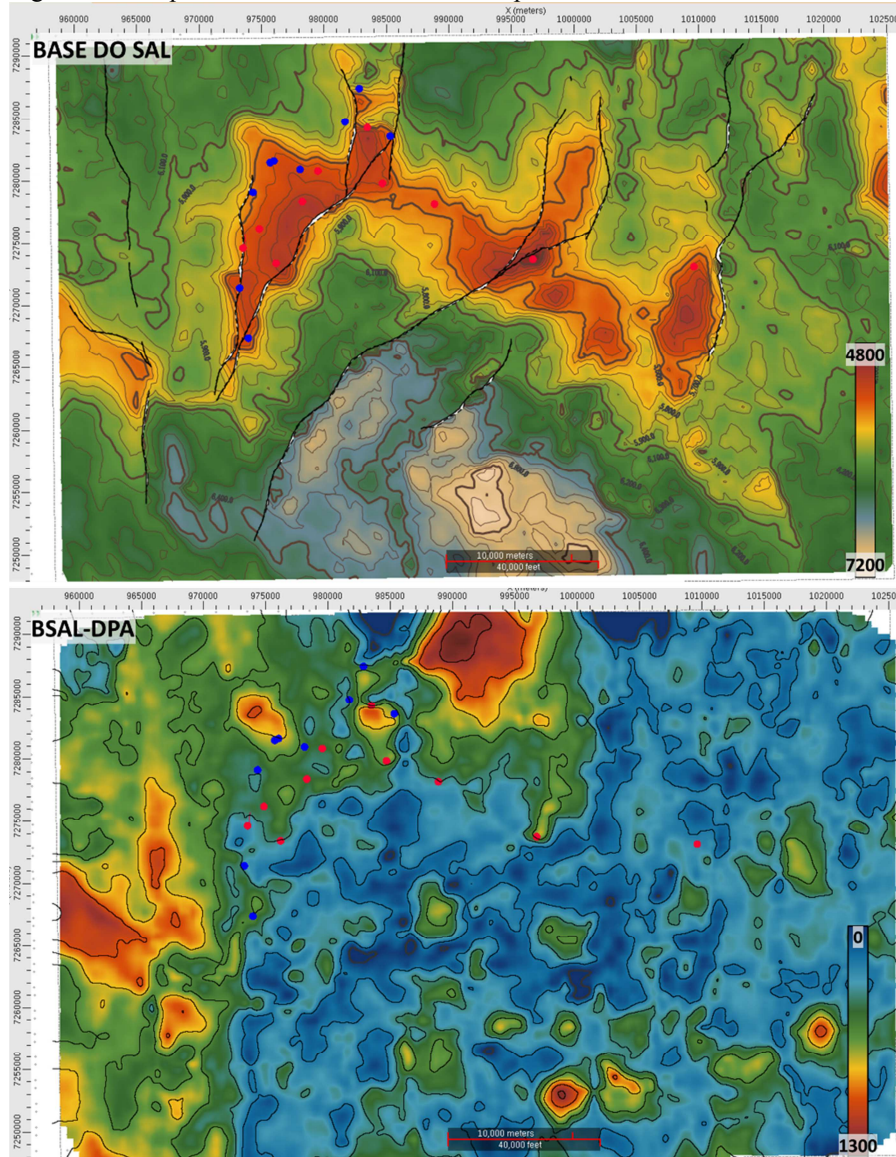
O horizonte sísmico do embasamento foi interpretado abaixo do refletor mais basal de alta amplitude na porção oeste do dado e, em direção a leste, onde já não ocorre uma reflexão contínua e bem iluminada, a interpretação corresponde ao início da fácies sísmica opaca de refletores não coerentes, que representam o embasamento acústico da bacia. Dessa forma, a confiabilidade da interpretação do embasamento na porção sudeste do dado é bastante inferior à da área restante. O impacto da perda de sinal na porção sudeste do dado é destacado no mapa de amplitudes RMS (Figura 30), calculado entre o horizonte referente à DPA e o horizonte do embasamento. Na porção central norte, ocorre um depocentro, onde em sua porção mais profunda ainda ocorrem sedimentos e o topo do embasamento não é imageado pelo dado, o qual termina em 10 quilômetros de profundidade.

Figura 26 - Mapas estruturais dos horizontes embasamento e DPA e mapa de isópacas entre os dois horizontes



Fonte: O Autor, 2023.

Figura 27 - Mapa estrutural da base do sal e isópacas entre base do sal e DPA



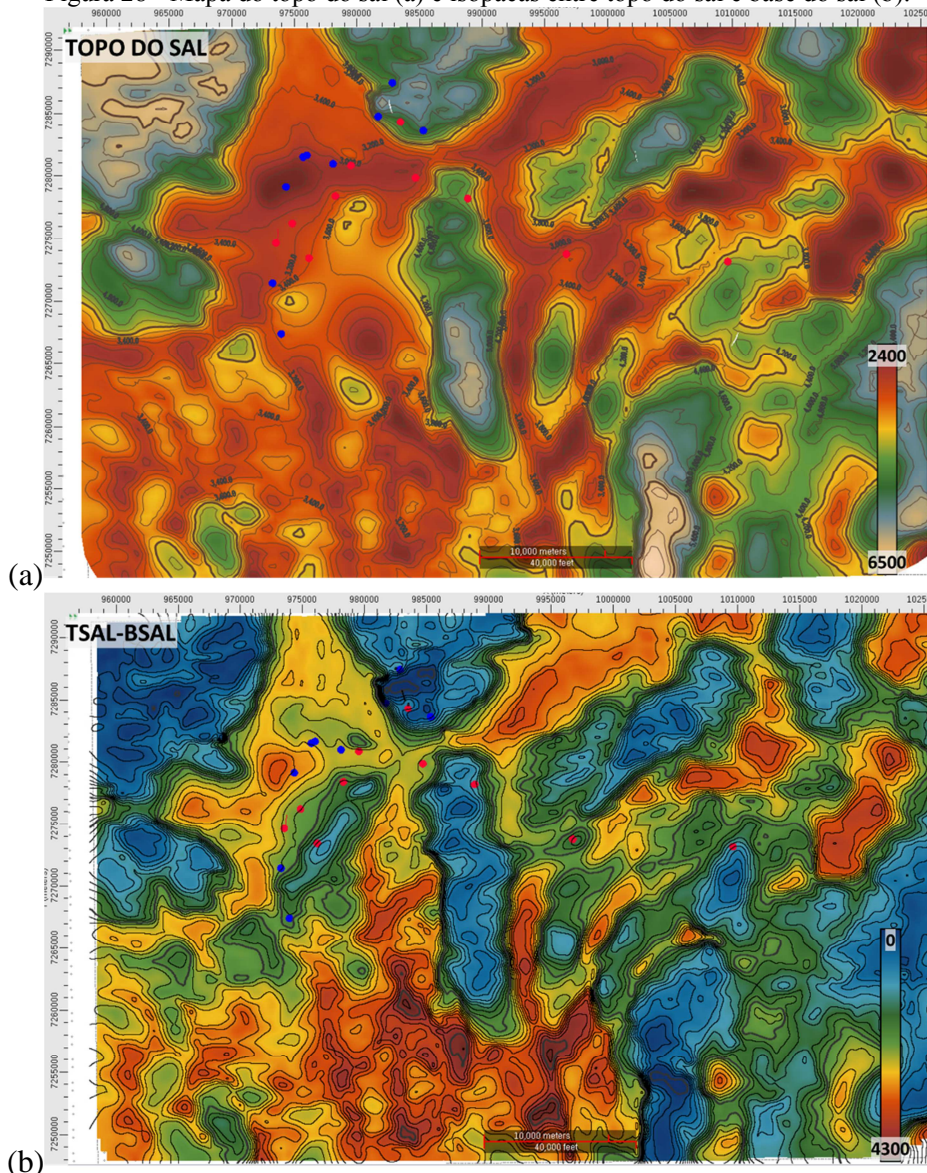
Fonte: O Autor, 2023.

O horizonte foi interpretado com base nos marcadores dos poços fornecidos e extrapolado a partir deles. Em boa parte do dado, a DPA corresponde a um refletor forte e contínuo subparalelo, sotoposto a fácies sísmicas granulares a caóticas representantes dos carbonatos da Formação Barra Velha. O caráter contínuo desse horizonte, principalmente junto aos baixos estruturais do embasamento se deve à presença de rochas ígneas, descritas por RANCAN et al. (2018) e datadas em $125,5 \pm 0,7$ Ma, pelo método $^{40}\text{Ar}/^{39}\text{Ar}$, intercaladas aos sedimentos da Formação Itapema. São rochas toleíticas concentradas nas depressões entre os altos estruturais do embasamento. Apresentam sismofácies lenticulares a plano paralelas, estas interpretadas pelos autores como lava-deltas, responsáveis por preencher as depressões rifte. Nas seções sísmicas é possível observar a presença dessas fácies sísmicas, caracterizadas

por reflexões de alta amplitude e continuidade lateral dos refletores, na porção central do dado, entre os poços 3-BRSA-1353D-RJS e 3-BRSA-1267A-RJS e no baixo estrutural do embasamento entre os poços 7-MRO-6DB e 7-MRO-3-RJS.

O mapa estrutural de base do sal corresponde ao topo da Formação Barra Velha, onde ocorrem os reservatórios carbonáticos de petróleo. Essa formação é caracterizada pela ocorrência de calcários microbiais intercalados a folhelhos, sendo calcários estromatolíticos e laminitos microbiais, localmente dolomitizados (MOREIRA et al., 2007). Essa camada, que corresponde às Sequências K44 e K46-48, é caracterizada por sismofácies caóticas a granulares, refletores de baixa amplitude e pouca continuidade lateral, possui espessura média de 300 metros, com espessamento em direção a oeste, onde alcança espessura superior a 1000 metros, assim como também no depocentro norte.

Figura 28 - Mapa do topo do sal (a) e isópacas entre topo do sal e base do sal (b).



Fonte: O Autor, 2023.

É possível notar uma mudança na malha estrutural a partir da análise do mapa estrutural do topo do sal, com predomínio de falhas com orientação ENE, na porção sudoeste do dado, passando a orientação NNE a N na porção norte e nordeste. Essa mudança estrutural pode estar relacionada à Zona de Transferência de Libra (ZTL), de direção NW-SE, descrita por RANCAN et al.(2018) e ZHAO et al. (2019). No topo da Formação Barra Velha ocorrem as intrusões ígneas, de ocorrência estimada em 80 Ma (ZHAO et al., 2019), descritas como diabásio alcalino e gabro. Os autores estimam que as falhas limitadoras dos altos estruturais, tiveram papel importante na migração e acumulação do magma, o que é corroborado pela proximidade com as falhas dos poços que amostraram essas rochas, tais como 3-BRSA-1356D-RJS, 3-BRSA-1345-RJS, 3-BRSA-1305A-RJS, entre outros.

O topo do sal mostra paredões com largura superior a dez quilômetros e orientação preferencial norte sul, na porção sul, passando a ENE, na porção norte do dado. Nas regiões de paredões de sal, onde a camada superior é geralmente composta por sedimentos siliciclásticos, o refletor correspondente ao topo de sal é de alta amplitude, devido ao alto contraste de impedância entre os sedimentos siliciclásticos sobrepostos e os sedimentos evaporíticos. Nas regiões de minibacias, onde ocorrem os carbonatos da Formação Guarujá, a interpretação é um pouco mais desafiadora, uma vez que essas rochas possuem alta impedância acústica e, portanto, são caracterizadas por refletores contínuos de alta amplitude. Em geral, a diferenciação é feita a partir da análise das fácies sísmicas correspondentes à camada de evaporitos.

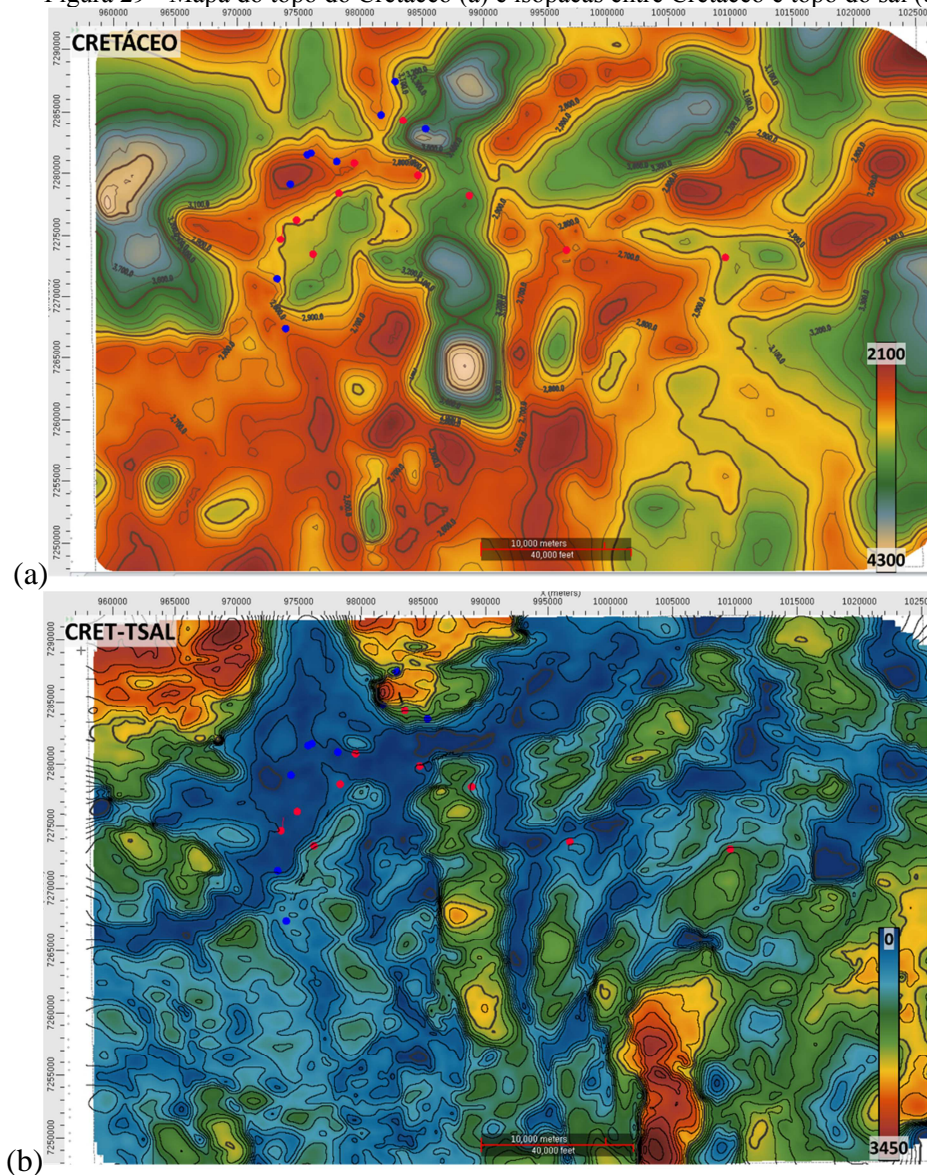
Na seção arbitrária de direção L-W, a camada exhibe sismofácies livre de reflexões no paredão de sal, na região do campo de Mero, enquanto na região central a camada de evaporitos, denominada Sequência K50 por MOREIRA et al. (2007), geralmente composta por halita e anidrita, é caracterizada pela presença de sal estratificado, exibindo reflexões de alta amplitude e extensa continuidade lateral e presença de esparsos refletores contínuos limitados de alta impedância, interpretados como intrusões ígneas, particularmente na região do poço 4-BRSA-1346-RJS. A nordeste, a camada de sal apresenta sismofácies caótica, também marcada por intrusões ígneas causadas pela suíte vulcânica situada na região, relacionada ao evento magmático do Santoniano-Campaniano.

Nessa região, a camada de sal possui em média 2000m de espessura, alcançando 4000 m na porção sul do dado. Janelas de sal, definidas como regiões onde a camada de sal é interior a algumas dezenas de metros ou ausente, podem estar presentes próximas aos poços 3-BRSA-1322-RJS e 9-MRO-1-RJS, a nordeste do Campo de Mero, e numa pequena região a sudeste do dado, como pode ser observado na

Figura 28. As janelas de sal

representam locais de possível migração e perda do petróleo gerado, representando, dessa forma, são locais de selo comprometido.

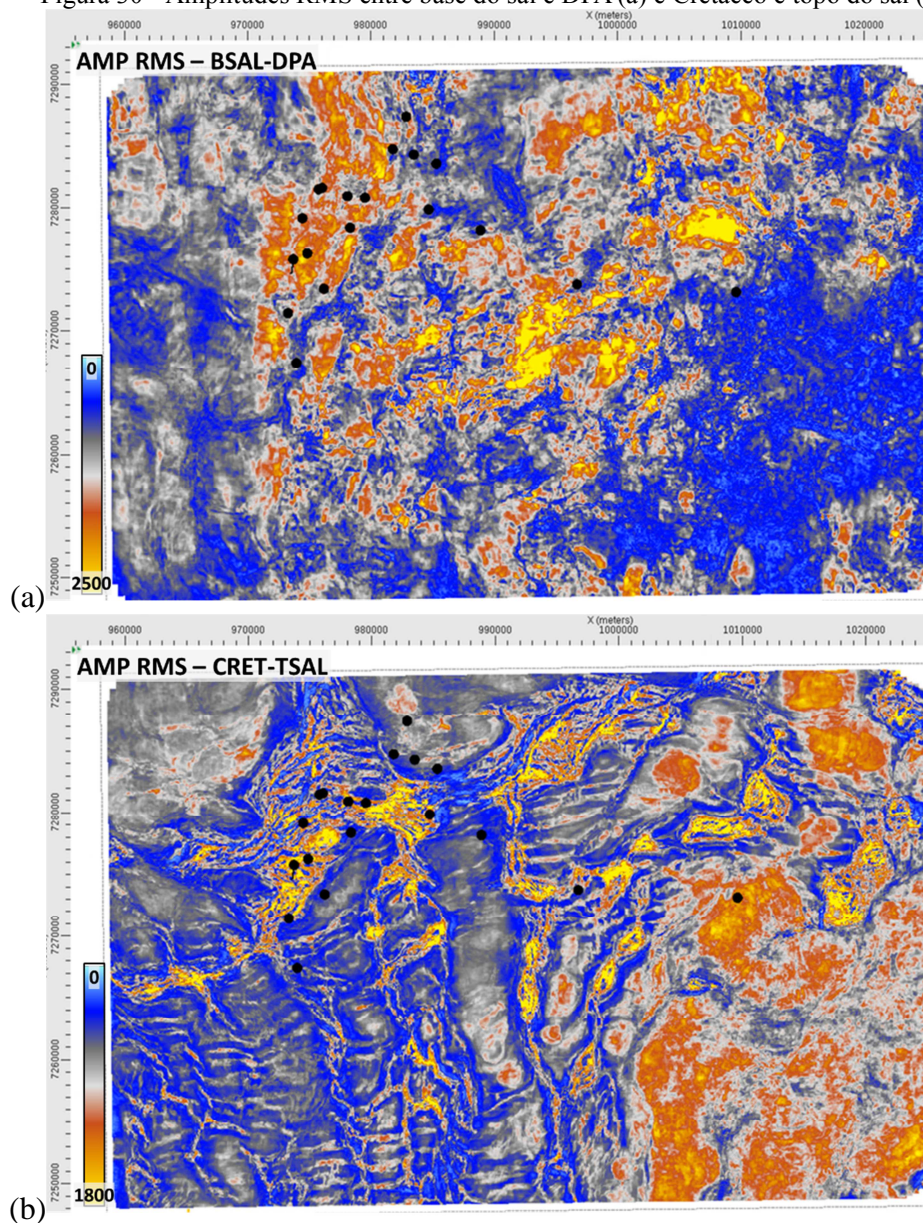
Figura 29 - Mapa do topo do Cretáceo (a) e isópacos entre Cretáceo e topo do sal (b)



Fonte: O Autor, 2023.

Na porção noroeste do dado, um paredão de sal de orientação ENE e extensão superior a 40 km, limita três minibacias sedimentares, localizadas no noroeste, norte e centro do dado (Figura 29) e concentraram a maior parte dos sedimentos acumulados nesse período, com espessura da camada superando 3000 m. Uma quarta minibacia é observada na porção sudeste do dado, limitada pelos diápiros de direção N a NE, neste local se concentrou boa parte do vulcanismo do Santoniano-Campanian. A presença dessas minibacias está associada à halocinese e destacam a influência dessa movimentação no período.

Figura 30 - Amplitudes RMS entre base do sal e DPA (a) e Cretáceo e topo do sal (b)



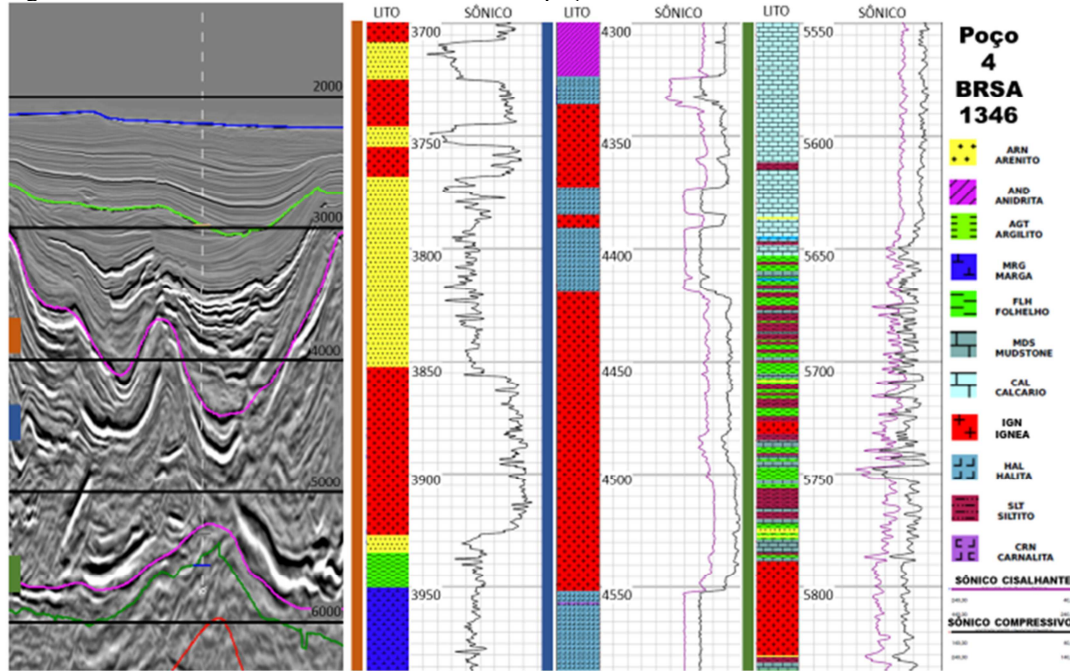
Fonte: O Autor, 2023.

Em geral, os refletores nessa camada apresentam alta continuidade e alta frequência dominante. Na porção basal são correspondentes aos carbonatos da Formação Guarujá, gradando para intercalações de arenitos e folhelhos relacionados à Formação Itanhaém, característicos por serem encaixados em baixos deposicionais gerados e controlados pela tectônica salífera (MOREIRA et al., 2007). O horizonte que corresponde ao topo dessa sequência é a discordância relacionada ao evento Cretáceo-Paleógeno, e foi interpretado a partir dos marcadores dos poços e por sua característica fácies sísmica de refletores convergentes.

A Figura 31 exibe uma linha sísmica nos arredores do poço 4-BRSA-1346, único poço da região a amostrar rochas ígneas nas porções, pré-, intra- e pós-sal. Na seção pós-sal, as

rochas ígneas possuem o caráter típico de intrusões em sedimentos siliciclásticos, descrito por diversos autores (BISCHOFF et al., 2021; EIDE et al., 2018; MAGEE et al., 2015, 2021), apresentando alta amplitude e formato de pires. Em outros locais, são observadas feições sísmicas que indicam extravasamento do magma, tais como cones vulcânicos com reflexões paralelas no entorno, correspondentes a derrames basálticos.

Figura 31 - Trecho de linha sísmica no entorno do poço 4-BRSA-1346.



Legenda: Barras verticais laranja, azul e verde representam a posição dos trechos do poço representados à direita da sísmica

Fonte: O Autor, 2023.

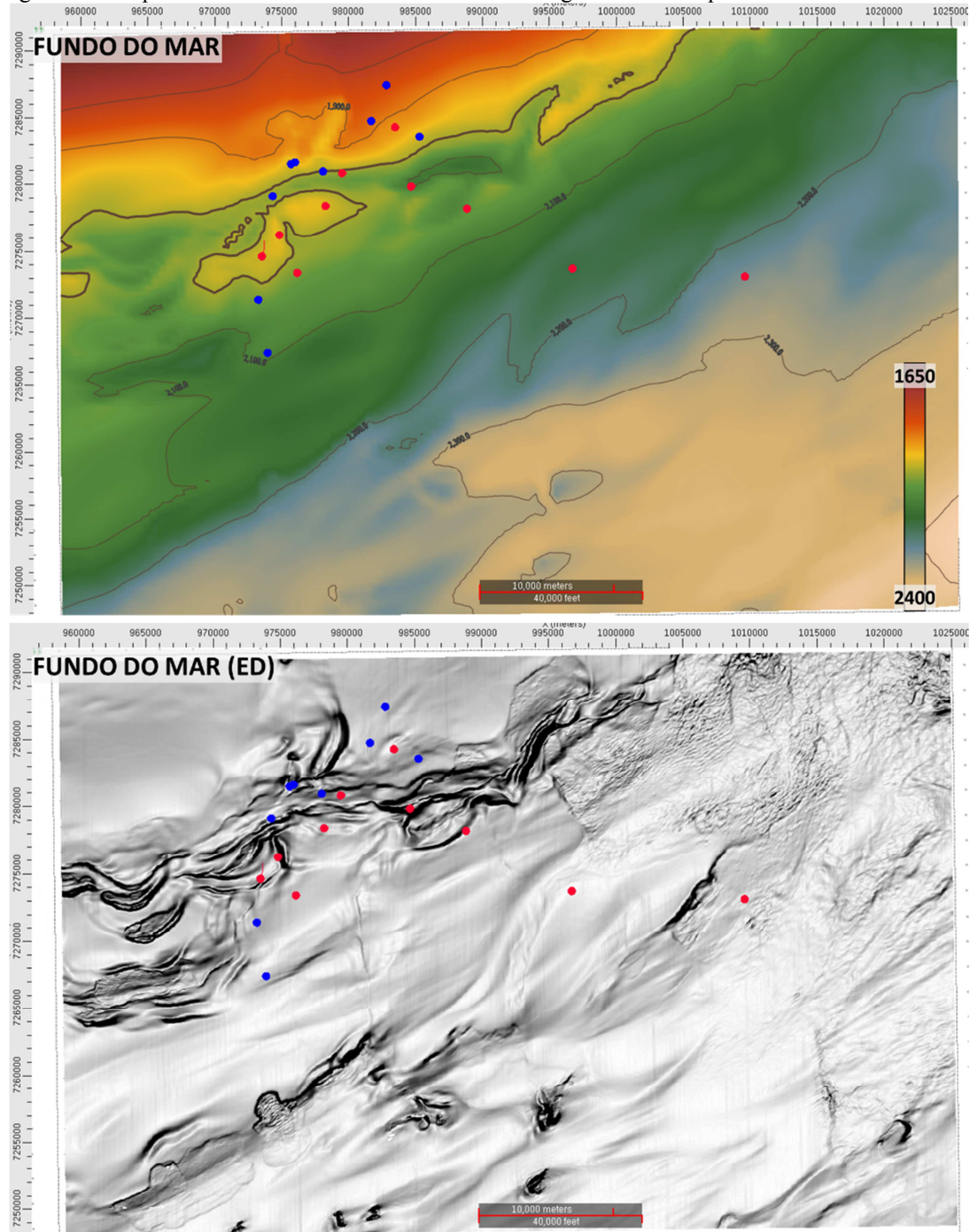
Na seção intrassal superior, as intrusões ígneas são plano paralelas próximas ao topo e à base do sal, com algumas reflexões de alto ângulo na porção intermediária da camada. A seção pré-sal exhibe uma forte perda de sinal provocada pela absorção e dispersão das lâminas lenticulares das intrusões sobrepostas. O formato dessas intrusões age como uma lente convexa que dispersa o sinal sísmico descendente para regiões externas, causando uma perda de iluminação abrupta nas camadas inferiores e consequentes fácies sísmicas granulares a livres de reflexão. Além da perda de iluminação sísmica, o perfil sônico do poço destaca forte intercalação de siltitos e folhelhos, de alta velocidade, provavelmente, devido à compactação sofrida, que resultam em poucos contrastes marcados de impedância. Ou seja, ainda que não houvesse o problema de iluminação na região, as rochas vulcânicas basais poderiam ser de difícil observação na sísmica.

Toda a seção cenozoica nessa região é marcada por refletores de baixa amplitude, de fácies sísmica paralela a subparalela, típica de folhelhos. Os paredões de sal, na porção norte do dado, limitam a sul a maior concentração de deposição de sedimentos para o período, indicando halocinese ainda pervasiva. Essa movimentação parece persistir até a ocorrência da discordância do mioceno (não interpretada).

O horizonte de fundo-do-mar possui o refletor mais claro e contínuo em todo dado sísmico bem processado, à exceção apenas nos lugares onde a lâmina d'água é muito rasa (inferior a 100 m). Dessa forma, há alguns anos, a maioria dos intérpretes recorre à técnica de rastreamento automático para esse refletor. Para o intérprete basta interpretar manualmente poucas linhas para o software de interpretação buscar automaticamente por formas de onda parecidas no entorno e, devido à sua continuidade, em poucos segundos todo o dado sísmico é preenchido, muitas vezes sem a necessidade de interpolações.

Um dos maiores benefícios para esse tipo de interpretação é o retorno de informações que ela é capaz de trazer quando associado à extração de atributos como a amplitude instantânea e Edge Detection. Este atributo destaca mudanças abruptas no horizonte, indicando falhas, bordas de canais, ruídos, entre outros (BARNES, 2016). Para o fundo-do-mar, o Edge Detection é capaz, ainda, de identificar pockmarks, definidas por CHOPRA; MARFURT (2007) zonas de gás, ou seja, indicadores da presença de hidrocarbonetos. Neste caso, essas estruturas não foram identificadas, entretanto, na porção nordeste do dado, uma mudança abrupta na textura da superfície do mar ocorre, passando de marcadamente lisa e homogênea a alta rugosidade. Outras feições de destaque são os lineamentos de direção NE, que destacam a posição dos paredões de sal na subsuperfície, indicando influência da halocinese até o recente.

Figura 32 - Mapa estrutural do fundo do mar e atributo “Edge Detection” aplicado



Fonte: O autor, 2023.

2.2 Análise de Poços

A ANP forneceu dados referentes a 19 poços perfurados na região do 3D de Libra e que já eram de domínio público à época da solicitação, em dezembro de 2021. Os dados incluem perfis petrofísicos e compostos dos poços, à exceção dos poços de desenvolvimento de produção, os quais não possuem perfis compostos.

O poço 1-SHEL-5-RJS foi um poço perfurado em 2001, pela companhia Shell e classificado como seco sem indícios. Nesse período os poços eram perfurados com a intenção de identificar acumulações similares às da Bacia de Campos, ou seja, turbiditos intercalados a folhelhos da Formação Itajaí-Açu. Entretanto, apesar de ter amostrado extensas camadas de areia, a maior com aproximadamente 500 metros, o poço foi classificado como seco sem indício de petróleo, apontando falha do sistema petrolífero.

A presença de reservatório, com camadas de folhelhos sobrepostas, capazes de atuar como selos, associado com a ausência de indícios aponta falha de migração dos hidrocarbonetos, uma vez que, na região, a geração de hidrocarbonetos gerados em grandes volumes já é provada. Como o poço não alcançou a camada pré-sal, ele recebe o nome informal de “pendurado”.

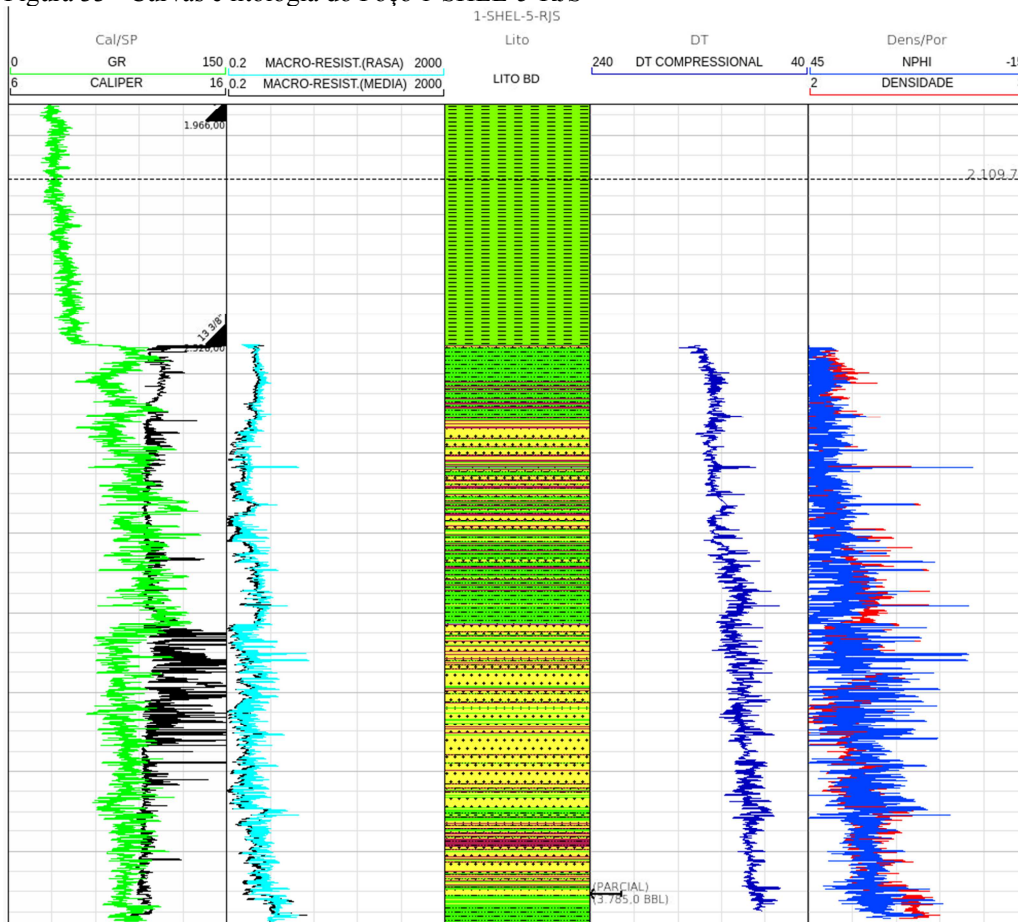
Tabela 1 - Poços fornecidos pela ANP e marcadores litoestratigráficos

Poço - Sigla Anp	Fmar	Cretáceo	Topo Sal	Base Sal	Esp Sal	DPA	Embasamento	Topo IG1	Base IG1	Posição	Topo IG2	Base IG2	Posição	Topo IG3	Base IG3	Posição	Esp Total Ígnea
1-SHEL-5-RJS	1919	2910			0												
2-ANP-2A-RJS	1985	2742	2872	5394	2522	5572											
3-BRSA-1255-RJS	1989	2790	3382	5196	1814	5307	5581~	5580	5736	Pré-Sal							156
3-BRSA-1267A-RJS	2183	2714	3198	4926	1728		5789~	4432	4495	Intra	5180	5476	Pré-Sal	5533	5808	Pré-Sal	634
3-BRSA-1305A-RJS	1974	2996	3113	5329	2216	5706		5343	5430	Pré-Sal							87
3-BRSA-1310-RJS	2070	3374	5090	5472	382			5360	5380	Intra	5547	5600	Pré-Sal	5645	5750	Pré-Sal	178
3-BRSA-1322-RJS	1937	3322	4980	5177	197		5730~	5195	5215	Pré-Sal	5500	5610	Pré-Sal	5630	5798	Pré-Sal	298
3-BRSA-1339A-RJS	2057	2930	3058	5301	2243	5614											
3-BRSA-1342A-RJS	2022	2731		5371	2518	5590											
3-BRSA-1343-RJS	2050	3185	3863	5240	1377	5357	5650	5170	5240	Intra	5535	5605	Pré-Sal	5630	5777		287
3-BRSA-1345-RJS	2111	2756	3345	5385	2040	5830		3425	3427	Intra							2
3-BRSA-1350-RJS	2122		2952	5266	2314	5649		5158	5224	Intra	5431	5482	Pré-sal				117
3-BRSA-1353D-RJS	2033		2487	5497	3010	5802											
3-BRSA-1356D-RJS	1997	2988	3178	5620	2442	5990		5620	5675	Pré-Sal							55
4-BRSA-1346-RJS	2277	3059	4244	5420	1176	5653	5788	3514	3768	Pós-Sal	3852	3927	Pós-Sal	4335	4550	Intra	544
7-MRO-3-RJS	2035	2702	2898	5393	2495	5673	5793	5765	5790	Pré-Sal	5793	5812	Pré-Sal				44
7-MRO-6DA-RJS	1982	2762	2886	5395	2509	5570		5342	5362	Pré-Sal							20
9-MRO-1-RJS	1868	2968	5169	5368	199			5144	5315	Pós-Sal							171
9-MRO-5-RJS	1991	3855	5093	5400	307												

Fonte: O Autor, 2023.

O poço 2-ANP-2A-RJS foi um poço perfurado a partir de uma parceria entre a Petrobras e a ANP e é considerado poço descobridor da acumulação de toda a região de Libra. Apenas recentemente a região foi formalmente dividida entre Campo de Mero e Bloco de Libra, o qual permanece em estágio de desenvolvimento para comercialidade. O poço foi perfurado em 2010 e obteve a classificação de descobridor de campo com óleo. Ambos 1-SHEL-5-RJS e 2-ANP-2A-RJS não perfuraram rochas ígneas e apenas seus marcadores litológicos foram utilizados nesse trabalho, para auxiliar a interpretação sísmica.

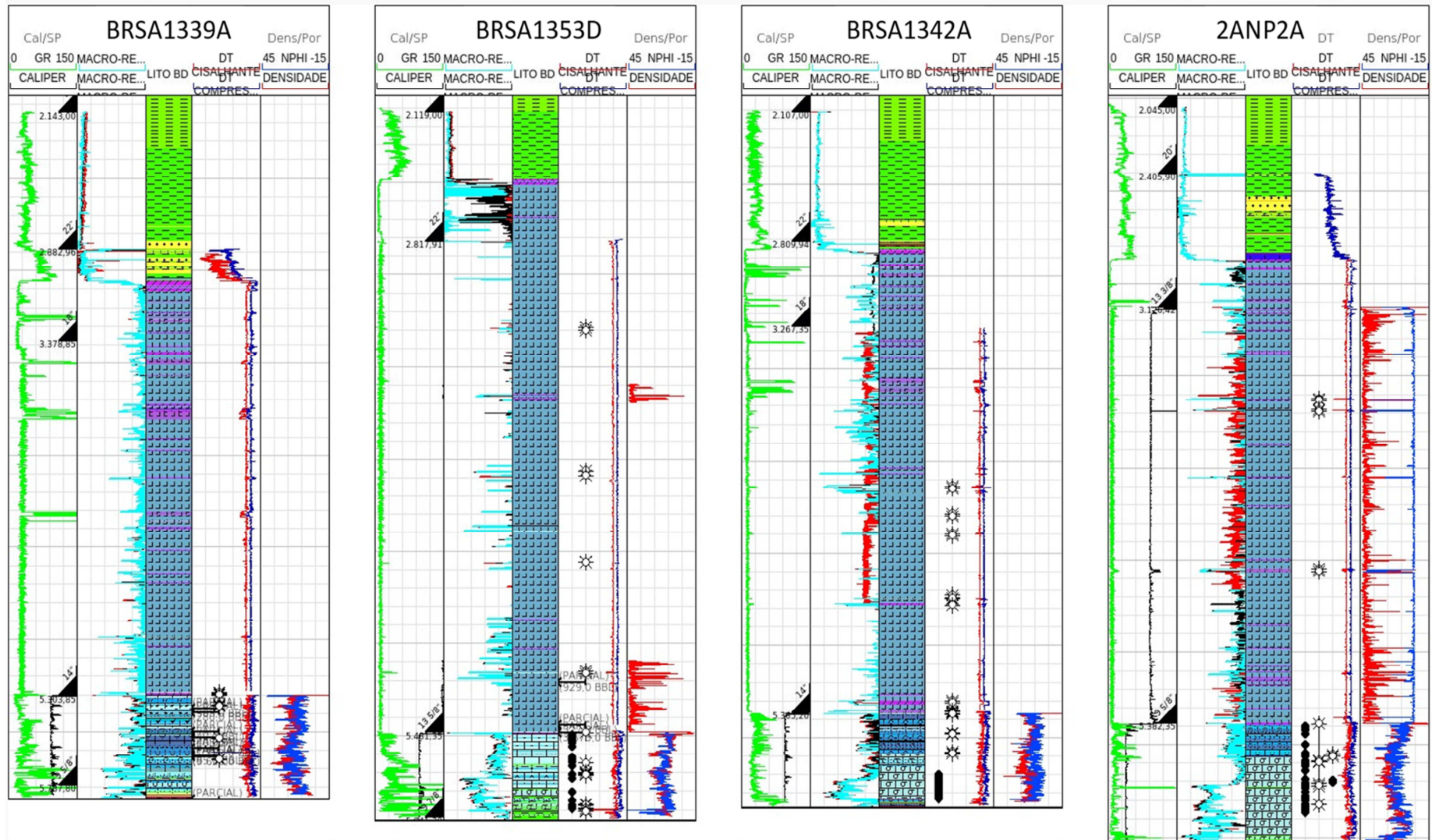
Figura 33 - Curvas e litologia do Poço 1-SHEL-5-RJS



Fonte: O Autor, 2023.

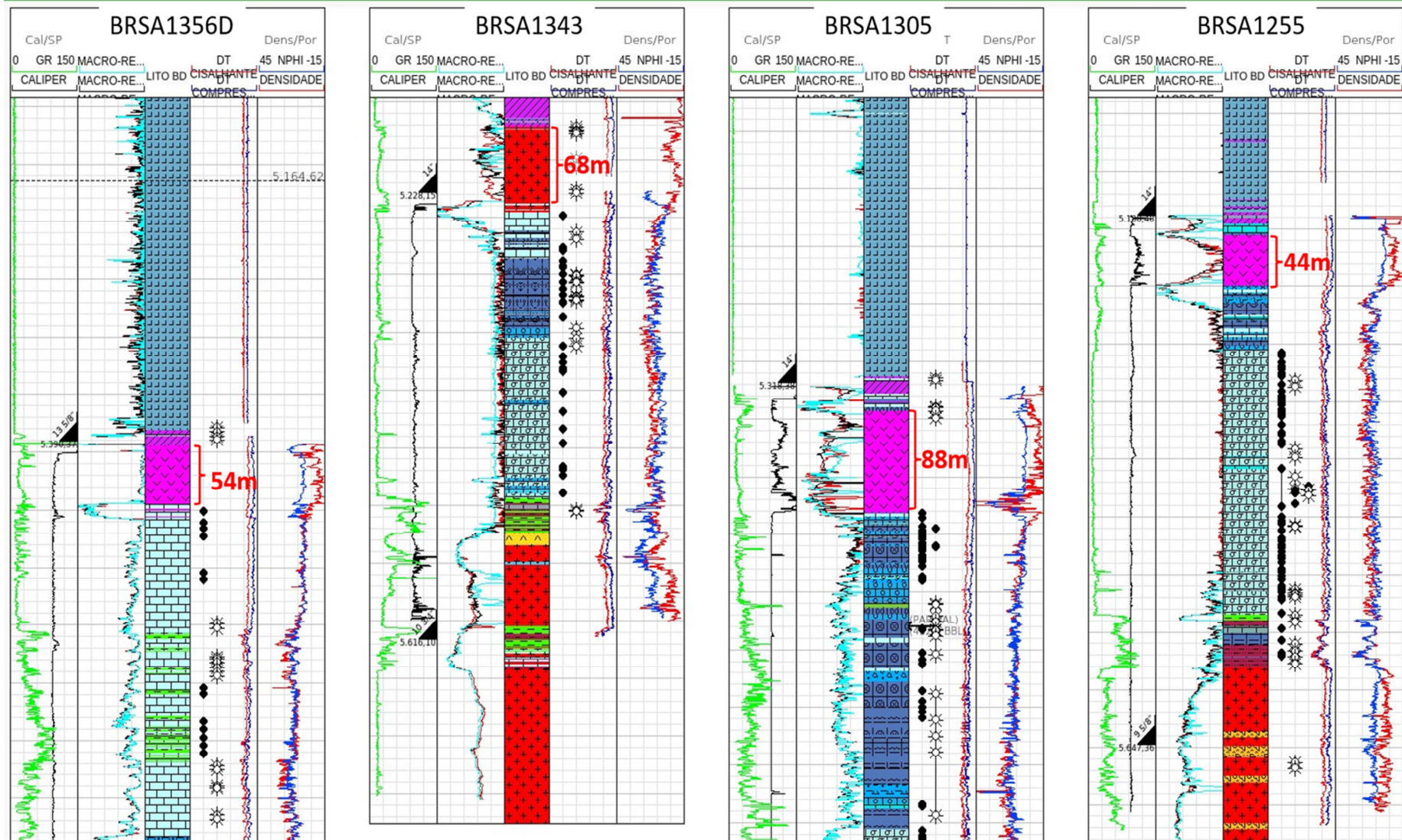
Além dos poços supracitados, os poços 3-BRSA-1339A-RJS, 3-BRSA-1353D-RJS, 3-BRSA-1342-RJS e 9-MRO-5-RJS (sem perfil composto), também não perfuraram rochas ígneas e seus marcadores litológicos foram aplicados à interpretação sísmica. Os quatro poços são exibidos na Figura 34, a qual mostra a ausência de intrusões ígneas no intervalo de estudo.

Figura 34 - Poços sem a presença de rochas ígneas



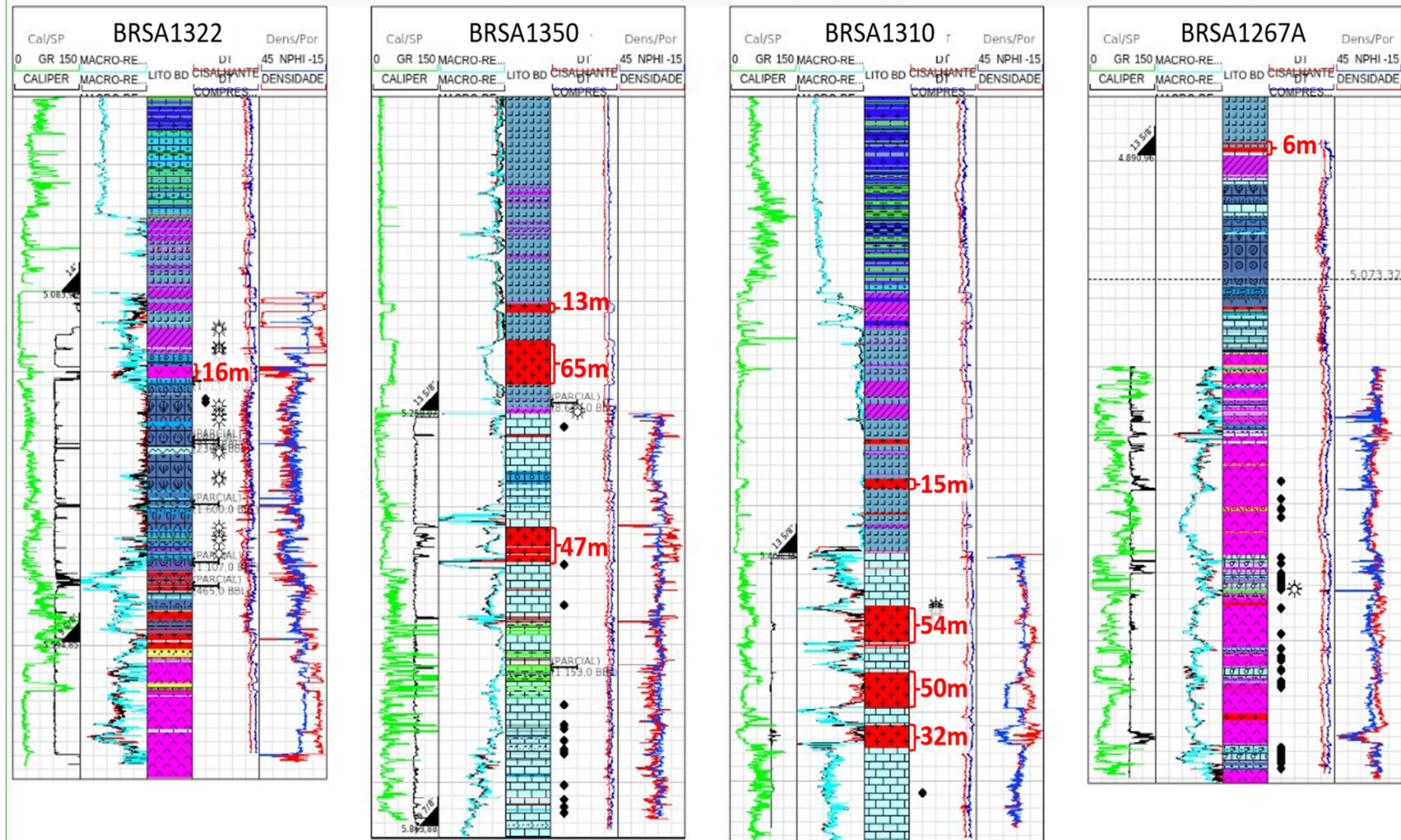
Fonte: O Autor, 2023.

Figura 35 - Poços com intrusões ígneas de espessura relevante próximas à base do sal



Fonte: O Autor, 2023.

Figura 36 - Poços com intrusões ígneas de espessura menor na base do sal



Fonte: O Autor, 2023.

Os poços 3-BRSA-1356D-RJS, 3-BRSA-13543-RJS, 3-BRSA-1305A-RJS, 3-BRSA-1255-RJS e 3-BRSA-1350-RJS amostraram rochas ígneas intrusivas de espessura superior a 40 metros junto à base do sal, região de interesse deste trabalho. Na profundidade em que estas camadas intrusivas ocorrem, é possível que tais espessuras não sejam suficientes para que suas reflexões sejam discerníveis, uma vez que pode ocorrer o chamado efeito de “tuning” (afinação).

O efeito de “tuning” ocorre quando duas reflexões sísmicas passam a interferir construtiva ou destrutivamente, formando apenas uma reflexão (ou destruindo ambas as reflexões) e a espessura mínima necessária para que os eventos sejam individualizados é chamada espessura de afinação. É possível estimar a espessura de afinação utilizando modelagens, onde as características petrofísicas das rochas e o conteúdo de frequências do dado sísmico são utilizados como parâmetros para modelagem, realizada, neste estudo, a partir do repositório GitHub da Society of Exploration Geophysicists (SEG, HALL, 2014).

As informações presentes nos perfis dos poços, referentes aos perfis de sônico compressivo e cisalhante foram utilizadas para calcular as velocidades das ondas P e S, respectivamente, nos intervalos de interesse, bem como os valores do perfil de densidade, os quais foram utilizados na parametrização do modelo de cunha, os valores utilizados estão dispostos no Quadro 2. Além disso, no modelo também são inseridas as informações a respeito do conteúdo de frequências do dado sísmico, a partir da modelagem de uma ondaleta Ricker com aplicação de filtro passa-bandas de modo a representar o dado sísmico naquela profundidade. Para obter esses valores é realizada a estimação do espectro de frequências, diretamente no dado sísmico, janelado na profundidade de interesse, o qual pode ser observado na Figura 37. O espectro de frequências foi obtido na linha interceptada pelo poço 3-BRSA-1305A-RJS, e a janela de extração é concomitante à janela dos atributos sísmicos utilizados como dado de entrada, descritos na próxima seção.

Quadro 2 - Valores utilizados para modelagem

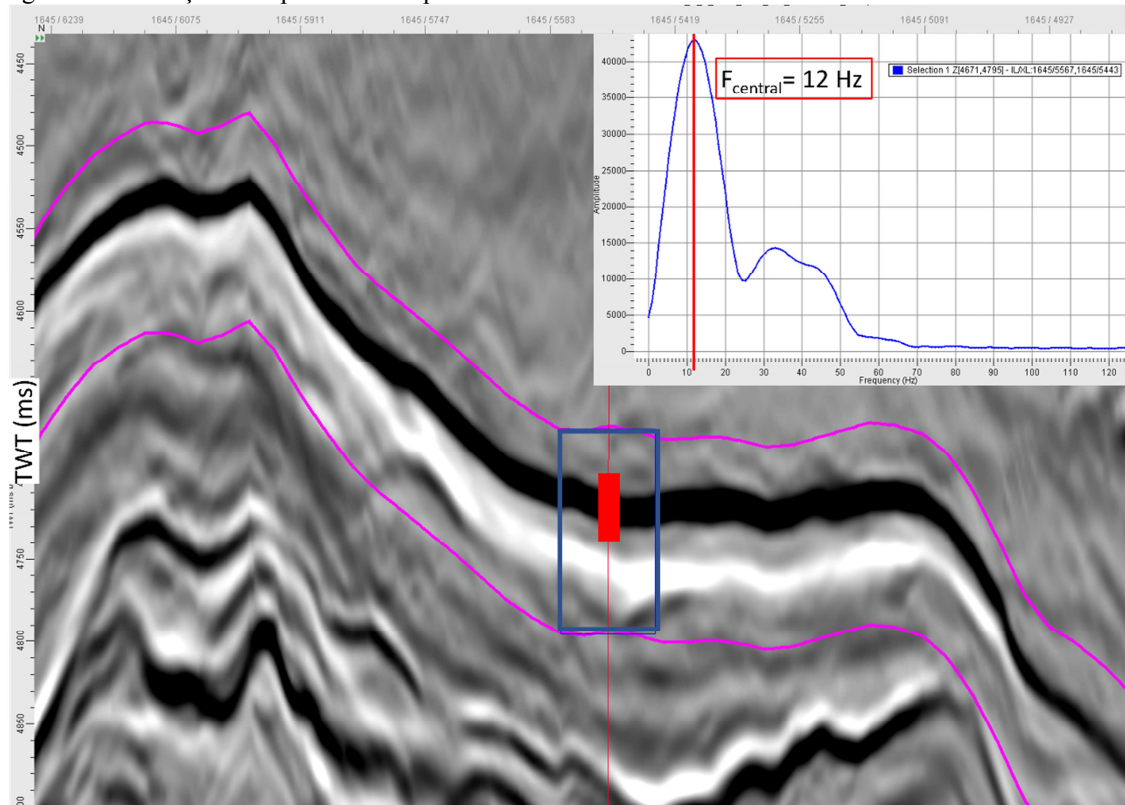
LITOLOGIA	V_P(M/S)	V_S(M/S)	DENSIDADE(G/CM³)
HALITA	4350	2540	2,16
ÍGNEA	5500	3050	2,9
CARBONATO	5080	2900	2,55

Fonte: O Autor, 2023.

No espectro de frequências é possível observar claramente o pico em 12 Hz, o qual recebe o nome de frequência central do dado naquele ponto de interesse. Além da frequência

central, são necessários quatro pontos de referência para o filtro passa-bandas, aplicado à ondaleta para que esta seja representativa da profundidade da região de interesse. Os pontos são referentes às frequências de corte mínimo e máximo (F1 e F4) e as frequências centrais que definem o final da rampa de corte mínimo e início da rampa de corte máximo (F2 e F3).

Figura 37 - Extração de espectro de frequência do dado sísmico



Legenda: No centro da imagem, a janela utilizada para cálculo do espectro de frequências do dado sísmico. Em rosa, os limites do dado utilizado nas etapas de extração de fácies sísmicas. No canto superior direito o espectro com a frequência central de 12 Hz destacada. Em vermelho a posição da intrusão ígnea na seção sísmica, em tempo.

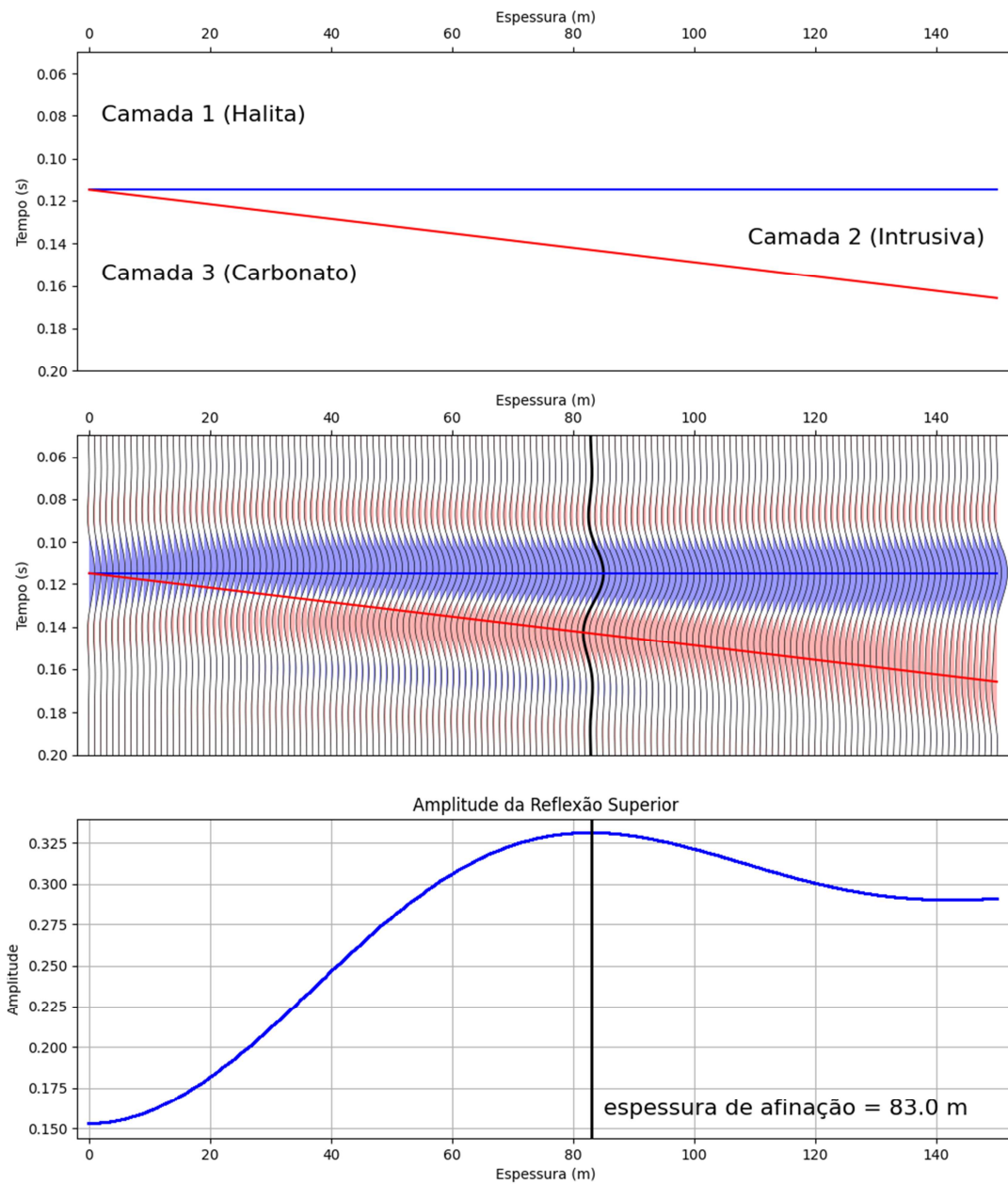
Fonte: O Autor, 2023.

Como a curva do espectro de frequências possui gradiente elevado em seu início, foram utilizados valores mais baixos para F1 e F2, 0 e 3 Hz, respectivamente. Após a frequência central, o conteúdo de frequências decai rapidamente até aproximadamente 25 Hz, a partir da qual, ocorre um moderado aumento do conteúdo de frequências, que pode estar associado ao ruído do dado. Devido à incerteza em relação à causa desse aumento do conteúdo de frequências, foi escolhida uma rampa um pouco menos acentuada para os pontos F3 e F4, os quais receberam os valores de 20 e 30 Hz respectivamente.

A Figura 38 exibe o resultado da modelagem de cunha. O primeiro quadro mostra a disposição da cunha, onde a primeira camada é de Halita, seguida da cunha referente à

intrusão ígnea e da camada posterior de carbonato, representando, portanto, o contexto geológico deste estudo. A espessura de afinção encontrada pela modelagem foi de 83 metros. Esse resultado implica que apenas a camada ígnea no poço 3-BRSA-1305A-RJS, de 88 metros, teria uma reflexão sísmica individualizada, entretanto, como é possível observar na Figura 37, ela não é observada, o que pode indicar que a espessura não tem grande distribuição lateral.

Figura 38 - Modelagem de cunha sedimentar adaptada para intrusão ígnea



Fonte: O Autor, 2023.

O resultado da modelagem não implica, no entanto, que os dados selecionados para este estudo não poderão ser utilizados, uma vez que, como apontado em EIDE et al. (2018); RABEL et al. (2018), mesmo camadas subsísmicas podem perturbar as reflexões sísmicas de modo que estas possam ser observadas em atributos estruturais, por exemplo. Estas perturbações podem, inclusive, ser observadas nesta modelagem, em particular nas espessuras inferiores a 20 metros, onde nota-se um deslocamento para baixo da reflexão, devido à interferência destrutiva entre as duas reflexões.

A modelagem mostra também que o refletor referente à camada de rocha carbonática possui amplitude muito sutil, o que destaca o desafio de diferenciação entre a rocha ígnea não-reservatório e o carbonato potencial reservatório. Entretanto, devido à interferência construtiva entre o lobo de inferior da reflexão da intrusão ígnea e o lobo principal da reflexão da camada carbonática, pode surgir uma aparente reflexão de impedância negativa com comprimento de onda elevado logo abaixo da reflexão da rocha ígnea. Este efeito pode ser observado na modelagem no intervalo de espessura da cunha entre 40 e 70 metros e na Figura 24, entre os poços 3-BRSA-1343-RJS e 3-BRSA-1255-RJS. Nesse intervalo ocorre um refletor contínuo, de alta amplitude negativa e elevado comprimento de onda, o qual parecer ser resultado desse fenômeno e, futuramente, pode ser utilizado como indicativo da presença de intrusão ígnea.

A análise dos perfis sônico e de densidade dos poços destaca a importância da modelagem sísmica apresentada, uma vez que esses perfis, particularmente o perfil sônico, mostram uma certa homogeneidade entre as camadas de sal e as intrusivas, com maiores variações nos perfis de densidade, o qual se mostra maior responsável pelas reflexões sísmicas observadas, uma vez que o coeficiente de reflexão é diretamente proporcional ao contraste de impedâncias entre as camadas (a impedância de uma rocha é o produto entre velocidade sísmica e densidade da rocha).

O perfil da ferramenta caliper destaca a presença de fraturas nas rochas ígneas e nos poços 3-BRSA-1343-RJS, 3-BRSA-1305-RJS e 3-BRSA-1350-RJS indica intervalos permoporosos mesmo nas rochas ígneas, o que pode indicar possíveis reservatórios não-convencionais. O poço 3-BRSA-1343-RJS mostra um intervalo junto à camada de sal de 68 metros com perfil caliper homogêneo e não permoporoso, enquanto a camada de rocha ígnea mais basal apresenta um perfil de alta variação e aparente intervalo permoporoso.

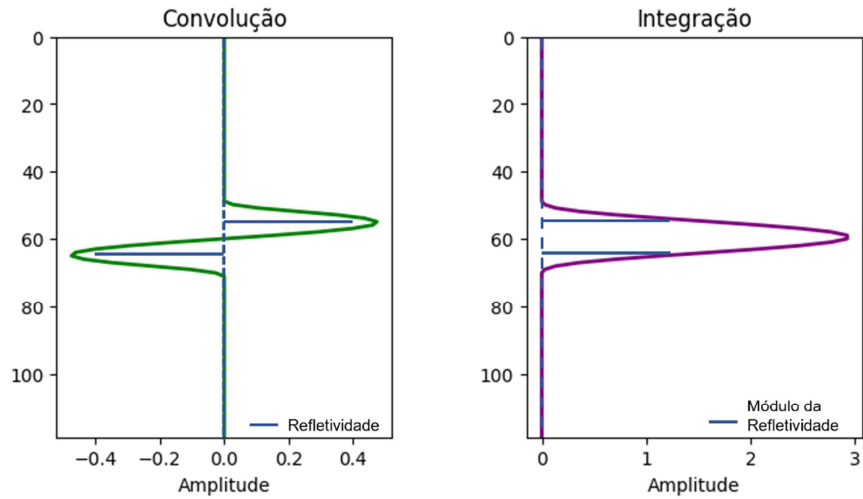
2.3 Atributos Sísmicos

A seleção de atributos para a classificação automática de fácies pode ter um impacto maior nos resultados finais do que a escolha do algoritmo não-supervisionado (BARNES; LAUGHLIN, 2002; ZHAO et al., 2015; RODEN; SMITH; SACREY, 2015). Como observado na seção anterior, o objetivo deste estudo é ainda mais desafiador, ao tentar destacar a presença de camadas intrusivas no limite da resolução sísmica disponível. É preciso lembrar, também, que, além da seleção automática de fácies para observar as intrusões ígneas, outro objetivo é que todo o processo seja realizado com as ferramentas disponíveis na suíte de aplicativos do geocientista de interpretação. Desta forma, processamentos com o intuito de aumentar o conteúdo de frequências do dado não foram utilizados, uma vez que estes demandam alto poder computacional, acima do disponível em estações de interpretação.

O processamento sísmico moderno utiliza, tipicamente, ondaletas de fase zero, ou seja, o pico da ondaleta corresponde ao pico da reflexão sísmica, em outras palavras, o refletor corresponde ao limite entre as camadas sedimentares e não à camada em si. Nos casos em que são comparadas as camadas com os poços, isso faz com que exista um pequeno deslocamento entre a posição do refletor e a posição da camada no poço. Apesar de pequeno, esse deslocamento dificulta a análise visual dos dados, especialmente quando se trabalha no limite de sua resolução. Para sanar esse problema é realizada a integração do dado sísmico, comumente realizada junto aos processamentos de aumento de conteúdo de frequências. A

Figura 39 é uma representação esquemática do que ocorre quando um traço sísmico, resultante da convolução entre um traço de refletividade e uma ondaleta de fase zero, é integrado, passando a representar a camada. Como a integração acentua as baixas frequências do dado, um filtro estrutural também foi aplicado. O dado integrado com filtro estrutural, a partir do dado em profundidade de migração RTM, foi o dado de entrada utilizado para o cálculo de todos os atributos utilizados.

Figura 39 - Traço sísmico da convolução com a refletividade (esquerda), Integração (direita)

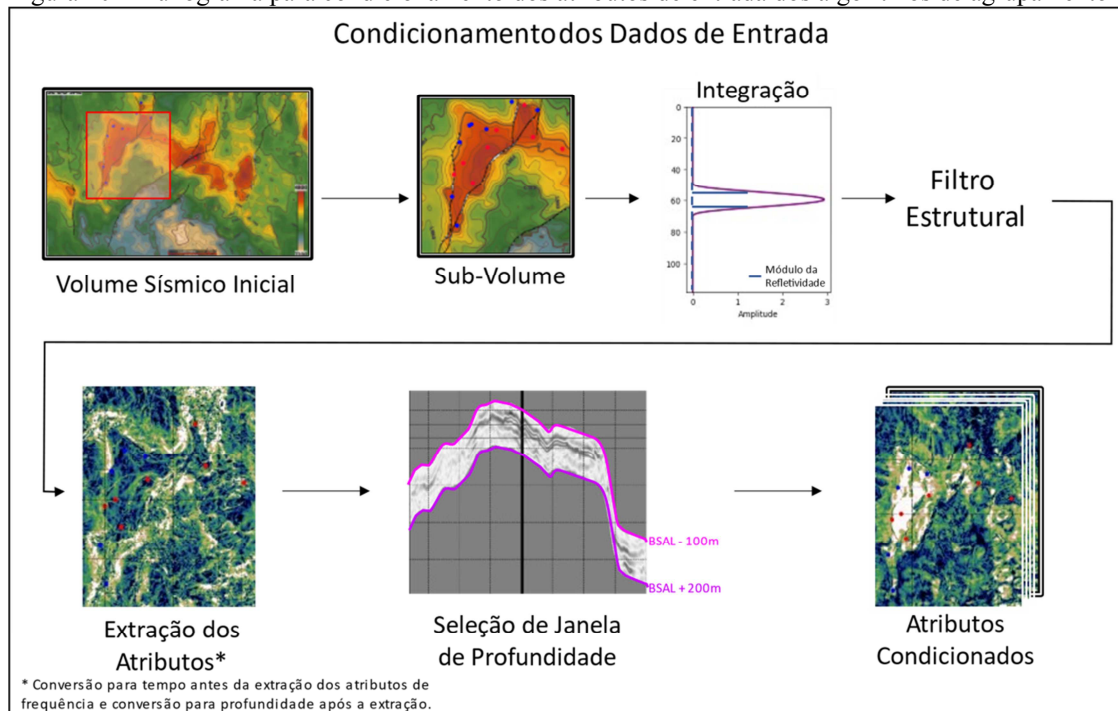


Fonte: O Autor, 2023.

Para reduzir a exigência de memória nas etapas seguintes, foi também escolhido um polígono dentro do volume total e que corresponde apenas à região da estrutura do Campo de Mero, além de três poços à leste (3-BRSA-1322-RJS, 3-BRSA-1350-RJS, 3-BRSA-1310-RJS), incluídos por também apresentarem rochas ígneas no intervalo de interesse, mas ainda não situados na área da estrutura de Libra. Foi também aplicada uma máscara no dado, utilizando como referência o horizonte de base do sal, excluindo todo o dado anterior ao horizonte de sal menos 100 metros e posterior ao horizonte mais 200 metros. Com isso, a janela utilizada foi de apenas 300 metros. Essa prática também tem o objetivo de diminuir a quantidade de amostras a serem agrupadas nos algoritmos, aumentando a possibilidade de agrupamentos bem-sucedidos. Todo o fluxo de preparação dos atributos pode ser observado na

Figura 40.

Figura 40 - Fluxograma para condicionamento dos atributos de entrada dos algoritmos de agrupamento



Fonte: O Autor, 2023.

Foram estabelecidos critérios para a seleção dos atributos de modo a destacar as relações de amplitude das rochas no intervalo de interesse e sua disposição estrutural, ou seja, foram selecionados desde atributos clássicos como Fase Instantânea e Envelope até atributos modernos que destacam texturas do dado sísmico, como os GLCM. Também foram selecionados atributos calculados a partir de decomposição espectral, quando as frequências do dado são discretizadas em diferentes volumes e os atributos calculados a partir destes. Os atributos baseados na frequência do dado, tais como fase instantânea, frequência média, decomposições espectrais, entre outros, foram convertidos para tempo antes de serem estimados e, posteriormente, convertidos de volta para profundidade.

Ao todo, foram gerados 28 atributos sísmicos, dispostos no Quadro 3. O programa utilizado para o cálculo dos atributos foi o Attribute-Assisted Seismic Processing & Interpretation (AASPI), desenvolvido por um consórcio entre as Universidades de Oklahoma,

Universidade do Alabama, Universidade do Texas e SISMO. Toda a sua documentação e a teoria utilizada para cálculo dos atributos estão dispostas em sua página da internet (“AASPT”, 2022). É importante destacar que a maioria dos atributos não possui parametrização extensa, a qual se resume à seleção de uma janela temporal para cálculo. Como o objetivo do estudo encontra-se no limite da resolução sísmica, a janela selecionada sempre foi a menor possível com o intuito de não reduzir ainda mais a resolução vertical do dado.

Quadro 3 - Lista dos atributos selecionados. D.E.: Decomposição Espectral

Amplitude	Magnitude do Mergulho	Homogeneidade GLCM
Largura de Banda Média	Desordem	Média GLCM
Frequência Média	Razão de Energia	Variância GLCM
Energia Coerente 10 Hz (D.E.)	Razão de Energia 10 Hz (D.E.)	Frequência Instantânea
Energia Coerente 20 Hz (D.E.)	Razão de Energia 20 Hz (D.E.)	Fase Instantânea
Energia Coerente 30 Hz (D.E.)	Razão de Energia 30 Hz (D.E.)	Convergência de Refletores
Cosseno da Fase Instantânea	Envelope	Sweetness
Mergulho Direção Crossline	Correlação GLCM	Energia Total
Mergulho Direção Inline	Dissimilaridade GLCM	
Curvatura	Entropia GLCM	

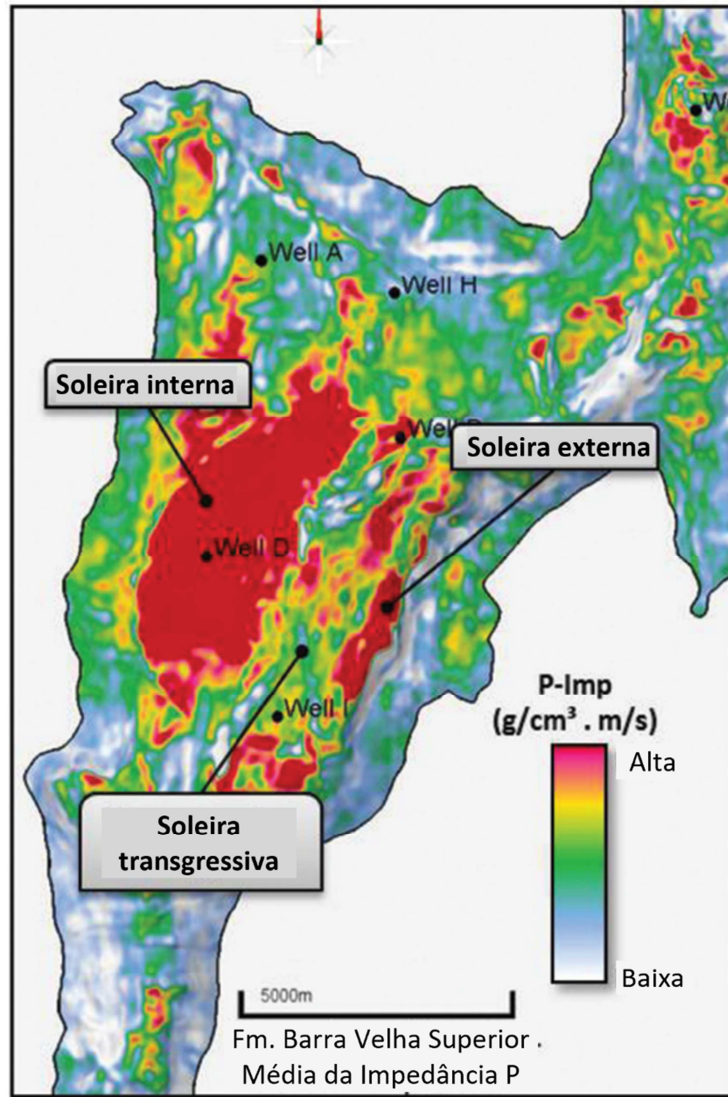
Fonte: O Autor

A

Figura 41, extraída de PENNA et al. (2019), mostra a provável extensão das soleiras no intervalo de interesse do estudo. Ela pode ser considerada uma espécie de gabarito para os resultados almejados, o que significa que pode ser usada para indicar atributos com maior potencial de destacar a presença das rochas ígneas. As Figura 42 a **Erro! Fonte de referência não encontrada.** mostram mapas instantâneos de

todos os atributos extraídos a partir do horizonte de base do sal. Eles são instantâneos porque foi utilizada apenas a amostra na profundidade referente ao horizonte, em detrimento a uma janela de cálculo, usualmente utilizada para representar atributos. A tabela de cores foi selecionada de modo a facilitar a observação dos pontos referentes aos poços utilizados no estudo.

Figura 41 - Disposição das soleiras na área de interesse



Fonte: Adaptado de PENNA et al. (2019).

Ao observar a

Figura 41, o formato aproximadamente circular da soleira interna, denominada desta forma pelos autores, é a estrutura que chama mais atenção e é possível

observar seu traçado em diversos atributos, por exemplo, magnitude do mergulho, dissimilaridade GLCM, incluindo até a amplitude, entre outros. Esta é mais uma evidência de que, ainda que a reflexão isolada da intrusão não seja visível, a estrutura em si pode ser observada.

Os atributos GLCM dividem o traço sísmico em níveis de cinza, o que possibilita a extração de informações de texturas presentes nos dados. Em geral, a classificação de fácies sísmicas é dada a partir da análise de texturas, conforme introduzido por MITCHUM JR; VAIL; SANGREE (1977) e seu uso foi determinado, neste trabalho, a partir da limitação já citada das reflexões limitadoras dos pacotes litológicos não serem suficientes para determinar a presença de intrusões ígneas na base do sal.

É importante destacar, no entanto, que a visualização apenas em mapa dos atributos esconde um efeito indesejado, a perda de resolução vertical, particularmente em atributos de coerência, como curvatura e as decomposições espectrais de coerência e razão de energia (Figura 43). Portanto, apesar de parecerem adequados na visão mapa, eles podem inserir ruídos verticais que podem reduzir a capacidade de agrupamento dos algoritmos discutidos nas próximas seções, onde também serão discutidas as abordagens utilizadas para sanar esse problema.

A partir da comparação dos mapas de atributos, na Figura 42 e do mapa da disposição de soleiras na área de interesse, apresentado por PENNA et al. (2019), um conjunto dos seis atributos (Quadro 4) com maior semelhança visual foi selecionado para compor a segunda abordagem ao problema, representando a escolha dirigida dos atributos para a etapa de seleção automática das fácies sísmicas, realizada com os algoritmos de agrupamento. Este conjunto de atributos não passará, portanto, pela etapa de seleção e extração de facetas, uma vez que seu número reduzido não justifica a aplicação de técnicas de redução de dimensionalidade.

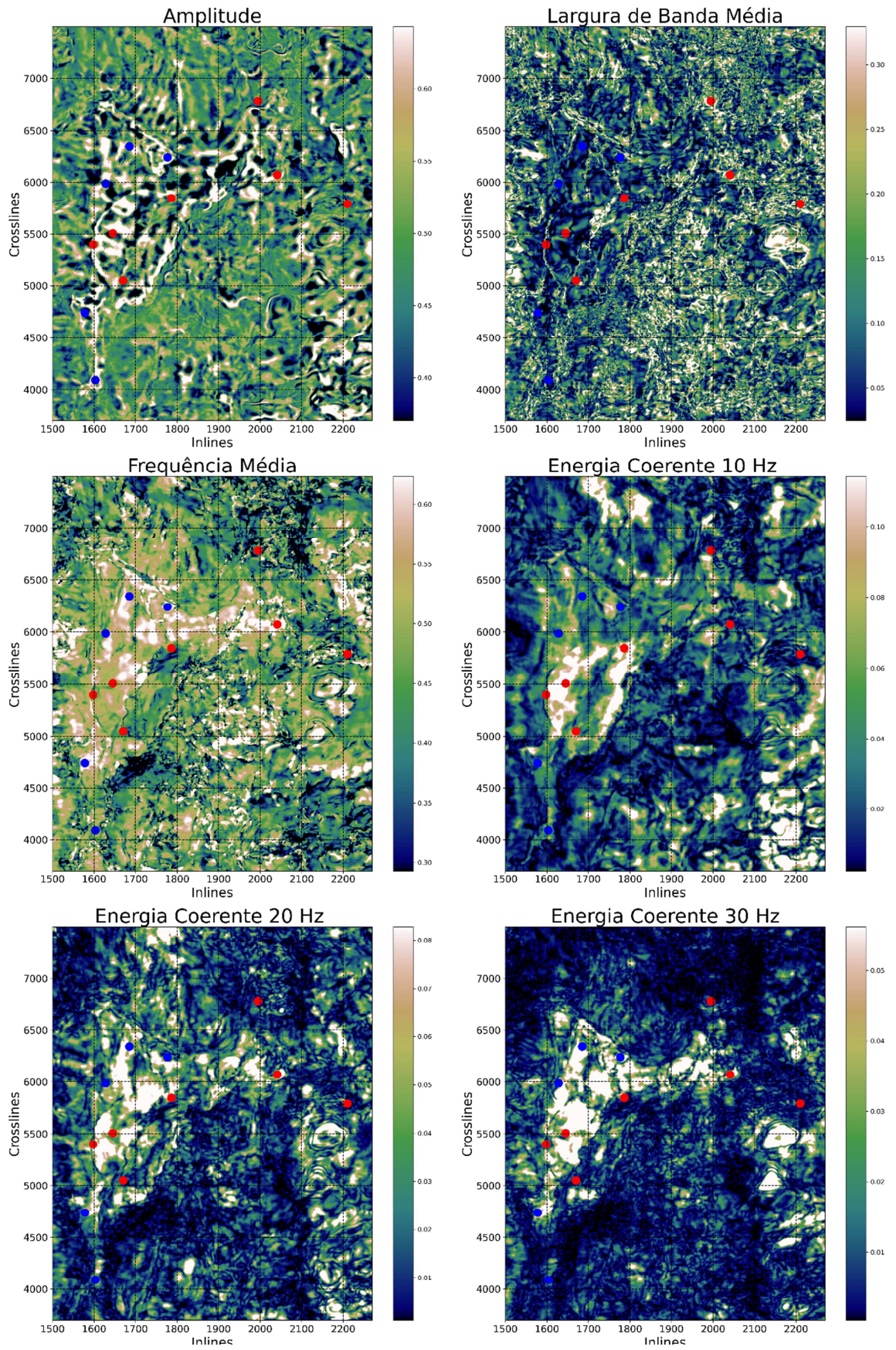
Tabela 2: Atributos selecionados por escolha dirigida

Quadro 4: Atributos selecionados por escolha dirigida

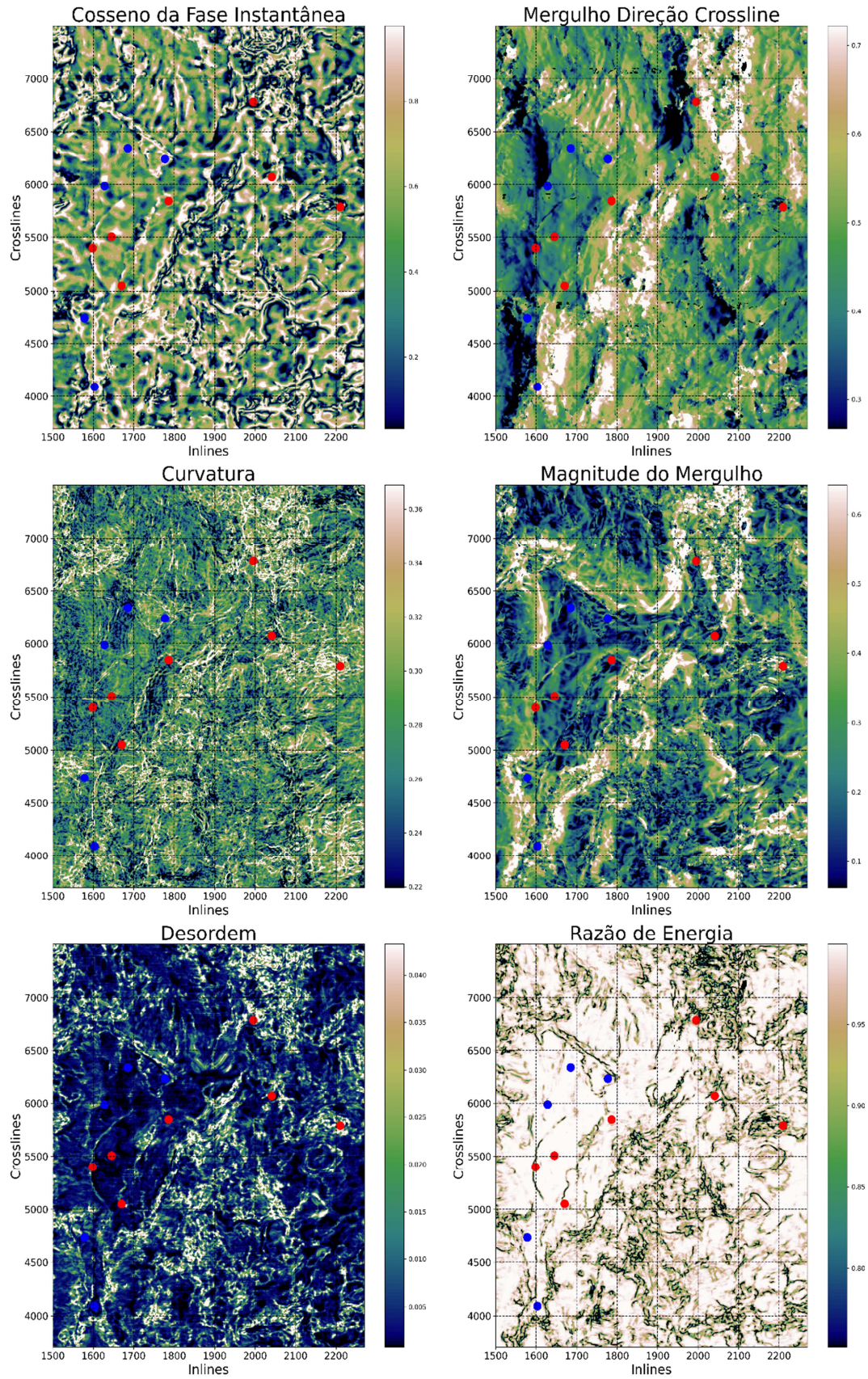
Largura de Banda Média	Envelope	Fase Instantânea
Magnitude do Mergulho	Dissimilaridade GLCM	Sweetness

Fonte: O Autor, 2023.

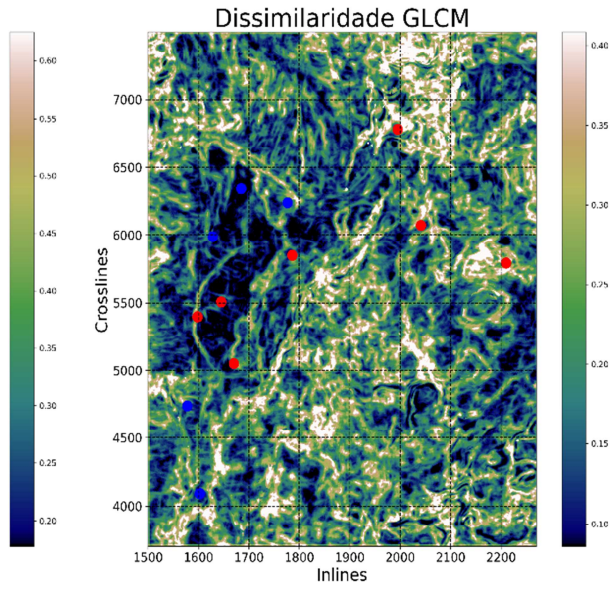
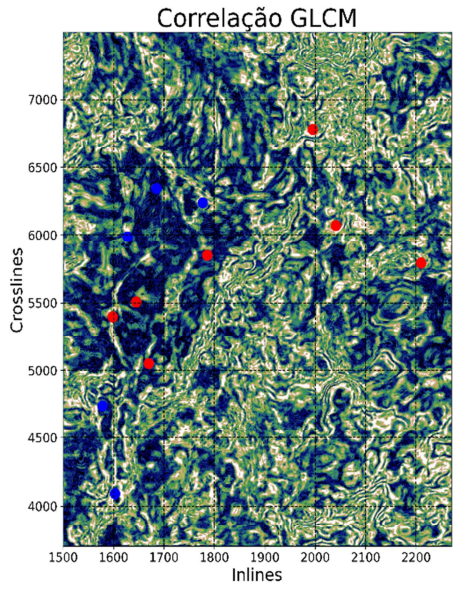
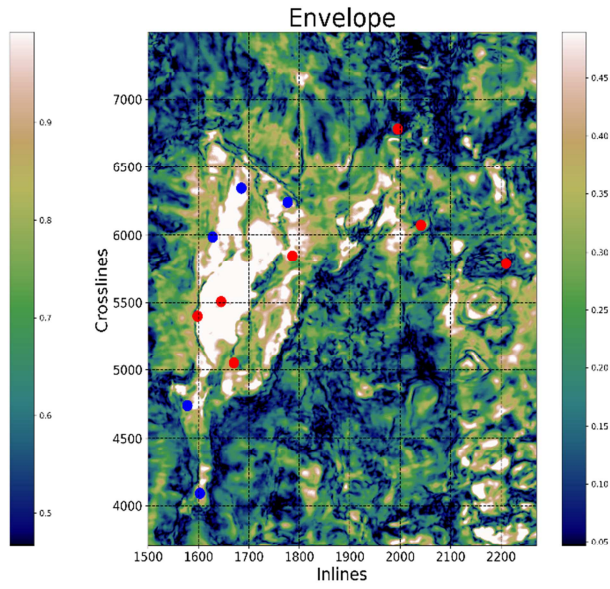
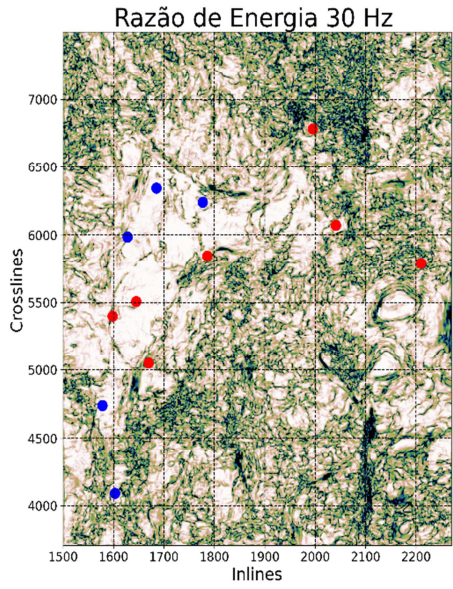
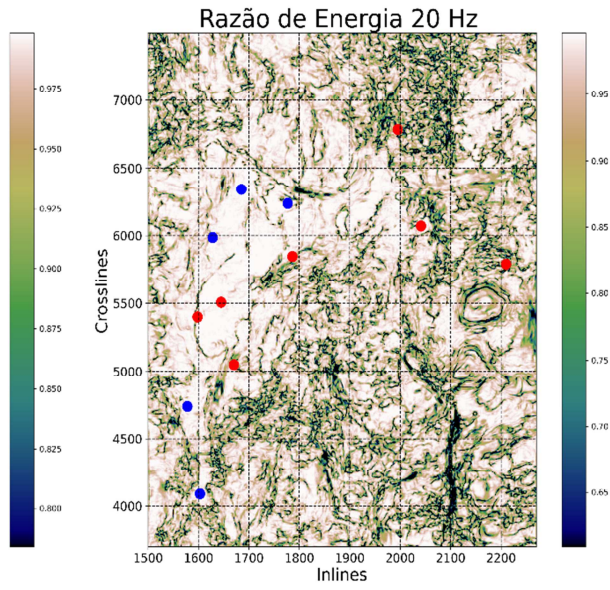
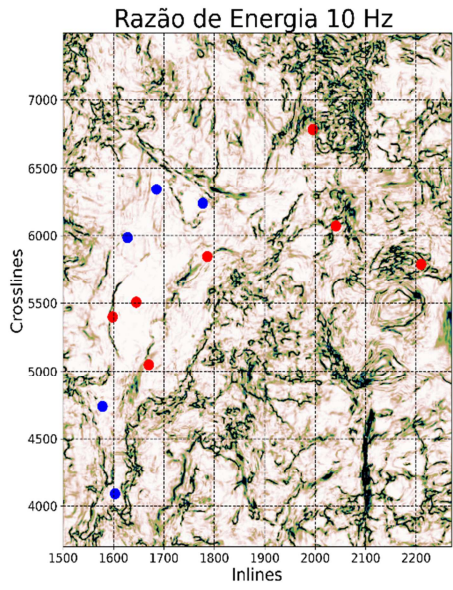
Figura 42 - Atributos sísmicos gerados (continua)



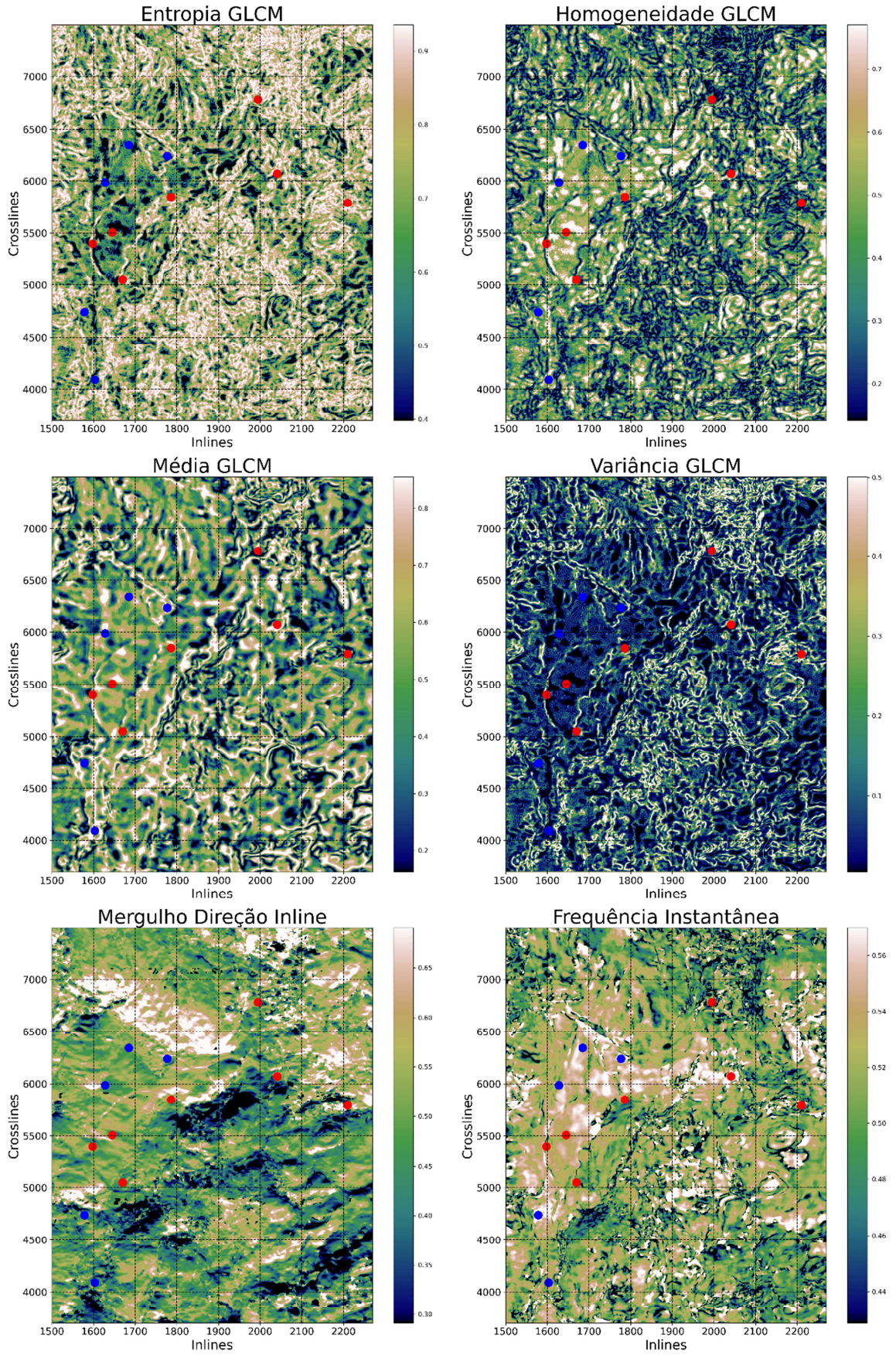
(continuação)



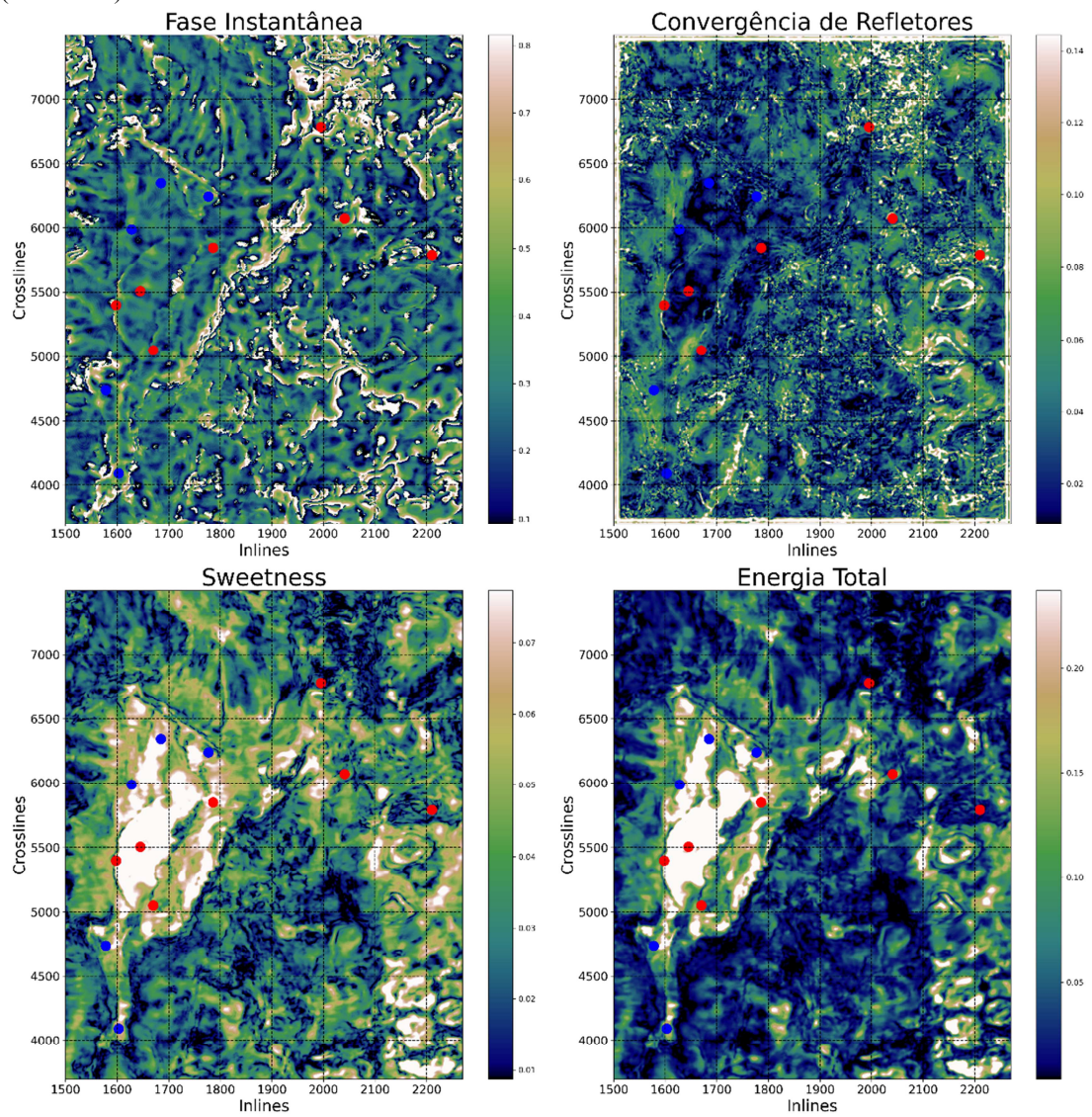
(continuação)



(continuação)



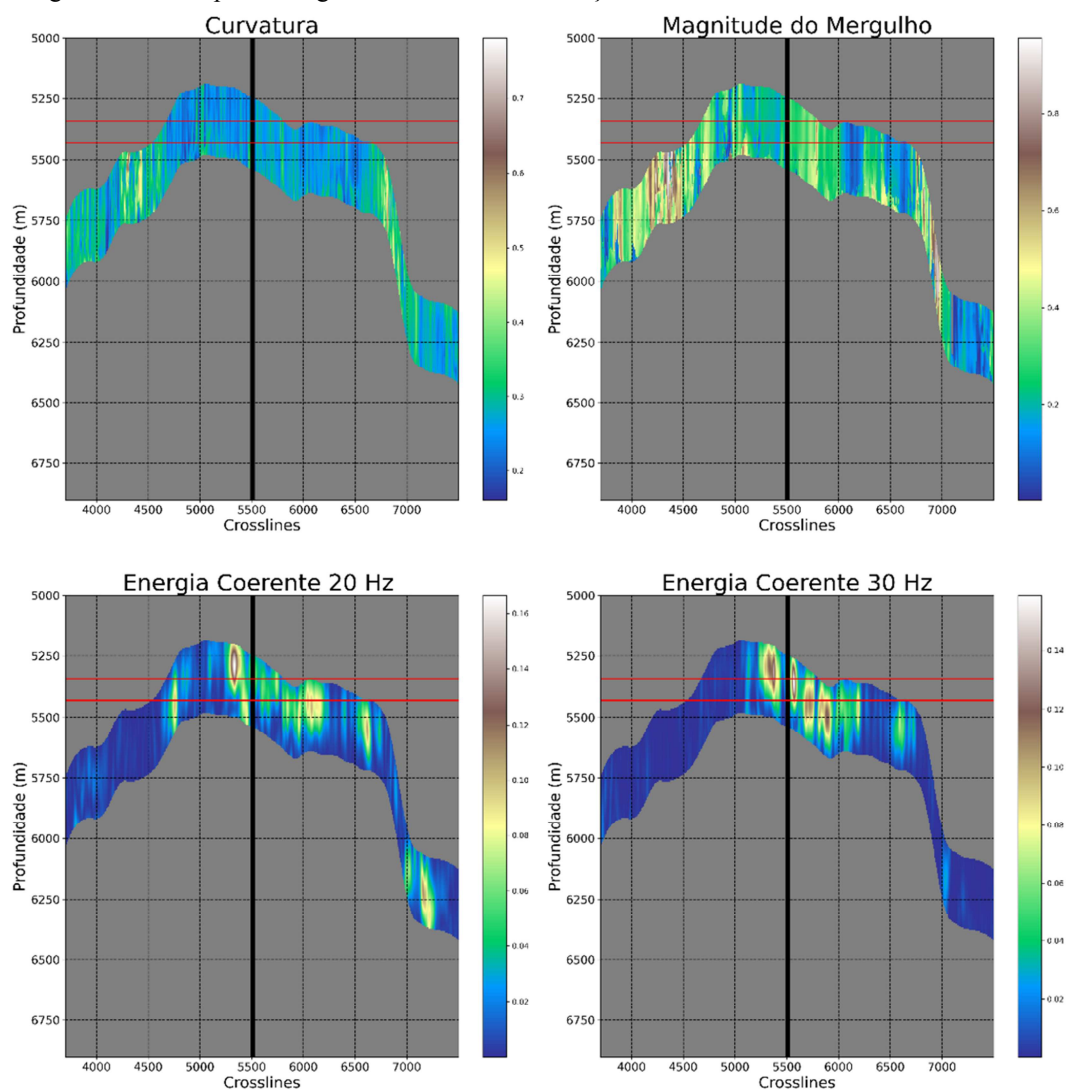
(conclusão)



Legenda: Pontos em vermelho mostram os poços com camadas de rochas ígneas, em azul os que não amostraram estas rochas.

Fonte: O Autor, 2023.

Figura 43 - Exemplos de alguns atributos com resolução vertical reduzida



Fonte: O Autor, 2023.

2.4 Técnicas de agrupamento

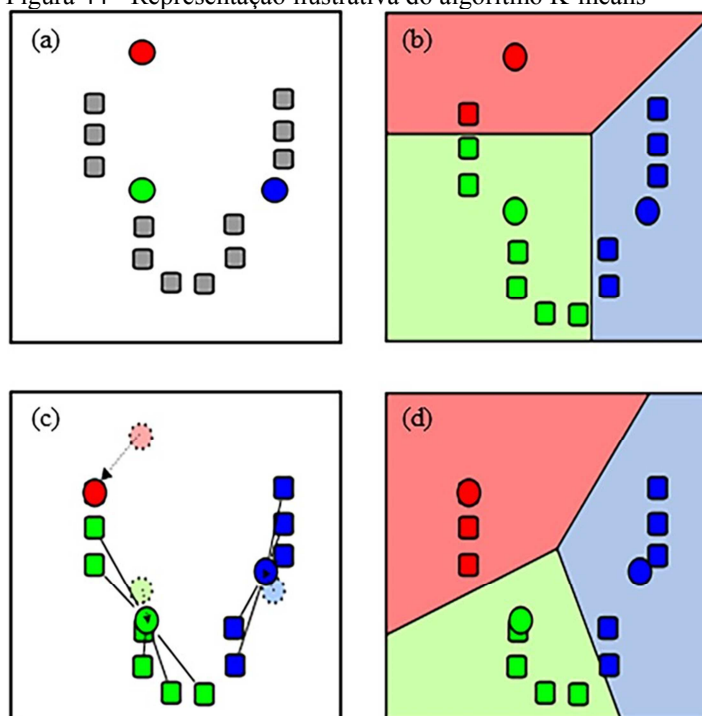
2.4.1 Algoritmo K-Means

A técnica de agrupamento particionado mais popular é a *K-Means* (DE AMORIM, 2016), demonstrada por MACQUEEN, (1967), desenvolvida inicialmente no campo de processamento de sinais, seu objetivo é dividir os dados em um número K de agrupamentos

com a menor distância euclidiana entre as amostras e seus centroides (DI; SHAFIQ; ALREGIB, 2018). É um algoritmo iterativo, e funciona da seguinte forma (AMR, 2020):

- a) Inicia escolhendo um número K de pontos aleatórios e define-os como centroides dos agrupamentos;
- b) Atribui cada ponto ao centroide mais próximo de modo a formar K agrupamentos;
- c) Calcula um novo centroide para os agrupamentos formados;
- d) Como os centroides foram atualizados, retorna ao passo 2 para reatribuir as amostras para seus novos agrupamentos. Se o centroide não se moveu muito, o algoritmo convergiu e para.

Figura 44 - Representação ilustrativa do algoritmo K-means

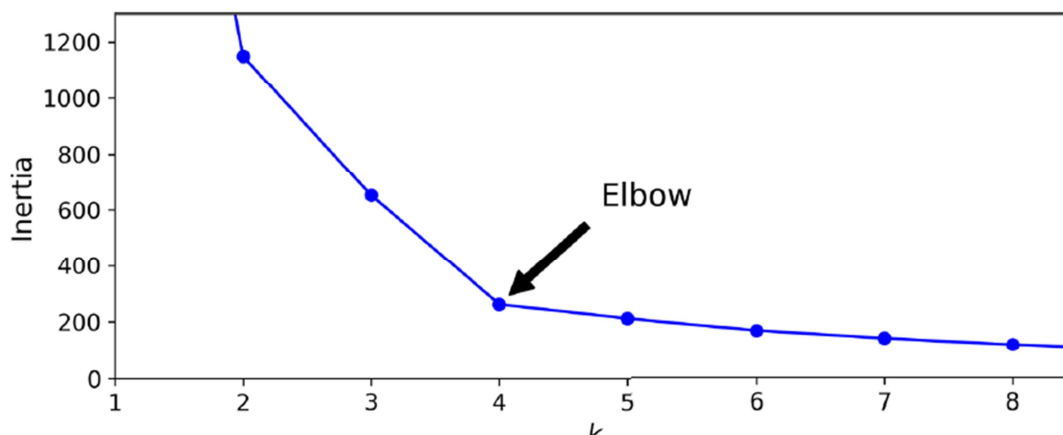


Fonte: DI; SHAFIQ; ALREGIB, 2018;

O número de agrupamentos em K-Means é um parâmetro inicial do algoritmo, a cada iteração, o número de amostras de cada agrupamento pode variar, bem como seus centroides, entretanto, o número K não varia. Para determinar o número de componentes, costuma se medir a inércia que, inicialmente, não é uma boa métrica de aproveitamento um vez que esta sempre decresce como o aumento de número de grupos (GERON, 2022) e um número alto resulta em superclassificação das amostras. No entanto, é possível utilizar a inércia como uma ferramenta gráfica, o chamado Método do Cotovelo (THORNDIKE, 1953), o qual identifica o

valor de K onde ocorre a inflexão passando de redução acentuada da inércia para relativa estabilidade, apresentando decréscimo menos acentuado (Figura 45).

Figura 45 - Inércia plotada em função do número de grupos e o ponto de inflexão chamado “cotovelo”



Fonte: GERON, 2022.

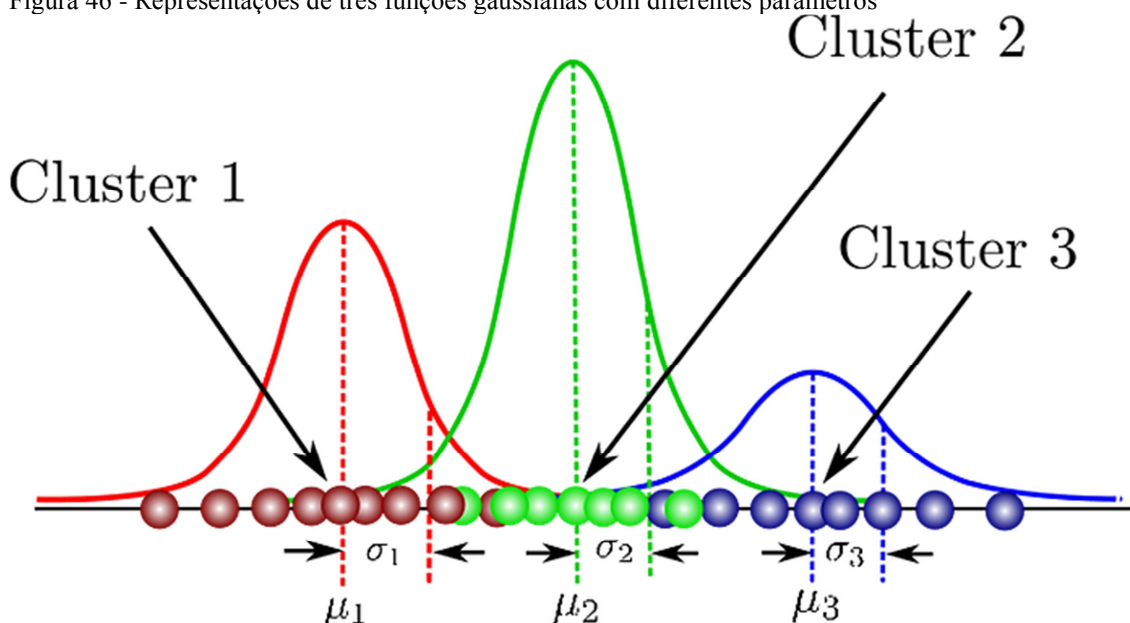
2.4.2 Algoritmo Modelos de Mistura Gaussiana

Enquanto o algoritmo K-Means atribui instâncias a clusters com atribuições definidas (“hard clustering”), o GMM usa atribuições não-definidas (“soft clustering”), estimando a probabilidade de que uma instância pertença a cada. Em outras palavras, o K-Means define um único grupo para cada amostra, enquanto o GMM estima a probabilidade de cada amostra para vários grupos, sendo o primeiro, o mais provável. Durante a etapa de expectativa do EM, o algoritmo calcula as probabilidades com base nos parâmetros do cluster atual. Na etapa de maximização, os clusters são atualizados usando todas as instâncias do conjunto de dados, com cada instância ponderada pela probabilidade de pertencer a um cluster específico (BISHOP, 2006; GERON, 2022).

Uma Mistura Gaussiana é uma função composta por várias distribuições gaussianas, cada uma identificado por $k \in \{1, 2, \dots, K\}$, onde K é o número de clusters do conjunto de dados. Cada gaussiana k na mistura é composto pelos seguintes parâmetros: uma média μ que define seu centro, uma covariância Σ que define sua largura, ou seja, as dimensões de um elipsóide em um cenário multivariado e uma probabilidade de mistura π que define quão grande ou pequena será a função gaussiana (CARRASCO, 2020). Essa relação pode ser observada graficamente na Figura 46.

O algoritmo GMM é implementado a partir da aplicação de um outro algoritmo, chamado algoritmo de maximização de expectativa (“Expectation-Maximization”, EM), um modo iterativo de estimar as probabilidades máximas dos parâmetros do modelo quando o dado possui variáveis desconhecidas, neste caso, o número de clusters e suas dimensões (μ , Σ). O algoritmo escolhe valores aleatórios para os dados desconhecidos e estima um novo conjunto de dados. Estes valores são atualizados recursivamente até o algoritmo alcançar convergência gaussiana (CARRASCO, 2020).

Figura 46 - Representações de três funções gaussianas com diferentes parâmetros



Fonte: CARRASCO, 2020.

Nos casos em que há muitas dimensões ou muitos clusters, o algoritmo EM pode ter problemas para alcançar a solução ótima. É possível reduzir a dificuldade da tarefa limitando o número de parâmetros a serem estimados pelo algoritmo, para isso, se pode limitar os tipos de formas e orientações que os clusters possuem, o que pode ser alcançado impondo limites às matrizes de covariância. O parâmetro tipo de covariância se refere à forma que as distribuições gaussianas ocorrem, que podem ser esféricas, quando os clusters possuem formato esférico, mas podem possuir diâmetros distintos, diagonais, quando possuem formato elipsoidal de qualquer tamanho mas os eixos do elipsoide devem ser paralelos aos eixos de coordenada, amarrados, quando os clusters possuem a mesma forma elipsoidal, tamanho e orientação, ou completos quando cada cluster possui qualquer forma, tamanho e orientação (GERON, 2022).

Para encontrar o número ideal de componentes, se costuma utilizar o Critério de Informação Bayesiano (WALLET; HARDISTY, 2019; GERON, 2022), entretanto estes autores divergem no método de cálculo e na interpretação dos valores. Para GERON (2022), quanto menor o valor de BIC melhor a parametrização. Para (WALLET; HARDISTY (2019), quanto maior o valor, melhor.

2.4.3 Algoritmo Mapas Auto-organizáveis de Kohonen

Os SOMs reduzem as dimensões dos dados produzindo um mapa 2D que agrupa itens de dados semelhantes com base em suas semelhanças. A análise identifica estruturas organizacionais naturais nos dados, permitindo a visualização de clusters e padrões que podem não ser aparentes em dados de amplitude convencionais (RODEN; SMITH; SACREY, 2015).

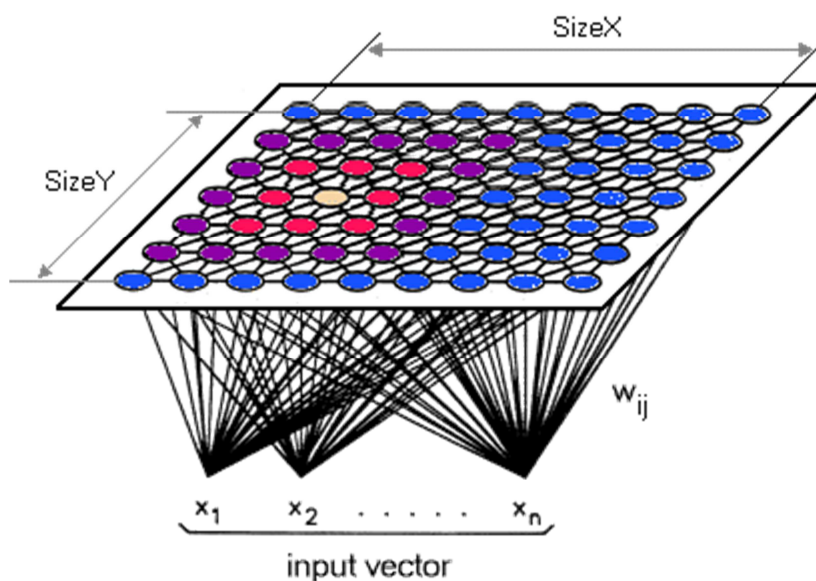
O algoritmo SOM mapeia o dado de entrada em uma variedade geométrica (espaço topológico que lembra um espaço euclidiano, com a propriedade de que cada ponto possui vizinhança que é homeomorfa a um subconjunto aberto do espaço n-dimensional euclidiano) chamado “mapa auto-organizado”. Esse mapa é definido por um conjunto de vetores protótipos, geralmente arranjados em mapas de estrutura hexagonal ou retangular que preservam sua relação com a vizinhança original, de modo que os vetores protótipos vizinhos representam vetores de dados semelhantes. O número de vetores protótipos no mapa 2D determina a eficácia e a generalização do algoritmo (ZHAO et al., 2015).

Em outras palavras, os SOMs possuem uma estrutura de grade 2D linear (Figura 47), na forma de camada única de neurônios. Todos os nós desta grade estão conectados diretamente ao vetor de entrada, mas não entre si, ou seja, os nós não conhecem os valores de seus vizinhos e apenas atualizam o peso de suas conexões em função das entradas fornecidas. A grade em si é o mapa de características que se organiza a cada iteração em função da entrada dos dados de entrada. Assim, após o agrupamento, cada nó possui sua própria coordenada (i,j) , o que permite calcular a distância euclidiana entre 2 nós por meio do teorema de Pitágoras (EKLAVYA, 2019). O algoritmo SOM funciona da seguinte forma (EKLAVYA, 2019):

- a) Inicialize o peso de cada nó w_{ij} para um valor aleatório;
- b) Selecione um vetor de entrada aleatório x_k ;

- c) Repita os seguintes pontos 4 e 5 para todos os nós do mapa;
- d) Calcule a distância euclidiana entre o vetor de entrada $x(t)$ e o vetor de peso w_{ij} associado ao primeiro nó, onde $t, i, j = 0$;
- e) Acompanhe o nó que produz a menor distância t ;
- f) Encontre a Unidade de Melhor Encaixe (Best Matching Unit, BMU) geral, ou seja, o nó com a menor distância de todos os calculados.
- g) Determine a vizinhança topológica $\beta_{ij}(t)$ seu raio $\sigma(t)$ da BMU no mapa de Kohonen
- h) Repita para todos os nós na vizinhança da BMU: atualize o vetor de peso w_{ij} do primeiro nó na vizinhança da BMU adicionando uma fração da diferença entre o vetor de entrada $x(t)$ e o peso $w(t)$ do neurônio.
- i) Repita toda esta iteração até atingir o limite de iteração escolhido $t=n$.

Figura 47 - Um modelo de mapa de Kohonen, em amarelo o neurônio vencedor



Fonte: EKLAVYA, 2019.

O passo 1 é chamado de fase de inicialização, os passos 2 a 9 representam a fase de treinamento. Para avaliar a qualidade do mapa de características costuma-se utilizar dois indicadores: qualidade de aprendizado e qualidade de projeção. O indicador de qualidade de aprendizagem é determinado através da medição do erro de quantização (EQ). O indicador de qualidade da projeção é determinado através da medição do erro topográfico (ET). Se os valores de EQ e ET forem pequenos, o mapa de características será avaliado com boa qualidade. Muitos estudos mostraram que a qualidade do mapa de características é muito

afetada pelos parâmetros iniciais da rede, incluindo tamanho do mapa, números de treinamento e raio da vizinhança (ANH TU, 2020)

Em geral, o tamanho da grade é estimado a partir da técnica de tentativa e erro, a partir de aplicação de diferentes parâmetros até alcançar valores pequenos o suficiente de EQ e ET (ANH TU, 2020). Além do tamanho da grade, a quantidade de treinamentos realizada, mencionada no passo 9 do algoritmo, o tamanho da vizinhança topológica, mencionada no passo 7, e a taxa de aprendizado, utilizada no cálculo da fração da diferença no passo 8, determinam a qualidade do resultado final, sendo estes, portanto, os principais parâmetros a serem ajustado durante a execução do algoritmo.

3 RESULTADOS

A partir dos atributos selecionados, exibidos na seção anterior, foram realizadas etapas de seleção de facetas, para eliminar atributos redundantes, extração de facetas, para extrair as características de maior variância para análise de atributos e, por fim, utilizados três algoritmos de agrupamento diferentes, K-Means, SOMs e GMMs, para a classificação automática de fácies sísmicas.

3.1 Seleção de facetas

A seleção de facetas foi realizada a partir da correlação de Spearman, para identificar atributos redundantes e evitar viés no algoritmo de agrupamento. Os atributos sísmicos possuem diferentes valores, desde adimensionais (como amplitude) até valores de frequência, ângulo, entre outros. Comparar valores numéricos altamente variáveis por meio de algoritmos pode não resultar em bons resultados.

Para contornar essa questão, uma técnica de normalização estatística padrão é aplicada. Essa técnica assegura que os valores tenham a mesma distribuição numérica, ao subtrair cada valor pela média da distribuição e dividi-lo pelo desvio padrão correspondente. Após a normalização, calcula-se a correlação de Spearman entre todos os atributos.

A opção da correlação Spearman em detrimento à correlação Pearson, mais comumente utilizada, se dá porque a correlação de Pearson avalia a relação linear entre duas variáveis, enquanto a correlação de Spearman avalia a tendência geral de aumento ou queda dos valores de cada variável, denominada relação monotônica, que inclui relações não-lineares, mais comuns entre os atributos sísmicos.

Uma correlação de Spearman igual a zero indica a ausência de uma tendência geral consistente de aumento ou diminuição dos valores nos dados. Uma correlação de -1 indica uma relação monotônica negativa perfeita entre as variáveis, em que uma variável diminui à medida que a outra aumenta, seguindo um padrão inversamente relacionado. Por outro lado, uma correlação de 1 indica uma correlação positiva perfeita entre as variáveis. No caso da autocorrelação, os valores devem ser sempre 1, como observado na diagonal do mapa de calor na

Figura 48.

Como o objetivo da etapa de seleção de facetas é remover dados redundantes, a correlação é utilizada para definir o quanto os atributos são similares, quantitativamente. Correlações fortemente positivas e fortemente negativa indicam alta similaridade e foram selecionados os valores de -0,6 e +0,6 como limites de seleção. Ou seja, entre os atributos que possuem correlação menor que -0,6 e maior que 0,6, apenas um deles deve ser selecionado de modo a reduzir sua redundância para as próximas etapas. O Quadro 5 indica as relações encontradas para essa parametrização, onde a primeira refere-se ao atributo avaliado e as colunas seguintes os

atributos que foram considerados de alta similaridade, como, por exemplo, no caso da Amplitude que possui alta correlação com Cosseno da Fase Instantânea e Média GLCM, 0,9 e 0,93, respectivamente, conforme os valores observados na

Figura 48.

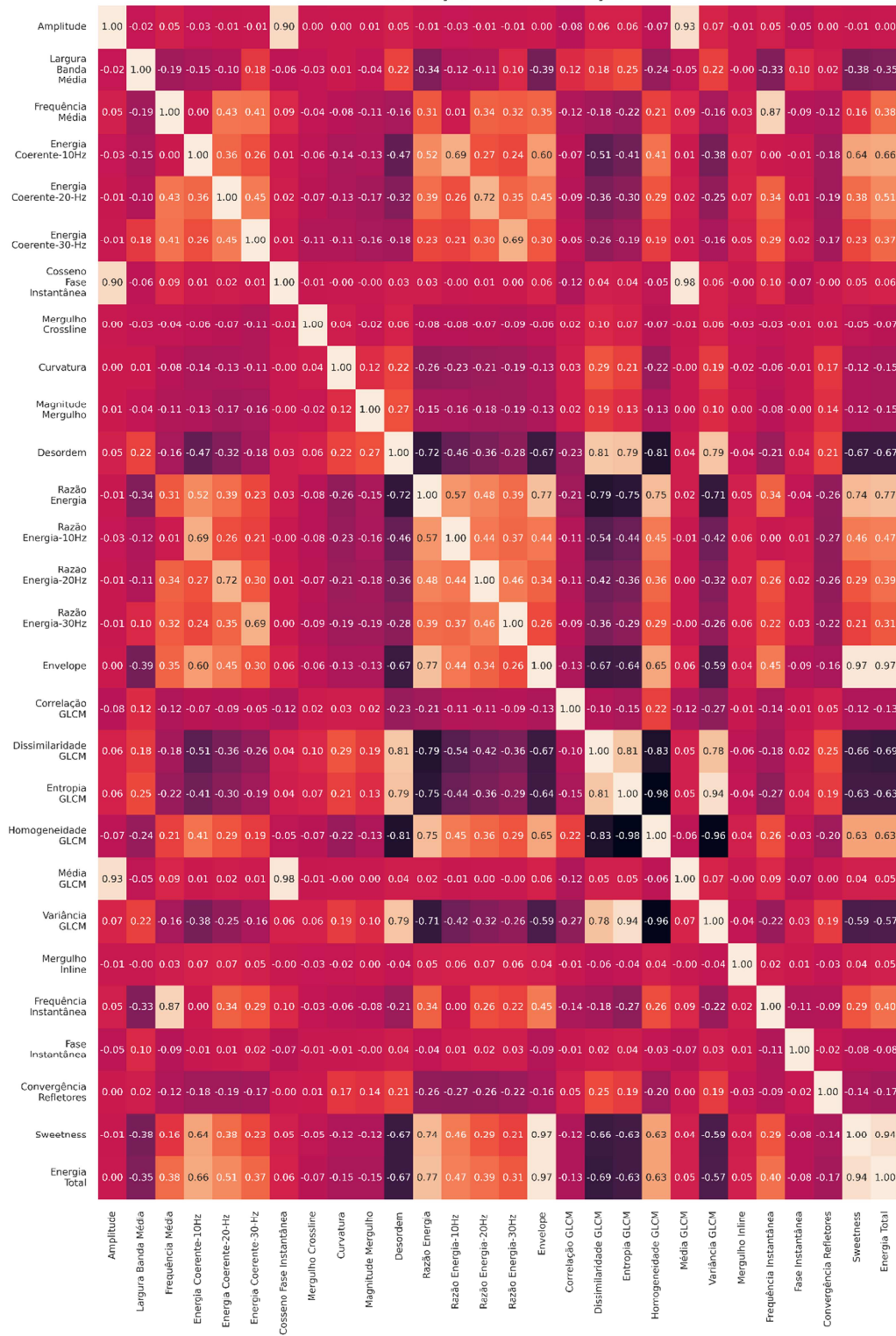
O mapa de calor, apresentado na

Figura 48, permite extrair outras informações acerca da similaridade entre todos os atributos. Ainda que os limites seleção aplicados sejam uma métrica interessante para a seleção de atributos, a não correlação também deve ser considerada, ou seja, atributos que não apresentam correlação com os outros, valores próximos a zero, devem ser priorizados, de modo a aumentar a variância entre os atributos selecionados para a próxima etapa.

Conforme mencionado na seção anterior, alguns dos atributos possuem resolução vertical muito baixa e, apesar de se mostrarem adequados em visualizações de mapas, aumentam o risco de inserir ruídos verticais aos resultados dos algoritmos de agrupamento. A resolução vertical reduzida significa a perda de relação com as camadas litológicas, entretanto, alguns desses atributos são importantes para a demarcação dos limites da soleira observada na região de estudo ou são atributos que não possuem correlação e, caso forem retirados, diminuiriam a variância entre os atributos. Em específico, os atributos de Mergulho Crossline, Mergulho Inline, Curvatura e Magnitude do Mergulho, foram mantidos. Para evitar que esses atributos inserissem os ruídos verticais mencionados, se optou por balancear os resultados, mantendo alguns atributos que representam melhor a geologia, mas que possuem alta similaridade com outros atributos, tais como “Sweetness” e Energia Total.

O objetivo dessa etapa não é apenas, portanto, identificar os atributos com maior correlação, mas também identificar aqueles que não se assemelham a nenhum outro. A escolha final dos atributos combinou a seleção quantitativa da correlação de Spearman com uma seleção qualitativa baseada não só nos mapas dos atributos como também na relação destes com a geologia. Após essa etapa, o número de atributos foi reduzido de 28 a 15, os quais constam no Quadro 6 e serão utilizados na etapa seguinte de extração de facetas.

Figura 48 - Correlação Spearman entre todos os atributos.



Fonte: O Autor, 2023.

Quadro 5 - Atributos com alta correlação

Amplitude	Cosseno da Fase Instantânea	Média GLCM						
Frequência Média	Frequência Instantânea							
Energia Coerente 10 Hz	Razão de Energia 10 Hz	Envelope	Sweetness	Energia Total				
Energia Coerente 20 Hz	Razão de Energia 20 Hz							
Energia Coerente 30 Hz	Razão de Energia 30 Hz							
Cosseno da Fase Instantânea	Amplitude	Média GLCM						
Desordem	Razão de Energia	Envelope	Dissimilaridade GLCM	Entropia GLCM	Homogeneidade GLCM	Variância GLCM	Sweetness	Energia Total
Razão de Energia	Desordem	Envelope	Dissimilaridade GLCM	Entropia GLCM	Homogeneidade GLCM	Variância GLCM	Sweetness	Energia Total
Razão de Energia 10 Hz	Energia Coerente 10 Hz							
Razão de Energia 20 Hz	Energia Coerente 20 Hz							
Razão de Energia 30 Hz	Energia Coerente 30 Hz							
Envelope	Energia Coerente 10 Hz	Desordem	Razão de Energia	Dissimilaridade GLCM	Entropia GLCM	Homogeneidade GLCM	Sweetness	Energia Total
Dissimilaridade GLCM	Desordem	Razão de Energia	Envelope	Entropia GLCM	Homogeneidade GLCM	Variância GLCM	Sweetness	Energia Total
Entropia GLCM	Desordem	Razão de Energia	Envelope	Dissimilaridade GLCM	Homogeneidade GLCM	Variância GLCM	Sweetness	Energia Total
Homogeneidade GLCM	Desordem	Razão de Energia	Envelope	Dissimilaridade GLCM	Entropia GLCM	Variância GLCM	Sweetness	Energia Total
Média GLCM	Amplitude	Cosseno da Fase Instantânea						
Variância GLCM	Desordem	Razão de Energia	Dissimilaridade GLCM	Entropia GLCM	Homogeneidade GLCM			
Frequência Instantânea	Frequência Média							
Sweetness	Energia Coerente 10 Hz	Desordem	Razão de Energia	Envelope	Dissimilaridade GLCM	Entropia GLCM	Homogeneidade GLCM	Energia Total
Energia Total	Energia Coerente 10 Hz	Desordem	Razão de Energia	Envelope	Dissimilaridade GLCM	Entropia GLCM	Homogeneidade GLCM	Sweetness

Fonte: O Autor, 2023

Quadro 6 - Atributos selecionados após seleção de facetas

Amplitude	Curvatura	Correlação GLCM
Largura de Banda Média	Magnitude do Mergulho	Entropia GLCM
Frequência Média	Desordem	Fase Instantânea
Mergulho Direção Crossline	Razão de Energia	Convergência de Refletores
Mergulho Direção Inline	Envelope	Sweetness

Fonte: O Autor, 2023.

3.2 Extração de Facetas

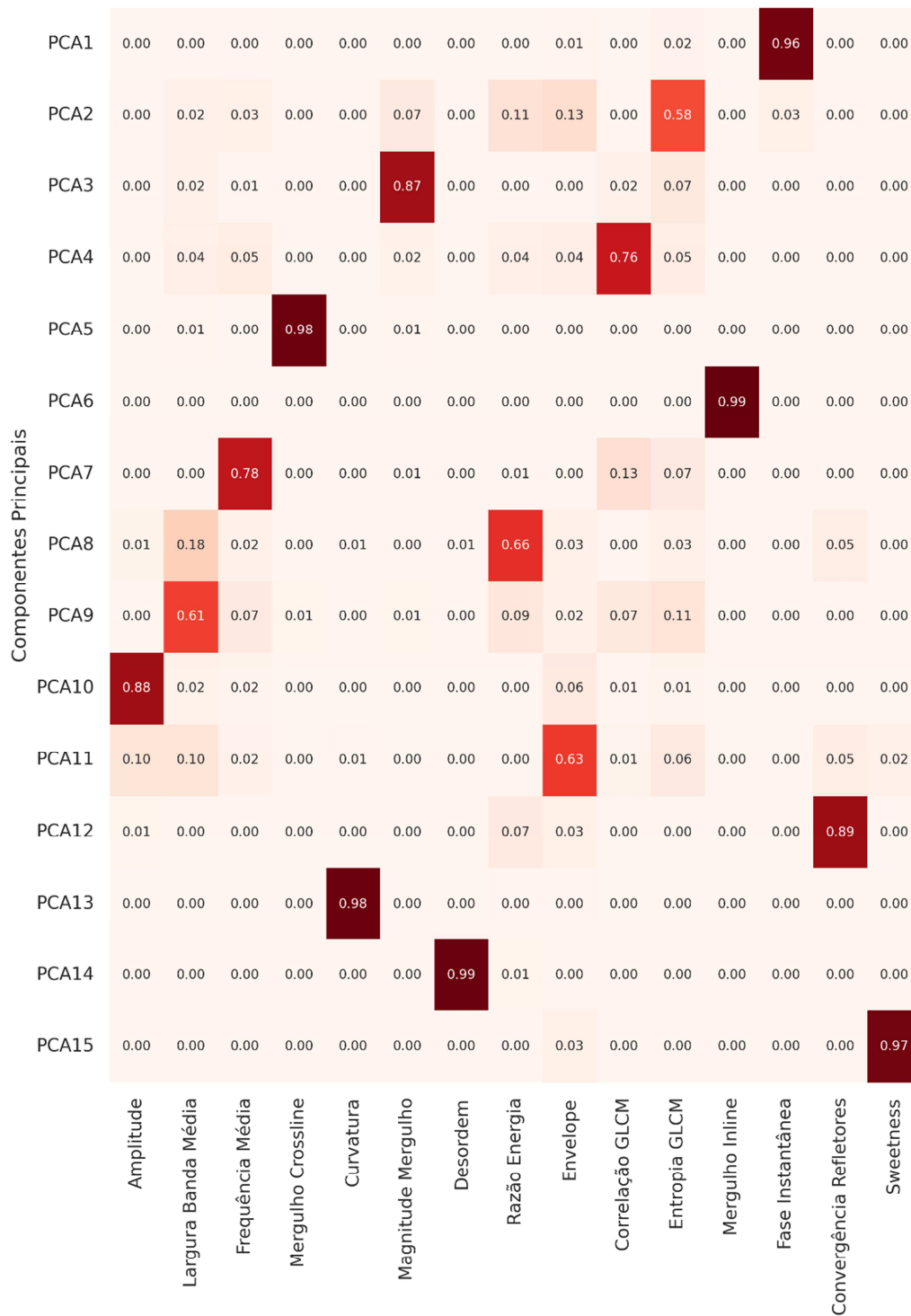
A técnica PCA foi aplicada como método de extração de facetas. Como entrada, foram utilizados os quinze atributos sísmicos selecionados durante a etapa anterior, de seleção de facetas e o algoritmo foi executado de modo a gerar o mesmo número de componentes principais, cujo número máximo é o número de atributos de entrada.

Como já foi mencionado, a PCA age criando variáveis, os componentes principais, não correlacionadas que maximizam a variância entre cada atributo de entrada a partir de uma combinação linear. Dessa forma, cada componente principal representa um tipo de mistura entre os atributos de entrada e pode ser considerado um novo atributo. Devido a seu caráter matemático, não há uma definição fixa para o significado de cada componente, cada aplicação do método deve ser tratada de forma diferente.

A Figura 49 exibe a contribuição de cada atributo para cada componente principal. Neste caso, chama a atenção o fato de que muitos componentes principais são compostos majoritariamente por apenas um atributo (PCAs 1, 5, 6, 10, 12, 13, 14 e 15), o que pode ser considerada uma indicação de aplicação bem-sucedida da seleção de facetas, atributos com alta variabilidade e provável baixa redundância foram selecionados.

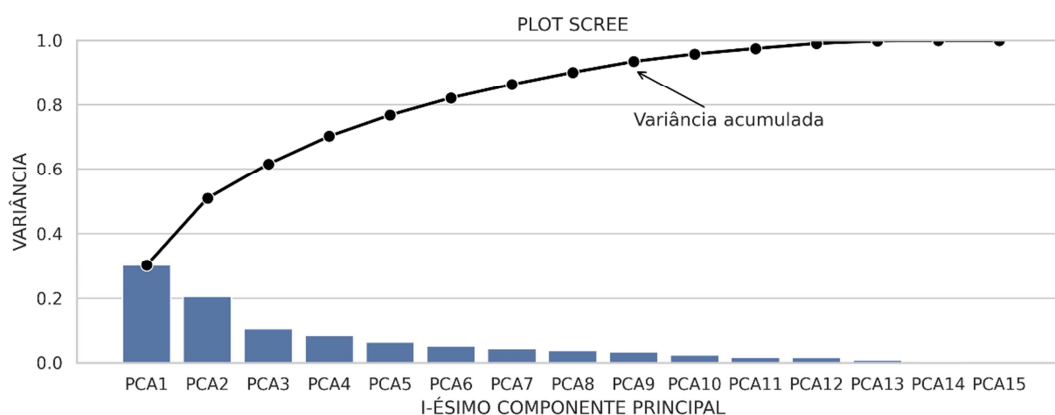
A técnica PCA organiza automaticamente os componentes em ordem decrescente de variância (dispersão dos valores da variável em relação à sua média), como pode ser observado no gráfico da Figura 50, cujas barras mostram a variância de cada componente e a linha mostra a acumulação da variância. O primeiro componente possui uma variância de 0,3, enquanto a do último é praticamente nula.

Figura 49 - Composição dos componentes principais a partir dos atributos de entrada



Fonte: O Autor, 2023.

Figura 50 - Variância dos componentes principais - Gráfico Scree



Fonte: O Autor, 2023.

3.3 Técnicas de agrupamento

As técnicas de agrupamento selecionadas foram K-Means, GMMs e SOMs, as duas primeiras por possuírem implementação e parametrização similares, já a técnica SOM por possuir um histórico mais amplo de uso para classificação automática de fácies e estar disponibilizada em programas comerciais para este propósito.

É importante destacar, no entanto, que todas as implementações foram feitas utilizando rotinas de autoria própria na linguagem de programação Python, utilizando bibliotecas públicas, Scikit-Learn para K-Means e GMMs (PEDREGOSA et al., 2011) e MiniSOM (VETTIGLI, 2018), para SOMs. A decisão de uso de rotinas de implementação própria foi feita de modo a possuir total controle das técnicas utilizadas em todas as etapas do processo, tais como normalização dos dados e seleção de todos os parâmetros disponíveis, haja visto que alguns aplicativos comerciais não disponibilizam a metodologia de implementação utilizada em algumas etapas do processo, ou disponibilizam apenas um tipo de agrupamento.

Após a escolha dos atributos sísmicos, estes são exportados no formato SEG-Y, padrão para dados sísmicos. É utilizada a biblioteca SegYIO, disponibilizada no GitHub pela empresa Equinor, para lidar com a entrada dos dados no ambiente de processamento. A biblioteca transforma cada arquivo em um objeto com as três dimensões. Nesta etapa também é realizada a normalização dos dados, mencionada anteriormente. Todos os algoritmos compartilham essa etapa preliminar.

Finalizado o carregamento de todos os atributos, estes são agrupados em uma única tabela multidimensional, da biblioteca Pandas (MCKINNEY, 2010; TEAM, 2023), onde as três dimensões espaciais são definidas nos índices do objeto e os atributos são dispostos em diferentes colunas. Essa é uma das etapas mais delicadas de todo o processo, uma vez que a disposição espacial do dado não pode ser perdida, sendo importante assegurar que os resultados observados correspondem espacialmente aos dados de entrada. Esse procedimento assegura a reestruturação do dado após as etapas de agrupamento.

A tabela gerada passa a ser o dado de entrada para a etapa de extração de facetas, onde foi aplicada a técnica PCA, que projeta os dados em novas dimensões, denominadas componentes principais. Como descrito no item anterior, dos quinze componentes principais, apenas os dez primeiros foram selecionados para a etapa de agrupamento. A Figura 51 exibe o fluxograma simplificado do processo.

Figura 51 - Fluxograma do processo de agrupamento



Fonte: O Autor, 2023.

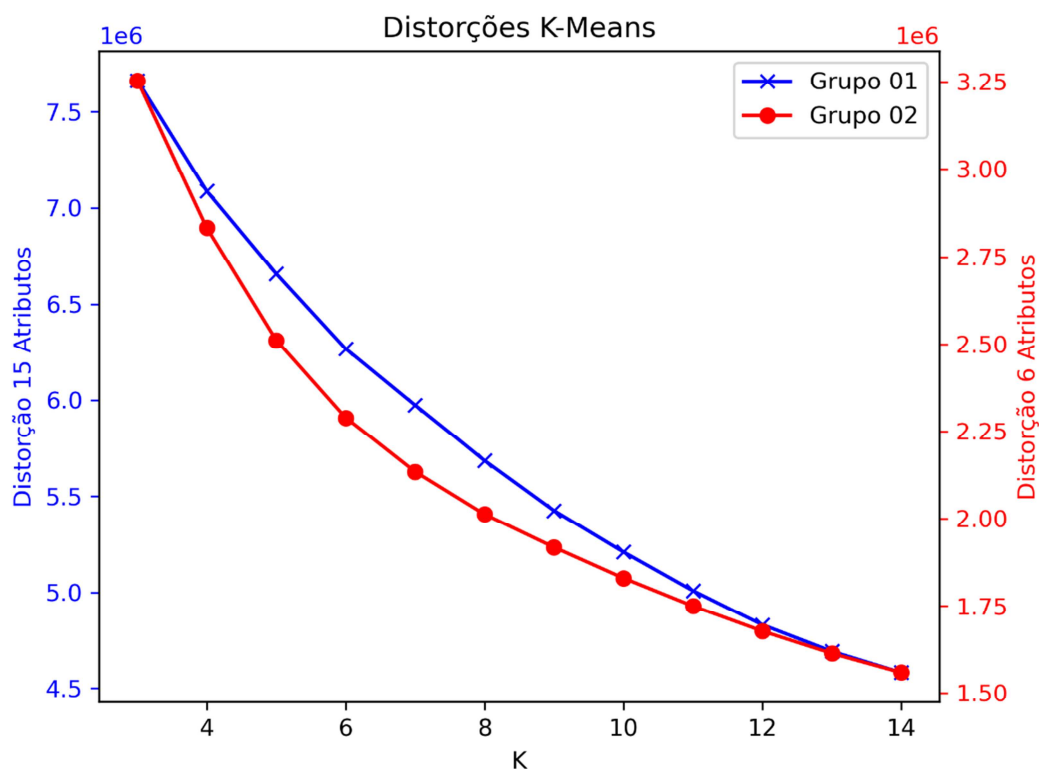
Como dois grupos de atributos distintos foram selecionados, a partir deste ponto, o grupo de 15 atributos oriundos da seleção de facetas será denominado Grupo 01 e o grupo selecionado por escolha dirigida, composto por apenas seis atributos, será denominado Grupo 02. Em relação ao processo aqui descrito, o Grupo 02 não possui apenas a etapa de seleção de componentes principais.

3.3.1 Algoritmo K-Means

O K-Means é um algoritmo muito utilizado para agrupamento devido à sua velocidade de execução e fácil implementação. Seu parâmetro é o número de classes. A primeira etapa é escolher o número de classes, em geral, a parte mais desafiadora do algoritmo. Uma prática comum é calcular a denominada inércia (ou distorção) do dado para um intervalo de “k” componentes que corresponde ao intervalo esperado de grupos. As distorções para cada componente e para os dois grupos, pode ser observada na

Figura 52.

Figura 52 - Testes do cotovelo para ambos os grupos.



Fonte: O Autor, 2023.

Ao analisar os gráficos de distorções, nota-se imediatamente que, para nenhum dos dois casos, existe um ponto de inflexão, a partir do qual a distorção passa a reduzir em menor magnitude, ou seja, não há o esperado cotovelo. Entretanto, os eixos de distorções são independentes, para o Grupo 01, as distorções variaram entre aproximadamente $4,5 \times 10^6$ e $7,5 \times 10^6$, para o Grupo 02, estas variaram entre $1,5 \times 10^6$ e $3,25 \times 10^6$. Como a maior diferença entre os dois testes é o número de atributos, pode-se inferir que, quanto maior a dimensionalidade, pior o resultado do algoritmo.

Como o teste do cotovelo foi inconclusivo, se optou por executar o K-Means em um intervalo de classes, entre 3 e 10 e avaliar visualmente seu resultado. Como o objetivo é fazer uma classificação automática de fácies sísmicas e como neste curto intervalo de cálculo não ocorrem tantas fácies, usar números maiores não faria sentido.

No primeiro teste, a distorção foi calculada realizando agrupamentos de 3 a 15 classes, com intervalo 2. É fácil notar que, nesse intervalo, não ocorre uma queda brusca na distorção, como se o hipotético braço não estivesse dobrado. Portanto, um segundo teste foi realizado de 2 a 30 classes, mais uma vez com intervalo de duas classes, para acelerar o processo. Mais uma vez, o resultado foi inconclusivo, colocar um número alto de classes pode tornar o resultado não interpretável.

A Figura 53 mostra o mapa de amplitude instantânea ao longo do horizonte da base do sal e as linhas referentes aos poços. As linhas horizontais vermelhas nas seções indicam a posição da camada de rochas ígneas nos poços, representada pela linha vertical preta. Além disso, as Figuras Figura 54 a Figura 57 exibem os resultados dos testes de K-means com 7 a 10 componentes.

Na Figura 53, observa-se que a polaridade nas linhas referentes aos poços 3-BRSA-1356D-RJS e 3-BRSA-1343-RJS é negativa (representada pela cor branca), enquanto na linha referente ao poço 3-BRSA-1305A-RJS ocorre uma variação de polaridade positiva, seguida por uma camada delgada de polaridade negativa e novamente polaridade positiva ao fim do intervalo. Essa variação provavelmente ocorre devido ao efeito de tuning entre as camadas, o que pode influenciar os resultados do agrupamento.

Cada figura apresenta o resultado para o mesmo número de componentes, com o Grupo 01 exibido acima e o Grupo 02 abaixo. Não foram mostrados outros resultados com valores de distorção maiores e que extrapolam as expectativas para as fácies sísmicas no intervalo de estudo. A análise concentra-se nos resultados obtidos com sete a dez componentes. Os mapas em cada figura foram calculados a partir do mapa de base do sal, com uma janela de 10 amostras (50 metros) para baixo. O cluster vencedor foi determinado pela mediana na janela, ou seja, o cluster que possuía mais amostras no intervalo. É importante ressaltar que os resultados do K-means não compartilham os mesmos rótulos nas diferentes rodadas, portanto, as cores representativas são iguais apenas dentro de cada figura individual. Essa variação nos rótulos pode gerar confusões na interpretação dos resultados. Como discutido nos capítulos anteriores, as soleiras não estão dispostas ao longo de toda a área, são corpos limitados geralmente bastante delgados, portanto, se espera que o cluster que as representa, ou seja, ele não pode estar distribuído ao longo de toda a seção ou todo o mapa.

Os resultados do Grupo 02 exibem clusters com alta continuidade lateral, destacando os refletores mais contínuos e diferenciando-os com base em sua polaridade. Alguns clusters menores representam regiões onde os refletores são menos contínuos ou representam ruídos nos dados. Vale ressaltar que, nesse grupo, o dado de amplitude não foi utilizado.

Uma exceção no Grupo 02 é o resultado com 10 componentes, o qual apresenta mais clusters destacando ruídos do que clusters contínuos. No entanto, o cluster azul, mais contínuo, prevalece, especialmente na zona de interesse. Apesar de apresentar distorções menores, os resultados do Grupo 02 mostram que, para este objetivo, ainda que os atributos tenham sido escolhidos de forma dirigida, sua pequena quantidade dificulta a diferenciação dos eventos e reitera a alta redundância dos dados sísmicos e seus atributos, tornando interessante o uso de mais atributos.

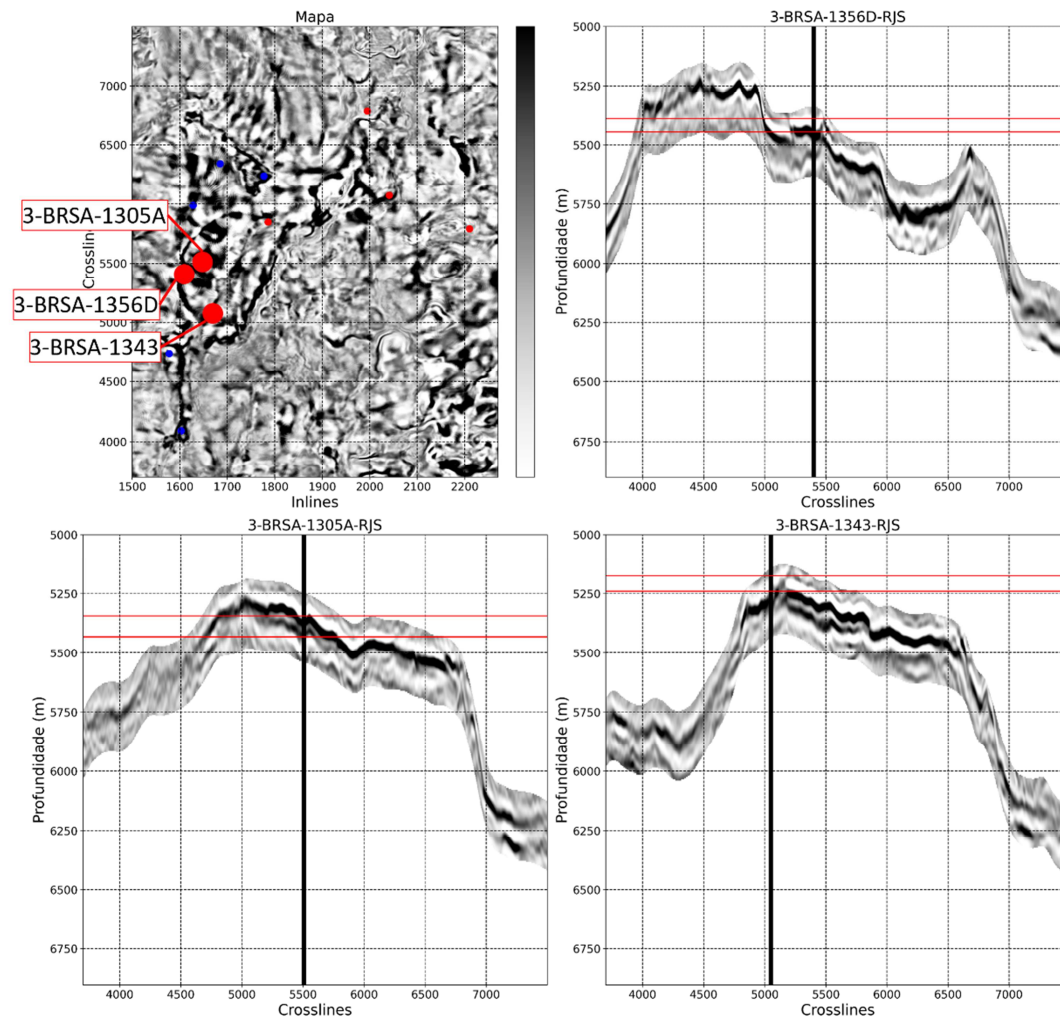
Os resultados do Grupo 01 mostram uma maior diferenciação em relação aos resultados do Grupo 02. No entanto, uma análise mais profunda revela que eles também são influenciados pelo dado de amplitude e geralmente correspondem a mudanças sutis na amplitude dos dados, apresentando clusters bastante delgados que acompanham os refletores.

Os resultados do Grupo 01 para 9 e 10 componentes exibem clusters com algumas características que lembram as esperadas para intrusões ígneas, para o grupo 01, o cluster azul escuro e para o grupo 02, o cluster azul claro. Porém, particularmente no resultado de 10 componentes, o cluster azul claro é muito distribuído ao longo da área de estudo. Neste caso, o uso da mediana na janela de cálculo pode não ser o ideal, uma vez que nas seções, esta distribuição parece ser muito mais limitada. No grupo 01, o cluster azul escuro é evidentemente menos distribuído, porém ele se repete em outras profundidades. É possível que um refinamento na aplicação do K-Means, utilizando uma abordagem controlada por amostras vizinhas, diminua esses erros. Entretanto, sua aplicação mostrou-se muito desafiadora e além do escopo da proposta deste estudo.

Essas observações destacam o desafio da interpretabilidade e ressaltam a importância de uma análise cuidadosa e crítica dos resultados obtidos por meio de técnicas de classificação não supervisionada, levando em consideração as particularidades geológicas da área de estudo e os desafios na interpretação dos resultados. dos resultados da classificação não supervisionada. Por exemplo, a presença de camadas muito delgadas, como soleiras, pode ser erroneamente inferida a partir de clusters que representam camadas delgadas e menos contínuas. No entanto, ao comparar os clusters em diferentes linhas e considerando a localização das camadas de interesse com base nos poços, fica claro que o mesmo cluster não ocorre em nenhuma das vezes. Essa diferença pode ser atribuída a desalinhamentos entre o

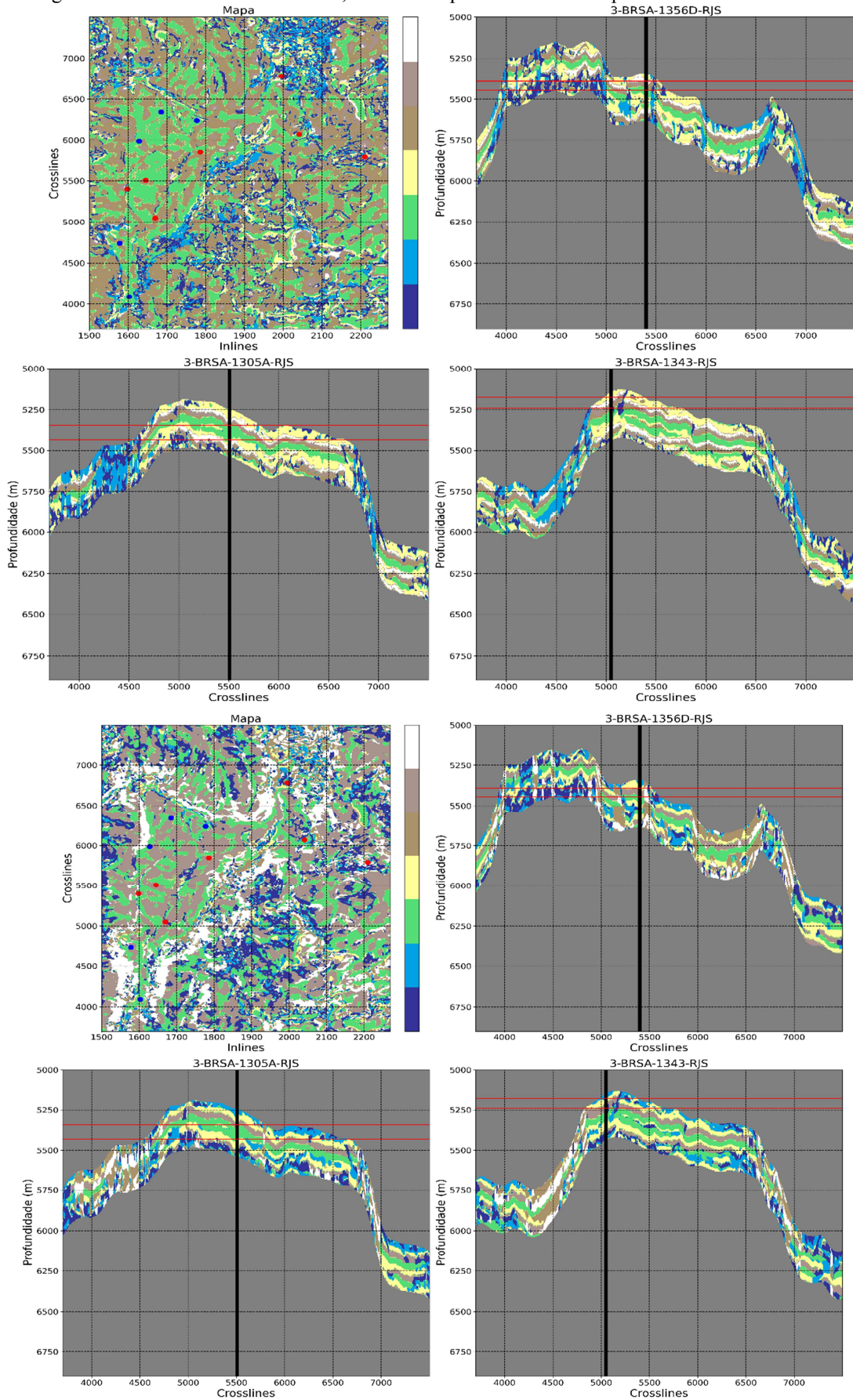
dado original e a geologia naquela profundidade específica, que não são incomuns na ordem de dezenas a centenas de metros. Em geral, nenhum algoritmo de aprendizado fornece um resultado 100% bem-sucedido, portanto, não se deve esperar resultados perfeitos para o K-Means.

Figura 53 - Mapa e Linhas do volume de amplitudes



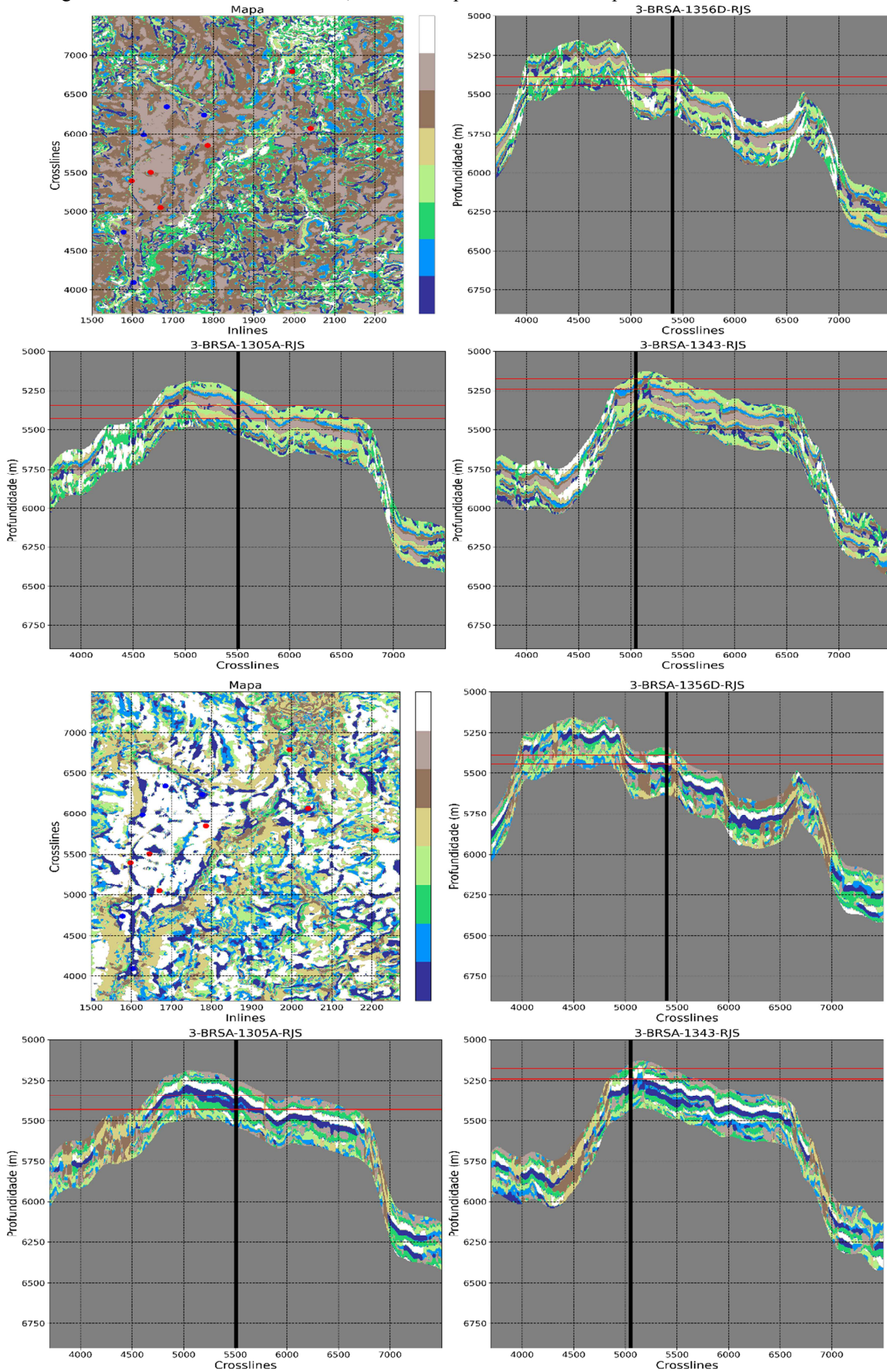
Fonte: O Autor, 2023

Figura 54 - Resultado K-Means K=7, Acima: Grupo 01. Abaixo: Grupo 02



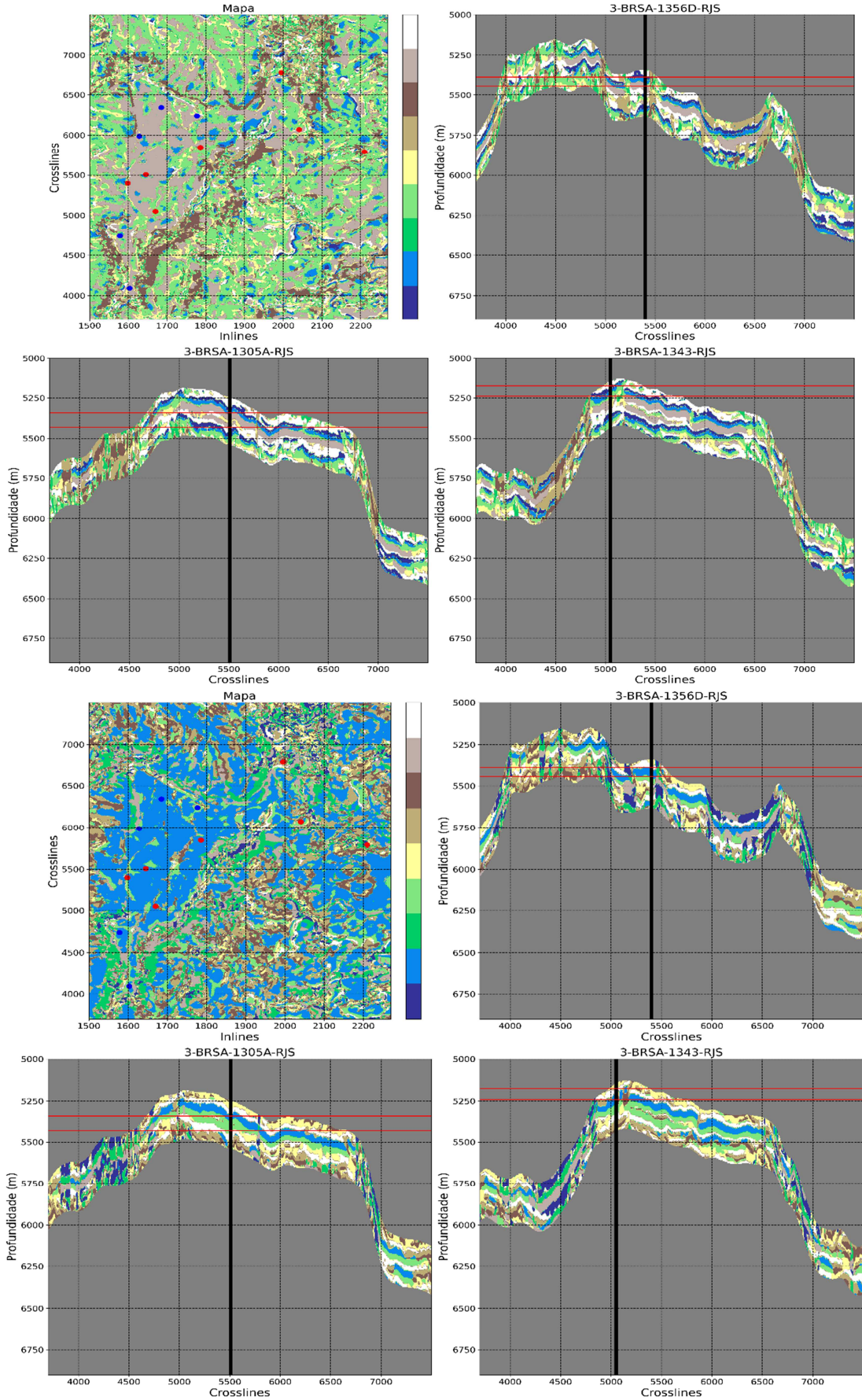
Fonte: O Autor, 2023.

Figura 55 - Resultado K-Means K=8, Acima: Grupo 01. Abaixo: Grupo 02



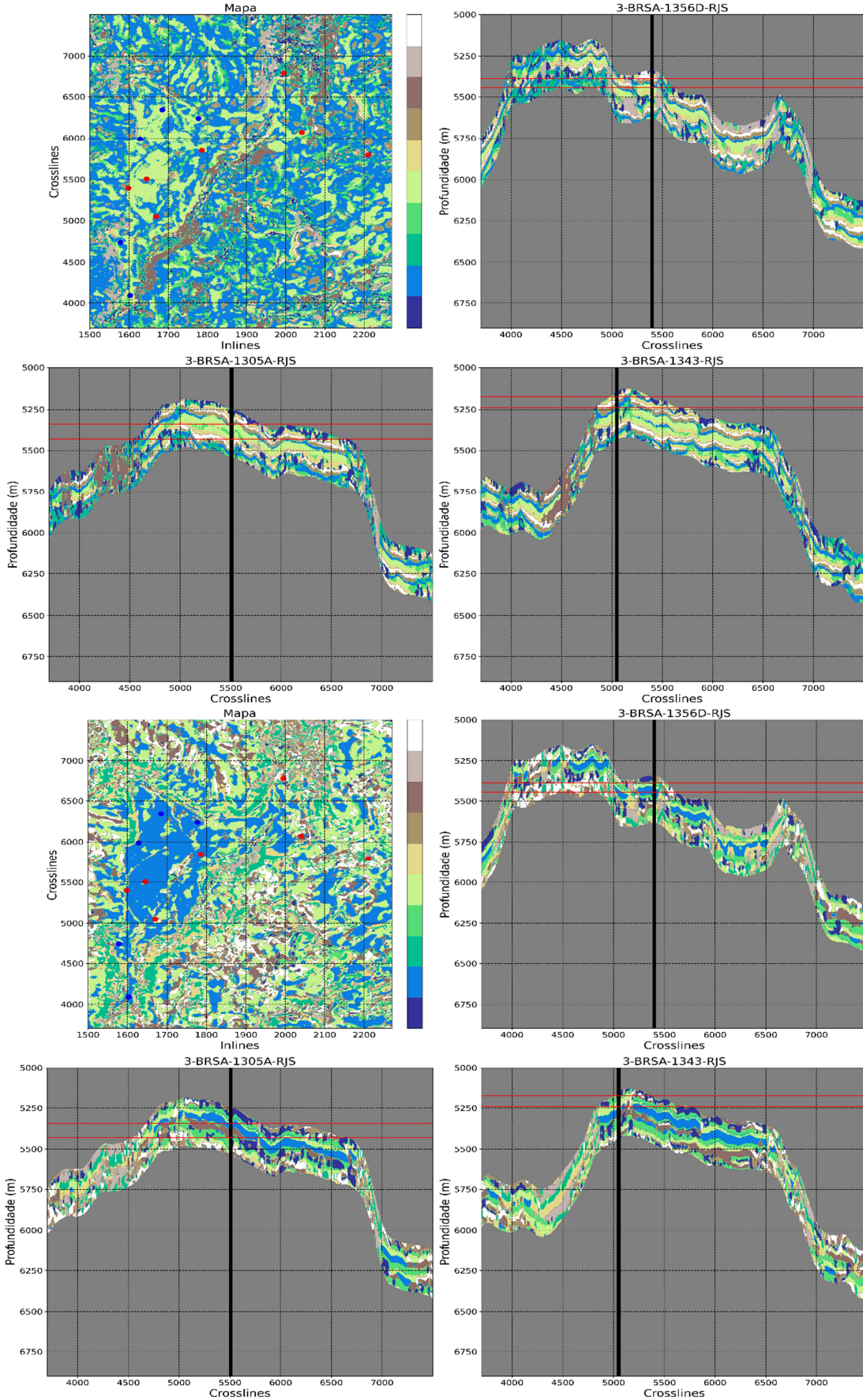
Fonte: O Autor, 2023.

Figura 56 - Resultado Means K=9, Acima: Grupo 01. Abaixo: Grupo 02



Fonte: O Autor, 2023.

Figura 57 - Resultado K-Means K=10. Acima: Grupo 01. Abaixo: Grupo 02



Fonte: O Autor, 2023

3.3.2 Modelos de Mistura Gaussiana

A aplicação dos modelos de mistura gaussiana na classificação automática de fácies sísmicas é bastante limitada, HAN et al. (2011) aplicam a técnica em dados sintéticos e apenas WALLET; HARDISTY (2019) mostram resultados em dados reais, identificando canais em sistemas turbidíticos na Bacia de Canterbury, Nova Zelândia, associando os GMMs à técnica SOM. HAN et al. (2011) defendem a aplicação da técnica assumindo que os dados sísmicos de cada fácies geológica seguem uma distribuição de probabilidade gaussiana, portanto, uma região qualquer de um dado sísmico, composto por diversas fácies, seria um exemplo de modelo de mistura gaussiana.

Ambos os trabalhos, no entanto, utilizam implementações adaptadas do algoritmo. Aqui, mostro os resultados da técnica de modo a ilustrar seu potencial como classificador automático de fácies, usando apenas a implementação simplificada pública, disponibilizada através da biblioteca Scikit-Learn.

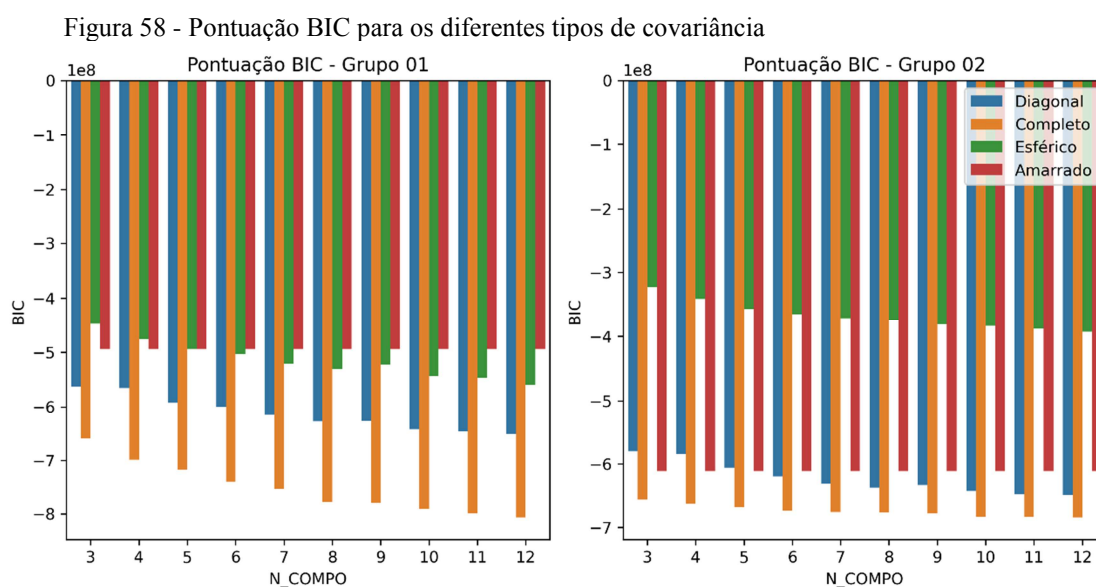
Devido à natureza do dado sísmico e à alta dimensionalidade inerente ao processo, é natural concluir que, neste caso o tipo de covariância deve ser completo, uma vez que se espera que os atributos possuam diversas distribuições. Entretanto, é possível avaliar quantitativamente a capacidade do algoritmo para cada tipo de covariância no dado a partir da métrica de Critério de Informação Bayesiana (“Bayesian information criterion”, BIC), o qual, em teoria, penaliza os modelos que possuem mais parâmetros, ou seja, mais clusters. A pontuação BIC é influenciada tanto pela probabilidade do dado se encaixar ao modelo como pelo seu número de parâmetros. Quanto menor seu valor, melhor o encaixe.

A Figura 58 mostra os resultados de pontuação BIC para os dois grupos. Todos os resultados negativos são bem distintos dos apresentados por WALLET; HARDISTY (2019), onde a pontuação é negativa para até 2 clusters e tende a zero ou levemente positiva com o aumento do número de clusters. Esta discrepância pode estar relacionada à implementação da biblioteca utilizada, como já apontado em seção anterior. Como GERON (2022) descreve a implementação da biblioteca Scikit-Learn, aqui utilizo como melhor estimativa o menor valor.

No Grupo 01, o modo amarrado rapidamente estabiliza, durante o processo este tipo de covariância consistentemente completava mais rapidamente, o que pode ser um indicativo de que o algoritmo estava convergindo para uma solução local, não ótima. Os outros tipos, no entanto, mostram persistente diminuição com o aumento de componentes, o que vai contra a teoria de que a métrica penaliza o aumento de parâmetros. Aumentar indiscriminadamente o

número de componentes não é recomendável, uma vez que o custo computacional aumenta consideravelmente a cada componente extra.

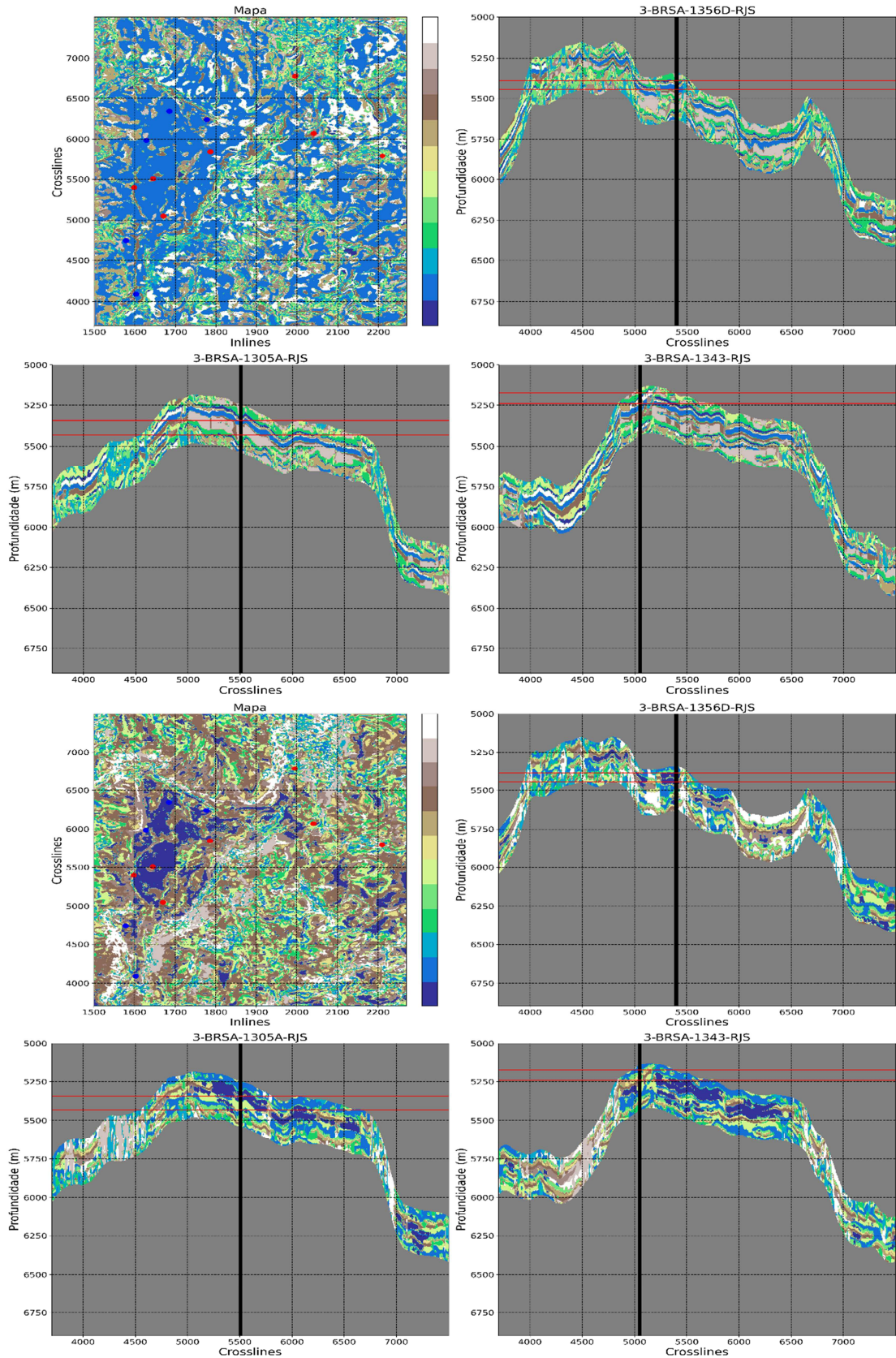
O Grupo 02 apresenta resultados muito distintos do Grupo 01, com valores BIC para tipo de covariância amarrado mais próximos do tipo completo, entretanto são estáveis para todos os números de componentes, reiterando a possibilidade de que a convergência tenha se dado em uma solução local. Assim como o outro grupo, este apresenta menores pontuações para o tipo completo.



Fonte: O Autor, 2023.

Para ilustrar os resultados, foram selecionados 12 componentes e tipos de variância diagonal e completa, uma vez que possuem os menores resultados de BIC. A Figura 59 mostra os resultados para o Grupo 01 e, apesar de não parecer haver classes correlacionadas apenas à presença das soleiras, o potencial do algoritmo como classificador automático de fácies sísmicas é notável, com a covariância completa apresentando resultados superiores quando comparada aos resultados da covariância diagonal. Esse resultado é esperado, uma vez que se espera que os atributos sísmicos possuem todo o tipo de distribuição. Ao mesmo tempo, os resultados não parecem acompanhar tanto as fácies sísmicas correspondentes apenas à amplitude, como no caso do K-Means, o GMM parece ter potencial de destacar anomalias sísmicas.

Figura 59 - Resultados para GMM 12 componentes: Covariância Completa (acima); Covariância Diagonal (abaixo)



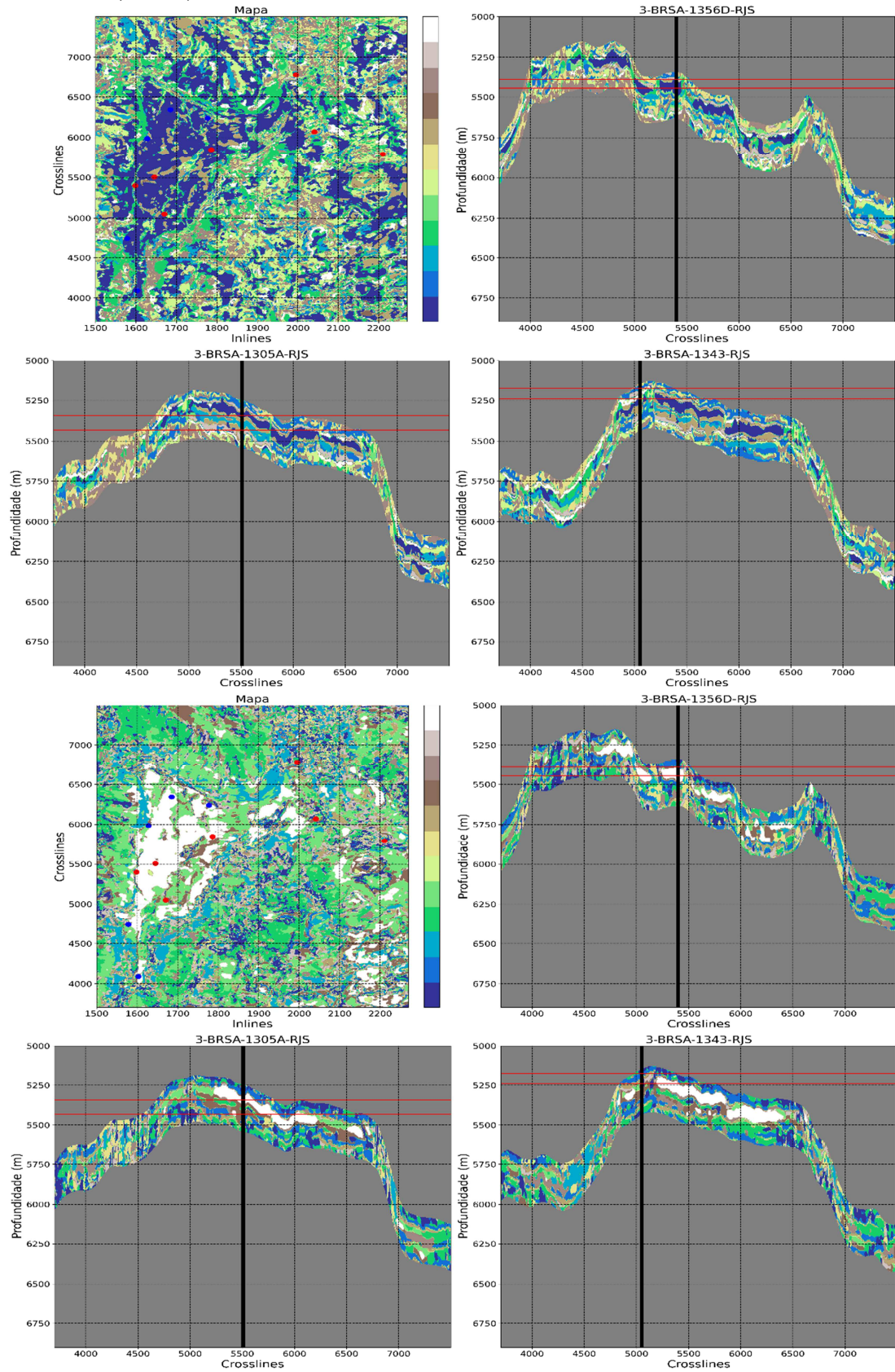
Fonte: O Autor, 2023.

A Figura 60 apresenta os resultados para o Grupo 02. Mais uma vez em contraste com o resultado do K-Means, a classificação de fácies se mostrou muito distinta, não ocorre a atribuição de classes contínuas ao longo de todo o dado, pelo contrário, apenas uma classe possui este aspecto em ambos os resultados, de covariância completa e diagonal, enquanto outras classes se mostram bastante ruidosas. Mais uma vez o resultado parece indicar potencial de identificação de anomalias para o algoritmo.

Em contraste aos resultados do Grupo 01, não parece haver muita diferença entre o resultado de covariância completa e diagonal, é provável que isto seja resultado da dimensionalidade reduzida influenciando a performance da covariância diagonal, o que indicaria que quando aplicado a dados de baixa dimensionalidade, a parametrização pode ser feita de modo a priorizar performance, ou seja, tipos de covariância que chegam à convergência mais rapidamente podem ser utilizados.

Os resultados indicam que o algoritmo tem alto potencial para a classificação automática de fácies sísmicas, apesar de possuir maior sensibilidade a ruídos. Esta sensibilidade pode, no entanto, ser empregada na busca por anomalias.

Figura 60 - Resultado GMM 12 componentes Covariância Completa (Acima) Covariância Diagonal (Abaixo)




Fonte: O Autor, 2023.

3.3.3 Mapas Auto-organizáveis (SOMs)

Em contraste com os métodos K-Means e GMMs, a técnica SOM possui parametrização mais extensa e com resultados muito distintos entre si. A técnica cria um mapa do dado multidimensional em um espaço com menos dimensões, em geral, apenas duas. O tamanho desse mapa determina o número de possíveis grupos a serem encontrados durante a classificação. Este é, portanto, o primeiro parâmetro a ser escolhido. Um mapa de tamanho 5 x 5, terá até 25 neurônios, ou seja, até 25 grupos, porém nem todos os neurônios costumam ter amostras denominadas vencedoras.

Cada neurônio do mapa possui um vetor de pesos, um grupo de valores (ou pesos) que representam a posição do nó no espaço de entrada, sua dimensionalidade depende do número de dimensões do dado de entrada. Esse vetor é atualizado automaticamente durante a etapa de treinamento, que consiste nas iterações realizadas de modo a encontrar a melhor representação em relação ao dado de entrada, a partir da similaridade (ou distância) entre aquele ponto no dado de entrada e os vetores de peso de todos os nós. As atualizações dos vetores de peso dependem de dois parâmetros: sigma e taxa de aprendizado.

A escolha dos valores de sigma e da taxa de aprendizado é uma etapa crucial para a otimização dos resultados. Sigma representa o tamanho da vizinhança em torno do neurônio vencedor e determina a extensão da influência que os nós vizinhos possuem sobre a atualização dos pesos. A taxa de aprendizado determina a magnitude das atualizações dos pesos, determina o quanto os vetores de peso do neurônio vencedor e seus vizinhos serão ajustados.

Cada iteração durante o processo de treinamento é chamada de época, para cada época pode ser avaliada a qualidade do mapa, a partir de dois indicadores: a qualidade do aprendizado e a qualidade da projeção. A qualidade do aprendizado é determinada a partir do erro de quantização, a qualidade de projeção é definida pelo erro topográfico. A partir do gráfico da  Figura 61, o ponto onde os erros estabilizam passa a ser a quantidade de épocas de treinamento aplicadas na seleção dos neurônios vencedores, cujos resultados serão observados a seguir. A partir dos resultados dos testes de erros de quantização e erro topográfico, foram selecionados como melhores valores de parametrização, dispostos no

Para o Grupo 02, não foram testados valores de épocas acima de 500 para evitar a tendência de superestimação dos resultados. Valores ainda maiores para o tamanho da grade

também poderiam resultar em estimações incorretas, uma vez que não se espera uma altíssima quantidade de rótulos de modo a também manter a interpretabilidade dos resultados.

Quadro 7: Parametrização SOM para cada grupo de atributos

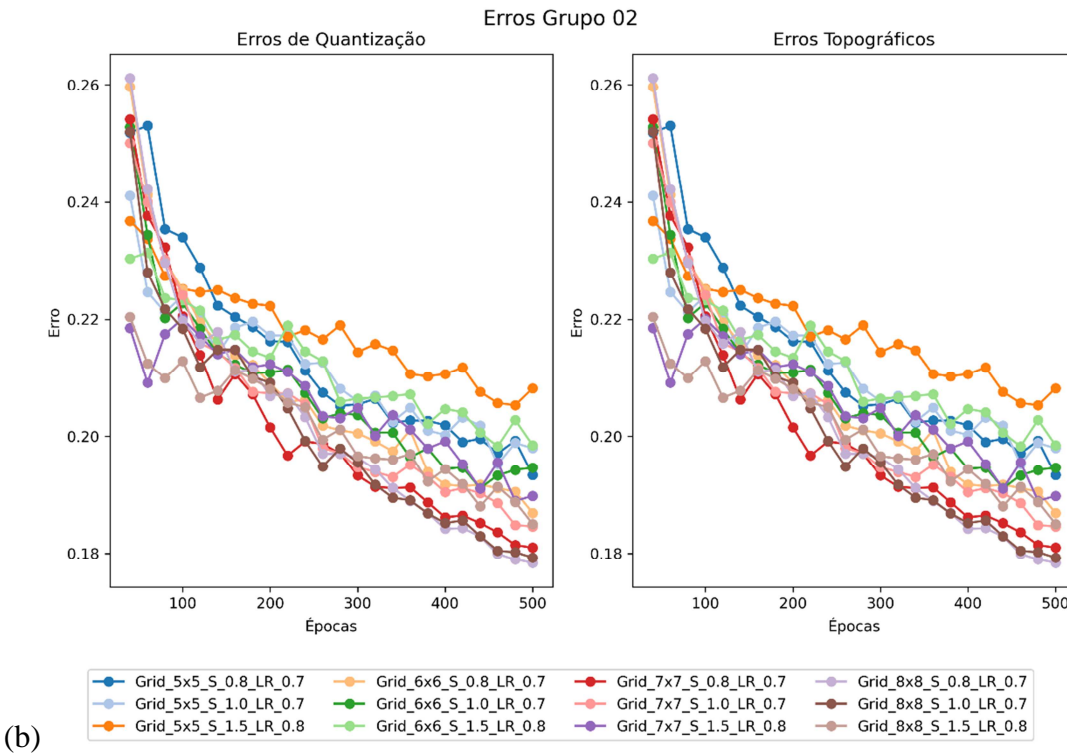
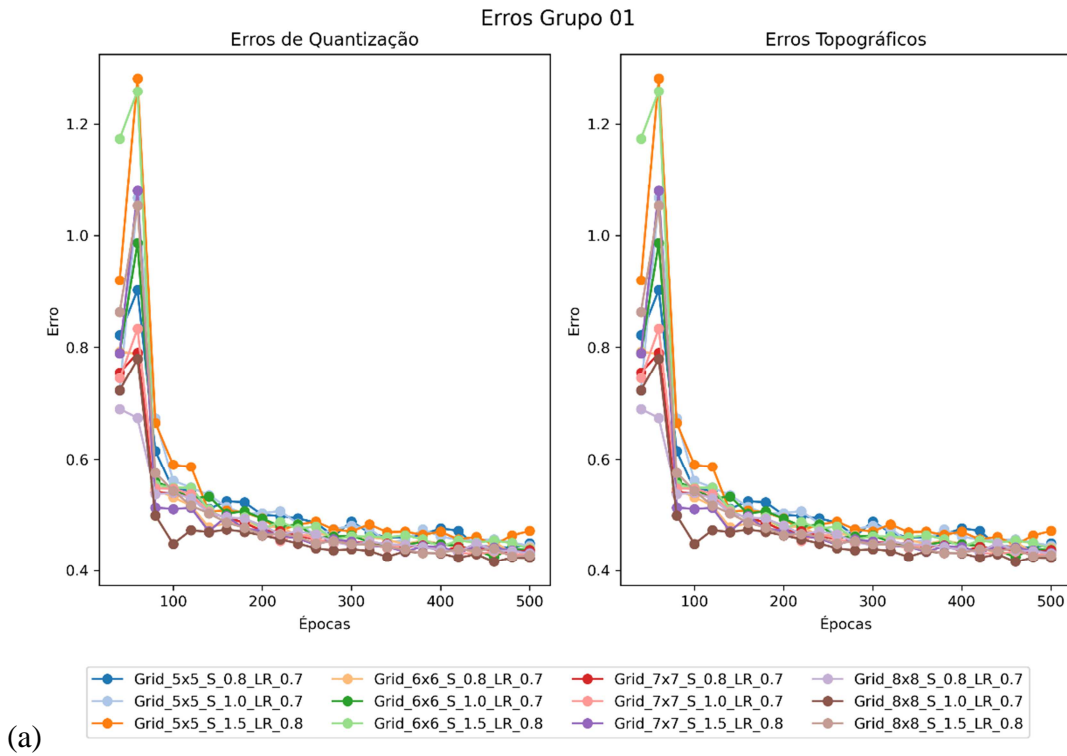
Grupo 01	Épocas (número de iterações de treinamento): 100
	Tamanho da grade: 8x8 (64 nós)
	Sigma (tamanho da vizinhança): 1,0
	Taxa de Aprendizado: 0,7
Grupo 02	Épocas (número de iterações de treinamento): 500
	Tamanho da grade: 8x8
	Sigma (tamanho da vizinhança): 0,8
	Taxa de Aprendizado: 0,7

Fonte: O Autor, 2023.

A Figura 62 mostra os resultados do algoritmo utilizando as parametrizações selecionadas pelos testes de EQ e ET. Apesar do mapa não destacar as soleiras em toda sua extensão, as seções mostram resultados muito próximos ao que se espera para as soleiras, com pouca extensão lateral e até mesmo vertical. Nas seções, as soleiras parecem estar exibidas a partir do cluster de cor amarela, de aspecto muito delgado e sem apresentar continuidade lateral, sendo restrito a apenas alguns pontos, em particular na seção do poço 3-BRSA-1356D-RJS. Na seção do poço 3-BRSA-1305A-RJS, próximo ao ponto em que ocorre o topo da soleira no poço, ocorre o mesmo cluster de coloração amarela, ainda que este não tenha a espessura real observada pelo poço. Por fim, na seção do poço 3-BRSA-1343-RJS, uma inspeção mais atenta permite observar, bem próximo à linha vertical do poço, ocorrência extremamente delgada e limitada desse cluster.

Como mencionado anteriormente, não se espera que os métodos de agrupamento tenham aproveitamento absoluto, no sentido de que é muito improvável que o algoritmo destaque exatamente a disposição da camada litológica, com sua extensão lateral e vertical, entretanto existe o potencial de indicar sua presença. Em termos de caráter geológico, portanto, este é o melhor resultado visto até o momento.

Figura 61 - Erros de Quantização e Topográficos para o Grupo 01 (a) e Grupo 02 (b)



Fonte: O Autor, 2023.

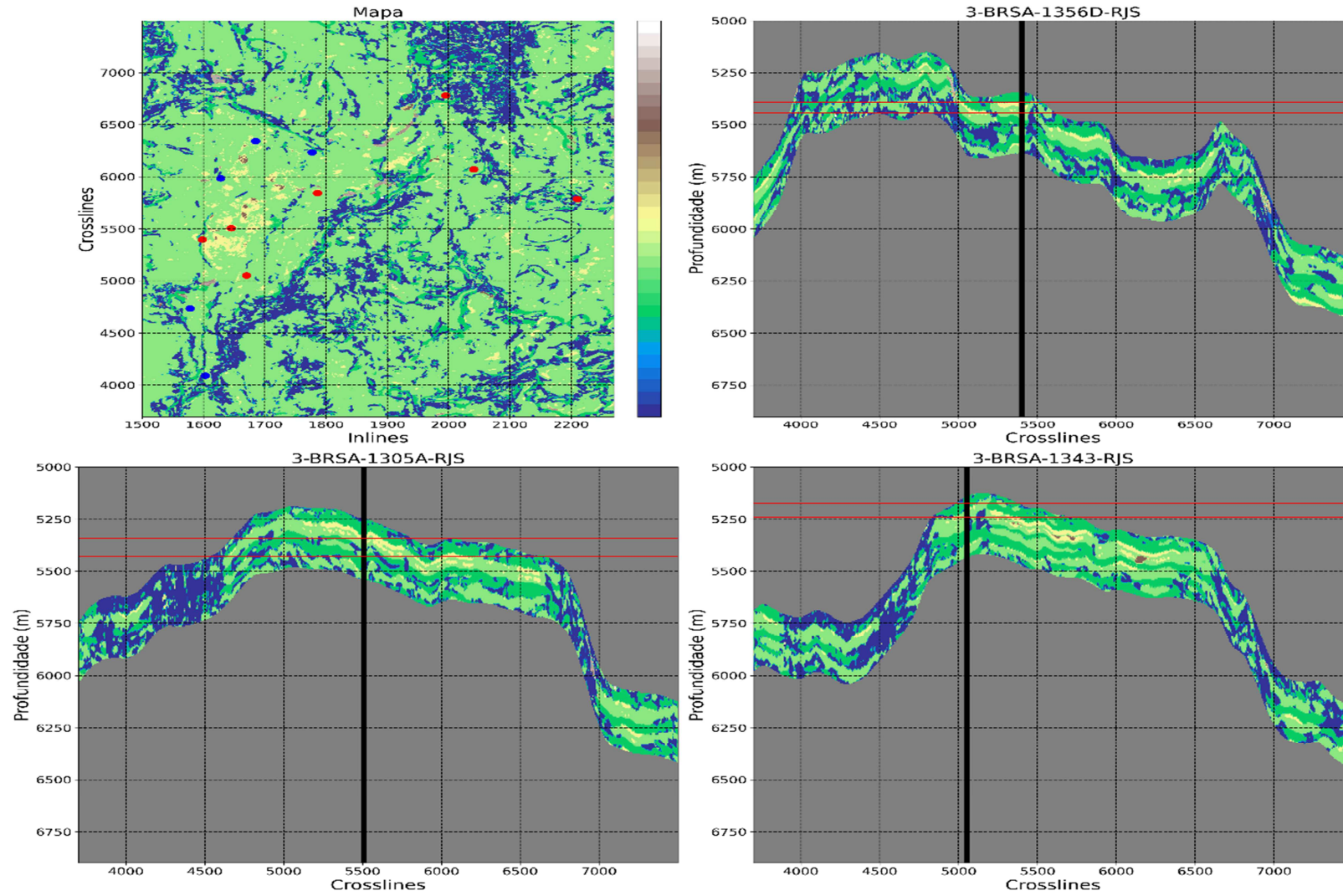
A Figura 63 exibe o resultado de outro teste com parametrização similar, com mesmo tamanho de grande, mesmo número de épocas, porém com sigma de valor 1,5 e taxa de aprendizado 0,8. Seu objetivo é ilustrar as diferenças e o impacto que pequenas

alterações na parametrização resultam no resultado do algoritmo, reiterando a importância da etapa de testes de erros. O cluster representado pela cor azul escura poderia ser interpretado como o referente à soleira, entretanto possui muito maior dispersão ao longo do dado e alta continuidade lateral.

A Figura 64 exibe o resultado para a melhor parametrização definida pelos testes de EQ e ET, para o Grupo 02, neste caso, 500 épocas, tamanho da grade 8x8, sigma 0,8 e taxa de aprendizado 0,7. O aumento considerável de iterações, quando comparado ao Grupo 01, resultou em uma alta homogeneização dos grupos, com classes principais muito contínuas, principalmente quando correlatas aos refletores originais, este resultado também pode ser atribuído, no entanto, à pequena quantidade de atributos e ao fato de estes não terem passado por etapas de redução de dimensionalidade. A Figura 65, mostra algumas classes contínuas enquanto outras parecem destacar ruídos, ou zonas de fácies sísmicas caóticas, porém, sem a alta homogeneização observada para o resultado definido como ótimo pelos testes de EQ e ET.

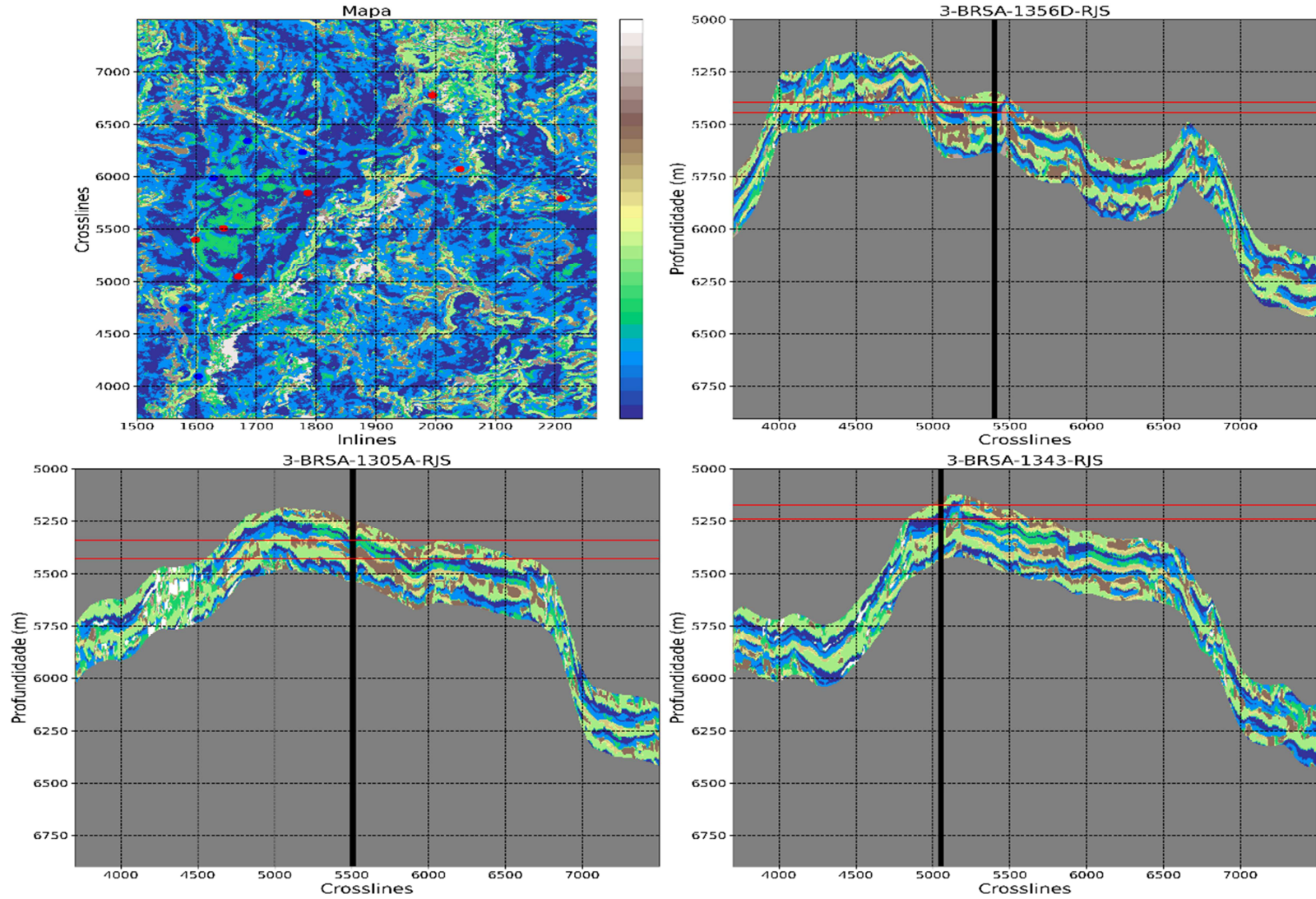
Os resultados do algoritmo SOM reiteram a popularidade deste, uma vez que apresentou o melhor resultado entre os três algoritmos, entretanto, também demonstram a importância da sua correta parametrização e da importância na escolha dos dados de entrada.

Figura 62 - Resultados do algoritmo SOM do Grupo 01 para os parâmetros selecionados a partir dos testes de EQ e ET



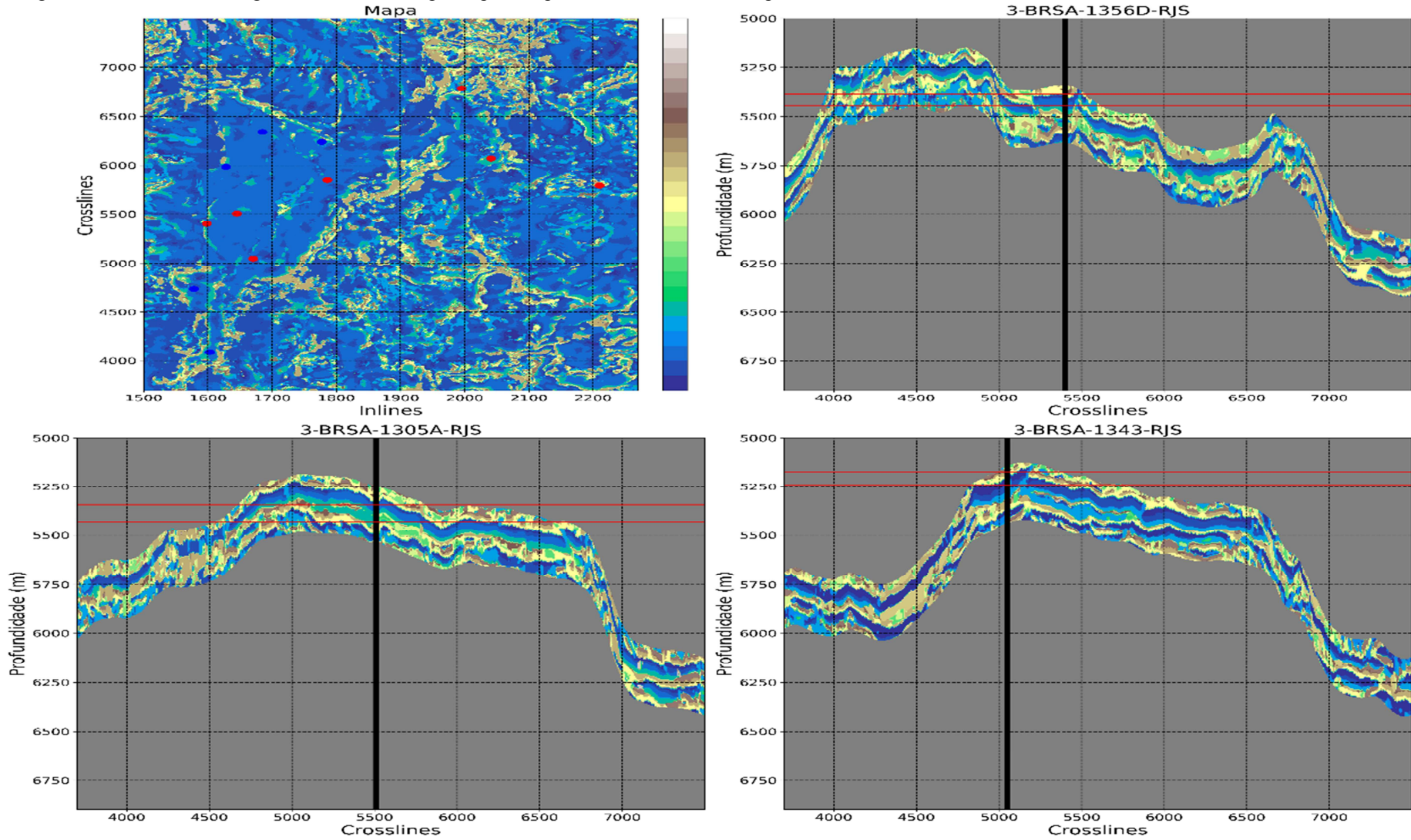
Legenda: Grade 8x8, Sigma 1,0 e Taxa de Aprendizagem 0,7.
 Fonte: O Autor, 2023

Figura 63 - Resultado do algoritmo SOM do Grupo 01 para os parâmetros não oriundos dos testes de EQ e ET



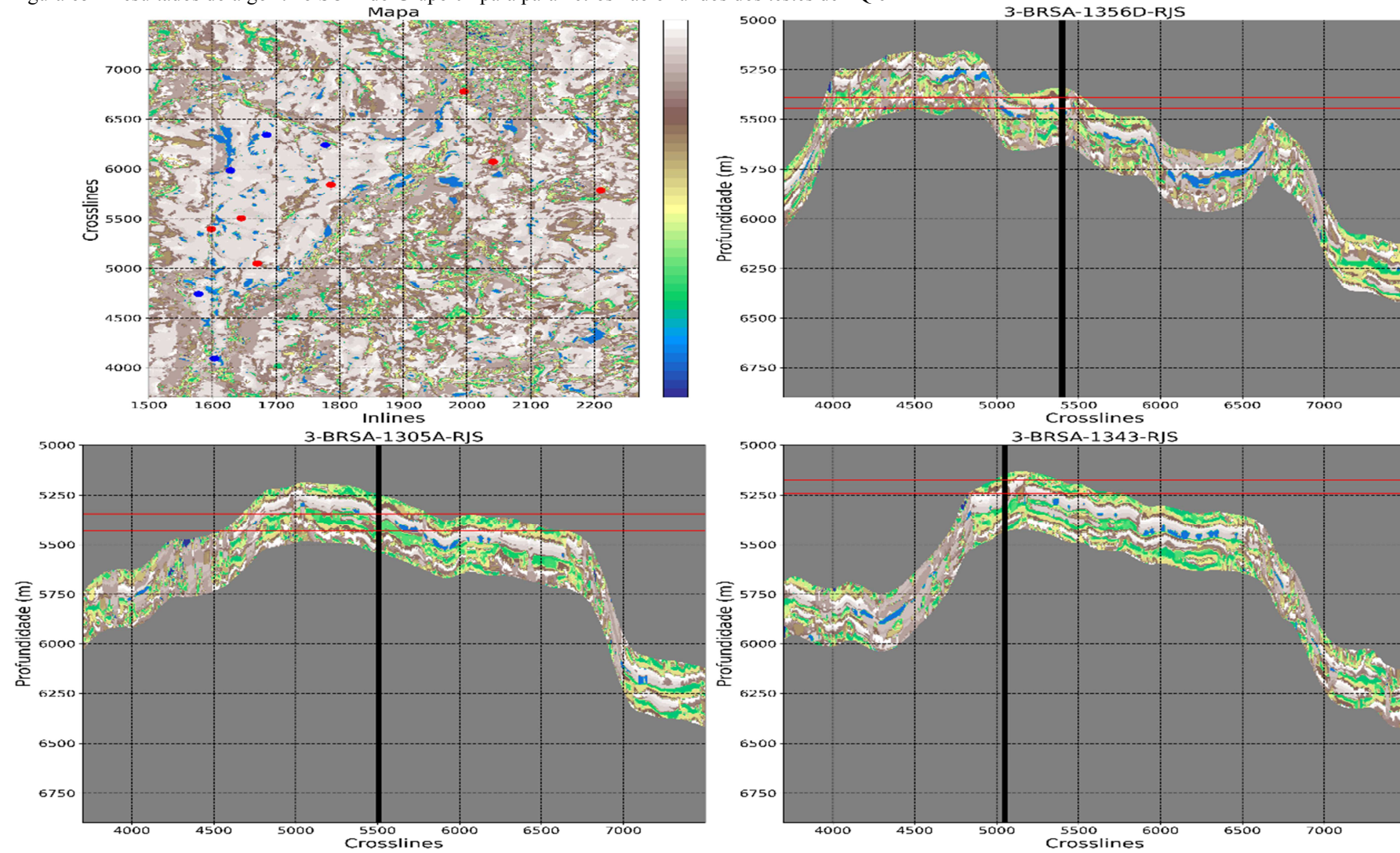
Legenda: Parâmetros: 100 épocas, Grade 8x8, Sigma 1,5 e Taxa de Aprendizado 0,8
 Fonte: O Autor, 2023

Figura 64 - Resultados do algoritmo SOM do Grupo 02 para os parâmetros selecionados a partir dos testes de EQ e ET:



Legenda: Grade 8x8, Sigma 0,8 e Taxa de Aprendizado 0,7.
 Fonte: O Autor, 2023.

Figura 65 - Resultados do algoritmo SOM do Grupo 02 para parâmetros não oriundos dos testes de EQ e ET



Legenda: Grade 8x8, Sigma 1,0 e Taxa de Aprendizagem 0,7.

Fonte: O Autor, 2023.

CONSIDERAÇÕES FINAIS

Inúmeras bacias sedimentares apresentam a ocorrência de rochas ígneas intercaladas com suas camadas sedimentares. Foram apresentados diversos exemplos de afloramentos, dados sísmicos e poços, em diferentes bacias sedimentares com o intuito de ilustrar e caracterizar as diferentes ocorrências ígneas, particularmente intrusões, mas também alguns exemplos relacionados a vulcanismo extrusivo, de modo a expandir o conhecimento a respeito de sua disposição geométrica e das fácies sísmicas características das intrusões e suas rochas encaixantes, além de destacar desafios para a imagem sísmica e os principais perfis de poços utilizados na caracterização de intrusões ígneas.

Em geral, a característica que permite maior distinção desses elementos com as camadas a seu entorno é a alta amplitude gerada pelo elevado coeficiente de reflexão. Entretanto, o sinal sísmico pode sofrer perdas e as rochas encaixantes podem possuir características petrofísicas similares às das intrusões ígneas, diminuindo a capacidade de identificação das soleiras.

Os diversos trabalhos indicam certa homogeneidade na ocorrência das soleiras. As intrudidas em sedimentos relativamente rasos costumam possuir capacidade de intrusão e transgressão das camadas sedimentares, culminando com o típico formado de pires (“saucer-shape”), enquanto as intrudidas em sedimentos mais profundos possuem a capacidade de transgressão dos sedimentos muito limitadas, devido às altas pressões exercidas pelas camadas sobrepostas.

Em geral, a característica que permite maior distinção desses elementos com as camadas em seu entorno é a alta amplitude gerada pelo elevado coeficiente de reflexão. Entretanto, o sinal sísmico pode sofrer perdas e as rochas encaixantes podem possuir características petrofísicas similares às das intrusões ígneas, diminuindo a capacidade de identificação das soleiras.

A detecção e visualização das intrusões ígneas em grandes profundidades é, portanto, um desafio ao intérprete e vários fatores devem ser levados em consideração: a seção sedimentar sobreposta, as rochas encaixantes, a profundidade total, o conteúdo de frequências do dado sísmico, as técnicas de construção do modelo de velocidade e os algoritmos de migração utilizados. Modelagens sísmicas mostram que camadas delgadas podem causar interferência sísmica suficiente para indicar a presença de intrusões, ainda que uma reflexão específica não seja observada.

A análise petrofísica de poços permite descrever intrusões e, recentemente, o advento dos perfis de imagem possibilita, inclusive, a caracterização de derrames basálticos e o potencial dessas camadas como reservatórios. Perfis clássicos, como o cáliper, auxiliam na estimação de fraturas, enquanto perfis sônico e de densidade contribuem para modelagens sísmicas mais realísticas e, por consequência, de resultados mais confiáveis.

As intrusões ocorridas na base do sal, associadas ao magmatismo Santoniano-Campaniano (80 Ma) são muito delgadas e se confundem com as reflexões dos sedimentos encaixados. A modelagem sísmica realizada com dados de poços da área de Mero, mostra que rochas intrusivas com 84 metros de espessura já são capazes de apresentarem reflexão sísmica própria, mesmo estando encaixadas em rochas com alta impedância acústica e em grande profundidade. Estudos sugerem ainda que, mesmo camadas abaixo desse limite podem gerar interferência sísmica suficiente para indicar sua presença.

Neste trabalho, foram empregados três algoritmos, K-Means, GMMs e SOMs, para realizar a classificação automática de fácies sísmicas. Os dois principais objetivos foram aplicar técnicas de agrupamento com o intuito de encontrar fácies sísmicas correlatas à presença de soleiras ígneas e realizar todo o processo sem utilizar técnicas avançadas de processamento e inversão, de modo que um geocientista possa executar a tarefa diretamente em sua estação de trabalho.

A área de estudo apresentou soleiras com espessuras variando de alguns metros a aproximadamente 100 metros. Modelagens indicaram que apenas soleiras com mais de 84 metros poderiam ser observadas diretamente. No entanto, estudos indicam que mesmo soleiras com espessura abaixo desse limite, conhecidas como subsísmicas, podem causar interferência sísmica suficiente para serem observadas em alguns atributos.

Inicialmente, foram selecionados 28 atributos de diferentes tipos. Esses atributos foram utilizados nas etapas de redução de dimensionalidade, por meio da seleção de facetas e extração de facetas. A técnica de seleção de facetas utilizada foi a correlação de Spearman, juntamente com a observação qualitativa dos atributos, resultando na exclusão de 13 atributos. Já a técnica de extração de facetas aplicada foi a Análise de Componente Principal (PCA), reduzindo o número de atributos de 15 para 10.

Além dos atributos selecionados pelas técnicas de redução de dimensionalidade, foi escolhido um segundo grupo de atributos com base na observação dos limites da soleira investigada nos mapas dos atributos extraídos do horizonte de base de sal. Esse grupo de atributos foi denominado Grupo 02. O Grupo 01 é, portanto, o grupo de atributos definidos pelas etapas de redução de dimensionalidade.

Os grupos de atributos selecionados foram utilizados como entrada para os algoritmos de agrupamento. O algoritmo K-Means mostrou resultados inconclusivos ao utilizar o teste do cotovelo para determinar o número de classes. No entanto, em conjunto com algoritmos que utilizem o conceito de vizinhos próximos, o K-Means pode apresentar melhores resultados. Observou-se que o Grupo 02 se mostrou altamente correlato à sísmica de amplitude, mesmo não sendo utilizado como dado de entrada. Isso indica o potencial do Grupo 02 quando aplicado em conjunto com algoritmos de interpretação automática ou para criação de “*geobodies*” em camadas geológicas com espessura suficiente para serem observadas.

O algoritmo GMM demonstrou ser muito sensível a ruídos, apresentando resultados significativamente diferentes do K-Means, especialmente quando comparados aos resultados do Grupo 02. Isso sugere que a técnica é ainda mais sensível aos dados de entrada selecionados. As métricas normalmente utilizadas para definir as parametrizações não foram conclusivas, assim como no K-Means. É importante ressaltar que esta é uma das primeiras aplicações da técnica na classificação automática de fácies sísmicas, e uma análise dedicada dos resultados e o desenvolvimento de métricas específicas para dados sísmicos podem melhorar consideravelmente os resultados obtidos com o GMM.

A técnica SOM foi empregada utilizando as métricas de Erros de Quantização e Erros Topográficos para determinar sua parametrização ideal para cada grupo. Foi observado um benefício em utilizar um script próprio em Python, em detrimento do uso de soluções em aplicativos comerciais, pois alguns desses aplicativos não disponibilizam uma parametrização extensa que facilite sua aplicação. No entanto, muitas vezes a técnica de parametrização empregada é desconhecida, o que pode justificar grandes diferenças de resultados entre diferentes aplicativos comerciais. Os SOMs demonstraram o maior potencial entre os algoritmos testados, resultando na classe mais próxima da esperada para as soleiras.

Os resultados também evidenciaram o desafio da interpretabilidade dos resultados das classificações não supervisionadas. Com o conhecimento a priori limitado, torna-se desafiador interpretar corretamente os resultados dessa técnica. As classificações não supervisionadas são altamente dependentes dos dados de entrada e das parametrizações, ao mesmo tempo em que faltam métricas confiáveis para a definição desses parâmetros. Em contraste, a métrica empregada na técnica SOM mostrou resultados mais satisfatórios.

Embora não tenha sido objetivo deste trabalho abordar métricas de custo computacional, pode-se mencionar que o K-Means é o algoritmo mais rápido, seguido pelo SOM. O GMM possui rápida convergência para os tipos de covariância esférico e amarrado, mas o tempo de processamento aumenta exponencialmente com o aumento do número de componentes. Em

qualquer caso, todos os algoritmos foram concluídos em questão de horas, mesmo considerando que os testes tenham sido realizados em um supercomputador dedicado. Portanto, uma estação de trabalho comum de um geocientista possui poder de processamento suficiente para realizar esses processos utilizando plugins entre o software de interpretação e o ambiente de processamento Python, já comuns nas principais soluções comerciais.

Os resultados obtidos indicam que é possível utilizar não apenas o SOM, mas também os outros algoritmos testados para a classificação automática de fácies. No entanto, devido à pequena espessura das soleiras e seu caráter subsísmico, a identificação das classes sísmicas referentes às mesmas foi bastante limitada.

Em suma, este estudo explorou o uso de algoritmos de agrupamento para a estimação automática de fácies sísmicas. Os resultados obtidos fornecem noções importantes sobre a aplicação desses algoritmos, suas limitações e seu potencial. A continuidade desse trabalho pode envolver a otimização das parametrizações, a utilização de algoritmos que levem em consideração a distribuição das amostras vizinhas, o desenvolvimento de métricas específicas para dados sísmicos e a aplicação dos algoritmos em outros conjuntos de dados, permitindo uma melhor compreensão das fácies sísmicas e seu contexto geológico.

REFERÊNCIAS

- AASPI. Disponível em: <<https://mcee.ou.edu/aaspi/documentation.html>>. Acesso em: 22 jun. 2023.
- AHMAD, I. **40 Algorithms Every Programmer Should Know**. S.l.: Packt Publishing, 2020.
- ALTMAN, N.; KRZYWINSKI, M. The curse(s) of dimensionality. **Nature Methods**, v. 15, n. 6, p. 399–400, jun. 2018.
- ALVES, T. et al. A tectono-stratigraphic review of continental breakup on intraplate continental margins and its impact on resultant hydrocarbon systems. **Marine and Petroleum Geology**, v. 117, p. 104341, jul. 2020.
- AMR, T. **Hands-on machine learning with scikit-learn and scientific Python toolkits: a practical guide to implementing supervised and unsupervised machine learning algorithms in Python**. Birmingham Mumbai: Packt, 2020.
- ANH TU, L. Improving Feature Map Quality of SOM Based on Adjusting the Neighborhood Function. Em: ALMUSAED, A.; ALMSSAD, A.; TRUONG - HONG, L. (Eds.). **Sustainability in Urban Planning and Design**. [s.l.] IntechOpen, 2020.
- ANP, A. N. DO P. G. N. E B. **Resultados — Português (Brasil)**. Disponível em: <<https://www.gov.br/anp/pt-br/rodadas-anp/rodadas-concluidas/partilha-de-producao/1a-rodada-partilha-producao-pre-sal/resultados>>. Acesso em: 19 jul. 2022.
- BARNES, A. E. Redundant and useless seismic attributes. **GEOPHYSICS**, v. 72, n. 3, p. P33–P38, maio 2007.
- BARNES, A. E. **Handbook of Poststack Seismic Attributes**. [s.l.] Society of Exploration Geophysicists, 2016.
- BARNES, A. E.; LAUGHLIN, K. J. **Investigation of methods for unsupervised classification of seismic data**. SEG Technical Program Expanded Abstracts 2002. **Anais...** Em: SEG TECHNICAL PROGRAM EXPANDED ABSTRACTS 2002. Society of Exploration Geophysicists, jan. 2002. Disponível em: <<http://library.seg.org/doi/abs/10.1190/1.1817152>>. Acesso em: 21 jul. 2022
- BAYSAL, E.; KOSLOFF, D. D.; SHERWOOD, J. W. C. Reverse time migration. **GEOPHYSICS**, v. 48, n. 11, p. 1514–1524, nov. 1983.
- BELTRAO, R. L. C. et al. **SS: Pre-salt Santos basin - Challenges and New Technologies for the Development of the Pre-salt Cluster, Santos Basin, Brazil**. All Days. **Anais...** Em: OFFSHORE TECHNOLOGY CONFERENCE. Houston, Texas: OTC, 4 maio 2009. Disponível em: <<https://onepetro.org/OTCONF/proceedings/09OTC/All-09OTC/Houston,%20Texas/35954>>. Acesso em: 4 jul. 2022

- BIARI, Y. et al. Structure and evolution of the Atlantic passive margins: A review of existing rifting models from wide-angle seismic data and kinematic reconstruction. **Marine and Petroleum Geology**, v. 126, p. 104898, abr. 2021.
- BIOUCAS-DIAS, J. M. et al. Hyperspectral Remote Sensing Data Analysis and Future Challenges. **IEEE Geoscience and Remote Sensing Magazine**, v. 1, n. 2, p. 6–36, jun. 2013.
- BISCHOFF, A. et al. Seismic Geomorphology, Architecture and Stratigraphy of Volcanoes Buried in Sedimentary Basins. Em: NÉMETH, K. (Ed.). **Updates in Volcanology - Transdisciplinary Nature of Volcano Science**. [s.l.] IntechOpen, 2021.
- BISHOP, C. M. **Pattern recognition and machine learning**. New York: Springer, 2006.
- BROWN, A. R. Seismic attributes and their classification. **The Leading Edge**, v. 15, n. 10, p. 1090–1090, out. 1996.
- BROWN, A. R. **Interpretation of three-dimensional seismic data**. Seventh edition ed. Tulsa, Oklahoma, U.S.A: Published jointly by the American Association of Petroleum Geologists and the Society of Exploration Geophysicists, 2011.
- CARLOTTO, M. A. et al. Libra: A Newborn Giant in the Brazilian Presalt Province. Em: MERRILL, R. K.; STERNBACH, C. A. (Eds.). **Giant Fields of the Decade 2000–2010**. [s.l.] The American Association of Petroleum Geologists, 2017.
- CARRASCO, O. C. **Gaussian Mixture Models Explained**. Disponível em: <<https://towardsdatascience.com/gaussian-mixture-models-explained-6986aaf5a95>>. Acesso em: 30 jun. 2023.
- CARROLL, R. D. The determination of the acoustic parameters of volcanic rocks from compressional velocity measurements. **International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts**, v. 6, n. 6, p. 557–579, nov. 1969.
- CHABOUREAU, A.-C. et al. Paleogeographic evolution of the central segment of the South Atlantic during Early Cretaceous times: Paleotopographic and geodynamic implications. **Tectonophysics**, v. 604, p. 191–223, set. 2013.
- CHEN, Q.; SIDNEY, S. Seismic attribute technology for reservoir forecasting and monitoring. **The Leading Edge**, v. 16, n. 5, p. 445–448, maio 1997.
- CHOPRA, S.; MARFURT, K. J. Seismic attributes — A historical perspective. **GEOPHYSICS**, v. 70, n. 5, p. 3SO-28SO, set. 2005.
- CHOPRA, S.; MARFURT, K. J. **Seismic Attributes for Prospect Identification and Reservoir Characterization**. [s.l.] Society of Exploration Geophysicists and European Association of Geoscientists and Engineers, 2007.
- CHOPRA, S.; MARFURT, K. J. **Seismic facies analysis using generative topographic mapping**. SEG Technical Program Expanded Abstracts 2014. **Anais...** Em: SEG TECHNICAL PROGRAM EXPANDED ABSTRACTS 2014. Denver, Colorado: Society of Exploration Geophysicists, 5 ago. 2014a. Disponível em: <<https://library.seg.org/doi/10.1190/segam2014-0233.1>>. Acesso em: 22 jul. 2022

- CHOPRA, S.; MARFURT, K. J. **Churning seismic attributes with principal component analysis**. SEG Technical Program Expanded Abstracts 2014. **Anais...** Em: SEG TECHNICAL PROGRAM EXPANDED ABSTRACTS 2014. Denver, Colorado: Society of Exploration Geophysicists, 5 ago. 2014b. Disponível em: <<https://library.seg.org/doi/10.1190/segam2014-0235.1>>. Acesso em: 21 jul. 2022
- COLÉOU, T.; POUPON, M.; AZBEL, K. Unsupervised seismic facies classification: A review and comparison of techniques and implementation. **The Leading Edge**, v. 22, n. 10, p. 942–953, out. 2003.
- CORREIA, U. M. DA C. et al. Geometric characterization of igneous intrusions: 3-D seismic insights from the Campos Basin, SE Brazil. **Marine and Petroleum Geology**, v. 102, p. 725–739, abr. 2019.
- DE AMORIM, R. C. A Survey on Feature Weighting Based K-Means Algorithms. **Journal of Classification**, v. 33, n. 2, p. 210–242, jul. 2016.
- DE LUCA, P. et al. **What is the Role of Volcanic Rocks in the Brazilian Pre-salt?** . Em: 77TH EAGE CONFERENCE AND EXHIBITION 2015. Madrid, Spain: 1 jun. 2015. Disponível em: <<http://www.earthdoc.org/publication/publicationdetails/?publication=80643>>. Acesso em: 5 jul. 2022
- DE MATOS, M. C. et al. Integrated seismic texture segmentation and cluster analysis applied to channel delineation and chert reservoir characterization. **GEOPHYSICS**, v. 76, n. 5, p. P11–P21, set. 2011.
- DE MATOS, M. C.; OSORIO, P. L.; JOHANN, P. R. Unsupervised seismic facies analysis using wavelet transform and self-organizing maps. **GEOPHYSICS**, v. 72, n. 1, p. P9–P21, jan. 2007.
- DE OLIVEIRA, L. C.; RANCAN, C. C.; OLIVEIRA, M. J. **Sill emplacement mechanisms and their relationship with the Pre-Salt stratigraphic framework of the Libra Area (Santos Basin, Brazil)**. . Em: LASI 6 CONFERENCE. 2019. Disponível em: <<https://www.researchgate.net/publication/337683465>>
- DEEPA, P.; THILAGAVATHI, K. **Feature extraction of hyperspectral image using principal component analysis and folded-principal component analysis**. 2015 2nd International Conference on Electronics and Communication Systems (ICECS). **Anais...** Em: 2015 2ND INTERNATIONAL CONFERENCE ON ELECTRONICS AND COMMUNICATION SYSTEMS (ICECS). Coimbatore, India: IEEE, fev. 2015. Disponível em: <<http://ieeexplore.ieee.org/document/7124989/>>. Acesso em: 30 set. 2022
- DEWETT, D. T.; PIGOTT, J. D.; MARFURT, K. J. A review of seismic attribute taxonomies, discussion of their historical use, and presentation of a seismic attribute communication framework using data analysis concepts. **Interpretation**, v. 9, n. 3, p. B39–B64, 1 ago. 2021.
- DI, H.; SHAFIQ, M.; ALREGIB, G. Multi-attribute *k* -means clustering for salt-boundary delineation from three-dimensional seismic data. **Geophysical Journal International**, v. 215, n. 3, p. 1999–2007, 1 dez. 2018.

DUDA, R. O.; HART, P. E. Pattern Classification and Scene Analysis. **The Library Quarterly**, v. 44, n. 3, p. 258–259, jul. 1974.

DY, J. G.; BRODLEY, C. E. Feature Selection for Unsupervised Learning. **Journal of Machine Learning Research**, v. 5, p. 845–889, ago. 2004.

EIDE, C. H. et al. Basin-scale architecture of deeply emplaced sill complexes: Jameson Land, East Greenland. **Journal of the Geological Society**, v. 174, n. 1, p. 23–40, jan. 2017.

EIDE, C. H. et al. Seismic interpretation of sill complexes in sedimentary basins: implications for the sub-sill imaging problem. **Journal of the Geological Society**, v. 175, n. 2, p. 193–209, mar. 2018.

EKLAVYA. **Kohonen Self-Organizing Maps**. Disponível em: <<https://towardsdatascience.com/kohonen-self-organizing-maps-a29040d688da>>. Acesso em: 30 jun. 2023.

FAROOQUI, M. Y. et al. Evaluating volcanic reservoirs. **Oilfield Review**, v. 21, n. 1, p. 36–47, 2009.

FLEXER, A. On the use of self-organizing maps for clustering and visualization. **Intelligent Data Analysis**, v. 5, n. 5, p. 373–384, 2001.

FORNERO, S. A. et al. Characterization of subaerial volcanic facies using acoustic image logs: Lithofacies and log-facies of a lava-flow deposit in the Brazilian pre-salt, deepwater of Santos Basin. **Marine and Petroleum Geology**, v. 99, p. 156–174, jan. 2019.

GAO, D. Application of three-dimensional seismic texture analysis with special reference to deep-marine facies discrimination and interpretation: Offshore Angola, west Africa. **AAPG Bulletin**, v. 91, n. 12, p. 1665–1683, dez. 2007.

GERON, A. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow concepts, tools, and techniques to build intelligent systems**. [s.l.: s.n.].

GORDON, A. C. et al. Magmatic cycles in Santos Basin (S.E. Brazil): Tectonic control in the temporal-spatial distribution and geophysical signature. **Journal of South American Earth Sciences**, p. 104111, nov. 2022.

GUDMUNDSSON, A.; LØTVEIT, I. F. Sills as fractured hydrocarbon reservoirs: examples and models. **Geological Society, London, Special Publications**, v. 374, n. 1, p. 251–271, 2014.

GUO, H.; LEWIS, S.; MARFURT, K. J. Mapping multiple attributes to three- and four-component color models — A tutorial. **GEOPHYSICS**, v. 73, n. 3, p. W7–W19, maio 2008.

HA, T. N. et al. An in-depth analysis of logarithmic data transformation and per-class normalization in machine learning: Application to unsupervised classification of a turbidite system in the Canterbury Basin, New Zealand, and supervised classification of salt in the Eugene Island minibasin, Gulf of Mexico. **Interpretation**, v. 9, n. 3, p. T685–T710, 1 ago. 2021.

- HALL, M. **Tuning Wedge Model**. SEG, , 2014. Disponível em: <<https://github.com/seg/tutorials-2014>>. Acesso em: 22 jun. 2023
- HALL-BEYER, M. **GLCM Texture: A Tutorial v. 3.0 March 2017**. PRISM, , 2017. Disponível em: <<https://prism.ucalgary.ca/handle/1880/51900>>
- HAN, M. et al. Application of EM algorithms for seismic facies classification. **Computational Geosciences**, v. 15, n. 3, p. 421–429, jun. 2011.
- HANCER, E.; XUE, B.; ZHANG, M. A survey on feature selection approaches for clustering. **Artificial Intelligence Review**, v. 53, n. 6, p. 4519–4545, ago. 2020.
- HANSEN, J. et al. Early Cenozoic saucer-shaped sills of the Faroe Islands: an example of intrusive styles in basaltic lava piles. **Journal of the Geological Society**, v. 168, n. 1, p. 159–178, jan. 2011.
- HONÓRIO, B. C. Z. et al. Independent component spectral analysis. **Interpretation**, v. 2, n. 1, p. SA21–SA29, 1 fev. 2014.
- HUGHES, G. On the mean accuracy of statistical pattern recognizers. **IEEE Transactions on Information Theory**, v. 14, n. 1, p. 55–63, jan. 1968.
- HUTTON, D. H. W. Insights into magmatism in volcanic margins: bridge structures and a new mechanism of basic sill emplacement – Theron Mountains, Antarctica. **Petroleum Geoscience**, v. 15, n. 3, p. 269–278, ago. 2009.
- HYVÄRINEN, A.; OJA, E. Independent component analysis: algorithms and applications. **Neural Networks**, v. 13, n. 4–5, p. 411–430, jun. 2000.
- INFANTE-PAEZ, L.; MARFURT, K. J. Seismic expression and geomorphology of igneous bodies: A Taranaki Basin, New Zealand, case study. **Interpretation**, v. 5, n. 3, p. SK121–SK140, 31 ago. 2017.
- INFANTE-PAEZ, L.; MARFURT, K. J. Using machine learning as an aid to seismic geomorphology, which attributes are the best input? **Interpretation**, v. 7, n. 3, p. SE1–SE18, 1 ago. 2019.
- ISLAM, MD. R.; AHMED, B.; HOSSAIN, MD. A. **Feature Reduction Based on Segmented Principal Component Analysis for Hyperspectral Images Classification**. 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE). **Anais...** Em: 2019 INTERNATIONAL CONFERENCE ON ELECTRICAL, COMPUTER AND COMMUNICATION ENGINEERING (ECCE). Cox's Bazar, Bangladesh: IEEE, fev. 2019. Disponível em: <<https://ieeexplore.ieee.org/document/8679394/>>. Acesso em: 30 set. 2022
- JIA, X.; KUO, B.-C.; CRAWFORD, M. M. Feature Mining for Hyperspectral Image Classification. **Proceedings of the IEEE**, v. 101, n. 3, p. 676–697, mar. 2013.
- KANGXU, R. et al. **Metamorphism of pre-salt limestones produced by Santonian-Campanian alkaline sills in the Libra Block, Santos Basin**. nov. 2019.

KARNER, G. D. et al. Chapter 9: Tectono-Magmatic Development of the Santos and Campos Basins, Offshore Brazil. Em: MELLO, M. R.; YILMAZN, P. O.; KATZ, B. J. (Eds.). **Memoir 124: The Supergiant Lower Cretaceous Pre-Salt Petroleum Systems of the Santos Basin, Brazil.** [s.l.] AAPG, 2021. p. 215–256.

KEAREY, P.; BROOKS, M.; HILL, I. **An introduction to geophysical exploration.** 3rd ed. Malden, MA: Blackwell Science, 2002. v. 140

KIM, Y.; HARDISTY, R.; MARFURT, K. J. Attribute selection in seismic facies classification: Application to a Gulf of Mexico 3D seismic survey and the Barnett Shale. **Interpretation**, v. 7, n. 3, p. SE281–SE297, 1 ago. 2019.

KOHONEN, T. Self-organized formation of topologically correct feature maps. **Biological Cybernetics**, v. 43, n. 1, p. 59–69, 1982.

KUMAR, B. et al. Feature extraction for hyperspectral image classification: a review. **International Journal of Remote Sensing**, v. 41, n. 16, p. 6248–6287, 17 ago. 2020.

LEE, J.; PARK, D.; LEE, C. Feature Selection Algorithm for Intrusions Detection System using Sequential Forward Search and Random Forest Classifier. **KSII Transactions on Internet and Information Systems**, v. 11, n. 10, 31 out. 2017.

LIU, Z. et al. A seismic facies classification method based on the convolutional neural network and the probabilistic framework for seismic attributes and spatial classification. **Interpretation**, v. 7, n. 3, p. SE225–SE236, 1 ago. 2019.

LUBO-ROBLES, D. et al. Quantifying the sensitivity of seismic facies classification to seismic attribute selection: An explainable machine-learning study. **Interpretation**, v. 10, n. 3, p. SE41–SE69, 1 ago. 2022.

LUBO-ROBLES, D.; MARFURT, K. J. Independent component analysis for reservoir geomorphology and unsupervised seismic facies classification in the Taranaki Basin, New Zealand. **Interpretation**, v. 7, n. 3, p. SE19–SE42, 1 ago. 2019.

MACQUEEN, J. **Some Methods for Classification and Analysis of Multivariate Observations.** Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. **Anais...** Em: 5TH BERKELEY SYMP. MATH. STATIST. PROBABILITY. California: University of California Press, 1967.

MAGEE, C. et al. Controls on the expression of igneous intrusions in seismic reflection data. **Geosphere**, v. 11, n. 4, p. 1024–1041, 2015.

MAGEE, C. et al. Magma Plumbing Systems: A Geophysical Perspective. **Journal of Petrology**, v. 59, n. 6, p. 1217–1251, 1 jun. 2018.

MAGEE, C. et al. Salt–magma interactions influence intrusion distribution and salt tectonics in the Santos Basin, offshore Brazil. **Basin Research**, v. 33, n. 3, p. 1820–1843, jun. 2021.

MANTRIPRAGADA, K. et al. The effects of spectral dimensionality reduction on hyperspectral pixel classification: A case study. **PLOS ONE**, v. 17, n. 7, p. e0269174, 14 jul. 2022.

MARFURT, K. J. **Seismic Attributes as the Framework for Data Integration Throughout the Oilfield Life Cycle**. [s.l.] Society of Exploration Geophysicists, 2018.

MARK, N. J. et al. Igneous intrusions in the Faroe Shetland basin and their implications for hydrocarbon exploration; new insights from well and seismic data. **Marine and Petroleum Geology**, v. 92, p. 733–753, abr. 2018.

MARROQUÍN, I. D.; BRAULT, J.-J.; HART, B. S. A visual data-mining methodology for seismic facies analysis: Part 1 — Testing and comparison with other unsupervised clustering methods. **GEOPHYSICS**, v. 74, n. 1, p. P1–P11, jan. 2009.

MATOS, R. M. D. DE et al. The fundamental role of the Borborema and Benin–Nigeria provinces of NE Brazil and NW Africa during the development of the South Atlantic Cretaceous Rift system. **Marine and Petroleum Geology**, v. 127, p. 104872, maio 2021.

MATOS, R. M. D. DE et al. Oblique Core-Complex Type Detachment Systems in the Santos Basin. **SSRN Electronic Journal**, 2022.

MCKINNEY, W. **Data Structures for Statistical Computing in Python**. . Em: PYTHON IN SCIENCE CONFERENCE. Austin, Texas: 2010. Disponível em: <<https://conference.scipy.org/proceedings/scipy2010/mckinney.html>>. Acesso em: 28 jun. 2023

MIAO, J.; NIU, L. A Survey on Feature Selection. **Procedia Computer Science**, v. 91, p. 919–926, 2016.

MITCHUM JR, R. M.; VAIL, P. R.; SANGREE, J. B. Seismic stratigraphy and global changes of sea level: Part 6. Stratigraphic interpretation of seismic reflection patterns in depositional sequences: Section 2. Application of seismic reflection configuration to stratigraphic interpretation. Em: **Seismic Stratigraphy — Applications to Hydrocarbon Exploration**. AAPG memoir. [s.l.] AAPG Special Volumes, 1977.

MOHAN, A.; SAPIRO, G.; BOSCH, E. Spatially Coherent Nonlinear Dimensionality Reduction and Segmentation of Hyperspectral Images. **IEEE Geoscience and Remote Sensing Letters**, v. 4, n. 2, p. 206–210, abr. 2007.

MOHRIAK, W.; NEMČOK, M.; ENCISO, G. South Atlantic divergent margin evolution: rift-border uplift and salt tectonics in the basins of SE Brazil. **Geological Society, London, Special Publications**, v. 294, n. 1, p. 365–398, 2008.

MOHRIAK, W. U.; GORDON, A.; MELLO, M. R. Chapter 11: Origin and Petroleum System of the Cabo Frio High Between the Santos and Campos Basins: Reviewed Integration of Structural and Paleogeographic Reconstruction with the Oil and Gas Systems. Em: MELLO, M. R.; YILMAZ, P. O.; KATZ, B. J. (Eds.). **Memoir 124: The Supergiant Lower Cretaceous Pre-Salt Petroleum Systems of the Santos Basin, Brazil**. [s.l.] AAPG, 2021. p. 273–324.

MOHRIAK, W. U.; LEROY, S. Architecture of rifted continental margins and break-up evolution: insights from the South Atlantic, North Atlantic and Red Sea–Gulf of Aden conjugate margins. **Geological Society, London, Special Publications**, v. 369, n. 1, p. 497–535, 2013.

MOREIRA, J. L. P. et al. Bacia de Santos. **Boletim de Geociências da Petrobras**, v. 15, n. 2, p. 531–549, 2007.

MUIRHEAD, J. D. et al. Interconnected sills and inclined sheet intrusions control shallow magma transport in the Ferrar large igneous province, Antarctica. **Geological Society of America Bulletin**, v. 124, n. 1–2, p. 162–180, 1 jan. 2012.

NAPOLI, O. O. et al. Accelerating Multi-attribute Unsupervised Seismic Facies Analysis With RAPIDS. 2020.

NELSON, C. E. et al. Understanding the facies architecture of flood basalts and volcanic rifted margins and its effect on geophysical properties. p. 20, 2009.

PANDIT, A. A.; PIMPALE, B.; DUBEY, S. A Comprehensive Review on Unsupervised Feature Selection Algorithms. Em: SINGH TOMAR, G. et al. (Eds.). **International Conference on Intelligent Computing and Smart Communication 2019**. Algorithms for Intelligent Systems. Singapore: Springer Singapore, 2020. p. 255–266.

PEACE, A. L. et al. A review of Pangaea dispersal and Large Igneous Provinces – In search of a causative mechanism. **Earth-Science Reviews**, v. 206, p. 102902, jul. 2020.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, n. 85, p. 2825–2830, 2011.

PENNA, R. et al. Carbonate and igneous rock characterization through reprocessing, FWI imaging, and elastic inversion of a legacy seismic data set in Brazilian presalt province. **The Leading Edge**, v. 38, n. 1, p. 11–19, jan. 2019.

PETERSOHN, E. et al. Chapter 13: The Giant Mero Oil Field: Geological and Petrophysical Settings of One of the Largest Pre-Salt Fields of the Santos Basin, Southern Brazil. Em: MELLO, M. R.; YILMAZ, P. O.; KATZ, B. J. (Eds.). **Memoir 124: The Supergiant Lower Cretaceous Pre-Salt Petroleum Systems of the Santos Basin, Brazil**. [s.l.] AAPG, 2021. p. 351–374.

PLANKE, S. et al. Seismic characteristics and distribution of volcanic intrusions and hydrothermal vent complexes in the Vøring and Møre basins. **Geological Society, London, Petroleum Geology Conference series**, v. 6, n. 1, p. 833–844, 2005.

PLANKE, S. et al. **Seismic imaging and petroleum implications of igneous intrusions in sedimentary basins constrained by outcrop analogues and seismic data from the Neuquén Basin and the NE Atlantic**. Em: 10º CONGRESO DE EXPLORACIÓN Y DESARROLLO DE HIDROCARBUROS. Mendoza, Argentina: Instituto Argentino del Petróleo y el Gas, 2018.

PROUSSEVITCH, A. A.; SAHAGIAN, D. L. Recognition and separation of discrete objects within complex 3D voxelized structures. **Computers & Geosciences**, v. 27, n. 4, p. 441–454, maio 2001.

QI, J. et al. Semisupervised multiattribute seismic facies analysis. **Interpretation**, v. 4, n. 1, p. SB91–SB106, 1 fev. 2016.

- QI, J. et al. Seismic attribute selection for machine-learning-based facies analysis. **GEOPHYSICS**, v. 85, n. 2, p. O17–O35, 1 mar. 2020.
- RABELL, O. et al. From field analogues to realistic seismic modelling: a case study of an oil-producing andesitic sill complex in the Neuquén Basin, Argentina. **Journal of the Geological Society**, v. 175, n. 4, p. 580–593, jul. 2018.
- RANCAN, C. C. et al. **Rochas Ígneas do Bloco de Libra, Bacia de Santos**. Anais do 49o Congresso Brasileiro de Geologia. **Anais...** Em: 49O CONGRESSO BRASILEIRO DE GEOLOGIA. Rio de Janeiro, Brazil: SBGeo, 2018. Disponível em: <<http://cbg2018anais.siteoficial.ws/resumos/7327.pdf>>
- RASCHKA, S.; MIRJALILI, V. **Python machine learning: machine learning and deep learning with Python, scikit-learn, and TensorFlow 2**. Third edition ed. Birmingham Mumbai: Packt, 2019.
- REUBER, K.; MANN, P. Control of Precambrian-to-Paleozoic orogenic trends on along-strike variations in Early Cretaceous continental rifts of the South Atlantic Ocean. **Interpretation**, v. 7, n. 4, p. SH45–SH69, 1 nov. 2019.
- RICCOMINI, C.; SANT’ANNA, L. G.; FERRARI, A. L. Evolução geológica do rift continental do sudeste do Brasil. Em: **Geologia do continente sul-americano: : evolução da obra de Fernando Flávio Marques de Almeida**. [s.l.: s.n.]. p. 383–405.
- RODAMEL, C.; SHAN, J. Principal component analysis for hyperspectral image classification. **Surveying and Land Information Science**, v. 62, n. 2, p. 115–123, 2002.
- RODEN, R.; SMITH, T.; SACREY, D. Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps. **Interpretation**, v. 3, n. 4, p. SAE59–SAE83, 1 nov. 2015.
- ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, n. 6, p. 386–408, 1958.
- ROY, A. et al. Generative topographic mapping for seismic facies estimation of a carbonate wash, Veracruz Basin, southern Mexico. **Interpretation**, v. 2, n. 1, p. SA31–SA47, 1 fev. 2014.
- SALOMON, E.; PASSCHIER, C.; KOEHN, D. Asymmetric continental deformation during South Atlantic rifting along southern Brazil and Namibia. **Gondwana Research**, v. 51, p. 170–176, nov. 2017.
- SAMUEL, A. L. Some Studies in Machine Learning Using the Game of Checkers. **IBM Journal of Research and Development**, v. 3, n. 3, p. 210–229, jul. 1959.
- SÁNCHEZ-MAROÑO, N.; ALONSO-BETANZOS, A.; TOMBILLA-SANROMÁN, M. Filter Methods for Feature Selection – A Comparative Study. Em: YIN, H. et al. (Eds.). **Intelligent Data Engineering and Automated Learning - IDEAL 2007**. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. v. 4881p. 178–187.

SCHOFIELD, N. et al. Regional magma plumbing and emplacement mechanisms of the Faroe-Shetland Sill Complex: implications for magma transport and petroleum systems within sedimentary basins. **Basin Research**, v. 29, n. 1, p. 41–63, fev. 2017.

SCHOFIELD, N. et al. Overpressure transmission through interconnected igneous intrusions. **AAPG Bulletin**, v. 104, n. 2, p. 285–303, fev. 2020.

SCHOFIELD, N. J. et al. Sill morphology and comparison of brittle and non-brittle emplacement mechanisms. **Journal of the Geological Society**, v. 169, n. 2, p. 127–141, mar. 2012.

SCHOFIELD, N.; STEVENSON, C.; RESTON, T. Magma fingers and host rock fluidization in the emplacement of sills. **Geology**, v. 38, n. 1, p. 63–66, jan. 2010.

SENGER, K. et al. Effects of igneous intrusions on the petroleum system: a review. **First Break**, v. 35, n. 6, 1 jun. 2017.

SHERIFF, R. E.; GELDART, L. P. **Exploration seismology**. [s.l: s.n.].

SHI, G. et al. Dimensionality Reduction of Hyperspectral Image Based on Local Constrained Manifold Structure Collaborative Preserving Embedding. **Remote Sensing**, v. 13, n. 7, p. 1363, 2 abr. 2021.

SINGH, A.; OJHA, M. Stable clustering of offshore downhole data using a combined k-means and Gaussian mixture modelling approach. **Marine Geophysical Research**, v. 43, n. 3, p. 31, set. 2022.

SINGH, Y. Lithofacies detection through simultaneous inversion and principal component attributes. **The Leading Edge**, v. 26, n. 12, p. 1568–1575, dez. 2007.

SMALLWOOD, J. R.; MARESH, J. The properties, morphology and distribution of igneous sills: modelling, borehole data and 3D seismic from the Faroe-Shetland area. **Geological Society, London, Special Publications**, v. 197, n. 1, p. 271–306, jan. 2002.

SOLORIO-FERNÁNDEZ, S.; CARRASCO-OCHOA, J. A.; MARTÍNEZ-TRINIDAD, J. FCO. A review of unsupervised feature selection methods. **Artificial Intelligence Review**, v. 53, n. 2, p. 907–948, fev. 2020.

TANER, M. T. et al. **Seismic attributes revisited**. SEG Technical Program Expanded Abstracts 1994. **Anais...** Em: SEG TECHNICAL PROGRAM EXPANDED ABSTRACTS 1994. Society of Exploration Geophysicists, jan. 1994. Disponível em: <<http://library.seg.org/doi/abs/10.1190/1.1822709>>. Acesso em: 21 jul. 2022

TANER, M. T. et al. **Reservoir Characterization By Calibration Of Selforganized Map Clusters**. . Em: 2001 SEG ANNUAL MEETING. 9 set. 2001. . Acesso em: 21 jun. 2023

TEAM, T. P. D. **pandas-dev/pandas: Pandas**. Zenodo, , 28 maio 2023. Disponível em: <<https://zenodo.org/record/7979740>>. Acesso em: 28 jun. 2023

THORNDIKE, R. L. Who belongs in the family? **Psychometrika**, v. 18, n. 4, p. 267–276, dez. 1953.

VERGARA, J. R.; ESTÉVEZ, P. A. A review of feature selection methods based on mutual information. **Neural Computing and Applications**, v. 24, n. 1, p. 175–186, jan. 2014.

VETTIGLI, G. **MiniSom: minimalistic and NumPy-based implementation of the Self Organizing Map**. , 2018. Disponível em: <<https://github.com/JustGlowing/minisom/>>

WALDEN, A. T. Non-Gaussian reflectivity, entropy, and deconvolution. **GEOPHYSICS**, v. 50, n. 12, p. 2862–2888, dez. 1985.

WALKER, R. J. et al. Igneous sills as a record of horizontal shortening: The San Rafael subvolcanic field, Utah. **Geological Society of America Bulletin**, p. B31671.1, 7 abr. 2017.

WALLET, B. C.; HA, T. N. A deep-learning method for latent space analysis of multiple seismic attributes. **Interpretation**, v. 9, n. 3, p. T945–T954, 1 ago. 2021.

WALLET, B. C.; HARDISTY, R. Unsupervised seismic facies using Gaussian mixture models. **Interpretation**, v. 7, n. 3, p. SE93–SE111, 1 ago. 2019.

WANG, Z. et al. Dimensionality reduction method for hyperspectral image analysis based on rough set theory. **European Journal of Remote Sensing**, v. 53, n. 1, p. 192–200, 1 jan. 2020.

WATSON, D. et al. Failure to predict igneous rocks encountered during exploration of sedimentary basins: A case study of the Bass Basin, Southeastern Australia. **Marine and Petroleum Geology**, v. 99, p. 526–547, jan. 2019.

WEST, B. P. et al. Interactive seismic facies classification using textural attributes and neural networks. **The Leading Edge**, v. 21, n. 10, p. 1042–1049, out. 2002.

XU, G.; HAQ, B. U. Seismic facies analysis: Past, present and future. **Earth-Science Reviews**, v. 224, p. 103876, jan. 2022.

XU, R.; WUNSCH II, D. Survey of Clustering Algorithms. **IEEE Transactions on Neural Networks**, v. 16, n. 3, p. 645–678, maio 2005.

ZHAO, J. et al. **Fault Activity and its Influences on Distribution of Igneous Rocks in Libra Block, Santos Basin: Semi-Quantitative to Quantitative Assessment of Fault Activity Based on High-Resolution 3D Seismic Data**. Day 2 Wed, October 30, 2019. **Anais...** Em: OFFSHORE TECHNOLOGY CONFERENCE BRASIL. Rio de Janeiro, Brazil: OTC, 28 out. 2019. Disponível em: <<https://onepetro.org/OTCBRASIL/proceedings/19OTCB/2-19OTCB/Rio%20de%20Janeiro,%20Brazil/181845>>. Acesso em: 9 jul. 2022

ZHAO, T. et al. A comparison of classification techniques for seismic facies recognition. **Interpretation**, v. 3, n. 4, p. SAE29–SAE58, 1 nov. 2015.

ZHENG, X.; YUAN, Y.; LU, X. Dimensionality Reduction by Spatial–Spectral Preservation in Selected Bands. **IEEE Transactions on Geoscience and Remote Sensing**, v. 55, n. 9, p. 5185–5197, set. 2017.

ZHU, D. et al. Adaptive Gaussian mixture model and convolution autoencoder clustering for unsupervised seismic waveform analysis. **Interpretation**, v. 10, n. 1, p. T181–T193, 1 fev. 2022.

APÊNDICE A – Rotinas dos algoritmos

Os algoritmos de agrupamento foram aplicados a partir de rotinas escritas na linguagem de programação Python, a partir de bibliotecas públicas disponibilizadas para o público geral. Em geral, todos os algoritmos se iniciam com o carregamento dos dados SEGYY:

```
#!/usr/bin/env python3

import segyio
import numpy as np
import pandas as pd
import xarray as xr
from sklearn.mixture import GaussianMixture
from sklearn.preprocessing import MinMaxScaler
from sklearn.decomposition import PCA
from sklearn.model_selection import GridSearchCV
import matplotlib.pyplot as plt
import seaborn as sns
import random
from joblib import Parallel, delayed

aviso = 'INICIALIZANDO CARREGAMENTO DE DADOS'
print(f"\n{'#' * len(aviso)}\n{aviso}\n{'#' * len(aviso)}\n")

with open("./atrib_300/atrib_FS.txt", encoding='UTF-8') as f:
    lines = f.read().split("\n")

with open("./atrib_300/nomes_FS.txt", encoding='ISO 8859-1') as g:
    nomes = g.read().split("\n")

lines = lines[0:-1]
nomes = nomes[0:-1]

ilines = np.arange(1500, 2271)
xlines = np.arange(3699, 7500, 4)
profs = np.arange(5000, 6901, 5)

dfs = {}
```

```

scaler = MinMaxScaler()

for i,line in enumerate(lines):
    with segyio.open('./atrib_300/' + line, iline=193, xline=197) as
    segy:
        traces = segyio.cube(segy)

        # Cria Xarray
        cube_data = xr.DataArray(traces, coords=[ilines, xlines,
        profs], dims=['iline', 'xline', 'prof'])

        # Converte xarray em DataFrame
        cube_df = cube_data.to_dataframe(name=nomes[i])

        # Normaliza DataFrame
        nonzero_mask = cube_df != 0
        nonzero_values = cube_df[nonzero_mask]
        normalized_values =
scaler.fit_transform(nonzero_values.values.reshape(-1, 1))
        cube_df[nonzero_mask] = normalized_values

        dfs[nomes[i]] = cube_df
        print(f"DFS {nomes[i]} gravado!!")

concat_df = pd.concat([dfs[n] for n in nomes], axis=1)
concat_df = concat_df.loc[(concat_df != 0.0).any(axis=1)]

indice_concat = concat_df.dropna().index

```

Essa rotina é comum aos três algoritmos e tem o intuito de transformar o dado SEGYY em um dataframe, utilizado pela biblioteca pandas, que facilita reunir todos os atributos em um único objeto multidimensional, o objeto “concat_df”.

A partir daí se inicia o processo de PCA e execução dos algoritmos, é importante ressaltar que, devido a restrições de memória, cada algoritmo é executado separadamente, à exceção da etapa de PCA que é comum a todos, quando esta é aplicada (Grupo 01). A partir daí é realizada a geração e exportação de imagens.

```
pca0 = PCA() #EXECUTA PCA
```

```

fit_pca0 = pca0.fit_transform(concat_df.dropna())

aviso = 'PCA FINALIZADO'
print(f"\n{'#' * len(aviso)}\n{aviso}\n{'#' * len(aviso)}\n")

# Select os primeiros 10 componentes de PCA
pca_components = fit_pca0[:, :10]

# INICIALIZAÇÃO DO KMEANS

for k in kah:
    kmns = KMeans(n_clusters=k, random_state=1) #INICIALIZAÇÃO
    labels = kmns.fit_predict(pca_components) #PREDIÇÃO

    # Create a new DataFrame with aligned indices
    cluster_df = pd.DataFrame(index=indice_concat)
    cluster_df['Cluster'] = labels

    cluster_labels_path =
f"./KMNS/ATRIB_FS_PCA5_K_{k}_components.csv"
    cluster_df['Cluster'].to_csv(cluster_labels_path, index=True) #
SALVA RESULTADOS EM TABELA

# INICIALIZAÇÃO DO GMM

# CRIAÇÃO DE UMA FUNÇÃO DO GMM PARA EXECUÇÃO PARALELIZADA
def train_gmm(covariance, n_components, X):
    #Inicializa GMM
    gmm = GaussianMixture(n_components=n_components,
covariance_type=covariance, init_params='k-means++')
    print(f"Inicializando {covariance}, nc {n_components}")
    #Predição GMM
    cluster_labels = gmm.fit_predict(X)

    cluster_labels_path =
f"./GMM/ATRIB_FSEL_{covariance}_NC_{n_components}.csv"

    # Cria tabela e exporta resultados
    cluster_df = pd.DataFrame(index=indice_concat)
    cluster_df['Cluster'] = cluster_labels

```

```

cluster_df[['Cluster']].to_csv(cluster_labels_path, index=True)
print(f"Finalizado {covariance}, nc {n_components}")

# Paralelização do GMM usando joblib
results = Parallel(n_jobs=4, max_nbytes='100G')(
    delayed(train_gmm)(covariance, n_components, pca_components)
    for n_components in compo
    for covariance in covariances
)

# CRIAÇÃO DE UMA FUNÇÃO DO SOM PARA EXECUÇÃO PARALELIZADA
def train_som(sigma, learning_rate, grid_size, X):
    # INICIALIZA SOM
    som = MiniSom(grid_size[0], grid_size[1], X.shape[1],
sigma=sigma, learning_rate=learning_rate)

    # TREINA SOM
    som.train_random(X, 500)

    output_filename =
f'./SOM/ATRIB_6ATRIB_500_{grid_size[0]}_{grid_size[1]}_s_{sigma}_lr_
{learning_rate}.csv'

    # Create a list to store the cluster labels and their dimensions
    cluster_data = []

    # ITERAÇÃO A CADA AMOSTRA PARA IDENTIFICAR NEURON VENCEDOR E
    POSIÇÃO
    for sample_index in range(X.shape[0]):
        # VALORES DO INDICE
        iline, xline, prof = concat_df.index[sample_index]

        # VALORES DA AMOSTRA
        sample_values = X[sample_index]

        # CALCULA POSIÇÃO DO NEURON VENCEDOR
        winner_position = som.winner(sample_values)

        # Calcula RÓTULO
        cluster_label = winner_position[0] * grid_size[1] +
winner_position[1]

```



```
        # ADICIONA RESULTADO À TABELA
        cluster_data.append({'iline': iline, 'xline': xline,
'prof': prof, 'Cluster': cluster_label})

# CRIA TABELA PARA EXPORTAÇÃO
cluster_df = pd.DataFrame(cluster_data)

# SALVA RESULTADO COMO ARQUIVO CSV
cluster_df.to_csv(output_filename, index=False)
```