



Universidade Federal Fluminense

Instituto de Geociências

Departamento de Geologia e Geofísica

Programa de Pós-Graduação em Dinâmicas dos Oceanos e
da Terra

Jeniffer Alves Nobre

**Caracterização Quantitativa das Rochas Geradoras da
Formação Pimenteiras: Predição e Modelagem 3D do
Carbono Orgânico Total Integrando Condicionamento
Sísmico, Inversão e Machine Learning**

Jeniffer Alves Nobre

**Caracterização Quantitativa das Rochas Geradoras da
Formação Pimenteiras: Predição e Modelagem 3D do
Carbono Orgânico Total Integrando Condicionamento
Sísmico, Inversão e Machine Learning**

Trabalho de conclusão de curso submetido ao Programa de Pós-Graduação em Dinâmica dos Oceanos e da Terra da Universidade Federal Fluminense como parte dos requisitos para a obtenção do título de Doutor.

Orientador: Prof. Dr. Antonio Fernando Menezes Freire

Niterói, RJ
2025

Jeniffer Alves Nobre

**Caracterização Quantitativa das Rochas Geradoras da
Formação Pimenteiras: Predição e Modelagem 3D do Carbono
Orgânico Total Integrando Condicionamento Sísmico, Inversão
e Machine Learning**

Trabalho de conclusão de curso submetido ao Programa de Pós-Graduação em Dinâmica dos Oceanos e da Terra da Universidade Federal Fluminense como parte dos requisitos para a obtenção do título de Doutor.

Aprovada em 03 de dezembro de 2025.

Comissão Examinadora:

Prof. Dr. Antonio Fernando Menezes Freire – Orientador (GIECAR/GGO/UFF)

Prof. Dr. Wagner Moreira Lupinacci (GIECAR/GGO /UFF)

Prof. Dr. Cleverson Guizan Silva (GGO /UFF)

Prof. Dr. Esteban Alfaro Sabogal (GGO /UFF)

Dr. Danilo Jotta Ariza Ferreira (TOTAL)

Dr. Roberto Salvador Francisco d'Ávila (PETROBRAS)

Niterói, RJ
2025

Ficha Catalográfica

Ficha catalográfica automática - SDC/BIG
Gerada com informações fornecidas pelo autor

N754c Nobre, Jeniffer Alves
Caracterização Quantitativa das Rochas Geradoras da
Formação Pimenteiras: Predição e Modelagem 3D do Carbono
Orgânico Total Integrando Condicionamento Sísmico, Inversão
e Machine Learning / Jeniffer Alves Nobre. - 2025.
346 f.

Orientador: Antonio Fernando Menezes Freire.
Tese (doutorado)-Universidade Federal Fluminense, Instituto
de Geociências, Niterói, 2025.

1. Source rock characterization. 2. Total organic carbon. 3.
Seismic inversion. 4. Machine learning. 5. Produção
intelectual. I. Freire, Antonio Fernando Menezes, orientador.
II. Universidade Federal Fluminense. Instituto de
Geociências. III. Título.

CDD - XXX

Bibliotecário responsável: Debora do Nascimento - CRB7/6368

Agradecimentos

Agradeço profundamente à minha família, que desde sempre me incentivou e me mostrou o valor da educação e do conhecimento. Obrigada por me apoiarem incondicionalmente e por demonstrarem orgulho em cada uma das minhas escolhas. Dedico esse trabalho a vocês pais, Adriana Alves e Francisco Nobre, que se dedicaram tanto a nós e muitas vezes nos deram aquilo que nem tinham. Aos meus irmãos Mydelane Nobre e Brenno Nobre, meus orgulhos e pilares e as minhas lindas tias Luanna Aves e Deusiana Alves que se dedicaram e cuidaram de mim com tanto carinho no momento que mais precisei.

Aos professores do Departamento de Geologia e Geofísica, minha sincera gratidão pelo conhecimento compartilhado e pela dedicação em despertar em mim a admiração pela pesquisa e pelas geociências. À Universidade Federal Fluminense, agradeço pela estrutura e pelo privilégio de ter recebido uma formação de excelência.

Em especial, ao meu orientador, Prof. Dr. Antonio Fernando Menezes Freire, expresso minha mais profunda admiração e gratidão. Obrigada por ser um guia inspirador, sempre generoso ao compartilhar seu conhecimento e motivar seus alunos a explorar todo o seu potencial com paciência, sabedoria e entusiasmo. Suas aulas e orientações transformaram o aprendizado em uma verdadeira jornada de descobertas.

Aos colegas do Projeto Parnaíba, foi um imenso prazer compartilhar o início desta trajetória com vocês. As trocas de experiências, o companheirismo e as amizades construídas marcaram profundamente esta caminhada.

Um agradecimento especial ao amigo Fábio Júnior por ter me ajudado enormemente, não só disponibilizando o código em python para a geração da inversão sísmica, mas também colocando a mão na massa comigo e acreditando no potencial deste trabalho Suas palavras de apoio foram valiosas.

Agradeço à Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) pela disponibilização dos dados técnicos públicos, fundamentais para o desenvolvimento deste trabalho. À minha empresa, SLB, deixo meu sincero reconhecimento pelo suporte tecnológico, pelas ferramentas disponibilizadas e pelo incentivo constante ao crescimento profissional e acadêmico.

Sou imensamente grata a todos os amigos da faculdade, do trabalho e da vida, que me inspiraram, me deram forças e acreditaram que chegaríamos até aqui. Um agradecimento especial ao meu parceiro de vida e amor, Matheus Ferraz, por estar sempre ao meu lado, torcendo, apoiando e criando um ambiente de serenidade para que tudo pudesse acontecer de forma leve.

Às minhas queridas amigas Adna Vasconcelos, Fernanda Alves, Fernanda Gois, Luiza Fonseca Katiane Souza, obrigada por me acompanharem, me impulsionarem e pela paciência durante os períodos em que precisei me ausentar. Aos meus amigos Bruno Macedo e Davi Lima Almeida por me ouvirem sempre e me lembrarem que eu era capaz. Agradeço profundamente ao Danilo Ferreira pela ajuda técnica, pelas revisões dos textos que foram fundamentais, e por ser uma pessoa e profissional admirável que sempre terei como inspiração.

Por fim, agradeço a Deus, pela dádiva da vida, pelas bênçãos concedidas e por me guiar mesmo quando tudo parecia impossível.

RESUMO

Esta tese de doutorado propõe um fluxo metodológico integrado para a caracterização quantitativa e modelagem tridimensional do teor de Carbono Orgânico Total (COT) em rochas geradoras, com foco na Formação Pimenteiras, localizada na Bacia do Parnaíba, Brasil. O estudo combina técnicas geofísicas, geoquímicas e computacionais, com o objetivo de aprimorar a compreensão espacial das propriedades das rochas geradoras em um sistema petrolífero atípico fortemente influenciado por intrusões ígneas. A pesquisa teve início com a definição de eletrofácies derivadas dos perfis de poço, por meio de *machine learning* não supervisionado (*K-Means*) e classificação Bayesiana, o que permitiu identificar variações verticais significativas nos intervalos de folhelhos, associadas a mudanças de impedância acústica e conteúdo orgânico. Em seguida, modelos de *machine learning* supervisionado, *Random Forest*, *XGBoost*, *CatBoost*, *Multi-layer Perceptron* e *Support Vector Machine*, foram treinados com dados laboratoriais de COT provenientes de poços distribuídos pela bacia, alcançando coeficientes de correlação superiores a 0.7. Esses modelos possibilitaram estimar curvas contínuas de COT em poços sem análises laboratoriais, ampliando o controle espacial da variável. O dado sísmico foi pré-condicionado a partir de uma abordagem híbrida que combina os métodos *Dip-Steered Median Filter* (DSMF) e *Spectral Blueing* (SSB), resultando em um aumento expressivo da relação sinal/ruído e em ganho de resolução vertical de aproximadamente 38%. A análise espectral confirmou a ampliação do conteúdo de altas frequências, fator essencial para a detecção de camadas delgadas. A amarração sísmica aos poços apresentou correlação elevada entre o traço sísmico e o sintético e menor valor residual, validando a melhoria obtida com o pré-condicionamento. Três métodos de inversão sísmica, determinístico (*model-based*), colorido e estocástico, foram aplicados para a obtenção de volumes de impedância acústica relativa e absoluta. A inversão estocástica apresentou o melhor desempenho, preservando variações locais sutis e apresentando forte correlação com zonas de maior teor de COT. A modelagem tridimensional do COT foi realizada por meio do algoritmo *Embedded Model Estimator* (EMBER), que combina geoestatística e *machine learning* para representar a variabilidade espacial não estacionária. O modelo 3D resultante mostrou coerência geológica e estatística onde foi possível definir os intervalos contínuos e ricos em COT. A análise de incerteza (P10, P50 e P90) indica maior variabilidade nas regiões distantes do controle de poços e nas áreas afetadas por soleiras ígneas. O fluxo de trabalho desenvolvido demonstrou elevada capacidade preditiva e adaptabilidade a contextos geológicos complexos. Este estudo representa um avanço significativo na caracterização de rochas geradoras em bacias intracratônicas, ao transpor metodologias tradicionalmente aplicadas à caracterização de reservatórios para o domínio das rochas fonte. Além disso, estabelece uma estrutura quantitativa e reprodutível para a integração entre dados de poços, sísmica e *machine learning* na avaliação de sistemas petrolíferos atípico ou de evolução térmica complexa.

Palavras-chave: Caracterização de rochas geradoras; Carbono orgânico total; Bacia do Parnaíba; Formação Pimenteiras; Inversão sísmica; Machine learning; EMBER.

ABSTRACT

This doctoral research proposes an integrated methodological workflow for the quantitative characterization and three-dimensional modeling of Total Organic Carbon (TOC) in source rocks, with a focus on the Pimenteiras Formation of the Parnaíba Basin, Brazil. The study combines geophysical, geochemical, and computational techniques to improve the spatial understanding of source rock properties in an atypical petroleum system influenced by igneous intrusions. The research begins with the definition of electrofacies derived from well-log data using unsupervised machine learning (K-Means) and Bayesian classification, which demonstrated significant vertical variations in shale intervals related to changes in acoustic impedance and organic content. Supervised machine learning models, Random Forest, XGBoost, CatBoost, Multi-layer Perceptron e Support Vector Machine, were then trained with TOC laboratory data from wells distributed across the basin, achieving correlation coefficients above 0.8 in most cases. These models allowed the prediction of continuous TOC curves in wells lacking laboratory analyses. The seismic dataset was preconditioned using a hybrid approach that combines Dip-Steered Median Filter (DSMF) and Spectral Blueing (SSB), resulting in an improved signal-to-noise ratio and a vertical resolution gain of approximately 38%. The frequency spectrum analysis confirmed the enhancement of high-frequency content, which was critical for the detection of thin layers. Subsequent well-to-seismic ties demonstrated better correlation with synthetic seismograms, validating the quality improvement of the preconditioned dataset. Three seismic inversion techniques, deterministic (model-based), colored, and stochastic, were applied to obtain relative and absolute acoustic impedance volumes. The stochastic inversion produced the most detailed results, preserving thin-layer variability and correlating well with zones of higher TOC. The integrated modeling of TOC was carried out using the Embedded Model Estimator (EMBER) algorithm, which combines geostatistics and machine learning to represent non-stationary spatial variability. The resulting 3D TOC model showed strong geological coherence, where it was possible to define continuous intervals rich in COT. Uncertainty analysis (P10, P50, P90) indicating greater uncertainty near intrusive zones and in areas distant from well control. The proposed workflow demonstrated high predictive capability and adaptability to complex geological settings. It advances the state of the art in source rock studies by transferring methodologies traditionally applied to reservoir characterization into the domain of petroleum system source analysis. Moreover, it establishes a quantitative and reproducible framework for integrating well, seismic, and machine learning data in the evaluation of unconventional or thermally complex petroleum systems.

Keywords: Source rock characterization; Total organic carbon; Parnaíba Basin; Pimenteiras Formation; Seismic inversion; Machine learning; EMBER.

Sumário

1. Introdução à Tese	26
1.1. INTRODUÇÃO	26
1.2. MOTIVAÇÃO E JUSTIFICATIVA	27
1.3. HIPÓTESES DE PESQUISA	28
1.4. OBJETIVOS	28
1.4.1. Objetivo Geral.....	28
1.4.2. Objetivos Específicos.....	28
1.5. ASPECTOS INOVADORES E CONTRIBUIÇÕES	28
1.6. ESTRUTURA DA TESE	29
2. Contexto Geológico	32
2.1. BACIA DO PARNAÍBA	32
2.2. ARCABOUÇO TECTÔNICO	32
2.2.1. Lineamento Transbrasiliano (LTB).....	34
2.3. ESTRATIGRAFIA DA BACIA DO PARNAÍBA	35
2.4. FORMAÇÃO PIMENTEIRAS	38
2.5. INTRUSÕES ÍGNEAS E O SISTEMA PETROLÍFERO ATÍPICO DA BACIA DO PARNAÍBA.....	40
3. Rochas Geradoras	45
3.1. INTRODUÇÃO	45
3.2. ORIGEM DA MATÉRIA ORGÂNICA.....	46
3.2.1. Fontes de Matéria Orgânica	46
3.3. FORMAÇÃO E TRANSFORMAÇÃO DO QUEROGÊNIO	51
3.4. TEOR DE CARBONO ORGÂNICO TOTAL.....	55
3.5. CARACTERIZAÇÃO GEOQUÍMICA DA FORMAÇÃO PIMENTEIRAS	56
4. Classificação de Fácies e Estudo de Viabilidade.....	60
4.1. INTRODUÇÃO	60
4.2. FUNDAMENTAÇÃO	60
4.2.1. Clusterização <i>K-Means</i>	61
4.2.3. Classificação Bayesiana.....	64
4.3. MÉTODO	68
4.3.1. Definição de Fácies via <i>K-Means</i>	68
4.3.2. Estudo de Viabilidade via Classificação Bayesiana.....	78
4.4. RESULTADOS E DISCUSSÕES.....	80
4.4.1. Integração dos Métodos K-means e Classificação Bayesiana no Estudo de Viabilidade	80

4.4.2. Interpretação dos <i>Clusters</i> em Eletrofácies.....	84
4.5. CONCLUSÃO	92
5. Machine Learning Aplicado a Estimativa de COT	94
5.1. INTRODUÇÃO	94
5.2. FUNDAMENTAÇÃO	96
5.2.1. Técnicas de <i>Machine Learning</i>	96
5.2.3. Métricas de Desempenho	110
5.2.4. Importância das Variáveis	113
5.4. MÉTODO	114
5.4.1. Seleção do conjunto de dados	115
5.4.2. Pré-condicionamento dos dados	117
5.4.3. Análise exploratória dos Dados	119
5.4.4. Transformação dos dados.....	123
5.4.5. Separação dos Dados	127
5.4.6. Treinamento dos modelos	127
5.4.7. Estimativa no conjunto de dados e <i>ensemble</i> dos modelos selecionados	131
5.4.8. Pós-processamento dos dados preditos	131
5.5. RESULTADOS	133
5.5.1 Treinamento dos modelos <i>machine learning</i>	133
5.5.3. Estimativa de COT no poço de validação	138
5.6. DISCUSSÕES	147
5.7. CONCLUSÃO	152
6. Pré-condicionamento do Dado Sísmico	155
6.1. INTRODUÇÃO	155
6.2. FUNDAMENTAÇÃO	156
6.2.1. DIP STEERED MEDIAN FILTER.....	156
6.2.2. SPECTRAL BLUEING	158
6.3. MÉTODO	160
6.3.1. DIP STEERED MEDIAN FILTER.....	160
6.3.2. SPECTRAL BLUEING	162
6.5. RESULTADOS E DISCUSSÕES	174
6.7. CONCLUSÃO	181
7. Análise do Espectro de Frequência e Estudo de Resolução Sísmica	183
7.1. INTRODUÇÃO	183
7.2. METODOLOGIA	183

7.2.1 Espectro de frequência.....	183
7.2.2. Continuous Wavelet Transform (CWT).....	184
7.2.3. Estudo de resolução sísmica	185
7.3. RESULTADOS	190
7.3.1. Análise do espectro de frequência.....	190
7.3.2. Espectrograma CWT.....	191
7.3.3. Estimativa do ganho de resolução no dado condicionado	193
7.5. CONCLUSÃO	199
8. Amarração Sísmica-Poço e a Construção do Grid 3D.....	201
8.1. INTRODUÇÃO	201
8.2. FUNDAMENTAÇÃO	201
8.2.1. Escalas e <i>Upscaling</i> de Velocidades	202
8.2.2. Calibração de Perfis de Velocidade.....	202
8.2.3. Extração e Ajuste da <i>Wavelet</i>	203
8.2.4. Construção do Sismograma Sintético	205
8.3. MÉTODO	206
8.3.1. Dado Sísmico Original.....	210
8.3.2. Dado Sísmico após o <i>Dip Steered Median Filter</i>	214
8.3.1. Dado Sísmico após o <i>Spectral Blueing</i>	216
8.4. RESULTADOS E DISCUSSÕES.....	219
8.4.1. Amarração Sísmica-Poço	219
8.4.2. Interpretação Sísmica.....	222
8.4.3. Construção do Grid 3D	224
8.5. CONCLUSÃO	225
9. Inversões Sísmicas.....	227
9.1. INTRODUÇÃO	227
9.2. FUNDAMENTAÇÃO	228
9.2.1. Inversão Colorida.....	228
9.2.2. Inversão Determinística	229
9.2.3. Inversão Estocástica.....	234
9.2.4. Modelagem Geostatística.....	238
9.3. MÉTODO	241
9.3.1. Inversão Colorida.....	241
9.3.2. Construção do modelo de baixa frequência	246
9.3.3. Inversão Determinística	248

9.3.4. Inversão Estocástica.....	251
9.4. RESULTADOS	256
9.4.1. Inversão Colorida.....	256
9.4.2. Inversão Determinística	262
9.4.3. Inversão Estocástica.....	270
9.5. DISCUSSÕES	288
9.6. CONCLUSÕES	290
10. Modelagem 3D do Carbono Orgânico Total	292
10.1. INTRODUÇÃO	292
10.2. FUNDAMENTAÇÃO	294
10.2.1. Contexto e evolução conceitual	294
10.2.2. Formulação em campos aleatórios condicionais (CRF)	295
10.2.3. Estimção da distribuição condicional via <i>Quantile Random Forests</i>	296
10.2.4. Modelos embutidos (<i>embedding</i>) e validação cruzada	296
10.2.5. Envelope e simulação condicional.....	297
10.3. MÉTODO	298
10.3.1. Definição das variáveis de entrada.....	299
10.3.2. Etapas de aplicação do EMBER	302
10.4. RESULTADOS	303
10.5. DISCUSSÕES	318
10.6. CONCLUSÃO	321
11. Conclusões	324
11.1 SÍNTESE DOS PRINCIPAIS RESULTADOS	324
11.2 CONTRIBUIÇÕES E AVANÇOS METODOLÓGICOS	325
11.3 LIMITAÇÕES E PERSPECTIVAS FUTURAS	326
11.4 CONSIDERAÇÕES FINAIS.....	326
12. Referências	328

Índice de Figuras

Capítulo 2 32

Figura 2.1. Mapa de localização da Bacia do Parnaíba, onde destaca-se a profundidade do embasamento, a geometria típica de uma sinéclise (ovalada) e os limites da bacia que representam o principal controle estrutural da variação do depocentro. O ringfence do Parque dos Gaviões, principal produtora de gás, também está evidenciada no mapa. Modificado de Góes et al., 1990 (Miranda et al 2018). 32

Figura 2.2. Seção geológica representando o arcabouço estrutural da Bacia do Parnaíba e seu preenchimento sedimentar segundo a direção WNW-ESE modificado de Góes et al., 1990 (Castro et al. 2014). 33

Figura 2.3. Carta estratigráfica da Bacia do Parnaíba segundo Vaz et al. (2007). 36

Figura 2.4. Interpretação do Rodrigues (1995) quando aos intervalos de folhelhos radioativos do trato de sistemas transgressivos na Formação Pimenteiras no poço 1-PAF-7-MA. 40

Figura 2.5. Exemplo da formação de um sistema petrolífero atípico cujo petróleo se origina de uma rocha geradora que atinge a maturação por meio do calor fornecido por uma intrusão ígnea (Magoon e Dow, 1994). 40

Figura 2.6. Seção geológica esquemática mostrando o modelo para o principal mecanismo de trapeamento para acumulação comercial devido aos saltos das soleiras de diabásio que selam o reservatório ao criar estruturas de fechamento em quatro vias. Além disso, também é destacado as soleiras planas mais finas na base da seção responsáveis por desencadear a maturação da rocha geradora (Miranda et al. 2018). 41

Figura 2.7. Esquema da rede de distribuição 3D de magma que flui de níveis profundos para rasos na crosta. Nesta imagem também pode-se observar dos diferentes formatos que os corpos intrusivos podem assumir (Jerram, D. Pertord, N., 2011). 43

Figura 2.8. Esboço mostrando a configuração das unidades de fácies de soleira identificadas em seções sísmicas interpretadas por Planke, S.; Rasmussen, T.; Rey, S. S.; Myklebust, R. (2005). 44

Capítulo 3 45

Figura 3.1. (a) Ambientes deposicionais favoráveis à formação de rochas geradoras e reservatórios, como depressões marinhas com baixa circulação, que favorecem a preservação da matéria orgânica. (b) Migração de petróleo após soterramento e maturação. Recifes formam armadilhas estratigráficas, enquanto arenitos associados a falhas compõem armadilhas estruturais (Bjorlykke, 2010). 48

Figura 3.2. Degradação da matéria orgânica em ambientes óxicos e anóxicos. Em condições anóxicas, a preservação da matéria orgânica é favorecida pela redução da bioturbação e da atividade bacteriana aeróbica, comuns em ambientes oxigenados. (Allen & Allen, 2013). 50

Figura 3.3. Representação esquemática da evolução do querogênio no diagrama de van Krevelen, no qual são indicados os estágios de diagênese, catagênese e metagênese, juntamente com os principais

produtos gerados em cada fase do processo de maturação térmica da matéria orgânica (Tissot & Welte, 1984).....	53
Figura 3.4. Esquema geral da formação de hidrocarbonetos em função do soterramento, conforme Tissot e Welte (1978). Com o aumento da temperatura devido ao soterramento progressivo, o material orgânico passa por estágios distintos: imaturo, geração de óleo, craqueamento do óleo (estágio de gás úmido) e, por fim, geração de gás seco. A figura apresenta as distribuições típicas de n-alcenos, cicloalcenos e compostos aromáticos em três fases dessa evolução.	54
Figura 3.5. Classificação dos folhelhos radioativos interpretados por René Rodrigues (1995) em dois poços da Bacia do Parnaíba.....	57
Figura 3.6. Resultados obtidos por Mussa (2020) em poços da região norte da bacia, evidenciando a variação do teor de COT de maneira geral e em profundidade, bem como sua relação com as soleiras de diabásio.....	58

Capítulo 4 60

Figura 4.1. Representação esquemática do funcionamento do algoritmo <i>K-means</i> . (a) Conjunto de dados bidimensionais com três agrupamentos potenciais; (b) seleção inicial de três centroides e atribuição preliminar dos pontos aos respectivos <i>clusters</i> ; (c) e (d) etapas intermediárias do processo iterativo, nas quais os rótulos de <i>cluster</i> e as posições dos centroides são atualizados a cada iteração; (e) resultado final da clusterização, mostrando a configuração estável alcançada após a convergência do algoritmo (Jain, A.K 2009).....	63
Figura 4.2. Exemplo ilustrativo de uma partição do espaço amostral em subconjuntos mutuamente exclusivos, utilizado para explicar a aplicação do Teorema de Bayes na classificação probabilística (Bussab, W. D. O., & Morettin, P. A. (2010)).....	65
Figura 4.3. Exemplo de funções de densidade de probabilidade condicionais a diferentes classes, ilustrando a probabilidade de observação de um valor específico da variável x , dado que a amostra pertence à categoria ω . Essas distribuições representam verossimilhanças associadas às hipóteses de classificação, constituindo a base para a aplicação do Teorema de Bayes no processo decisório (Duda <i>et al.</i> (2001)).	66
Figura 4.4. Probabilidades a posteriori calculadas para valores específicos de probabilidades a priori $P\omega_1$ e $P\omega_2$, considerando as densidades de probabilidade condicionais de classe apresentadas na Figura 3 (Duda <i>et al.</i> (2001)).....	67
Figura 4.5. Fluxo de trabalho integrando as técnicas <i>K-means</i> e classificação Bayesiana.	68
Figura 4.6. Suite básica dos perfis do poço 2-CP-1-MA após o pré-condicionamento. O comportamento das curvas sugere a existência de diferentes eletrofácies, e por isso os são candidatos a compor o conjunto de dados para a clusterização não supervisionada.....	70
Figura 4.7. Dados disponíveis no poço 1-OGX-16-MA após o pré-condicionamento. A curva de <i>gamma-ray</i> diferencia de forma evidente os folhelhos das soleiras de diabásio, e o comportamento geral dos perfis elétricos refletem a heterogeneidade presente no intervalo da Fm. Pimenteiras.	71
Figura 4.8. Conjunto de perfis medidos e preconditionados no poço 1-OGX-22-MA, onde ocorreu intrusões não somente na porção superior da formação, mas também na base da mesma. A interpretação	

dos perfis mostra um comportamento semelhante as demais acusando uma possível diversidade de eletrofácies a serem investigadas.....	72
Figura 4.9. Exibição do mesmo conjunto de curvas para o poço 3-OGX-57-MA, os quais foram condicionados e selecionados para testes de seleção de variáveis importantes para a clusterização <i>K-means</i>	73
Figura 4.10. Pairplot das variáveis GR, NPHI, DTCO e PEFZ após o pré-processamento e padronização (z-score). As distribuições individuais são mostradas na diagonal principal, enquanto os crossplots representam as relações bivariadas entre os atributos. Os dados estão coloridos de acordo com os diferentes poços analisados.	75
Figura 4.11. Exemplificação do método de <i>Elbow</i> e sua interpretação voltada para a escolha do número ideal de clusters. (Fonte: https://dchandra.com/machine%20learning/2018/12/16/K-means-Clustering-Algorithm-using-scikit-learn.html).	76
Figura 4.12. Pseudocódigo detalhando as etapas realizadas na clusterização <i>K-means</i> executadas utilizando a linguagem <i>python</i>	78
Figura 4.13. Etapas realizadas na classificação Bayesiana 1D no formato pseudocódigo o qual foi utilizando a linguagem <i>python</i>	79
Figura 4.14. Aplicação do método de <i>Elbow</i> a partir do conjunto de dados selecionados, indicando o número ideal de <i>clusters</i> , o qual está sendo apontado pela seta preta, a partir do qual a curva começa a estabilizar.....	80
Figura 4.15. Relação entre <i>gamma ray</i> e impedância acústica para diferentes testes de número de <i>clusters</i> , ilustrando a separação progressiva dos agrupamentos representando as diferentes eletrofácies.	81
Figura 4.16. Funções de densidade de probabilidade dos <i>clusters</i> definidos pelo <i>K-means</i> , mostrando a distribuição dos valores de GR e AI para cada agrupamento.	82
Figura 4.17. Funções de densidade e probabilidades <i>a posteriori</i> obtidas pela Classificação Bayesiana, evidenciando as diferenças e sobreposições entre os <i>clusters</i>	83
Figura 4.18. Matriz e barra de confusão da Classificação Bayesiana, apresentando os índices de acerto e os principais casos de confusão entre as eletrofácies representadas ainda pelos <i>clusters</i> obtidos do algoritmo <i>K-means</i>	84
Figura 4.19. Resultado das fácies classificadas pelo <i>K-means</i> e pela classificação Bayesiana, junto aos perfis utilizados no processo.	86
Figura 4.20. Resultado das fácies classificadas pelo <i>K-means</i> e pela classificação Bayesiana, junto aos perfis utilizados no processo.	87
Figura 4.21. Resultado das fácies classificadas pelo <i>K-means</i> e pela classificação Bayesiana, junto aos perfis utilizados no processo.	88
Figura 4.22. Resultado das fácies classificadas pelo <i>K-means</i> e pela classificação Bayesiana, junto aos perfis utilizados no processo.	89
Figura 4.23. Distribuições de Tório (Th), Potássio (K) e Urânio (U) para as diferentes fácies definidas no poço 2-CP-1-MA, reforçando a interpretação dada a clusterização obtida. É possível avaliar as variações das contribuições espectrais de <i>gamma ray</i> em profundidade que destaca principalmente os altos valores de U em alguns intervalos na porção superior da formação. Os <i>crossplots</i> da relação entre U e Th, e U e K, evidenciam as diferentes concentrações desses elementos por fácies.	90

Figura 4.24. Distribuições dos elementos Th, K e U para as fácies identificadas no poço 1-OGX-16-MA, reforçando a consistência da interpretação da clusterização. A análise espectral do GR evidencia variações significativas em profundidade, com destaque novamente, para concentrações elevadas de U em intervalos específicos da porção superior da formação, associadas a Fácies 2 como pode ser melhor demonstrado pelos crossplots entre U e Th, e entre U e K, contribuindo para a diferenciação entre folhelhos. 91

Capítulo 5 94

Figura 5.1. Fluxo conceitual proposto por Polikar 2012 para combinação de diferentes algoritmos de *machine learning*. 96

Figura 5.2. Esquema representativo do funcionamento do algoritmo Random Forest, ilustrando a construção de múltiplas árvores de decisão e a combinação dos resultados por média ou voto majoritário (Fonte: <https://www.scaler.com/topics/random-forest-in-r/>) 98

Figura 5.3. Representação do processo de *gradient boosting*, mostrando a adição sequencial de árvores para reduzir os erros residuais e aprimorar o modelo preditivo (Chen & Guestrin, 2016) 101

Figura 5.4. Arquitetura típica de uma rede neural Multi-Layer Perceptron (MLP), composta por camadas de entrada, ocultas e saída (modificado de Mandal *et al.*, 2021)..... 105

Figura 5.5. Ilustração do conceito de hiperplano ótimo em Support Vector Machine (SVM), evidenciando a maximização da margem entre classes (Fonte: <https://towardsdatascience.com/support-vector-machine-with-scikit-learn-a-friendly-introduction-a2969f2ff00d/>) 108

Figura 5.6. Fluxo de trabalho adotado para a estimativa do COT nos poços da Formação Pimenteiras. 114

Figura 5.7. Localização dos poços utilizados neste estudo, destacando em azul a área delimitada pela aquisição sísmica 3D. 115

Figura 5.8. Perfis geofísicos de GR, DT e RES, após o pré-condicionamento, em poços externos à área de estudo, juntamente com as amostras de COT disponíveis..... 118

Figura 5.9. Perfis geofísicos de GR, DT e RES, após o pré-condicionamento, nos poços da área de estudo, com destaque para o poço 2-CP-1-MA, único com medições de COT. 119

Figura 5.10. Histogramas das amostras de COT comparando a distribuição estatística para cada poço. 120

Figura 5.11. Análise conjunta de gráficos crossplot entre as variáveis e variável alvo coloridas por poços..... 122

Figura 5.12. Correlação linear (Pearson) e não linear (Spearman) entre os perfis de poço (GR, DT e RES) e o COT, evidenciando baixa correlação direta. 122

Figura 5.13. Relação entre variáveis de entrada (GR, DT e RES) coloridas de acordo com o valor de COT. 123

Figura 5.14. Distribuição dos valores do perfil de resistividade antes e após a transformação logarítmica aplicada..... 125

Figura 5.15. Distribuição dos dados após padronização com *StandardScaler*, demonstrando a homogeneização das escalas entre variáveis. 126

Figura 5.16. Esquema do método de validação cruzada (cross-validation) para otimização dos hiperparâmetros (Fonte: https://scikit-learn.org/stable/modules/cross_validation.html).....	128
Figura 5.17. Exemplo de validação cruzada K-Fold aplicada a um conjunto de dados, no qual este é dividido em subconjuntos destinados ao treinamento, à validação e ao teste do modelo. (Fonte: https://scikit-learn.org/stable/modules/cross_validation.html).....	129
Figura 5.18. Coeficientes de correlação e determinação por poço e por método.....	133
Figura 5.19. Crossplot correspondente as correlações entre o COT predito e medido por cada método de <i>machine learning</i> e pelo <i>ensemble</i> dos métodos de melhor performance. Os gráficos representam o conjunto de dados completo (teste e treino).	134
Figura 5.20. Exemplo das curvas estimadas por métodos em dois poços do conjunto de dados de treinamento.	135
Figura 5.21. Métricas estatísticas correspondente ao conjunto de dados de teste para avaliação da performance de cada método utilizado.	136
Figura 5.22. Avaliação da performance para o método <i>Random Forest</i> para o conjunto de dados teste, onde é observado os coeficientes de correlação evidenciando o ponto de maior erro, os valores residuais por amostra e a distribuição de densidade do residual.	137
Figura 5.23. Avaliação da performance para o método <i>XGBoost</i> para o conjunto de dados teste, onde é observado os coeficientes de correlação evidenciando o ponto de maior erro, os valores residuais por amostra e a distribuição de densidade do residual.....	137
Figura 5.24. Avaliação da performance para o método <i>Catboost</i> para o conjunto de dados teste, onde é observado os coeficientes de correlação evidenciando o ponto de maior erro, os valores residuais por amostra e a distribuição de densidade do residual.....	138
Figura 5.25. Visão geral sobre a performance dos métodos selecionados (RF, XGB e CB) e a combinação dos mesmos. É exibido os valores de coeficientes de correlação e determinação por método, bem como, os gráficos crossplots individualizados e o combinado.....	139
Figura 5.26. Visão geral sobre a performance dos métodos selecionados (RF, XGB e CB) e a combinação dos mesmos considerando os poços 1-OGX-101-MA e 1-OGX-93-MA os quais foram retirados do conjunto de dados de treinamento na etapa de análise exploratória. É exibido os valores de coeficientes de correlação e determinação por método, bem como, os gráficos crossplots individualizados e o combinado evidenciando o impacto gerado ao inserir dados não coerentes comparados o conjunto geral.....	140
Figura 5.27. Curvas de COT estimadas pelos métodos selecionados durante a análise de performance e o resultado da combinação dos três métodos.....	141
Figura 5.28. Painel de controle da correção da curva de COT estimada. Em a) curvas interpoladas nos intervalos das amostras de COT medidos no poço de validação (2-CP-1-MA), entre o dado real, predito e o predito corrigido. Em b) os valores residuais por amostra para o dados estimados antes e após a correção.	143
Figura 5.29. Curva de COT estimado após a correção (preto) e antes da correção (rosa) mostrando um melhor ajuste. O perfil de fácies é exibido como uma avaliação complementar para evidenciar a variação de COT por fácies. O perfil booleano corresponde a máscara feita pelo polígono vermelho no crossplot entre o COT estimado e medido. A máscara foi criada contemplando as amostras com valores superestimados pelo <i>machine learning</i>	144

Figura 5.30. Curvas de COT previstas nos poços da área de estudo.	145
Figura 5.31. Histogramas da distribuição de COT por fácies, destacando as suas diferentes concentrações, onde nota-se que ocorre uma maior sobreposição nas fácies 2 e 3, enquanto que para as fácies 4 há um predomínio de valores relativamente mais altos.	146
Figura 5.32. Exibição das variáveis utilizadas para estimar a curva de COT a partir do modelo <i>ensemble</i> , bem como do resultado obtido para o COT estimado junto a discretização de diferentes fácies definidas para o intervalo de estudo. Além disso, também é exibido o perfil de Caliper reforçando a qualidade dos poços, mostrando que para alguns intervalos é observado desmoronamento não tão importantes.	147
Capítulo 6	155
Figura 6.1. Visão geral do fluxo de trabalho de condicionamento aplicado aos dados sísmicos. .	156
Figura 6.2. Comparação entre os dados com e sem <i>steering cube</i> , representando os traços sísmicos com informações de inclinação e azimute (linha tracejada) (Tingdahl, Kristofer & de Groot, Paul., 2003)..	157
Figura 6.3. (a) Espectro do sinal branco, (b) espectro do sinal vermelho e (c) espectro do sinal azul proporcionais a $f\beta$. O espectro plano do sinal branco demonstra que a energia é distribuída igualmente entre todas as frequências. Em contraste, para os sinais vermelho e azul, a energia é predominantemente concentrada em frequências baixas e altas, respectivamente (Walden e Hosken, 1985).	159
Figura 6.4. Fluxo de trabalho da técnica Dip Steered Median Filter.....	160
Figura 6.5. Atributo de volume de <i>dip</i> sobre a amplitude sísmica com transparência, destacando as maiores variações de mergulho dos refletores sísmicos. Em a) o componente de mergulho na direção da <i>crossline</i> é exibido sobre uma <i>inline</i> e em b) o componente de mergulho na direção da <i>inline</i> é exibido sobre uma <i>crossline</i> . O mesmo é demonstrado em uma <i>timeslice</i>	161
Figura 6.6. Fluxo de trabalho da técnica Spectral Blueing (Blache-Fraser e Neep, 2004).	163
Figura 6.7. A definição de intervalos de tempo para a modelagem individual de cada operador, considerando a variação vertical da frequência do dado sísmico. A interpretação das principais sequências dentro da bacia serve como referência eficaz para a delimitação de cada zona.....	164
Figura 6.8. Dados de entrada utilizados para modelar o operador da primeira zona, onde em (a) pode ser observado o espectro de frequência dos quarenta traços sísmicos selecionados aleatoriamente, em (b) o espectro sísmico médio e o mesmo suavizado como mostrado pelo traço em preto, em (c) as séries de refletividade derivadas de cada poço, e por último, em (d) é exibido o espectro de frequência médio dos poços, o qual foi filtrado em 250 Hz, e também a curva de ajuste do espectro (em preto).	165
Figura 6.9. Construção do operador para a primeira zona, em a) os cortes de frequência baseado nos espectros do dado de entrada e saída, em b) o resultado do operador e em c) o espectro de frequência do sinal modelado.....	166
Figura 6.10. Comparação entre o dado sísmico original e após a aplicação do spectral blueing. É destacado no retângulo tracejado preto como é melhorado a continuação lateral dos refletores, bem como a resolução após a aplicação do fluxo do SBB. Também é evidenciado pelo círculo preto tracejado um efeito chamado de ringing (reverberação) que pode ocorrer nesse tipo de técnica.....	168

Figura 6.11. Modelagem do operador para a segunda zona. Em a) o espectro de frequência dos quarenta traços sísmicos em b) o espectro sísmico médio, em c) as séries de refletividade derivadas de cada poço, e em (d) o espectro de frequência médio dos poços com a curva de ajuste do espectro (em preto).	169
Figura 6.12. Construção do operador para a segunda zona, em a) os cortes de frequência baseado nos espectros do dado de entrada e saída, em b) o resultado do operador e em c) o espectro de frequência do sinal modelado.	170
Figura 6.13. Os resultados preliminares da convolução utilizando o operador modelado para a segunda zona indicam seções sísmicas que passam pelos poços. Isso permite uma avaliação da correlação entre os traços sísmicos extraídos dos poços e os refletores sísmicos. Além disso, permite uma comparação dos dados antes e depois do SSB, destacando as melhorias dentro do retângulo preto tracejado.	171
Figura 6.14. Modelagem do operador para a terceira zona. Em a) o espectro de frequência dos quarenta traços sísmicos em b) o espectro sísmico médio, em c) as séries de refletividade derivadas de cada poço, e em (d) o espectro de frequência médio dos poços com a curva de ajuste do espectro (em preto).	172
Figura 6.15. Construção do operador para a terceira zona, em a) os cortes de frequência baseado nos espectros do dado de entrada e saída, em b) o resultado do operador e em c) o espectro de frequência do sinal modelado.	172
Figura 6.16. Os resultados preliminares da convolução utilizando o operador modelado para a terceira zona são apresentados, mostrando seções sísmicas que passam pelos poços. Isso nos permite avaliar a correlação entre os traços sísmicos extraídos dos poços e os refletores sísmicos. Além disso, podemos comparar os dados antes e depois do SSB, destacando as melhorias indicadas pelo retângulo preto tracejado.	174
Figura 6.17. Resultado da aplicação da primeira etapa do condicionamento. É possível notar a atenuação significativa do ruído no dado após o DSMF em comparação com o dado sísmico original. A linha azul tracejada sobre o <i>time slice</i> identifica a localização da <i>inline</i> mostrada, e o triângulo azul identifica por onde passa a <i>time slice</i> exibida como referencial.	175
Figura 6.18. O volume de amplitude residual foi extraído da diferença entre os dados sísmicos original e o dado filtrado pelo DSMF, o que revelou principalmente uma redução de ruído e, em menor grau, a filtragem de alguns eventos sísmicos na direção NW-SE.	176
Figura 6.19. Comparação de dados sísmicos após o DSMF e o SSB em <i>time slices</i> e em seções sísmicas, ilustrando melhorias na resolução.	178
Figura 6.20: Interpretação das melhorias promovidas pela técnica <i>Spectral Blueing</i>	180
Figura 6.21. Análise dos dados sísmicos pré-condicionado (track 1) e original (track 3) juntos as fácies da Fm. Pimenteiras (track 2).	181

Capítulo 7 183

Figura 7.1. O limite de resolução segundo Rayleigh é atingido quando dois eventos sísmicos estão separados por um intervalo de tempo correspondente à distância entre o pico e o vale. Já o critério de Ricker define o limite de resolução como o intervalo de tempo igual a separação entre os pontos de inflexão da <i>wavelet</i> (Kallweit e Wood, 1982).	186
---	-----

Figura 7.2. Como observado por Widess (1973) a medida que a separação entre os picos diminui uma wavelet convolvida com dois picos de igual amplitude e polaridade oposta converge para a derivada da wavelet (Kallweit e Woods, 1982).....	187
Figura 7.3. Modelo simples de interferência, onde é apresentado em a) o modelo de cunha com reflexões de polaridades opostas, em b) o <i>crossplot</i> entre a amplitude do topo contra a espessura aparente, em c) o <i>crossplot</i> entre a espessura real contra a espessura aparente. Por fim em c) é exibido a <i>wavelet</i> utilizada para gerar os traços sísmicos (Simm e Bacon, 2014).	190
Figura 7.4. Comparação dos espectros de frequência dos dados de entrada e saída em cada etapa de pré-condicionamento. No painel a), os espectros de frequência dos dados originais são exibidos em verde, enquanto os dados pós-DSMF são mostrados em azul, ilustrando o efeito de filtragem em altas frequências. No painel b), a comparação o espectro dos dados após DSMF com os dados após SSB, representados em vermelho, o que demonstra a recuperação de energia em altas frequências. Por fim, o painel c) fornece uma comparação abrangente do espectro de frequência para todos os três conjuntos de dados sísmicos.	191
Figura 7.5. Espectrograma estimado pela técnica da Transformada <i>Wavelet</i> Contínua (CWT), em a) para o dado original em b) para o dado após o DSMF e em c) para o dado após o SSB.	192
Figura 7.6. <i>wavelets</i> estatísticas médias extraídas nas posições dos poços, em verde do dado original e em vermelho do dado após o SSB , juntamente com seus respectivos espectros de frequência.....	194
Figura 7.7. Camadas utilizadas no modelo de cunha e suas propriedades elásticas, incluindo valores de V_p , V_s e densidade.	195
Figura 7.8. Modelos de cunha com suas propriedades elásticas, que servirão de base para a geração do sismograma sintético.	196
Figura 7.9. Os resultados do sismograma sintético gerado para o modelo de cunha são apresentados com base nas frequências dos dados sísmicos antes e depois do fluxo de trabalho de pré-condicionamento na parte a). Na parte b), são apresentados gráficos que comparam as espessuras real e aparente em milissegundos e metros para os dados originais e aprimorados. Por fim, na parte c), a amplitude é exibida em relação às espessuras real e aparente, com ênfase específica na amplitude de <i>tuning</i>	197
Figura 7.10. Gráficos de densidade 2D dado pela distribuição de velocidade compressional contra a densidade por fácies, onde os maiores valores (cores quentes) representam a região de maior concentração dos dados.	198
Figura 7.11. Perfil de fácies e impedância acústica dos poços, a fim de demonstrar principalmente as espessuras de cada fácies encontradas.....	199
Capítulo 8	201
Figura 8.1. Esquema ilustrativo do processo de calibração entre o perfil sônico e o checkshot, destacando o cálculo do drift e o ajuste tempo–profundidade conforme o método descrito por Simm & Bacon (2014).	203
Figura 8.2. Representação esquemática da construção do sismograma sintético a partir da série de refletividade e da wavelet estimada, evidenciando o modelo convolucional clássico e a função de correlação cruzada utilizada para avaliação da qualidade da amarração (Simm & Bacon, 2014).	206

Figura 8.3. Fluxograma do processo de amarração sísmica–poço adotado neste estudo, adaptado de Simm & Bacon (2014).....	207
Figura 8.4. Perfis sônico (DT), densidade (RHOZ), gamma ray (GR) e caliper (CALI) após o condicionamento, destacando a remoção de pontos espúrios e a correção de intervalos afetados por instabilidade de parede nos poços 1-OGX-16-MA e 2-CP-1-MA.	208
Figura 8.5. Perfis sônico (DT), densidade (RHOZ), gama natural (GR) e caliper (CALI) após o condicionamento para os poços 1-OGX-22-MA e 3-OGX-57-MA, além de demonstrar a filtragem em frequência aplicada a todos os poços.....	208
Figura 8.6. Calibração do perfil sônico a partir dos dados de checkshot medidos nos poços 1-OGX-16-MA e 2-CP-1-MA.....	209
Figura 8.7. Amarração sísmica-poço para o poço 2-CP-1-MA utilizando o dado sísmico original...	211
Figura 8.8. Amarração sísmica-poço para o poço 1-OGX-16-MA utilizando o dado sísmico original.	212
Figura 8.9. Amarração sísmica-poço para o poço 1-OGX-22-MA utilizando o dado sísmico original.	213
Figura 8.10. Amarração sísmica-poço para o poço 3-OGX-57-MA utilizando o dado sísmico original.	213
Figura 8.11. Amarração sísmica-poço para o poço 2-CP-1-MA após aplicação do Dip-Steered Median Filter.....	214
Figura 8.12. Amarração sísmica-poço para o poço 1-OGX-16-MA após aplicação do Dip-Steered Median Filter.	215
Figura 8.13. Amarração sísmica-poço para o poço 1-OGX-22-MA após aplicação do Dip-Steered Median Filter.	215
Figura 8.14. Amarração sísmica-poço para o poço 3-OGX-57-MA após aplicação do Dip-Steered Median Filter.	216
Figura 8.15. Amarração sísmica-poço para o poço 2-CP-1-MA após aplicação do Spectral Blueing.	217
Figura 8.16. Amarração sísmica-poço para o poço 1-OGX-16-MA após aplicação do Spectral Blueing.	217
Figura 8.17. Amarração sísmica-poço para o poço 1-OGX-22-MA após aplicação do Spectral Blueing.	218
Figura 8.18. Amarração sísmica-poço para o poço 3-OGX-57-MA após aplicação do Spectral Blueing.	218
Figura 8.19. Amarração sísmica–poço para os poços 1-OGX-16-MA e 2-CP-1-MA mostrando o comportamento do refletor da Formação Pimenteiras.....	219
Figura 8.20. Amarração sísmica–poço para os poços 1-OGX-22-MA e 3-OGX-57-MA, evidenciando inversão de polaridade associada à presença de soleiras na base da Formação Pimenteiras.	220
Figura 8.21. Seção sísmica arbitrária orientada NE–SW passando pelos poços CP-1, OGX-16, OGX-22 e OGX-57, após a amarração dos poços à sísmica.....	221
Figura 8.22. Distribuição dos resíduos entre o sismograma sintético e o traço sísmico para o dado original e o dado pré-condicionado.	222
Figura 8.23. Mapeamento sísmico das superfícies de topo e base da Formação Pimenteiras.	223

Figura 8.24. Superfícies de topo e base da Formação Pimenteiras, interpoladas e suavizadas, evidenciando o efeito estrutural das intrusões ígneas.....	224
Figura 8.25. Ilustração da geometria do grid gerado para a zona do Pimenteiras bem como a disposição de camadas do grid.	225

Capítulo 9 227

Figura 9.1. Comparação entre o resultado da inversão colorida e o método sparse-spike conforme Lancaster & Whitcombe (2000). Observa-se que a Coloured Inversion produz resultados de qualidade semelhante, com significativa redução da complexidade e do tempo computacional.	229
Figura 9.2. Exemplo ilustrativo dos produtos gerados pela inversão geoestatística: (a) realização individual, (b) média das realizações e (c) desvio padrão. Observa-se que a média apresenta resposta suavizada em comparação à realização única, enquanto o desvio padrão aumenta nas regiões afastadas do controle dos poços, indicando a consistência da abordagem estocástica na reprodução da variabilidade e na honra aos dados de poço (adaptado de Lamy et al., 1999; Simm & Bacon, 2014).	235
Figura 9.3. Fluxograma do algoritmo de Simulação Gaussiana Sequencial (SGS) conforme Azevedo & Soares (2017). O diagrama ilustra a sequência de amostragem aleatória e a incorporação condicional dos traços simulados.	236
Figura 9.4. Exemplo de variograma anisotrópico utilizado no controle da continuidade lateral e vertical da impedância acústica. (Simm & Bacon, 2014).	236
Figura 9.5. Esquema da inversão geoestatística iterativa conforme Azevedo & Soares (2017), destacando o processo de simulação condicional e atualização iterativa do modelo até o atingimento do critério de correlação global.	237
Figura 9.6. Comparação esquemática entre o conteúdo espectral dos modelos de impedância derivados das inversões determinística e estocástica. O aumento do conteúdo de alta frequência temporal na solução estocástica é decorrente da incorporação do modelo de continuidade espacial durante a simulação geoestatística (adaptado de Dubrule, 2003; Azevedo, L., & Soares, A. (2017)).....	238
Figura 9.7. Modelos teóricos de variogramas e seus parâmetros principais (nugget, sill e range). Adaptado de Deutsch & Journel (1992).	240
Figura 9.8. Localização e distribuição dos poços utilizados na área de estudo.....	242
Figura 9.9. Fluxograma metodológico da inversão colorida (Lancaster & Whitcombe, 2000), mostrando as etapas de estimativa do expoente α , modelagem do operador e aplicação convolucional sobre o dado sísmico.	243
Figura 9.10. Determinação do expoente α a partir da regressão linear dos espectros log-frequência de impedância acústica dos poços, a partir da combinação do espectro médio dos traços e do espectro das curvas de impedância dos poços.....	244
Figura 9.11. Ajustes do espectro de amplitude da impedância do poço e da sísmica a partir de cortes de frequência para modelar o operador que concolvido com o dado sísmico resultará no espectro de amplitude (curva vermelho) mostrado no gráfico <i>Design operator</i>	245
Figura 9.12. Exemplo ilustrativo das limitações da interpolação direta por krigagem e vizinhos mais próximos (closest), mostrando padrões concêntricos e inconsistentes em torno dos poços.	246

Figura 9.13. Fluxograma da construção do modelo de baixa frequência, destacando as etapas de upscaling, modelagem de variograma, co-krigagem e filtragem passa-baixa.....	247
Figura 9.14. Variogramas verticais e laterais utilizados na co-krigagem para o modelo de baixa frequência.	248
Figura 9.15. Esquema geral do fluxo de trabalho da inversão determinística model-based por generalização de Tikhonov, indicando o uso do modelo de baixa frequência e da wavelet média....	249
Figura 9.16. Wavelets extraídas individualmente em cada poço e wavelet média resultante empregada no processo de inversão determinística.	250
Figura 9.17. Fluxograma geral da inversão estocástica.....	252
Figura 9.18. Controle de qualidade após reamostragem dos perfis de AI dos poços para o grid 3D, e comparação entre os histogramas dos valores medidos e reamostrados para validação do método escolhido.....	253
Figura 9.19. Modelo experimental do variograma vertical ajustado aos dados de impedância acústica dos poços, adotando função esférica com alcance de aproximadamente 25 m.....	253
Figura 9.20. Variogramas horizontais derivados do volume de impedância determinística e parâmetros de busca adotados para o cálculo da continuidade lateral.	254
Figura 9.21. Modelo inicial derivado da inversão determinística e os poços, evidenciando coerência lateral e continuidade geológica.	255
Figura 9.22. Seção arbitrária e <i>time slice</i> evidenciando o resultado da inversão colorida em comparação à amplitude sísmica original.....	258
Figura 9.23. Seções sísmicas interceptando os poços OGX-16 e CP-1, mostrando a definição aprimorada das intrusões ígneas e dos folhelhos adjacentes após a inversão colorida.	259
Figura 9.24. Seções sísmicas atravessando os poços OGX-22 e OGX-57, evidenciando a influência estrutural das soleiras ígneas e o impacto na continuidade dos refletores.	260
Figura 9.25. Detalhe das seções sísmicas destacando a correlação entre as fácies sedimentares e as variações de impedância acústica relativa na Formação Pimenteiras.	261
Figura 9.26. Seções sísmicas detalhadas mostrando o contraste de impedância associado às fácies sedimentares e às intrusões ígneas, com destaque para a fácies 3 no poço OGX-22.....	262
Figura 9.27. Modelo de baixa frequência obtido por co-krigagem e filtragem passa-baixa, utilizado como dado inicial para a inversão determinística.....	263
Figura 9.28. <i>Crossplot</i> entre a impedância acústica medida nos poços e a derivada da inversão determinística, mostrando os coeficientes de correlação individuais e médio.	264
Figura 9.29. Seção sísmica ilustrando o ajuste entre o modelo de baixa, a impedância reamostrada dos poços e o volume resultante da inversão determinística.	265
Figura 9.30. Comparação entre as curvas de impedância acústica dos poços e os resultados da inversão determinística, destacando desvios nas zonas de altos contrastes.....	266
Figura 9.31. Seção sísmica arbitrária e <i>time slice</i> evidenciando o ganho de resolução e heterogeneidade na distribuição de impedância na Formação Pimenteiras.....	267
Figura 9.32. Seção sísmica atravessando os poços CP-1 e OGX-16, evidenciando os contrastes internos no intervalo de folhelhos que podem está associados aos limites entre parassequências do Trato de Sistemas Transgressivo.....	268

Figura 9.33. Seção sísmica interceptando os poços OGX-22 e OGX-57, mostrando intervalos de baixa impedância correlacionados a altos valores de GR e à Superfície de Máxima Inundação no topo da Fom. Pimenteiras.....	269
Figura 9.34. Seção sísmica arbitrária e <i>time slice</i> comparando os resultados obtidos pela inversão colorida e pela inversão determinística.	270
Figura 9.35. <i>Crossplot</i> mostrando a correlação linear entre o modelo inicial utilizado na inversão estocástica e a impedância do poço.	272
Figura 9.36. Comparação entre o perfil de impedância acústica do poço teste, a inversão estocástica de maior correlação e a média das realizações com correlação > 0,7. O perfil do segundo <i>track</i> exibe o desvio padrão entre as simulações. Também é mostrado o <i>crossplot</i> entre impedâncias invertida e medidas, colorido pelo desvio padrão.	274
Figura 9.37. Avaliação quantitativa da inversão estocástica: <i>crossplot</i> entre impedâncias acústicas de poços e volume invertido, com tabela de coeficientes de correlação individuais e valor médio.	275
Figura 9.38. Seções sísmicas mostrando o resultado da inversão estocástica, e a comparação entre as curvas de impedância reamostradas dos poços com a inversão estocástica e o modelo inicial utilizado. Nota-se maior detalhamento vertical e preservação dos contrastes de impedância nas camadas finas para a solução estocástica, refletindo o ganho de resolução obtido.	276
Figura 9.39. Seções sísmicas mostrando o resultado da inversão estocástica, e a comparação entre as curvas de impedância reamostradas dos poços com a inversão estocástica e o modelo inicial utilizado. Nota-se maior detalhamento vertical e preservação dos contrastes de impedância nas camadas finas para a solução estocástica, refletindo o ganho de resolução obtido.	277
Figura 9.40. <i>Crossplot</i> e perfis de impedância mostrando a correlação entre a média das realizações estocásticas e a impedância acústica medida no poço teste (OGX-16).....	277
Figura 9.41. Controle de qualidade do volume médio da inversão estocástica, destacando a elevada correlação nas posições dos poços e a suavização do modelo médio em comparação às realizações individuais.	278
Figura 9.42. Comparação entre as realizações individuais e o volume médio da inversão estocástica, mostrando a preservação das tendências estruturais e o aumento da suavidade estatística no modelo médio.	279
Figura 9.43. Resultado tridimensional da inversão estocástica de maior correlação, mostrando a continuidade lateral das intrusões ígneas e os contrastes de impedância acústica na Formação Pimenteiras.	280
Figura 9.44. Volume médio das realizações da inversão estocástica, evidenciando a coerência espacial e a atenuação das flutuações locais de impedância.	280
Figura 9.45. Volume do desvio padrão das realizações estocásticas, mostrando menores valores próximos aos poços e aumento progressivo em áreas distantes, especialmente na porção superior da formação.	281
Figura 9.46. Seção <i>inline</i> atravessando o poço CP-1, mostrando os resultados da inversão estocástica com resolução aprimorada, ressaltando contrastes de impedância mais definidos em comparação à inversão determinística.	282

Figura 9.47. <i>Inline</i> passando pelo poço OGX-16 destacando a variação vertical e lateral da impedância acústica estimada pela inversão estocástica, ilustrando o ganho de detalhamento e a coerência com as fácies identificadas nos poços.	283
Figura 9.48. <i>Inline</i> passando pelo poço OGX-22 evidenciando o maior poder de resolução e a preservação de camadas delgadas no resultado estocástico.	284
Figura 9.49. <i>Inline</i> passando pelo poço OGX-57 destacando a variação vertical e lateral da impedância acústica estimada pela inversão estocástica, ilustrando o ganho de detalhamento e a coerência com as fácies identificadas nos poços.	285
Figura 9.50. Comparação integrada entre amplitude sísmica, inversão estocástica, média das realizações e inversão determinística. A inversão estocástica apresenta maior capacidade de resolver contrastes de camadas finas, enquanto a determinística produz um volume mais suavizado e geologicamente contínuo. Em termos quantitativos, a média das realizações mostra valores de impedância mais compatíveis com os medidos nos poços.	287

Capítulo 10 292

Figura 10.1. Fluxo de trabalho para predição de COT via EMBER: síntese das principais etapas do trabalho geral da tese.	299
Figura 10.2. Correlação entre COT e AI antes e após o upscaling (pontos coloridos por fácies), destacando a tendência não linear negativa ajustada por polinômio de 2º grau.	300
Figura 10.3. Volume de impedância acústica resultante da inversão estocástica (<i>training feature</i>): realce de camadas delgadas; associação entre baixas impedâncias e fácies mais ricas em COT; altas impedâncias compatíveis com intrusões ígneas (fácies 1).	301
Figura 10.4. Arcabouço tridimensional (grid 3D) utilizado nas simulações (50 × 50 m em x-y; 4 ms em z), configurado para compatibilizar a resolução entre volumes sísmicos e poços.	302
Figura 10.5. Avaliação no poço cego a partir da simulação do EMBER: (a) perfis de COT reamostrado vs. COT estimado (SIM5); (b) crossplot entre o COT do poço e COT estimado pela SIM5; (c) crossplot entre o COT do poço e COT estimado a partir do P50; (d) histogramas de COT (distribuição de todos os poços, P50 e SIM5); (e) histograma do <i>spread</i> (P90–P10).	306
Figura 10.6. Resultados da SIM5 em seção e mapas: AI (entrada), COT e incerteza (P90–P10). Evidenciam-se altos valores de COT nas mais baixas impedâncias (relação inversa AI–COT) e a dispersão espacial da incerteza.	308
Figura 10.7. Modelo 3D de COT (SIM5) em volume com cortes ortogonais (i, j) e três fatias do grid (K = 73, 89 e 110); destaque para a continuidade lateral dos níveis enriquecidos em COT e a heterogeneidade espacial associada às intrusões.	310
Figura 10.8. Volumes de probabilidade de ocorrência por classes de COT segundo Peters & Cassa (1994): P(COT > 1,0 %), P(0,5 % < COT < 1,0 %) e P(COT < 0,5 %); padrões coerentes com as eletrofácies identificadas nos poços e com a distribuição esperada de rochas ricas, moderadas e pobres em matéria orgânica.	311
Figura 10.9. Mapas de probabilidade por classe de COT (mesmas faixas da Figura 8) em três camadas representativas (K = 73, 89 e 110), evidenciando variação lateral, predominância do intervalo moderado (0,5–1,0 %) e concentração de baixos COT próximos às intrusões.	313

Figura 10.10. Seções dos modelos de COT passando pelo poço CP-1 mostrando o resultado da simulação de maior correlação (SIM5) e os cenários: provável (P50), pessimista (P10) e otimista (P90).	315
Figura 10.11. Seções dos modelos de COT passando pelo poço OGX-16 mostrando o resultado da simulação de maior correlação (SIM5) e os cenários: provável (P50), pessimista (P10) e otimista (P90).	316
Figura 10.12. Seções dos modelos de COT passando pelo poço OGX-22 mostrando o resultado da simulação de maior correlação (SIM5) e os cenários: provável (P50), pessimista (P10) e otimista (P90).	317
Figura 10.13. Seções dos modelos de COT passando pelo poço OGX-57 mostrando o resultado da simulação de maior correlação (SIM5) e os cenários: provável (P50), pessimista (P10) e otimista (P90).	317
Figura 10.14. Seção arbitrária atravessando os CP-1, OGX-16, OGX-22 e OGX-57 com o modelo de COT (com e sem interpretação), destacando três níveis enriquecidos em COT (N1, N2, N3) de boa continuidade lateral; correlação estratigráfica proposta com os folhelhos radioativos A, B e C de Rodrigues (1995).	318

Índice de Tabelas

Tabela 3.1. Parâmetros de COT que classificam seu potencial gerador de hidrocarbonetos segundo Peters & Cassa, 1994.	55
Tabela 4.1. Resumo das métricas estatísticas dos perfis candidatos a clusterização anterior a transformação dos dados.	74
Tabela 4.2. Estatísticas descritivas das variáveis de entrada após o pré-processamento (transformações logarítmicas e padronização), mostrando a homogeneização das escalas dos dados.	74
Tabela 4.3. Relação entre os <i>clusters</i> obtidos pelo algoritmo <i>K-means</i> e a reclassificação em eletrofácies, bem como a sua descrição interpretativa.	85
Tabela 5.1. Quantidade de amostras de COT disponíveis nos poços de fora e de dentro da área de estudo, utilizadas no treinamento e validação dos modelos	117
Tabela 5.2. Estatística descritiva dos perfis GR, DT e RES e das amostras de COT, após análise exploratória.	121
Tabela 5.3. Distribuição das amostras nos conjuntos de treinamento, teste e validação, indicando a quantidade total e a proporção em cada conjunto.	127
Tabela 5.4. Parâmetros utilizados para treinar o algoritmo Random Forest.	130
Tabela 5.5. Otimização de hiperparâmetros utilizados para treinar o algoritmo XGBoost e a demonstração dos melhores parâmetros encontrados.	130
Tabela 5.6. Parâmetros utilizados para a otimização e seus respectivos valores que melhor ajustaram o modelo.	130
Tabela 5.7. Parâmetros utilizados no ajuste do modelo Multi-Layer Perceptron.	131

Tabela 5.8. Configuração utilizadas no treinamento do modelo Support Vector Machine.	131
Tabela 5.9. Importância das variáveis nos métodos RF, XGB e CB.	138
Tabela 5.10. Coeficientes angular e linear obtidos na regressão e aplicada na correção da curva estimada de COT no poço de validação.	142
Tabela 5.11. Resumo comparativo dos trabalhos recentes na estimativa de COT por machine learning, detalhando métodos utilizados, variáveis de entrada, tamanho das amostras e desempenho.	151
Tabela 6.1. Parâmetros utilizados para modelar o operador da primeira zona. Em a) aos valores dos cortes de frequências e em b) é mostrado as correlações com os poços após a aplicação do operador com a sísmica na zona 1, e os valores dos coeficientes beta.	167
Tabela 6.2. Parâmetros utilizados para modelar o operador da segunda zona. Em a) aos valores dos cortes de frequências e em b) é mostrado as correlações com os poços após a aplicação do operador com a sísmica na zona 1, e os valores dos coeficientes beta.	170
Tabela 6.3. Parâmetros utilizados para modelar o operador da terceira zona. Em a) aos valores dos cortes de frequências e em b) é mostrado as correlações com os poços após a aplicação do operador com a sísmica na zona 1, e os valores dos coeficientes beta.	173
Tabela 7.1. Avaliação da resolução sísmica em dados antes e depois do fluxo de trabalho de condicionamento para as litologias típicas da Formação Pimenteiras na Bacia do Parnaíba.	199
Tabela 8.1. Valores de correlação cruzada entre o sismograma sintético e o traço sísmico para os diferentes volumes testados (Original, DSMF e SSB)	221
Tabela 8.2. Métricas estatísticas da análise residual entre poço e sísmica.	222
Tabela 8.3. Parametrização adotada para a construção do grid 3D.	225
Tabela 9.1. Parâmetros utilizados na modelagem do operador convolucional da inversão colorida, incluindo limites de frequência inferior (Low cut) e superior (High cut).	245
Tabela 9.2. Valores dos coeficientes de correlação poço-sísmica e dos expoentes α estimados para cada poço. O valor médio de α foi adotado para representar o comportamento espectral da impedância acústica na área de estudo.	246
Tabela 9.3. Parâmetros de extração das wavelets sísmicas e cálculo da wavelet média utilizada como operador de convolução na inversão determinística.	249
Tabela 9.4. Identificação e renomeação dos poços utilizados nas seções sísmicas apresentadas neste capítulo	257
Tabela 9.5. Resultados da inversão estocástica: coeficientes de correlação entre a impedância acústica dos poços e as 50 realizações simuladas. As células em verde destacam as realizações com correlação superior a 0,7, e as em vermelho indicam valores inferiores a esse limite.	273
Tabela 10.1. Parâmetros numéricos empregados nas simulações do EMBER.	303
Tabela 10.2 . Avaliação do desempenho das 20 simulações no pço cego.	304

1. Introdução à Tese

1.1. INTRODUÇÃO

A caracterização de rochas geradoras é um dos pilares da avaliação de sistemas petrolíferos, pois envolve compreender a capacidade de uma unidade geológica gerar e expulsar hidrocarbonetos em condições específicas de maturação térmica (Tissot & Welte, 1984; Magoon & Dow, 1994). Tradicionalmente, esse processo depende fortemente de análises geoquímicas laboratoriais, como as medições do Carbono Orgânico Total (COT) e da pirólise Rock-Eval, obtidas a partir de amostras de calha e testemunhos. Apesar de precisas, tais análises são custosas, destrutivas e especialmente limitadas, o que impõe restrições significativas à compreensão tridimensional das rochas geradoras em escala de bacia (Reis et al., 2022).

Nesse contexto, a integração entre dados geofísicos, geoquímicos e métodos computacionais tem se consolidado como uma alternativa metodológica eficiente para ampliar o entendimento sobre a distribuição e as propriedades das rochas geradoras (Løseth et al., 2011; Carcione & Avseth, 2015; Del Monte et al., 2018). A utilização de métodos sísmicos quantitativos, como inversões e análise de atributos, associada a modelos de machine learning, possibilita inferir propriedades como o COT de maneira contínua e com maior resolução vertical e lateral, reduzindo incertezas e custos exploratórios (Reis et al., 2022).

Este trabalho insere-se nesse cenário metodológico, propondo um fluxo de trabalho integrado que abrange desde o pré-condicionamento sísmico até a modelagem tridimensional do teor de COT, aplicado à Formação Pimenteiras, principal rocha geradora da Bacia do Parnaíba. A pesquisa busca adaptar metodologias frequentemente empregadas em reservatórios para o estudo de rochas geradoras, com ênfase em sua resposta sísmica, comportamento físico e potencial de geração de hidrocarbonetos.

A Bacia do Parnaíba constitui uma bacia intracratônica paleozoico-mesozóica da Plataforma Sul-Americana, formada sobre crosta continental estável não afetada pelos eventos orogênicos andinos e caribenhos (Almeida et al., 2000). O registro sedimentar dessa bacia compreende espessas sequências siliciclásticas e vulcano-sedimentares, cuja história deposicional remonta ao final do Ciclo Brasileiro/Pan-Africano. Dentre as unidades estratigráficas que a compõem, a Formação Pimenteiras

(Devoniano) destaca-se como o principal intervalo potencialmente gerador de hidrocarbonetos (Rodrigues, 1995; Vaz et al., 2007; Mussa et al., 2020).

Os folhelhos da Formação Pimenteiras são ricos em matéria orgânica, com teores de COT variando entre 1 e 5 %, podendo atingir até 6 % em zonas afetadas por intrusões ígneas (Rodrigues, 1995; Mussa et al., 2020). O querogênio predominante é uma mistura dos tipos II e III, refletindo a alternância de condições marinhas e continentais associadas a ciclos transgressivos e regressivos do Devoniano. A particularidade da Bacia do Parnaíba reside em seu sistema petrolífero atípico, no qual a maturação térmica da matéria orgânica não é controlada exclusivamente pelo soterramento, mas principalmente pelo aquecimento associado às intrusões ígneas (Miranda, 2014; Cioccarri & Mizusaki, 2019). Assim, a maturação ocorre de forma localizada e heterogênea, condicionada à proximidade com soleiras e diques.

Diversos autores têm enfatizado o caráter singular desse sistema e as incertezas ainda existentes quanto à influência térmica das intrusões e à distribuição espacial das zonas geradoras (Souza et al., 2017; Andrade et al., 2020). A compreensão detalhada dessas variações é fundamental para reduzir os riscos exploratórios, visto que, segundo Rudolph e Goulding (2017), grande parte das falhas em campanhas exploratórias globais decorre de interpretações imprecisas sobre a presença, maturação e qualidade da rocha geradora.

1.2. MOTIVAÇÃO E JUSTIFICATIVA

A presente tese nasce da necessidade de aprofundar a caracterização tridimensional das rochas geradoras em sistemas atípicos como o da Bacia do Parnaíba. Apesar dos avanços nas técnicas de caracterização de reservatórios, há uma lacuna notável na aplicação de métodos quantitativos voltados especificamente às rochas geradoras, especialmente em contextos onde a maturação térmica é controlada por processos não convencionais, como o contato térmico com intrusões ígneas.

A Formação Pimenteiras, por sua heterogeneidade litológica e térmica, representa um excelente laboratório natural para investigar a relação entre as propriedades físicas e o conteúdo orgânico. Regiões com teores médios de COT superiores a 2 % são restritas e definem as chamadas cozinhas geradoras, cuja delimitação espacial é essencial para o entendimento da geração e migração de hidrocarbonetos. Assim, compreender o comportamento sísmico dessas rochas e estimar seu conteúdo orgânico em três dimensões é de extrema relevância para a avaliação de risco exploratório e o planejamento de futuras campanhas.

1.3. HIPÓTESES DE PESQUISA

1. É possível estabelecer relações quantitativas entre a impedância acústica e o teor de COT, permitindo inferir o conteúdo orgânico a partir de propriedades sísmicas derivadas.
2. As variações laterais e verticais de fácies e eletrofácies nos folhelhos da Formação Pimenteiras podem ser identificadas por meio de dados de poço e inversão sísmica.
3. A modelagem tridimensional do COT baseada em machine learning assistida por geoestatística é capaz de representar adequadamente a heterogeneidade das rochas geradoras, ampliando a resolução e reduzindo as incertezas associadas ao sistema petrolífero atípico.

1.4. OBJETIVOS

1.4.1. Objetivo Geral

Desenvolver uma metodologia integrada para a caracterização sísmica e modelagem tridimensional do teor de Carbono Orgânico Total na Formação Pimenteiras, com base em dados de poços, inversão sísmica e algoritmos de machine learning.

1.4.2. Objetivos Específicos

1. Desenvolver um fluxo de pré-condicionamento sísmico visando aprimorar a relação sinal/ruído e a resolução vertical;
2. Identificar eletrofácies dentro do intervalo de folhelhos por meio de técnicas de machine learning não supervisionadas;
3. Estimar as curvas de COT em poços sem dados laboratoriais a partir de modelos supervisionados;
4. Integrar os resultados das inversões sísmicas com as estimativas de COT;
5. Gerar o modelo 3D do COT utilizando o método Embedded Model Estimator (EMBER), assistido por geoestatística.

1.5. ASPECTOS INOVADORES E CONTRIBUIÇÕES

O caráter inovador desta pesquisa reside em sua abordagem metodológica integrada, que aplica técnicas tradicionalmente voltadas à caracterização de reservatórios, como inversões sísmicas, filtragem espectral e modelagem estocástica, ao estudo de rochas geradoras. A combinação entre machine learning e geoestatística para a modelagem tridimensional do COT em uma bacia

intracratônica com sistema petrolífero atípico representa uma contribuição inédita no contexto da Bacia do Parnaíba.

Além disso, a tese incorpora etapas quantitativas de validação e avaliação de incerteza. Os modelos de COT foram validados com métricas estatísticas (R , R^2 , RMSE) e poço independente, e a modelagem final envolveu a geração de cenários P10, P50 e P90, permitindo quantificar o grau de confiança espacial dos resultados. O pré-condicionamento sísmico (Dip Steered Median Filter e Spectral Blueing) produziu ganhos de resolução de até 38 %, essenciais para detectar camadas finas e contrastes de folhelhos. A definição de eletrofácies via K-Means e classificação Bayesiana serviu como base para a integração entre poços, sísmica e inversões. Finalmente, a aplicação do algoritmo EMBER, combinando geoestatística e machine learning, permitiu representar de forma não estacionária a distribuição do COT e honrar a continuidade lateral das camadas ricas em matéria orgânica.

Esses avanços metodológicos reforçam a relevância deste trabalho para a redução de incertezas na avaliação de sistemas petrolíferos atípicos e para o aperfeiçoamento de ferramentas de interpretação quantitativa em geociências.

1.6. ESTRUTURA DA TESE

Este trabalho de doutorado possui caráter metodológico e envolve o desenvolvimento de um fluxo de trabalho aplicado em diferentes domínios do conhecimento. Por essa razão, o texto foi estruturado em capítulos que, de acordo com a complexidade e o rigor técnico de cada etapa, apresentam subdivisões detalhadas quanto à fundamentação teórica e um tópico de conclusão. De forma resumida, os capítulos tratam dos seguintes temas principais:

Capítulo 1 – Introdução à Tese: apresenta as motivações, hipóteses e principais objetivos que orientam a pesquisa. Além disso, fornece uma visão geral da estrutura do trabalho e de como cada tema será abordado nos capítulos seguintes.

Capítulo 2 – Contexto Geológico: descreve o enquadramento geológico regional da Bacia do Parnaíba, área de estudo desta pesquisa, com ênfase no detalhamento da Formação Pimenteiras, intervalo de maior interesse, e nas intrusões ígneas que afetam a bacia. São discutidos os efeitos térmicos e estruturais dessas intrusões sobre os folhelhos da formação, considerando seu papel no sistema petrolífero atípico da região.

Capítulo 3 – Rochas Geradoras: aborda os conceitos fundamentais relacionados às rochas geradoras, incluindo a origem, acumulação e transformação da matéria orgânica. O capítulo também

revisa os principais estudos sobre a caracterização geoquímica dos folhelhos da Formação Pimenteiras, fornecendo a base teórica necessária para a interpretação dos resultados obtidos nesta pesquisa.

Capítulo 4 – Classificação de Fácies e Estudo de Viabilidade: marca o início da etapa prática da tese, com a investigação das eletrofácies identificáveis a partir dos dados de poço. Demonstra-se a existência de variações significativas no intervalo de folhelhos da Formação Pimenteiras, evidenciadas por meio de um método de *machine learning* não supervisionado (K-means). Em seguida, aplica-se a classificação Bayesiana para avaliar a relação entre as diferentes eletrofácies e a impedância acústica, justificando sua utilização como base para a modelagem do COT.

Capítulo 5 – *Machine Learning* Aplicado à Estimativa de COT: considerando que apenas um dos quatro poços da área de estudo possui medições laboratoriais de COT, este capítulo dedica-se à estimativa das curvas de COT por meio de modelos de *machine learning* supervisionado. Os algoritmos foram treinados com dados de poços distribuídos em outras regiões da bacia, incorporados ao banco de dados deste trabalho, e validados estatisticamente para garantir confiabilidade nas previsões.

Capítulo 6 – Pré-condicionamento do Dado Sísmico: o dado sísmico constitui o elemento central desta pesquisa, pois serve de base tanto para a interpretação quantitativa quanto para a modelagem tridimensional. Assim, este capítulo apresenta a combinação de dois métodos, *Dip-Steered Median Filter* e *Spectral Blueing*, aplicada ao pré-condicionamento sísmico, visando aprimorar a relação sinal/ruído e aumentar a resolução vertical, essencial para o estudo de níveis estratigráficos de baixa espessura.

Capítulo 7 – Análise Espectral e Estudo de Resolução Sísmica: complementando o pré-condicionamento, este capítulo realiza análises quantitativas para verificar e quantificar os ganhos obtidos. São discutidas as modificações no espectro de frequência dos dados sísmicos e suas causas, bem como o impacto do ganho de resolução sobre a integração sísmica-poço.

Capítulo 8 – Amarração Sísmica–Poço e Construção do Grid 3D: apresenta os procedimentos que integram as escalas de poço (1D) e sísmica (3D). Realiza-se a amarração sísmica-poço para o dado sísmico pré-condicionado e sua comparação com a amarração original, além da reinterpretação do topo e da base da Formação Pimenteiras. Em seguida, é descrita a construção do *grid* tridimensional, que servirá de base para as etapas de inversão sísmica e modelagem do COT.

Capítulo 9 – Inversões Sísmicas: aborda três métodos distintos, Inversão Estocástica, Inversão Determinística (*model-based*) e Inversão Colorida, empregados para a obtenção da impedância acústica relativa e absoluta. A impedância absoluta destaca-se como dado tridimensional de alta resolução, capaz de realçar contrastes internos nos folhelhos e auxiliar na caracterização de níveis ricos em COT. Já a impedância relativa foi fundamental como dado de entrada para obtenção de resultados robustos nas inversões subsequentes.

Capítulo 10 – Modelagem 3D do Carbono Orgânico Total: representa o ápice metodológico desta tese. Neste capítulo, é desenvolvida a modelagem tridimensional do teor de COT por meio do método Embedded Model Estimator (EMBER), que combina *machine learning* e geoestatística. Integram-se as estimativas de COT obtidas nos poços com o volume de impedância acústica resultante da inversão sísmica, permitindo gerar um modelo 3D contínuo e quantificar as incertezas espaciais associadas (P10, P50 e P90).

Capítulo 11 – Conclusões: encerra a tese com uma síntese dos principais resultados, discussões e contribuições metodológicas de cada etapa, além de apresentar as limitações do estudo e sugestões para trabalhos futuros.

2. Contexto Geológico

2.1. BACIA DO PARNAÍBA

A Bacia do Parnaíba constitui uma extensa bacia sedimentar intracratônica situada no Nordeste do Brasil, abrangendo predominantemente os estados do Piauí, Maranhão e Tocantins, e, em menor extensão, os estados do Pará, Ceará e Bahia. Esta bacia cobre uma área superior a 600 mil km² e atinge uma espessura de aproximadamente 3.500 metros em seu depocentro, conforme descrito por Vaz *et al.* (2007). Seus limites são definidos por arcos regionais: ao noroeste, pela presença do Arco de Tocantins; ao norte, pelo Arco Ferrer-Urbano Santos; e, finalmente, ao sudeste, pelo Arco São Francisco, conforme ilustrado na Figura 2.1.

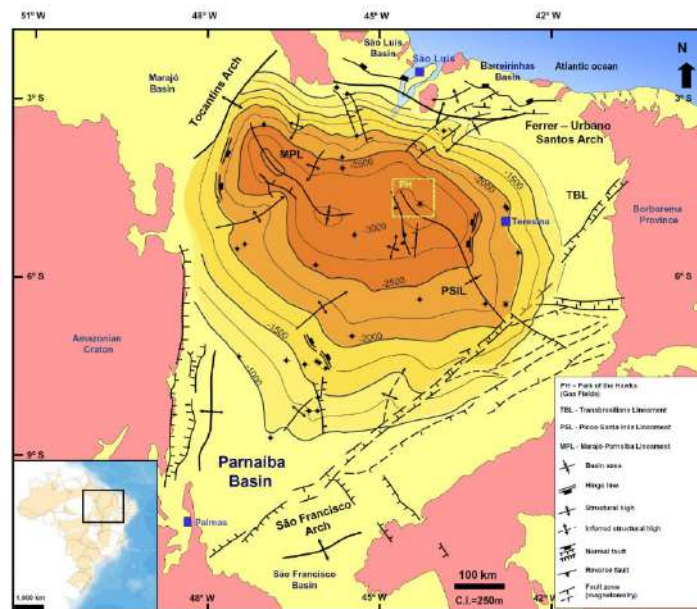


Figura 2.1. Mapa de localização da Bacia do Parnaíba, onde destaca-se a profundidade do embasamento, a geometria típica de uma sinéclise (ovalada) e os limites da bacia que representam o principal controle estrutural da variação do depocentro. O ringfence do Parque dos Gaviões, principal produtora de gás, também está evidenciada no mapa. Modificado de Góes et al., 1990 (Miranda et al 2018).

2.2. ARCABOUÇO TECTÔNICO

A Bacia do Parnaíba desenvolveu-se ao longo de uma bacia continental durante o Estágio de Estabilização da Plataforma Sul-Americana, conforme relatado por Almeida e Carneiro (2004). Através de correlações com os litotipos presentes nas faixas de dobramentos, maciços medianos e outras

formações complexas situadas nas suas bordas ou proximidades, infere-se que o substrato dessa bacia é composto por rochas metamórficas, ígneas e sedimentares, cujas idades variam amplamente, do Arqueano ao Ordoviciano. No entanto, é possível que predominem as rochas formadas entre o final do Proterozóico e o início do Paleozoico, período correspondente à consolidação dessa plataforma (Vaz *et al.* 2007)

De acordo com Vaz *et al.* (2007), a subsidência inicial que deu origem à Bacia do Parnaíba está possivelmente relacionada às deformações e eventos térmicos fini- e pós-orogênicos do Ciclo Brasileiro ou ao Estágio de Transição da plataforma, utilizando a terminologia de Almeida e Carneiro (2004). O arcabouço estrutural da bacia foi interpretado a partir de dados sísmicos, gravimétricos e magnetométricos, revelando a presença de estruturas grabeniformes no substrato, configuradas por blocos falhados de pequenos rejeitos, além de dobras e outras estruturas resultantes de intrusões ígneas mesozoicas nas camadas sedimentares (Figura 2.2).

A evolução da bacia foi significativamente controlada pelas direções dos eixos de deposição até o Eocarbonífero, influenciadas por falhas e fraturas geradas no embasamento. Entre os períodos Neocarbonífero e Jurássico, a bacia adquiriu a forma ovalada típica de uma sinéclise interior, com os depocentros deslocados para a parte central. Além dos fatores mencionados, o registro sedimentar também foi influenciado pela subsidência causada pelo estiramento litosférico, sobrecarga representada pela inserção de sedimentos e outros processos dinâmicos continentais (Vaz *et al.*, 2007).

As ascensões eustáticas desempenharam um papel crucial nas sucessivas inundações marinhas no Brasil durante o Fanerozóico. A eustasia foi, portanto, o fator primordial no controle dos ciclos transgressivos-regressivos e, conseqüentemente, das discordâncias que definem os limites das seqüências da Bacia do Parnaíba (Almeida e Carneiro, 2004).

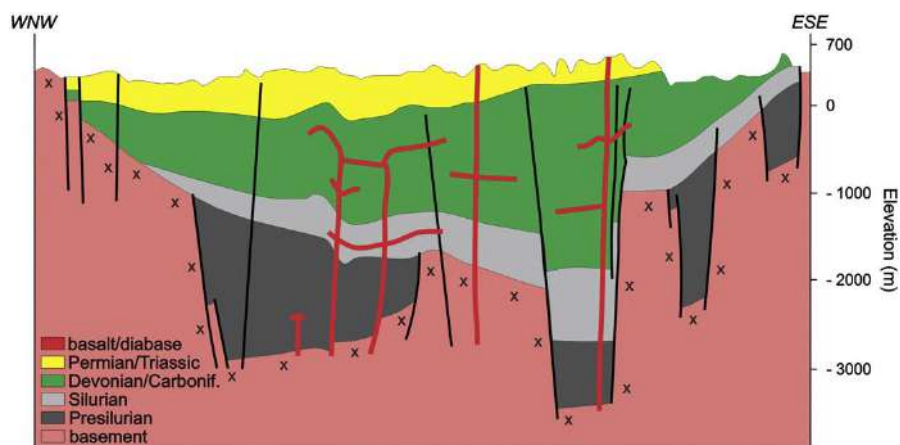


Figura 2.2. Seção geológica representando o arcabouço estrutural da Bacia do Parnaíba e seu preenchimento sedimentar segundo a direção WNW-ESE modificado de Góes *et al.*, 1990 (Castro *et al.* 2014).

2.2.1. Lineamento Transbrasiliano (LTB)

O embasamento da Bacia do Parnaíba é atravessado por três grandes lineamentos: o Transbrasiliano, o Picos–Santa Inês e o Tocantins–Araguaia (Ferraz, 2015). O Lineamento Transbrasiliano (LTB) é um sistema de falhas transcorrentes (strike-slip) localizado entre o Cráton Amazônico e a porção leste da Plataforma Sul-Americana. Este sistema é composto por vários segmentos de falhas, alguns dos quais apresentam setores levemente curvilíneos (Praxedes, 2015). O LTB foi caracterizado como uma faixa cataclasada com aproximadamente 2700 km de extensão, que atravessa o Brasil desde o Ceará até o Mato Grosso do Sul, estendendo-se até o Paraguai e a Argentina (Schobbenhaus *et al.*, 1975, *apud* Goulart, 2019). Diversos autores sugerem que o LTB representa a reativação de uma mega-sutura que desempenhou um papel crucial na formação do supercontinente Gondwana, entre o final do Proterozóico e o início do Paleozoico (Cordani e Sato, 1999; Cordani *et al.*, 2000; Almeida *et al.*, 2000; Brito Neves e Fuck, 2014).

O LTB é constituído por uma série de zonas de cisalhamento dúcteis que se estendem por áreas vastas, potencialmente alcançando o fundo da litosfera. O movimento na zona de cisalhamento provavelmente teve início logo após o fechamento do Oceano Goiano-Faruíano, aproveitando as zonas litosféricas enfraquecidas formadas durante colisões continentais. Todas as suturas colisionais relacionadas às faixas Brasileiro-Pan Africanas são acomodadas em algumas centenas de quilômetros do lineamento. Este lineamento é claramente visível no centro e nordeste do Brasil, formando uma série de anomalias magnéticas de baixa amplitude que podem ser rastreadas ao longo de todo o país, de nordeste a sudoeste. Na Bacia do Parnaíba, o LTB está associado ao principal depocentro de sedimentos do Paleozoico (Cordani *et al.*, 2013).

As reativações do LTB iniciaram-se no Cambriano (540 Ma), continuaram no Mesozoico (250 Ma) e no início do Cenozoico, há cerca de 65 milhões de anos (Chamani, 2015). Reativações mais recentes também foram observadas, evidenciadas por uma série de abalos sísmicos associados ao LTB, especialmente na região do Pantanal Mato-grossense (Farias, 2016, *apud* Goulart, 2019).

O Lineamento Picos–Santa Inês é uma faixa cataclasada disposta no sentido NW-SE, transversalmente ao Lineamento Transbrasiliano, com uma extensão de 160 km e um rejeito de 200 a 300 metros. Esta faixa reflete na morfologia contemporânea uma série de alinhamentos orientados para NW (Vetorazzi, 2012). Tanto o Lineamento Transbrasiliano quanto o Picos–Santa Inês constituem fraturas e falhas herdadas do embasamento, desempenhando um papel significativo desde a fase inicial da bacia e ao longo de toda a sua evolução, controlando as direções dos eixos deposicionais até o Eocarbonífero (Vaz *et al.*, 2007).

2.3. ESTRATIGRAFIA DA BACIA DO PARNAÍBA

Uma vez que o foco deste trabalho é a Formação Pimenteiras, será feito nessa seção um breve resumo dos principais ambientes descritos para as sequências estratigráficas da Bacia do Parnaíba, as quais podem ser encontradas em detalhes através dos trabalhos publicados como por Small (1914), Aguiar (1971), Lima & Leite (1978), Caputo & Lima (1984), Caputo (1984), Della Fávera (1990), Góes A. M. (1995) e mais recente, onde encontra-se a carta estratigráfica atualizada da Bacia do Parnaíba, em Vaz *et al.* (2007).

A Bacia do Parnaíba foi intensamente influenciada por processos magmáticos, responsáveis pelo emplacement de diques e soleiras em diferentes níveis estratigráficos (Ernesto et al., 2002). A atividade ígnea apresentou três fases distintas, com duração superior à observada em outras bacias paleozóicas brasileiras (Daly et al., 2018). O primeiro evento, de natureza vulcânica a subvulcânica félsica, ocorreu no Cambriano, anterior à formação da bacia, e está relacionado às rochas da Província Borborema e ao Grupo Jaibaras (Daly et al., 2018; Oliveira & Mohriak, 2003). As fases mesozoicas posteriores são registradas em afloramentos e subsuperfície, evidenciando reativação magmática significativa (Daly et al., 2018).

A Bacia do Parnaíba é composta por uma sucessão de rochas sedimentares e ígneas, dispostas em cinco supersequências a partir do período de sinéclise da bacia: Siluriana, Mesodevoniana-Eocarbonífera, Neocarbonífera-Eotriássica, Jurássica e Cretácea. Essas supersequências são delimitadas por discordâncias que se estendem por toda a bacia ou abrangem regiões extensas, conforme ilustrado na carta estratigráfica (Vaz *et al.*, 2007, Figura 2.3).

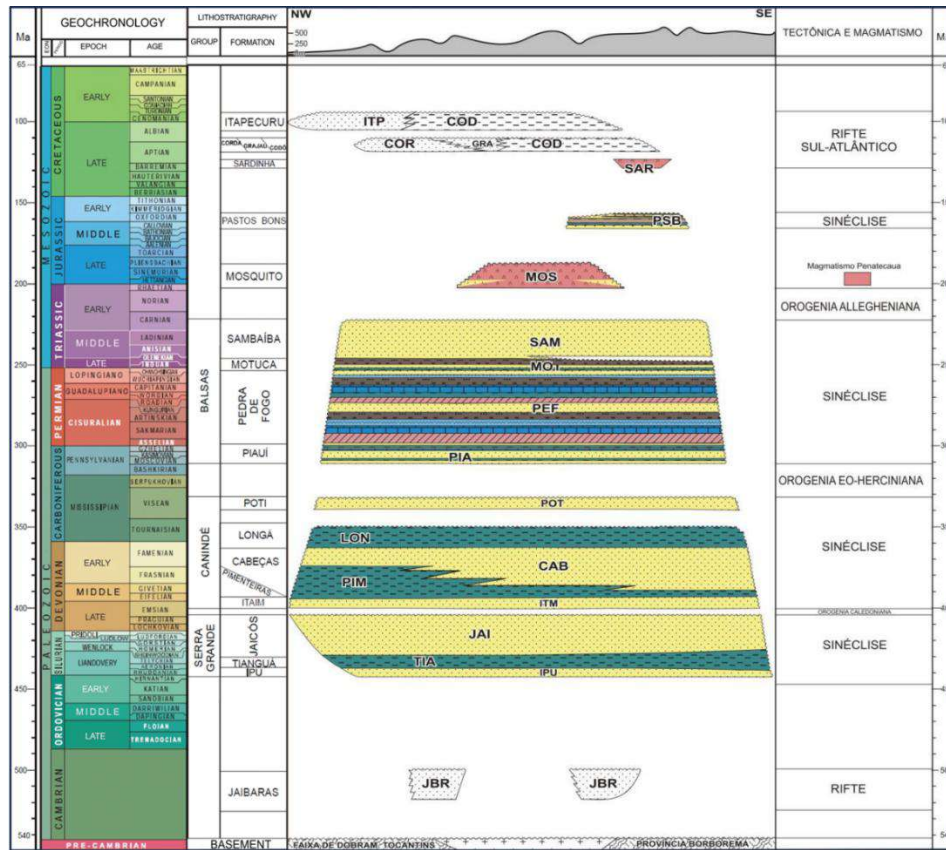


Figura 2.3. Carta estratigráfica da Bacia do Parnaíba segundo Vaz et al. (2007).

A sequência Siluriana da Bacia do Parnaíba, em contato direto com o embasamento pré-cambriano, corresponde litoestratigraficamente ao Grupo Serra Grande, inicialmente proposto por Small (1914) e redefinido por Caputo & Lima (1984), sendo subdividido em três formações: Ipú, Tianguá e Jaicós. A Formação Ipú, unidade mais basal foi depositada em ambientes glaciais proximais a frentes deltaicas (Caputo, 1984). Já a formação sobreposta a ela, a Formação Tianguá ocorreu em um ambiente de plataforma rasa (Góes e Feijó, 1994). Encerrando a sequência, a Formação Jaicós teve origem em ambientes de sistemas fluviais entrelaçados (Caputo, 1984).

A Sequência Mesodevoniana-Eocarbonífera da Bacia do Parnaíba é litoestratigraficamente representada pelo Grupo Canindé, subdividido nas formações Itaim, Pimenteiras, Cabeças, Longá e Poti (Caputo, 1984). A Formação Itaim, formação mais antiga dessa sequência, formou-se a partir de sistemas deltaicos e plataformais dominados por marés e tempestades (Vaz *et al.*, 2007). Em seguida tem-se a Formação Pimenteiras foi depositada em ambiente de plataforma rasa dominada por tempestades (Lima e Leite, 1978). Já a Formação Cabeças possui características de um ambiente glacial ou periglacial como descrito por Caputo (1984). A Formação Longá representa uma fase de transgressão marinha depositada em ambiente nerítico plataformal dominado por tempestades

(Albuquerque e Dequech, 1946). O ciclo sedimentar encerra-se com a Formação Poti, composta por depósitos típicos de ambientes deltaicos e de planícies de maré influenciadas ocasionalmente por tempestades (Lima e Leite, 1978).

A Sequência Neocarbonífera-Eotriássica da Bacia do Parnaíba é representada pelo Grupo Balsas, composto pelas formações Piauí, Pedra de Fogo, Motuca e Sambaíba (Góes *et al.*, 1990). A Formação Piauí é subdividida em duas sucessões indicando a transição de um ambiente fluvial com influências eólicas para marinha em clima semiárido a desértico (Lima e Leite, 1978). A Formação Pedra de Fogo é caracterizada por depósitos de ambiente marinho raso a litorâneo com planícies de sabkha sob ocasional influência de tempestades (Góes e Feijó, 1994). A Formação Motuca teve sua origem em um ambiente continental desértico controlado por sistemas lacustres (Vaz *et al.*, 2007). A sequência encerra-se com a Formação Sambaíba é composta de depósitos tipicamente formados por sistemas eólicos em ambientes desérticos, refletindo uma grande regressão marinha global no final do Permiano-início do Triássico (Caputo, 1984).

A Sequência Jurássica da Bacia do Parnaíba foi inicialmente atribuída ao Grupo Mearim, que englobava as formações Mosquito, Pastos Bons, Corda e Sardinha (Lisboa, 1914; Aguiar, 1959, *apud* Góes e Feijó, 1994). No entanto, Góes *et al.* (1992, *apud* Góes e Feijó, 1994) redefiniram o Grupo Mearim para incluir apenas as formações Pastos Bons e Corda, que ocorrem interdigitadas. Posteriormente, Vaz *et al.* (2007) redefiniu a Sequência Jurássica para incluir apenas a Formação Pastos Bons, enquanto a Formação Corda foi reinterpretada como parte da Sequência Cretácea. A Formação Pastos Bons é litologicamente dividida em três partes e foi interpretada como depositada em paleodepressões continentais e lacustrinas, com alguma contribuição fluvial, em um clima semiárido a árido (Vaz *et al.*, 2007).

A última sequência depositada na Bacia do Parnaíba é composta pelas Formações Codó, Corda, Grajaú e Itapecuru. A Formação Corda é caracterizada por estruturas cruzadas típicas de dunas eólicas, indicando deposição em um sistema desértico (Vaz *et al.*, 2007). A Formação Grajaú e a Formação Codó foram descritos por Rossetti *et al.* (2003) como ambos depositados em ambientes marinho raso, lacustre e flúvio-deltaico. Por último, a Formação Itapecuru é composta por depósitos que correspondem a seis ciclos deposicionais transgressivos, atribuídos a sistemas de vales estuarinos incisos e ambientes estuarino-lagunares afetados por ondas de grande escala, incluindo canais fluviais, lagunas, canais de maré e litorâneos (Rossetti *et al.*, 2003).

Na Bacia do Parnaíba, as rochas magmáticas são representadas por intrusivas (diques e soleiras) e extrusivas de composição básica, classificadas estratigraficamente em duas unidades: Formação

Mosquito e Formação Sardinha, (Vaz *et al.*, 2007). Em subsuperfície, diques e soleiras predominam na Sequência Mesodevoniana-Eocarbonífera, também ocorrendo na Sequência Siluriana e raramente na Neocarbonífera-Eotriássica. A Formação Mosquito, datada do Eocretáceo, é caracterizada por derrames basálticos com intercalações de arenitos, já a Formação Sardinha, do Eojurássico, é composta por derrames de basalto, variando de preto a roxo, com espessura média de 20 metros em afloramentos (Aguiar, 1971). Os eventos magmáticos na bacia estão ligados ao rifteamento Jurássico do supercontinente Pangeia, processo tectônico que resultou em uma nova configuração geológica no Brasil, marcando o início de uma fase tectônica que culminaria na abertura do Oceano Atlântico durante o período Cretáceo. Consequentemente, houve uma reativação e formação significativa de falhamentos, moldando a estrutura atual da bacia e influenciando diretamente os sistemas petrolíferos da região (Almeida & Carneiro, 2004; Zalán, 2004; Vaz *et al.*, 2007).

2.4. FORMAÇÃO PIMENTEIRAS

A Formação Pimenteiras, segundo Small (1914), é composta por folhelhos cinza-escuros a pretos, esverdeados e radioativos, com intercalações de siltitos e arenitos piritosos e pode atingir espessuras superiores a 500 metros (Goes *et al.*, 1990). Na interpretação de Campanha e Mabesoone (1976) e, posteriormente, Mabesoone (1977), essa formação faz parte de uma unidade regressiva depositada principalmente em planícies de maré e lagoas pouco profundas separadas por barreira em condições frias. Por outro lado, Ribeiro e Dardenne (1978) salientaram que a gênese da Formação Pimenteiras, bem como dos ironstones, corresponde à de uma zona de planície de maré em condições tropicais. Caputo (1984) interpretou a Formação Pimenteiras como um registro de uma transgressão global do Devoniano Superior, marcada por oscilações do nível do mar. Durante o estágio Frasniano, as camadas de folhelho laminado com alta radioatividade indicam o nível máximo do mar. Além disso, Caputo (1984) identificou no nordeste da Bacia do Parnaíba duas seções distintas: folhelho siltoso sobreposto por arenito de grão fino a muito fino. Essas seções, com até 70 metros de espessura cada, representam depósitos pró-leque e depósitos frontais em leque-delta, respectivamente. Nos afloramentos sudeste, as camadas de arenito grosso registram lobos de leques deltáico menores. A extensa bioturbação sugere que o mar interior da Pimenteiras provavelmente era sem marés, indicando baixa atividade durante o período. Durante as fases iniciais da transgressão sobre o cráton, como observado ao longo das margens ocidentais da Bacia do Parnaíba, a deposição costeira dominada por tempestades era prevalente. O clima durante a deposição da parte inferior da Formação Pimenteiras

pode ter sido ártico ou subártico, inferido pela natureza da fauna, escassez de calcário e ausência de leitos vermelhos, recifes e evaporitos (Copper, 1977; Klein, 1982).

Rodrigues (1995), em sua tese de doutorado, realizou um estudo geoquímico aprofundado, no qual interpretou três níveis de folhelhos radioativos da Formação Pimenteiras. A formação foi descrita como parte de uma sequência que se inicia com um trato de sistemas de mar baixo, representado pela Formação Itaim, seguido por um trato de sistemas transgressivos na parte inferior da Formação Pimenteiras, e culmina com um trato de sistemas de mar alto, abrangendo a parte superior da Formação Pimenteiras e a Formação Cabeças (Della Fávera, 1990). Este estudo detalhado será a referência principal a ser citada neste trabalho. Em seu estudo, Rodrigues (1995) detalha também a importância dos folhelhos radioativos na Bacia do Parnaíba, destacando três principais intervalos estratigráficos dentro da Formação Pimenteiras, correspondentes ao Devoniano (Figura 2.4), sobre os quais há indícios de geração de óleo ou gás a partir do efeito térmico. O primeiro intervalo, os folhelhos radioativos A, de idade Givetiano, são caracterizados por sua maior radioatividade, resistividade e menor densidade, encontrando-se na base do trato de sistemas transgressivos e apresentando espessura máxima de cerca de 20 metros. O segundo intervalo, os folhelhos radioativos B, situados aproximadamente 100 metros acima da superfície transgressiva numa sequência francamente transgressiva, com empilhamento retrogradantes, possuem espessuras entre 10 e 20 metros e foram depositados em condições mais profundas e distais, permitindo maior preservação de matéria orgânica. Os folhelhos radioativos B representam o primeiro nível importante de afogamento após a superfície transgressiva. O terceiro intervalo, os folhelhos radioativos C, datados do Frasniano, são os mais espessos, com até 40 metros, e marcam o momento de inundação marinha máxima do Devoniano, correlacionando-se com folhelhos geradores de hidrocarbonetos das bacias do Solimões e Amazonas.

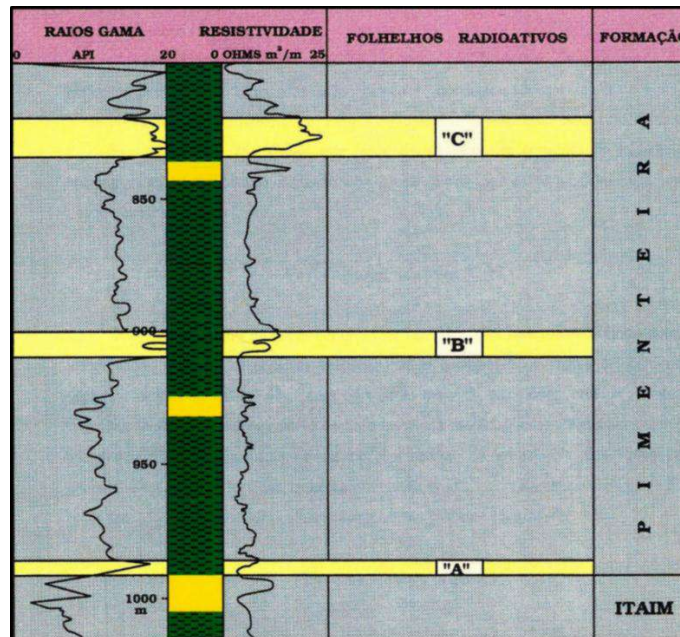


Figura 2.4. Interpretação do Rodrigues (1995) quando aos intervalos de folhelhos radioativos do trato de sistemas transgressivos na Formação Pimenteiras no poço 1-PAF-7-MA.

2.5. INTRUSÕES ÍGNEAS E O SISTEMA PETROLÍFERO ATÍPICO DA BACIA DO PARNAÍBA

Magoon e Dow (1994) definiram um sistema petrolífero típico como aquele em que a rocha geradora atinge a maturação térmica da matéria orgânica durante o processo de soterramento, devido à sobrecarga de rochas sobrepostas. Sendo assim, o sistema petrolífero é reconhecido como típico quando a rocha geradora é submetida à maturação em função do aumento da temperatura com a profundidade do soterramento. Por outro lado, um sistema petrolífero é considerado atípico quando os hidrocarbonetos são gerados por processos diferentes do gradiente geotérmico. Por exemplo, isso ocorre quando a rocha geradora atinge o grau de maturação devido a uma fonte de calor promovida por intrusões ígneas de diques e soleiras (Magoon e Dow, 1994) (Figura 2.5).

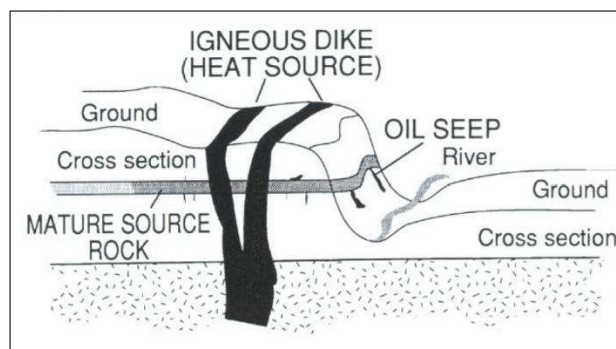


Figura 2.5. Exemplo da formação de um sistema petrolífero atípico cujo petróleo se origina de uma rocha geradora que atinge a maturação por meio do calor fornecido por uma intrusão ígnea (Magoon e Dow, 1994).

A Bacia do Parnaíba é caracterizada por sua relativa baixa profundidade, com um depocentro que atinge, no máximo, 3500 metros. A formação que contém o principal intervalo de potencial gerador localiza-se em profundidades variando entre 500 e 2500 metros. Tais profundidades são insuficientes para que as rochas geradoras alcancem a janela de maturação da matéria orgânica por meio do processo típico de aumento de temperatura devido ao soterramento e sobrecarga com a profundidade. Dessa maneira, a geração de hidrocarbonetos na bacia foi promovida por fontes de calor associadas a intrusões ígneas, que estão amplamente distribuídas em diferentes níveis estratigráficos. Essas intrusões, frequentemente, também desempenham o papel de trapas e selos dos reservatórios (Rodrigues, 1995; Miranda, 2018) (Figura 2.6).

Sendo assim, uma vez que, na natureza a formação do petróleo ocorre a baixas temperaturas (geralmente inferiores a 150°C) ao longo de um extenso período de tempo geológico, espera-se que esse processo também ocorra nas proximidades de intrusões devido às altas temperaturas presentes nessas áreas, embora em um período de tempo consideravelmente menor (Tissot & Welte, 1978). Esse seria um processo de geração intermediário em termos de tempo envolvido, situando-se entre os experimentos de laboratório, que utilizam altas temperaturas e curtos períodos, e o processo natural, caracterizado por baixas temperaturas e longos períodos de tempo geológico (Rodrigues, 1995).

Um mecanismo de geração semelhante foi relatado em várias bacias sedimentares, como as bacias do Paraná, Amazonas e Solimões no Brasil; Bacia de Neuquén na Argentina; Bacia do Noroeste da Austrália na Austrália; Bacia de Karoo na África do Sul; e Bacia de Liahoe na China (Milani & Zalán 1999; Eiras & Wanderley Filho 2003; Rodriguez Monreal *et al.* 2009; Aarnes *et al.* 2011; Holford *et al.* 2013, *apud* Miranda 2018).

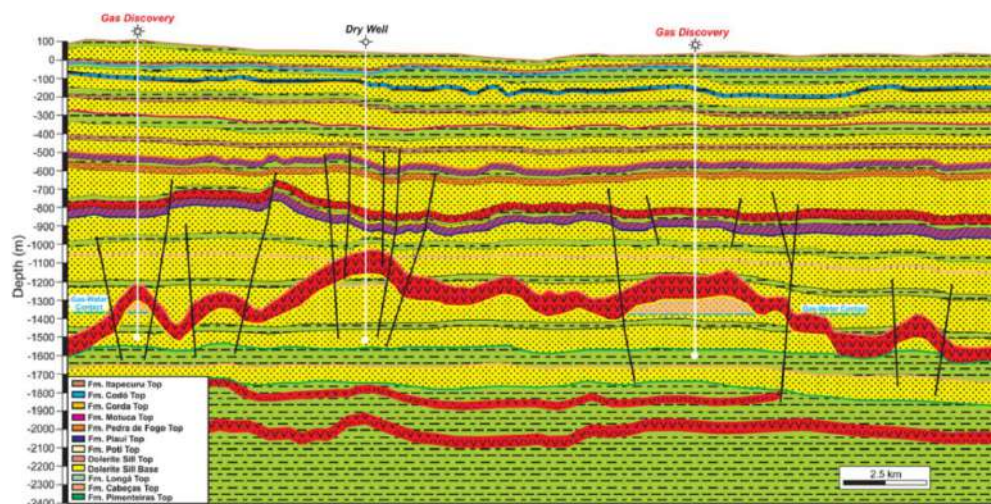


Figura 2.6. Seção geológica esquemática mostrando o modelo para o principal mecanismo de trapeamento para acumulação comercial devido aos saltos das soleiras de diabásio que selam o reservatório ao criar estruturas de fechamento em quatro vias. Além disso, também

é destacado as soleiras planas mais finas na base da seção responsáveis por desencadear a maturação da rocha geradora (Miranda et al. 2018).

De modo geral, acredita-se que o aumento da temperatura induzido por corpos intrusivos provoca alterações químicas em uma distância equivalente a uma ou duas vezes a espessura do próprio corpo intrusivo (Correia & Maury, 1975; Rodrigues & Quadros, 1975; Dow, 1977; Kendrick *et al.*, 1977; Rodrigues *et al.*, 1990, *apud* Rodrigues 1995). No entanto, para uma compreensão mais precisa dessas alterações, é essencial realizar estudos detalhados sobre a taxa de resfriamento e a diferença de temperatura entre o corpo intrusivo e as rochas encaixantes antes da intrusão. Além disso, a litologia das rochas encaixantes, incluindo sua composição mineralógica e orgânica, permeabilidade, porosidade e a presença de fluidos, bem como sua condutividade térmica e evolução térmica pré-intrusão, são fatores cruciais que devem ser considerados.

Estudos petrográficos conduzidos por Rodrigues (1995) indicaram que o processo de resfriamento das rochas intrusivas ocorreu de maneira lenta, evidenciado pelo bom desenvolvimento dos minerais como augita, plagioclásio e minerais opacos da matriz. Em contraste, o contato com a rocha sedimentar é caracterizado por concentrações de minerais opacos e pela presença de vidro vulcânico, que frequentemente se altera para minerais argilosos, sugerindo um processo de resfriamento mais rápido nessas áreas.

As soleiras são intrusões ígneas em forma de tabuleiro que se alinham principalmente de maneira paralela às camadas da rocha encaixante e são quase horizontais. Por outro lado, os diques são canais através dos quais o magma ascende, cortando os estratos de forma geralmente sub-vertical. Ambos, diques e soleiras, criam auréolas metamórficas de contato devido ao aquecimento localizado das rochas encaixante adjacentes (Kim Senge *et al.* 2017).

Em estudos de campo nas Montanhas Henry, Utah (EUA), Pollard e Johnson (1973) e Johnson e Pollard (1973) investigaram o mecanismo de alojamento das soleiras e a transição entre soleiras e lacólitos (*apud* Junior, I. T. *et al.* 2014). Segundo esses autores, quando um dique é reorientado para um plano horizontal, a intrusão se propaga mais facilmente dentro da rocha encaixante devido à concentração de tensões na ponta da soleira. Esse processo de propagação forma uma zona cataclásada na rocha encaixante como resultado da intrusão (Rubin, 1995). Além disso, Pollard e Johnson (1973) notaram que soleiras e lacólitos frequentemente possuem diques periféricos, observação que foi confirmada em modelos físicos análogos. À medida que as intrusões se espalham lateralmente, os corpos magmáticos tendem a elevar e dobrar a coluna sedimentar acima da soleira. Esse processo de intrusão resulta na formação de fraturas na periferia das soleiras e lacólitos, que são aproveitadas pelo

magma, criando os diques periféricos observados. Em alguns casos, esses diques podem se intrudir horizontalmente em níveis estratigráficos superiores, formando soleiras com geometria em degraus (Figura 2.7, Jerram, D. Pertord, N., 2011).

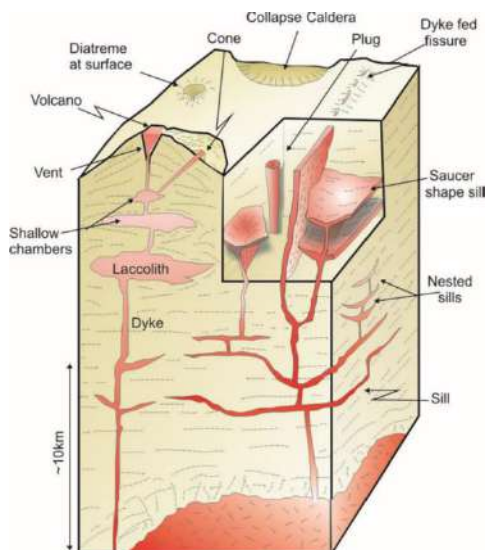


Figura 2.7. Esquema da rede de distribuição 3D de magma que flui de níveis profundos para rasos na crosta. Nesta imagem também pode-se observar dos diferentes formatos que os corpos intrusivos podem assumir (Jerram, D. Pertord, N., 2011).

Muitas soleiras aparecem paralelas ao acamamento, geralmente intrudindo espessos pacotes de folhelhos e ao longo dos contatos das formações Longá-Poti, Pimenteiras-Cabeças e Tianguá-Jaicós (Trosdorf Junior *et al.*, 2014). Seções sísmicas revelam que as soleiras apresentam diversas geometrias, incluindo formas plano-paralelas, de pires, planares transgressivas e em bloco falhado, assim como interpretadas e esquematizadas por Planke, S., *et al.* (2005) (Figura 2.8). O tipo mais comum de soleira na Bacia do Parnaíba é a plano-paralela, que pode se estender por dezenas a centenas de quilômetros, com espessuras variando de 2-5m a 250-300m (Trosdorf Junior *et al.*, 2014).

Além disso, o maior volume de rochas intrusivas se encontra nos folhelhos da Formação Pimenteiras, que também contêm os maiores teores de matéria orgânica da bacia (Trosdorf Junior *et al.*, 2014). A combinação de rochas pelíticas (acamamento e tensão estratificada horizontalmente) e altos teores de matéria orgânica, cujo contato com o magma em altas temperaturas gera grande produção de gás, aumentando a pressão de fluidos e facilitando a abertura de fendas na rocha encaixante, desempenhou um papel crucial no alojamento desses corpos (Trendall, 1994; Leaman, 1995, *apud* Trosdorf Junior *et al.*, 2014). Esses fatores podem ter proporcionado as condições necessárias para que uma mesma soleira se estendesse por mais de 100 km dentro do mesmo nível estratigráfico, como evidenciado por seções sísmicas e datações geocronológicas em poços (Trosdorf Junior *et al.*, 2014).

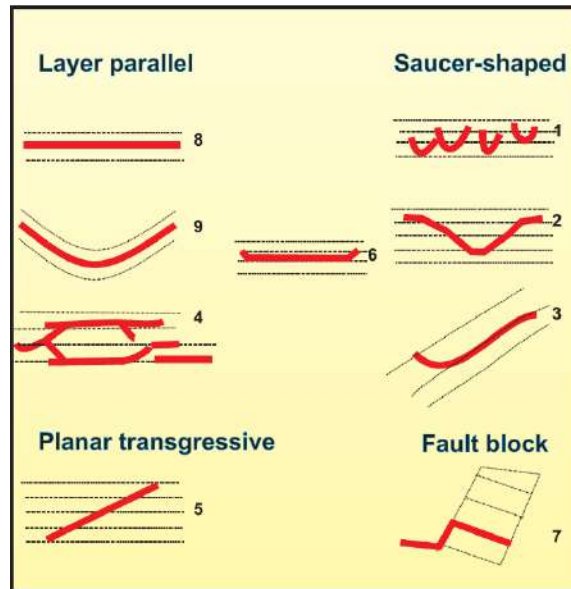


Figura 2.8. Esboço mostrando a configuração das unidades de fácies de soleira identificadas em seções sísmicas interpretadas por Planke, S.; Rasmussen, T.; Rey, S. S.; Myklebust, R. (2005).

3. Rochas Geradoras

3.1. INTRODUÇÃO

O conceito de sistema petrolífero, amplamente difundido na literatura da geologia do petróleo, estabelece que a geração e acumulação de hidrocarbonetos dependem da interação eficaz entre diversos elementos e processos geológicos, incluindo rocha geradora, reservatório, selo, armadilhas, e mecanismos de migração e preservação. Entre esses elementos, a rocha geradora ocupa um papel central, uma vez que sua presença é condição indispensável para a origem dos hidrocarbonetos. Sem uma rocha geradora efetiva, mesmo os melhores reservatórios ou selos não resultariam em acumulações comerciais de petróleo (Magoon & Dow, 1994).

A rocha geradora pode ser definida como uma rocha sedimentar que contém quantidades significativas de matéria orgânica e que, ao ser submetida a soterramento e aquecimento ao longo do tempo geológico, é capaz de gerar óleo e gás natural. Essa matéria orgânica, em especial o querogênio, precisa estar presente em concentração suficiente para justificar seu potencial gerador. Além disso, o tipo de querogênio presente influencia diretamente na natureza dos hidrocarbonetos formados, óleos associados ou gás seco, por exemplo (Tissot, B. P., & Welte, D. H., 1984; Gluyas & Swarbrick, 2004).

A gênese de uma rocha geradora eficaz envolve fatores ambientais que controlam tanto a produção quanto a preservação da matéria orgânica. Ambientes com alta produtividade orgânica, como zonas de ressurgência costeira, mares rasos, pântanos e lagos estagnados, favorecem o acúmulo de matéria vegetal e algas. Contudo, para que essa matéria orgânica seja preservada e não completamente reciclada no ciclo biológico, são necessárias condições anóxicas no fundo do corpo d'água, as quais inibem a decomposição bacteriana (Gluyas & Swarbrick, 2004). Esse equilíbrio entre produção, preservação e soterramento define a qualidade de uma rocha geradora.

O querogênio, principal componente da matéria orgânica insolúvel em solventes, constitui o precursor dos hidrocarbonetos. A avaliação do potencial de geração de uma rocha é comumente realizada por análises geoquímicas que incluem medições de Carbono Orgânico Total (COT), pirólise Rock-Eval, identificação do tipo de querogênio e determinação de sua maturidade térmica. Essas técnicas fornecem informações cruciais sobre a qualidade e o estágio evolutivo da rocha, sendo que a

maturidade é essencial para indicar se o querogênio já passou pelas transformações térmicas necessárias para a geração de óleo e/ou gás (Tissot, B. P., & Welte, D. H., 1984; Allen & Allen, 2013).

A classificação óptica do querogênio em grupos como exinita, vitrinita e inertinita também contribui para a previsão do tipo de hidrocarboneto gerado. Exinitas tendem a gerar óleos com gás associado, enquanto vitrinitas são geralmente precursoras de gás seco. Já as inertinitas, com origem em matéria orgânica altamente oxidada, possuem baixo ou nenhum potencial gerador (Allen & Allen, 2013).

A origem da matéria orgânica também varia conforme o ambiente sedimentar. Fitoplâncton marinho é a principal fonte nos oceanos, enquanto plantas terrestres dominam nos ambientes continentais. Estas últimas podem ser transportadas para ambientes marinhos ou lacustres, compondo um suprimento alóctone, ou podem se acumular localmente, formando depósitos autóctones como turfeiras em zonas de planície deltaica (Allen & Allen, 2013). A preservação desse material depende da rápida deposição em ambientes anóxicos, como ocorre em pântanos, mares restritos e lagos profundos, com aporte limitado de oxigênio e lenta circulação de água.

Assim, ao considerar a importância das rochas geradoras dentro do sistema petrolífero, é fundamental reconhecer não apenas sua capacidade de gerar hidrocarbonetos, mas também as condições geológicas e ambientais que controlam sua formação, preservação e evolução térmica ao longo do tempo. Essa compreensão é essencial para qualquer abordagem exploratória em bacias sedimentares.

3.2. ORIGEM DA MATÉRIA ORGÂNICA

3.2.1. Fontes de Matéria Orgânica

A origem da matéria orgânica presente nas rochas geradoras está intimamente ligada à composição e à produtividade da biomassa primária disponível nos diferentes ambientes deposicionais ao longo do tempo geológico. Desde a identificação de porfirinas como composto um "marcador biológico" em rochas sedimentares por Treibs (1934), acumulou-se um corpo substancial de evidências geoquímicas que confirma a origem biológica do petróleo, derivado da matéria orgânica soterrada em ambientes sedimentares (Tissot, B. P., & Welte, D. H., 1984, 1985; Huc, 2003).

A matéria orgânica viva é composta majoritariamente por quatro grupos de compostos: carboidratos, proteínas, lipídios e lignina. Dentre estes, apenas os lipídios e a lignina apresentam resistência suficiente à degradação para serem efetivamente preservados e incorporados aos

sedimentos. Os lipídios, abundantes em algas e também presentes em partes cerosas de plantas terrestres, são os principais precursores do petróleo. Por outro lado, a lignina, exclusiva de plantas terrestres, é pouco propensa à geração de óleo, mas constitui uma importante fonte de gás natural (Tissot, B. P., & Welte, D. H., 1984; Allen & Allen, 2013).

A composição da biomassa varia significativamente entre ambientes continentais e marinhos. Ecossistemas marinhos são dominados por fitoplâncton fotossintetizante, como diatomáceas, dinoflagelados e nanoplâncton, que respondem por mais de 90% da produção orgânica nos oceanos. Já os ambientes continentais apresentam uma maior contribuição de plantas terrestres e algas de água doce, cuja produtividade está fortemente condicionada por fatores climáticos e geográficos (Allen & Allen, 2013; Gluyas & Swarbrick, 2004).

As plataformas continentais tendem a concentrar maiores teores de carbono orgânico por área, especialmente em zonas como estuários, recifes e zonas intermareais. Nessas áreas, há acúmulo expressivo de matéria orgânica devido à alta produtividade e à deposição em ambientes com circulação restrita. Em contraste, os ambientes de mar aberto, como os “pelagic red clays” das bacias oceânicas profundas, apresentam baixos teores de matéria orgânica devido à menor produtividade primária e à alta oxigenação das águas (Huc, 2003; Magoon & Dow, 1994).

Além da produção autóctone de matéria orgânica, ambientes continentais também recebem contribuições alóctones significativas, como o transporte de matéria vegetal terrestre por rios para sistemas lacustres ou marinhos adjacentes. A turfa representa o depósito autóctone mais importante de matéria orgânica continental, formada sob condições anóxicas em pântanos ou lagos mal drenados e rica em tecido vegetal parcialmente decomposto. Sua preservação é favorecida pela baixa taxa de decomposição microbiana e pela subsidência contínua do substrato (Allen & Allen, 2013).

A natureza da vegetação terrestre também influencia a qualidade da matéria orgânica preservada. Enquanto os carvões paleozoicos são geralmente ricos em lignina e mais propensos à geração de gás, os carvões jurássicos e terciários podem conter frações significativas de componentes lipídicos, conferindo-lhes potencial para geração de óleo (Tissot, B. P., & Welte, D. H., 1984; Gluyas & Swarbrick, 2004).

É importante destacar os três principais ambientes sedimentares responsáveis pela deposição de camadas geradoras: lagos, deltas e bacias marinhas. Os lagos, especialmente em ambientes continentais, constituem locais privilegiados para a deposição de matéria orgânica de origem algácea e bacteriana, sobretudo quando há estratificação da coluna d'água e condições anóxicas no fundo. Lagos

geologicamente duradouros, com baixo aporte clástico, tendem a formar rochas geradoras espessas e ricas em matéria orgânica (Allen & Allen, 2013; Huc, 2003).

3.2.2. Ambientes Depositionais

Os ambientes deposicionais exercem influência direta tanto na quantidade quanto na qualidade da matéria orgânica acumulada nos sedimentos. A efetiva geração de hidrocarbonetos está condicionada à deposição de matéria orgânica sob condições que favoreçam sua preservação, além da contribuição de precursores lipídicos de alta qualidade (**Error! Reference source not found.**, Bjorlykke, 2010). Entre os principais ambientes geradores destacam-se sistemas lacustres, deltas e bacias marinhas restritas ou profundas (Huc, 2003; Allen & Allen, 2013).

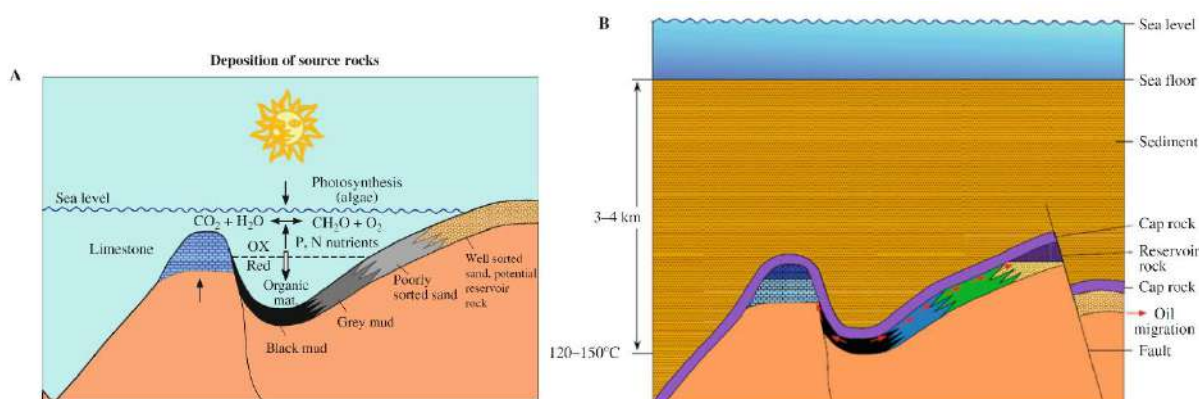


Figura 3.1. (a) Ambientes deposicionais favoráveis à formação de rochas geradoras e reservatórios, como depressões marinhas com baixa circulação, que favorecem a preservação da matéria orgânica. (b) Migração de petróleo após soterramento e maturação. Recifes formam armadilhas estratigráficas, enquanto arenitos associados a falhas compõem armadilhas estruturais (Bjorlykke, 2010).

Lagos (sistemas lacustres) são particularmente eficientes na acumulação de matéria orgânica, sobretudo nos climas tropicais e subtropicais. A produtividade primária elevada, aliada à estratificação térmica da coluna d'água e à presença de condições anóxicas no hipolímnio, favorece a preservação da matéria orgânica. Lagos tectônicos, caracterizados por subsidência prolongada e baixa diluição clástica, como os observados nas bacias do Eoceno do Oriente Médio e nas bacias brasileiras como a Formação Irati (Permiano) na Bacia do Paraná ou a Formação Piçarras (NeoBarremino) importante rocha geradora da Bacia de Santos, são classicamente associados à geração de óleo de alta qualidade (Peters et al., 2005; Huc, 2003).

Deltas, por sua vez, representam ambientes transicionais entre sistemas continentais e marinhos. Nesses sistemas, a interação entre fluxos fluviais, marés e ondas propicia zonas de alta produtividade e acúmulo de matéria orgânica, especialmente em ambientes de planície deltaica e baías de inundação. No entanto, o elevado aporte clástico pode diluir a matéria orgânica e dificultar sua preservação, a

menos que haja compartimentos anóxicos, como lagoas ou planícies de maré mal drenadas (Allen & Allen, 2013; Tyson, 1995). Um exemplo de rochas geradoras com essa origem podem ser encontrados na bacia do Canadá em folhelhos do grupo Mannville (Cretáceo inferior) (Higley, D. K., et al. 2005).

Bacias marinhas restritas, como mares epicontinentais ou bacias tectonicamente isoladas, constituem ambientes ideais para a formação de rochas geradoras, quando há alta produtividade biológica associada à restrição de circulação e à formação de zonas anóxicas ou subóxicas na coluna d'água. Exemplos clássicos incluem o Mar Negro moderno e as bacias marginais do Cretáceo (ex.: Formação La Luna, Venezuela). Essas condições permitem a deposição de folhelhos ricos em matéria orgânica do tipo II, com elevado potencial para geração de petróleo (Peters et al., 2005; Huc, 2003).

Ambientes de mar aberto, por outro lado, geralmente apresentam circulação vigorosa e maior oxigenação, o que favorece a decomposição da matéria orgânica. Entretanto, em zonas de alta produtividade biológica associadas à ressurgência oceânica (upwelling), como na costa do Peru ou da Namíbia, pode haver acúmulo local de sedimentos ricos em matéria orgânica sob condições subóxicas, formando depósitos pontuais de interesse gerador (Tyson, 1995; Gluyas & Swarbrick, 2004).

3.2.3. Condições de Preservação da Matéria Orgânica

A preservação da matéria orgânica nos sedimentos está fortemente condicionada à presença de ambientes anóxicos no momento da deposição. Esses ambientes reduzem significativamente a atividade de bactérias aeróbicas e organismos bioturbadores, que de outra forma degradariam a matéria orgânica antes do soterramento (Allen & Allen, 2013; Gluyas & Swarbrick, 2004). Assim, a previsão da presença de rochas geradoras depende, em grande parte, da identificação de contextos paleoambientais onde condições anóxicas eram prováveis.

De acordo com Allen e Allen (2013), condições anóxicas se desenvolvem quando a demanda por oxigênio, geralmente impulsionada pela degradação da matéria orgânica morta, excede o suprimento disponível. Ambientes de alta produtividade orgânica superficial, como zonas eufóticas de mares e lagos, tendem a gerar grandes volumes de matéria orgânica que, ao se depositarem no fundo, consomem rapidamente o oxigênio dissolvido na coluna d'água (**Error! Reference source not found.**). A limitada circulação de água oxigenada nesses ambientes favorece o estabelecimento de condições anóxicas, especialmente em mares restritos, bacias profundas com águas estagnadas ou em situações de estratificação térmica da coluna d'água (Allen & Allen, 2013; Huc, 2003).

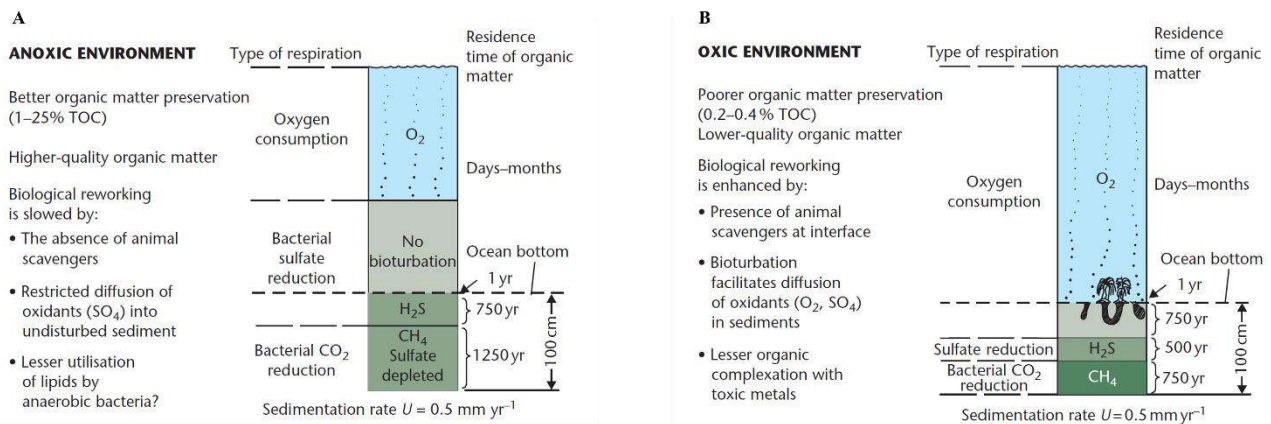


Figura 3.2. Degradação da matéria orgânica em ambientes óxicos e anóxicos. Em condições anóxicas, a preservação da matéria orgânica é favorecida pela redução da bioturbação e da atividade bacteriana aeróbica, comuns em ambientes oxigenados. (Allen & Allen, 2013).

Conforme discutido por Gluyas e Swarbrick (2004), essas condições são particularmente críticas nos primeiros centímetros de sedimento, onde ocorre a maior parte da bioturbação e degradação. Quando a concentração de oxigênio na água é inferior a 0,3 ml/L, a atividade da fauna bentônica é praticamente eliminada, permitindo a preservação de laminações e de matéria orgânica não alterada. Em tais ambientes, mesmo bactérias anaeróbicas, que continuam a degradar matéria orgânica apresentam atividade limitada devido à baixa difusividade de oxidantes no sedimento fino.

Segundo Huc (2003), essa degradação anaeróbica pode ser quase tão eficiente quanto a aeróbica. No entanto, em termos qualitativos, a preservação de compostos lipídicos, fundamentais para a formação de petróleo, tende a ser favorecida em ambientes anaeróbicos. Adicionalmente, a própria biomassa bacteriana em condições anóxicas pode contribuir significativamente para a carga orgânica do sedimento, especialmente em ambientes restritos e eutróficos.

Em termos de transporte da matéria orgânica para o fundo, Allen e Allen (2013) destacam que fatores como o tempo de trânsito da matéria na coluna d'água, o tamanho dos grãos do sedimento e a taxa de sedimentação controlam diretamente o grau de preservação. Taxas de sedimentação muito altas, embora possam proteger a matéria orgânica da oxidação ao reduzir seu tempo de exposição, geralmente resultam em diluição do conteúdo orgânico devido ao aporte mineral (Tissot, B. P., & Welte, D. H., 1984; Magoon & Dow, 1994). Esse é o caso típico de sequências pró-delta argilosas, onde mesmo que haja preservação relativa da matéria orgânica, sua concentração pode ser insuficiente para originar volumes comerciais de hidrocarbonetos líquidos.

Gluyas e Swarbrick (2004) ressaltam ainda que a permeabilidade dos sedimentos influencia a difusão de oxidantes. Sedimentos de granulação fina, típicos de ambientes de baixa energia, limitam a

difusão de oxigênio e demais oxidantes para os poros sedimentares, reduzindo a degradação microbiana e aumentando as chances de preservação da matéria orgânica. Já os sedimentos grossos, mais típicos de ambientes oxigenados, favorecem processos de oxidação e bioturbação.

A deposição de rochas geradoras, portanto, está fortemente relacionada à interação entre três fatores principais: (i) alta produtividade orgânica; (ii) limitação no suprimento de oxigênio (anoxia ou euxinia que inclui a presença de sulfeto de hidrogênio); e (iii) condições sedimentares favoráveis, como sedimentos finos e taxas de sedimentação moderadas (Allen & Allen, 2013; Huc, 2003). Em contrapartida, ambientes bem oxigenados podem ainda originar rochas geradoras em casos excepcionais, especialmente se a taxa de sedimentação for suficientemente alta para limitar o tempo de exposição à oxidação, embora essas rochas tendam a ter teores de carbono orgânico total mais baixos e maior geração de gás do que de óleo (Tissot, B. P., & Welte, D. H., 1984; Magoon & Dow, 1994).

3.3. FORMAÇÃO E TRANSFORMAÇÃO DO QUEROGÊNIO

3.3.1. Classificação do Querogênio

O querogênio representa a fração insolúvel da matéria orgânica presente em rochas sedimentares, sendo definido como um polímero de alto peso molecular, resistente a solventes orgânicos, ácidos e água. Essa substância é formada a partir da transformação de compostos orgânicos complexos, como aminoácidos, açúcares, ácidos húmicos e fúlvicos, durante o soterramento e compactação dos sedimentos (Bjørlykke, 2010). Nas primeiras dezenas a centenas de metros de profundidade, esses precursores passam por processos de polimerização, como a reação melanoide, e perdem grupos funcionais (ácidos, cetonas, aldeídos), o que resulta na diminuição da razão O/C e no aumento da razão H/C da matéria orgânica (Figura 3.3).

À medida que a profundidade de soterramento e a temperatura aumentam, o querogênio pode se decompor termicamente, gerando hidrocarbonetos líquidos e gasosos. Em bacias com soterramento mais lento, a geração de petróleo pode ocorrer por volta dos 100 °C, enquanto em ambientes com subsidência rápida, esse processo pode se iniciar em temperaturas superiores a 140 °C (Bjørlykke, 2010).

Do ponto de vista composicional, o querogênio é heterogêneo, podendo conter partículas morfologicamente reconhecíveis, como vitrinita e liptinita, derivadas de tecidos vegetais e algas, respectivamente. A separação do querogênio pode ser feita por digestão química (HCl e HF), métodos de densidade ou pirólise, técnica na qual a substância é aquecida entre 350–450 °C em atmosfera inerte, liberando compostos menores analisáveis por espectrometria de massas ou cromatografia gasosa.

Outros métodos analíticos empregados incluem espectroscopia IV, RMN e microscopia óptica com luz transmitida, refletida e ultravioleta (Bjørlykke, 2010; Gluyas & Swarbrick, 2004).

Com base em características químicas (relação H/C e O/C) e ópticas (refletância, fluorescência), o querogênio é convencionalmente classificado em quatro tipos principais, com diferentes potenciais geradores (Gluyas & Swarbrick, 2004):

- Tipo I (Liptinita): rico em lipídios e com alta relação H/C, é formado principalmente por algas e matéria sapropélica. Apresenta elevado potencial para geração de petróleo, sendo comum em ambientes lacustres. Sua fluorescência sob luz UV é intensa, e seu rendimento pode chegar a 80%. Apesar disso, é relativamente raro (Bjørlykke, 2010; Gluyas & Swarbrick, 2004).
- Tipo II (Exinita): de origem mista, derivado de fitoplâncton, zooplâncton e detritos vegetais membranosos, este tipo apresenta rendimentos médios de petróleo e gás (40–60%). Pode conter compostos sulfurados, sendo então denominado Tipo II-S. É o tipo de querogênio mais frequente em bacias marinhas (Gluyas & Swarbrick, 2004).
- Tipo III (Vitrinita): proveniente de material vegetal lenhoso (como lignina e celulose), possui baixa razão H/C e alto teor de oxigênio. É menos eficiente na geração de petróleo, sendo principalmente gerador de gás. Sua refletância aumenta com a maturação térmica, servindo como indicador de maturidade orgânica (Bjørlykke, 2010; Gluyas & Swarbrick, 2004).
- Tipo IV (Inertinita): composto por matéria orgânica altamente oxidada e carbonizada, como restos de vegetação previamente degradada. Apresenta baixíssimo teor de hidrogênio e, portanto, não possui potencial significativo de geração de hidrocarbonetos, sendo considerado "carbono morto" (Gluyas & Swarbrick, 2004).

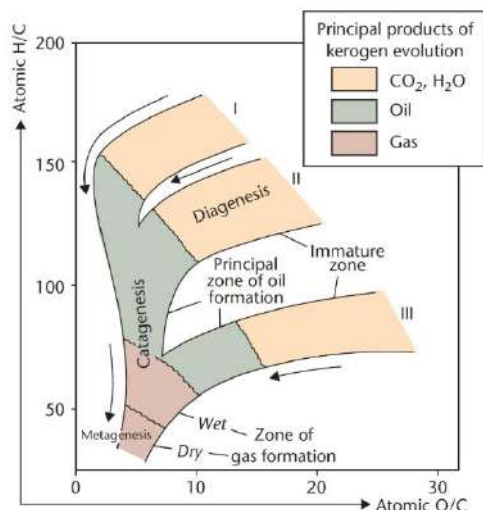


Figura 3.3. Representação esquemática da evolução do querogênio no diagrama de van Krevelen, no qual são indicados os estágios de diagênese, catagênese e metagênese, juntamente com os principais produtos gerados em cada fase do processo de maturação térmica da matéria orgânica (Tissot & Welte, 1984).

Além da composição e origem, o ambiente deposicional influencia fortemente a quantidade e a qualidade do querogênio. Ambientes anóxicos favorecem sua preservação, inibindo a degradação oxidativa e a bioturbação. Fatores como a taxa de sedimentação, aporte de material marinho versus terrestre e grau de retrabalhamento controlam o tipo de querogênio presente e, por consequência, o potencial gerador da rocha (Bjørlykke, 2010; Gluyas & Swarbrick, 2004).

A avaliação da quantidade de querogênio em rochas geradoras é comumente feita por meio da análise do Carbono Orgânico Total (TOC), expresso em porcentagem do peso da rocha. A qualidade, por sua vez, é medida por métodos de pirólise, que indicam o rendimento potencial de hidrocarbonetos (kg HC/tonelada de rocha), complementando os dados visuais e espectroscópicos (Gluyas & Swarbrick, 2004).

3.3.2. Evolução Térmica do Querogênio

A evolução térmica do querogênio nas rochas geradoras é um processo complexo que envolve a transformação de matéria orgânica em hidrocarbonetos sob condições geológicas específicas. Conforme discutido por Bjørlykke (2010), a conversão do querogênio em petróleo e gás requer temperaturas entre 80 e 150°C, o que corresponde a profundidades de 2 a 4 km com gradientes geotérmicos normais. Este processo é predominantemente impulsionado pela quebra de ligações carbono-carbono, resultando na formação de moléculas menores de hidrocarbonetos.

O conceito de maturidade térmica do querogênio é crítico, e pode ser avaliado através da refletividade da vitrinita, que indica o grau de transformação térmica (Bjørlykke, 2010). À medida que

a temperatura aumenta, a proporção de hidrocarbonetos leves também cresce, culminando na formação predominante de gás metano a temperaturas superiores a 170-180°C. Gluyas e Swarbrick (2004) enfatizam que a cinética da reação é controlada principalmente pela temperatura, com influências menores da pressão e da abundância do querogênio na rocha geradora. Eles observam que, em ambientes laboratoriais, é necessário atingir temperaturas significativamente mais altas para gerar petróleo, enquanto em bacias sedimentares naturais, o tempo geológico permite que a maturação ocorra em faixas de temperatura mais baixas.

Tissot e Welte (1984) descrevem um esquema evolutivo para a matéria orgânica, que inclui os estágios de diagênese, catagênese e metagênese (Figura 3.4). Durante a diagênese, a matéria orgânica começa a se transformar sob condições amenas, com a atividade microbiana desempenhando um papel crucial na decomposição e reorganização de biopolímeros em geopolímeros, que são precursores do querogênio. A catagênese é caracterizada por um aumento significativo de temperatura e pressão, levando à geração de petróleo líquido e gás úmido. Neste estágio, a matéria orgânica sofre grandes alterações, produzindo primeiro óleo e depois gás, principalmente metano. O estágio de metagênese, que antecede o metamorfismo, envolve transformações severas, onde o querogênio residual é convertido em carbono grafitico.

Allen e Allen (2013) oferecem um modelo cinético que descreve os estágios sucessivos da formação de petróleo, desde a diagênese até a metagênese, sem sobreposições significativas. Este modelo cinético prevê que o estágio imaturo precede a geração de petróleo, seguido pelo estágio de geração de óleo e gás e, finalmente, pelo estágio de metagênese, no qual o gás seco é gerado principalmente a partir de querogênio refratário.

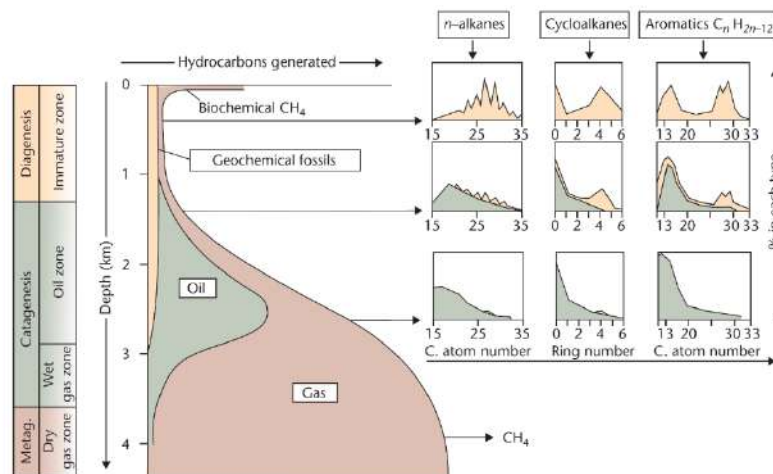


Figura 3.4. Esquema geral da formação de hidrocarbonetos em função do soterramento, conforme Tissot e Welte (1978). Com o aumento da temperatura devido ao soterramento progressivo, o material orgânico passa por estágios distintos: imaturo, geração de óleo,

craqueamento do óleo (estágio de gás úmido) e, por fim, geração de gás seco. A figura apresenta as distribuições típicas de n-alcenos, cicloalcanos e compostos aromáticos em três fases dessa evolução.

3.4. TEOR DE CARBONO ORGÂNICO TOTAL

O Carbono Orgânico Total (COT), ou Total Organic Carbon (TOC), é uma medida quantitativa da fração orgânica presente nas rochas sedimentares, sendo um parâmetro fundamental na avaliação do potencial gerador de hidrocarbonetos de rochas fonte. O COT representa a soma de todos os compostos orgânicos presentes em uma amostra, incluindo querogênio, betume e compostos residuais, sendo, portanto, um indicador direto da riqueza orgânica de uma rocha (Hunt, 1996; Tissot & Welte, 1984).

Para uma rocha ser considerada uma potencial geradora de óleo ou gás, ela deve conter uma quantidade mínima de matéria orgânica. De maneira geral, valores de COT superiores a 0,5% indicam rochas com potencial gerador, enquanto valores acima de 2% são considerados bons a excelentes como mostrados na Tabela 3.1 (Peters & Cassa, 1994). No entanto, o COT, isoladamente, não é suficiente para garantir o potencial de geração, sendo necessário considerá-lo em conjunto com a qualidade da matéria orgânica e seu grau de maturação térmica (Peters et al., 2007; Allen & Allen, 2013).

Tabela 3.1. Parâmetros de COT que classificam seu potencial gerador de hidrocarbonetos segundo Peters & Cassa, 1994.

COT (wt. %)	Potencial de Geração de Hidrocarbonetos
0 - 0.5	Pobre
0.5 - 1	Razoável
1 - 2	Bom
2 - 4	Muito Bom
> 4	Excelente

Diversas técnicas analíticas são utilizadas para a determinação do COT, sendo a combustão catalítica uma das mais comuns. Métodos indiretos, como o modelo proposto por Passey et al. (1990), permitem a estimativa do COT a partir de perfis geofísicos, especialmente em seções onde não há dados laboratoriais disponíveis. Essa abordagem tem sido amplamente aplicada na caracterização de formações não convencionais, como folhelhos ricos em matéria orgânica.

No contexto das bacias sedimentares brasileiras, Rodrigues (1995) ressalta que a ocorrência de hidrocarbonetos está geralmente associada a rochas geradoras com teores de carbono orgânico superiores a 1,5%. Em áreas mais prolíficas, esse valor tende a ultrapassar 2%, o que reforça a importância do teor de matéria orgânica como critério fundamental para avaliação do potencial gerador. A efetividade desse potencial depende do tipo de matéria orgânica presente (isto é, do tipo de

querogênio) e de sua evolução térmica ao longo da história geológica da bacia. Como enfatiza o autor, é a combinação entre a composição original da matéria orgânica e suas condições de preservação que define sua qualidade para a geração de hidrocarbonetos.

Allen e Allen (2013) ainda enfatizam que teores elevados de COT não garantem, por si sós, o potencial gerador. Isso porque, se a matéria orgânica presente for de natureza inerte, não haverá geração significativa de hidrocarbonetos, independentemente da quantidade. Por isso, além da quantificação do COT, é essencial avaliar a qualidade e a reatividade da matéria orgânica por meio de métodos como pirólise de rocha (Rock-Eval), análise da refletância da vitrinite e cromatografia gasosa.

Assim, o COT continua a ser um parâmetro essencial na geoquímica do petróleo, tanto para avaliação exploratória quanto para estudos de modelagem de geração e migração de hidrocarbonetos, desde que interpretado em conjunto com outros dados geoquímicos e petrofísicos.

3.5. CARACTERIZAÇÃO GEOQUÍMICA DA FORMAÇÃO PIMENTEIRAS

Diante da contextualização dos principais temas relacionados à rocha geradora, serão utilizados como referência dois importantes estudos que exploraram de forma aprofundada a Formação Pimenteiras, focando em sua caracterização geoquímica e origem deposicional. O primeiro estudo, realizado por Rodrigues (1995), e o segundo, mais recente, conduzido por Mussa (2020). Ambos destacam a importância das intrusões ígneas na geração de hidrocarbonetos e a considerável variação do Carbono Orgânico Total nos folhelhos dessa formação, tanto em profundidade quanto em diferentes regiões da Bacia do Parnaíba.

Rodrigues (1995) descreveu três intervalos de folhelhos radioativos (A, B e C) na Bacia do Parnaíba, cada qual apresentando características distintas com base em suas características litológicas, geoquímicas e estratigráficas, além da resposta geofísica observada em perfilagem de poços (Figura 3.5).

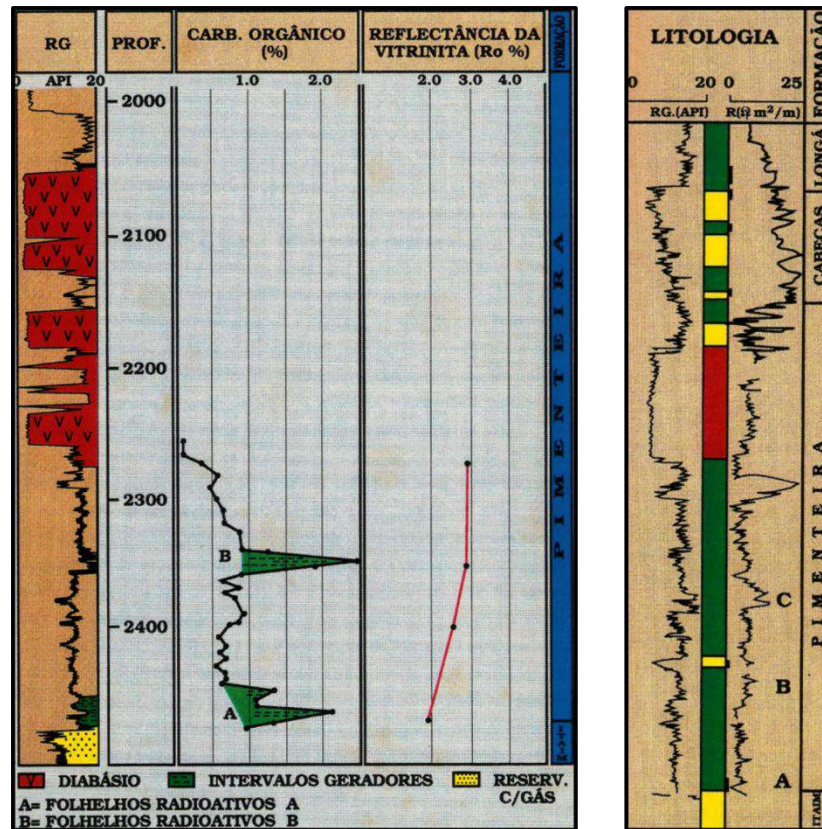


Figura 3.5. Classificação dos folhelhos radioativos interpretados por René Rodrigues (1995) em dois poços da Bacia do Parnaíba.

Os folhelhos radioativos A exibem concentrações de COT entre 1 e 3%, com os valores mais elevados concentrados na base do intervalo. Este autor sugere que a matéria orgânica é uma mistura heterogênea de tipos II e III, e destaca o efeito térmico das intrusões de diabásio como um fator determinante na maturação da matéria orgânica. Os folhelhos radioativos B apresentam teores de COT ligeiramente superiores, variando entre 1 e 3,5%, também com uma composição mista de matéria orgânica dos tipos II e III. Rodrigues observa que as porções mais espessas e com maior concentração de matéria orgânica estão situadas no norte da bacia, indicando condições favoráveis para a preservação da matéria orgânica. Por sua vez, os folhelhos radioativos C, correlacionados ao momento de inundação marinha máxima do Devoniano, possuem os teores mais elevados de COT, situando-se entre 1 e 5%. Rodrigues enfatiza que, apesar do potencial qualidade da matéria orgânica para geração de hidrocarbonetos, o efeito térmico das intrusões ígneas influencia significativamente a maturação térmica, especialmente em áreas mais senis.

A Formação Pimenteiras da Bacia do Parnaíba, conforme investigada por Mussa (2020), demonstra uma significativa riqueza em matéria orgânica, com teores de Carbono Orgânico Total (COT) variando de 1,54 a 16,60 wt.%. Esta variação reflete uma composição diversificada dos

folhelhos devonianos, que, segundo Rodrigues (1995), consistem em uma mistura de querogênios dos tipos II e III, resultantes dos ciclos transgressivos-regressivos do Devoniano. Mussa (2020) confirma que a origem da matéria orgânica é uma combinação de fontes marinha e continental. Em seu trabalho de doutorado, a análise de 32 poços, abrangendo 1077 amostras, revelou uma distribuição espacial heterogênea da Formação Pimenteiras (Figura 3.6), com profundidades do topo entre 500-2000 m e espessuras variando de 40 a 680 m. Mussa (2020) observa que as maiores espessuras ocorrem na parte centro-norte da bacia, com a profundidade máxima atingindo 2600 m na parte central.

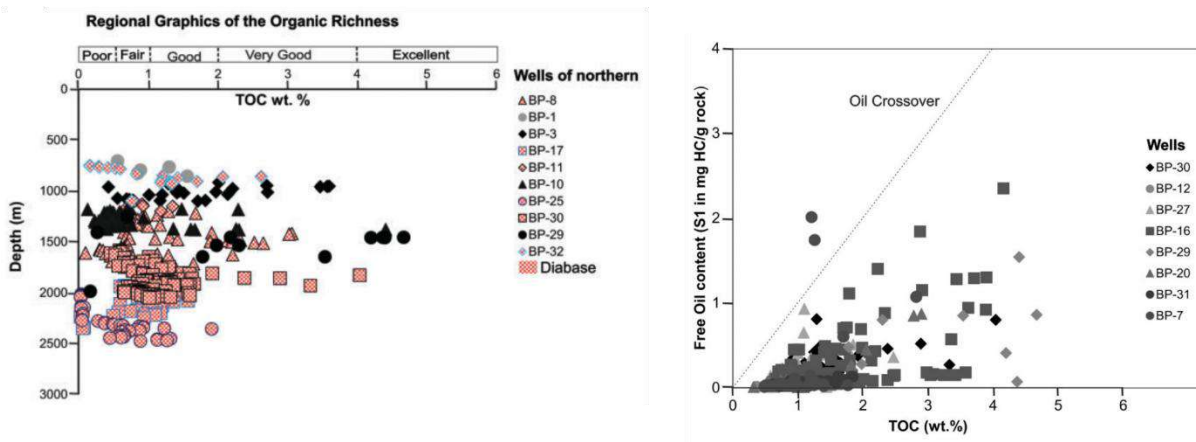


Figura 3.6. Resultados obtidos por Mussa (2020) em poços da região norte da bacia, evidenciando a variação do teor de COT de maneira geral e em profundidade, bem como sua relação com as soleiras de diabásio.

Os resultados encontrados por Mussa (2020) indicam que a maioria das amostras possui teores de COT entre 0,1 e 4,7 wt.%, classificados por Peters e Cassa (1994) como de bom a muito bom potencial de rocha geradora. No entanto, baixos teores de COT, entre 0,1 e 0,9 wt.%, foram observados nos poços ao sul e centro da bacia, classificados como de potencial de rocha geradora de ruim a razoável. As variações de COT na parte norte são classificadas de ruim a excelente, um padrão também observado na parte central. No sul, os teores de COT variam de ruim a bom, sugerindo que o acúmulo e preservação da matéria orgânica nas regiões norte e central estão ligados à anoxia do sistema, enquanto no sul, a preservação é favorecida por menores profundidades de sedimentação.

Mussa (2020) também destaca a presença de querogênio tipo IV, especialmente em poços com intrusões ígneas, sugerindo que o calor adicional pode ter causado o craqueamento do querogênio. As intrusões contribuem para a maturação da matéria orgânica, reduzindo os rendimentos de pirólise e os valores de índice de hidrogênio (HI). Estudos de Meyers e Simoneit (1999) indicam que folhelhos ricos em carbono orgânico próximos a intrusões ígneas apresentam teores reduzidos de hidrocarbonetos e alterações nos parâmetros geoquímicos. Dois poços na parte norte da bacia foram analisados para

tendências geoquímicas, mostrando que, em poços sem intrusões ígneas, o teor de COT diminui com a profundidade, variando de 0,2 a 4,7 wt.%, classificado como de fraco a excelente potencial de rocha geradora (Peters e Cassa, 1994).

Em síntese, Rodrigues (1995) e Mussa (2020) enfatizam a relevância das intrusões ígneas na maturação térmica da matéria orgânica na Formação Pimenteiras. Enquanto Rodrigues foca nos níveis de maior radioatividade associados a níveis geradores utilizando parâmetros geoquímicos, Mussa expande essa análise com uma avaliação abrangente do potencial de geração de hidrocarbonetos em múltiplos poços, variando em profundidade, região e a presença/ausência de intrusões ígneas. A integração desses estudos proporciona uma compreensão aprofundada dos processos que influenciam a origem e maturação da matéria orgânica da Formação Pimenteiras.

4. Classificação de Fácies e Estudo de Viabilidade

4.1. INTRODUÇÃO

O intervalo da Formação Pimenteiras, que pode atingir até 500 m de espessura, é predominantemente descrito em amostras de calha como folhelhos, variando entre cinza-escuro, preto e esverdeado. A interpretação dos perfis geofísicos evidencia variações de eletrofácies, caracterizando-a como uma formação heterogênea, fato explicado tanto por sua gênese, ciclos transgressivos-regressivos do Devoniano que influenciaram a heterogeneidade litológica e a preservação da matéria orgânica, quanto pelas alterações provocadas pelas intrusões ígneas (Mussa *et al.*, 2024). Com o objetivo de discriminar as eletrofácies observadas nos dados de poço e verificar se estas podem ser distinguidas por meio da inversão sísmica, além de avaliar sua relação com o teor de COT, aplicou-se inicialmente a classificação não supervisionada pela técnica *K-means*, a fim de definir os diferentes agrupamentos identificados pelo comportamento dos perfis de poço. Em seguida, os resultados dessa etapa foram utilizados na aplicação da classificação Bayesiana, de modo a avaliar se as fácies apresentam respostas distintas de impedância acústica, permitindo a viabilidade da inversão sísmica. Assim, neste capítulo são apresentadas as aplicações dessas duas técnicas, com o propósito de detalhar a formação de interesse e compreender a relação entre fácies, conteúdo orgânico e impedância acústica, objetivo central deste trabalho e aprofundado nos capítulos seguintes.

4.2. FUNDAMENTAÇÃO

As técnicas de aprendizado de máquina podem ser genericamente classificadas em métodos supervisionados e não supervisionados, de acordo com a disponibilidade de informações prévias sobre os dados analisados. No aprendizado supervisionado, o algoritmo é treinado a partir de um conjunto de dados rotulados (*labeled*), no qual cada observação está associada a uma classe ou valor de referência conhecido (Bishop, 2006; Hastie et al., 2009). O objetivo principal é aprender uma função de mapeamento capaz de generalizar esse relacionamento e realizar previsões ou classificações em novos

dados. Esse tipo de abordagem é amplamente empregado em problemas de classificação e regressão, nos quais existe conhecimento prévio sobre as categorias ou respostas desejadas.

Em contraste, o aprendizado não supervisionado é aplicado quando não há rótulos (*label*) ou classes previamente definidos. Nessa abordagem, os algoritmos buscam identificar padrões, estruturas internas ou relações de similaridade diretamente a partir da distribuição dos dados no espaço de atributos (Bishop, 2006; Hastie et al., 2009). Métodos de clusterização, como o K-means, inserem-se nesse contexto, sendo particularmente adequados para análises exploratórias e para a identificação de agrupamentos naturais em conjuntos de dados complexos, nos quais a definição prévia de classes não é viável ou desejável.

4.2.1. Clusterização *K-Means*

O algoritmo *K-means* é uma técnica de aprendizado não supervisionado, publicado pela primeira vez no trabalho de Iwayama, M., & Tokunaga, T. (1995), voltada para a clusterização de dados, cuja aplicação tem sido amplamente consolidada nas últimas décadas. Embora tenha sido proposto há mais de meio século, sua utilização permanece recorrente devido à simplicidade de implementação, à eficiência computacional e ao sucesso empírico obtido em diferentes contextos de análise de dados (Jain, A.K. 2009).

De modo geral, a clusterização consiste em agrupar observações em subconjuntos ou classes (*clusters*), de forma que objetos pertencentes a um mesmo grupo apresentem alta similaridade entre si, enquanto aqueles em grupos distintos exibam baixa similaridade. Nesse contexto, os *clusters* podem ser entendidos como regiões de maior densidade no espaço de atributos, separadas por regiões de menor densidade, sendo o objetivo central identificar estruturas ou padrões naturais em conjuntos de dados desprovidos de rótulos (Duda *et al.*, 2001). A análise de agrupamento insere-se, portanto, no domínio do aprendizado não supervisionado, distinguindo-se dos métodos supervisionados que dependem de informações previamente rotuladas (Jain, A. *et al.*, 1999).

Segundo a classificação proposta por Tukey (1977), os métodos de análise de dados podem ser organizados em duas abordagens principais: (i) exploratória ou descritiva, voltada à identificação de padrões e estruturas latentes sem hipóteses prévias, e (ii) confirmatória ou inferencial, direcionada à validação de hipóteses e modelos. Contudo, deve-se reconhecer que a definição de um "bom" agrupamento é, em certa medida, subjetiva, pois depende do contexto de aplicação e do conhecimento prévio do pesquisador sobre o domínio em estudo.

Matematicamente o algoritmo *K-means* é definido como um conjunto de n objetos representados por vetores de atributos $X = \{x_i\}, i = 1, \dots, n$, o *K-means* busca particioná-los em K grupos $C = \{c_k, k = 1, \dots, K\}$. O critério de otimização baseia-se na minimização da soma dos erros quadráticos entre cada ponto e o centróide de seu respectivo *cluster*. Para cada *cluster* c_k , define-se a média μ_k , e a função objetivo do algoritmo corresponde à soma das distâncias quadráticas de todos os pontos até seus centroides:

$$J(c_k) = \sum_{x_i \in c_k} \|x_i - \mu_k\|^2 \quad (4.1)$$

Assim, o algoritmo procura a partição que minimize a função custo global:

$$J(C) = \sum_{k=1}^K \sum_{x_i \in c_k} \|x_i - \mu_k\|^2 \quad (4.2)$$

O procedimento iterativo do *K-means* pode ser descrito em quatro etapas principais (Jain & Dubes, 1988) como também ilustrado pela Figura 4.1:

1. Seleção inicial do número de *clusters* K e definição de centroides iniciais (aleatórios ou segundo heurísticas).
2. Atribuição de cada ponto de dado ao *cluster* cujo centróide seja o mais próximo segundo uma métrica de distância, geralmente a euclidiana.
3. Recomputação dos centroides com base nos pontos atribuídos a cada *cluster*.
4. Repetição iterativa dos passos 2 e 3 até que não haja mais alterações significativas nos centroides ou até que seja atingido um critério de convergência.

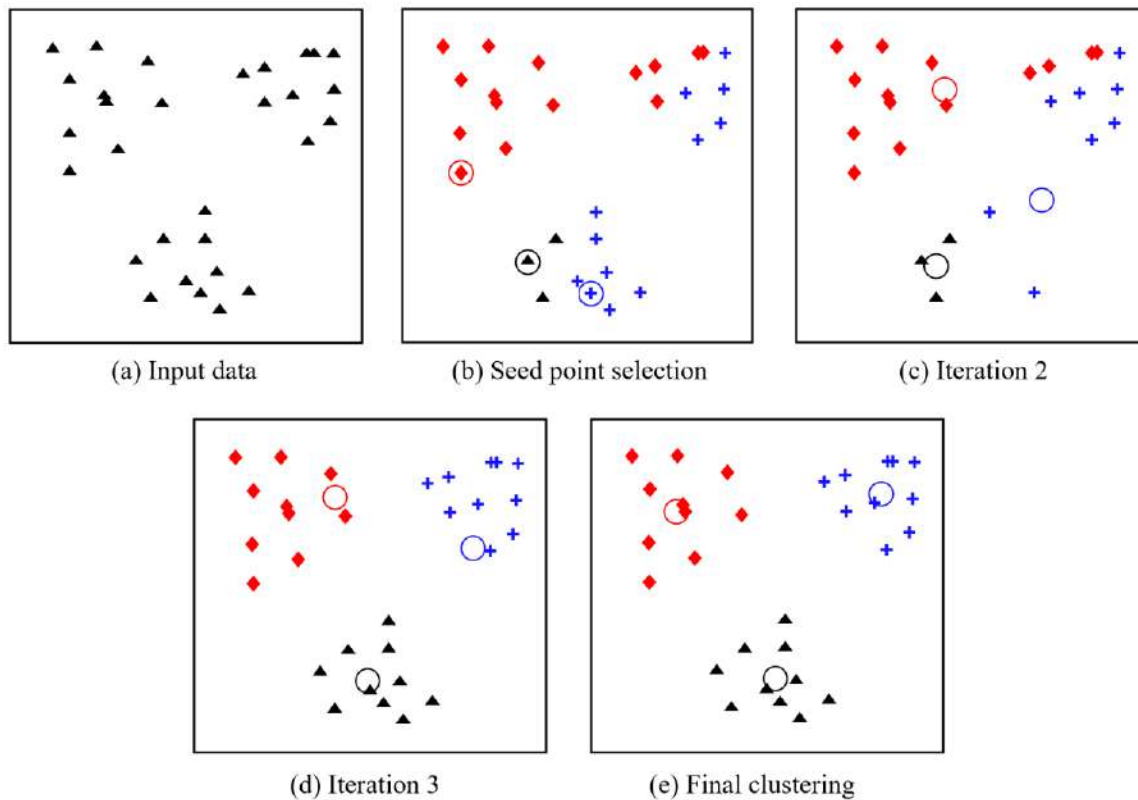


Figura 4.1. Representação esquemática do funcionamento do algoritmo *K-means*. (a) Conjunto de dados bidimensionais com três agrupamentos potenciais; (b) seleção inicial de três centróides e atribuição preliminar dos pontos aos respectivos *clusters*; (c) e (d) etapas intermediárias do processo iterativo, nas quais os rótulos de *cluster* e as posições dos centróides são atualizados a cada iteração; (e) resultado final da clusterização, mostrando a configuração estável alcançada após a convergência do algoritmo (Jain, A.K. 2009).

Apesar da popularidade e simplicidade do método de Lloyd para *K-means*, o algoritmo clássico possui limitações conhecidas. Um dos fatores mais críticos é a escolha do número K , para o qual não existe um critério universalmente ótimo. Diversas estratégias heurísticas têm sido propostas, como o método do Elbow (Thorndike, 1953), que envolve plotar a variância explicada por diferentes números de *clusters* e identificar o ponto "cotovelo", onde a taxa de variância diminui drasticamente e se estabiliza, sugerindo uma contagem de *clusters* apropriada para análise ou treinamento do modelo. Além disso, o método *K-means* é sensível à inicialização dos centróides, a escolha aleatória pode levar a soluções com qualidade arbitrariamente ruim. O método padrão pode convergir para mínimos locais muito distantes do ótimo global, sem garantias de aproximação.

A inicialização dos centróides pode ser dada de forma aleatória, determinística ou probabilística. O método mais robusto que melhor supera essa limitação é a inicialização chamada *K-means++* proposto por Arthur e Vassilvitskii (2006), no qual inicialmente o centróide é selecionado a partir um ponto do conjunto de dados escolhido de forma aleatória. A partir desse ponto, cada novo centróide é escolhido com probabilidade proporcional ao quadrado da distância mínima de cada ponto

ao centro já selecionado (distância D^2). Para cada ponto p , a probabilidade de ser escolhido como próximo centro é proporcional a $D(p)^2$, onde $D(p)$ é a menor distância de p a qualquer centróide já selecionado. Essa estratégia de *seeding*, denominada amostragem D^2 (D^2 seeding), assegura que centróides são espalhados no espaço de forma que regiões densas e fora dos centros já escolhidos tenham maior chance de serem representadas, evitando agrupamentos ruins iniciais. Experimentos conduzidos pelos autores evidenciam que o método melhora simultaneamente a acurácia e a velocidade de convergência, já que a inicialização mais cuidadosa reduz o número de iterações necessárias para estabilização do algoritmo.

4.2.3. Classificação Bayesiana

A classificação Bayesiana, ou decisão Bayesiana, fundamenta-se no teorema de Bayes, sendo aplicada para a determinação da probabilidade condicional de um evento dado um conjunto de evidências. Para compreender essa abordagem, é necessário introduzir o conceito de espaço amostral (Ω), definido como o conjunto de todos os valores possíveis de uma variável aleatória. Cada elemento desse conjunto é denominado ponto amostral, e os subconjuntos formados por esses pontos recebem o nome de eventos (Duda, R. O., & Hart, P. E., 2006).

O tratamento formal da probabilidade envolve três componentes principais: o espaço amostral, os eventos e uma função de probabilidade associada a cada evento. Essa função deve obedecer a três condições fundamentais, conhecidas como axiomas de Kolmogorov (Kolmogorov, A. N., 2018), formulados a partir da teoria dos conjuntos. Em termos práticos, para cada evento A pertencente ao espaço amostral Ω , atribui-se um valor $Pr(A)$, o qual representa a probabilidade de ocorrência desse evento. Para que essa atribuição seja matematicamente válida, devem ser satisfeitas as seguintes condições (DeGroot & Schervish, 2012):

O primeiro axioma afirma que a probabilidade de cada evento deve ser não negativa.

- i. Para cada evento A , $Pr(A) \geq 0$.

O segundo axioma afirma que se é certo que um evento ocorrerá, então a probabilidade desse evento é 1.

- ii. $Pr(\Omega) = 1$.

Quando dois eventos são mutuamente exclusivos, a probabilidade de ocorrência de pelo menos um deles corresponde à soma de suas probabilidades individuais. Essa característica, denominada propriedade aditiva, pode ser estendida não apenas a um número finito de eventos disjuntos, mas

também a sequências infinitas. Caso a aditividade fosse assumida apenas para conjuntos finitos, não haveria garantia de sua validade em situações infinitas. Por outro lado, ao pressupor sua validade para sequências infinitas, a propriedade se mantém automaticamente para qualquer conjunto finito. Essa generalização constitui a base do terceiro axioma da teoria da probabilidade.

iii. Para cada sequência infinita de eventos disjuntos A_1, A_2, \dots ,

$$Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} Pr(A_i) \quad (4.3)$$

Com base nos axiomas da probabilidade, podemos utilizar o exemplo dado por Bussab, W. D. O., & Morettin, P. A. (2010) ao considerar uma partição enumerável do espaço amostral como ilustrado pela Figura 4.2, isto é, um conjunto de subconjuntos mutuamente exclusivos cuja união possui probabilidade igual a 1. Seja $A \in \Omega$ um evento qualquer, para o qual se conhecem as probabilidades $P(C_i)$ e $P(C_i|A)$, com $i = 1, 2, \dots, n$.

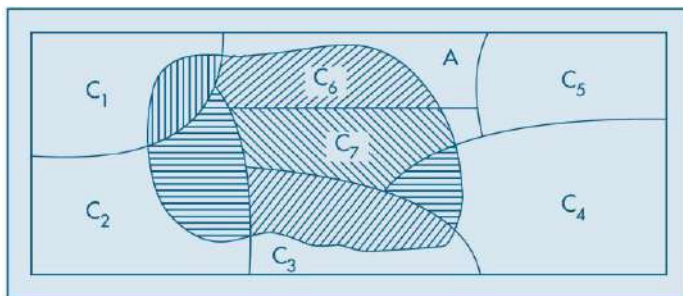


Figura 4.2. Exemplo ilustrativo de uma partição do espaço amostral em subconjuntos mutuamente exclusivos, utilizado para explicar a aplicação do Teorema de Bayes na classificação probabilística (Bussab, W. D. O., & Morettin, P. A. (2010)).

Nessa configuração, aplica-se o teorema de Bayes, que estabelece a probabilidade condicional de C_i dado A :

$$P(C_i|A) = \frac{P(C_i)P(A|C_i)}{\sum_{j=1}^n P(C_j)P(A|C_j)}, i = 1, 2, \dots, n \quad (4.4)$$

Nesse contexto, os conjuntos C_1, \dots, C_n podem ser interpretados como hipóteses mutuamente excludentes, sendo apenas uma verdadeira. A ocorrência do evento A altera a probabilidade inicial $P(C_i)$ (probabilidade a priori), produzindo a probabilidade condicional $P(C_i|A)$ (probabilidade a posteriori). Essa modificação ocorre multiplicando a probabilidade *a priori* por:

$$\frac{P(A|C_i)}{\sum_{j=1}^n P(C_j)P(A|C_j)} \quad (4.5)$$

onde $P(A|C_i)$ corresponde à verossimilhança associada à hipótese C_i . Assim, observa-se que $P(C_i|A) > P(C_i)$ sempre que $P(A|C_i) > P(A)$, sendo $P(A)$ calculado como média ponderada das verossimilhanças $P(A|C_j)$, cujos pesos são as probabilidades $P(C_j)$ que têm soma unitária. Como o numerador é sempre uma das parcelas do denominador $P(A)$, é feito o uso de um novo índice, j , na decomposição do mesmo.

Em palavras, conforme destacado por Duda *et al.* (2001), o teorema de Bayes pode ser expresso pela relação:

$$a\ posteriori = \frac{\text{verossimilhança} \times a\ priori}{\text{evidência}} \quad (4.6)$$

Uma vez que, a classificação Bayesiana constitui uma aplicação direta do Teorema de Bayes, sendo utilizada para estimar a probabilidade de um determinado padrão de variáveis estar associado a uma classe específica, é necessário que as classes sejam previamente definidas e que suas probabilidades iniciais (*a priori*) e funções de verossimilhança estejam disponíveis ou possam ser adequadamente estimadas.

Segundo Duda *et al.* (2001), esse problema pode ser ilustrado por meio de um exemplo simples, no qual se consideram duas classes, ω_1 e ω_2 , e uma única variável contínua $x \in [-\infty, \infty]$. As distribuições de x em cada classe podem ser representadas por funções densidade de probabilidade (PDF), denominadas verossimilhanças condicionais $p(x|\omega_j)$. Essas funções descrevem a probabilidade de se observar determinado valor de x , dado que o padrão pertence à classe ω_j , sendo normalizadas de modo que a área sob cada curva seja igual a 1 (Figura 4.3).

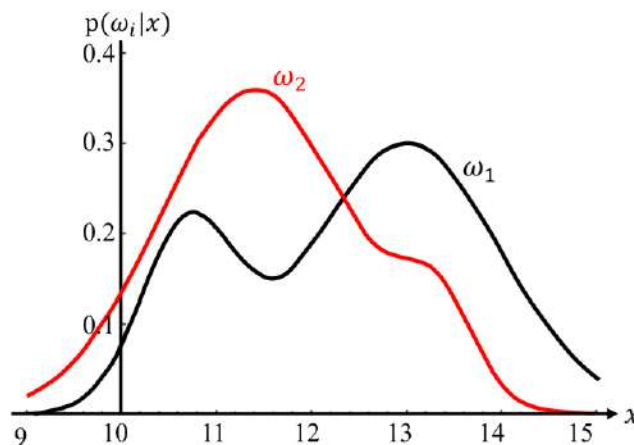


Figura 4.3. Exemplo de funções de densidade de probabilidade condicionais a diferentes classes, ilustrando a probabilidade de observação de um valor específico da variável x , dado que a amostra pertence à categoria ω . Essas distribuições representam

verossimilhanças associadas às hipóteses de classificação, constituindo a base para a aplicação do Teorema de Bayes no processo decisório (Duda *et al.* (2001)).

O Teorema de Bayes permite converter a probabilidade *a priori* $P(\omega_j)$ em uma probabilidade *a posteriori* $P(\omega_j|x)$, isto é, a probabilidade de o estado da natureza corresponder a ω_j , dado que se observou o valor x . Nessa formulação, o produto entre a probabilidade *a priori* e a verossimilhança assume papel central na determinação da probabilidade *a posteriori*, enquanto o termo de evidência $p(x)$ atua apenas como fator de normalização, garantindo que a soma das probabilidades seja igual a 1. A variação de $P(\omega_j|x)$, com x é ilustrada na Figura 4.4.

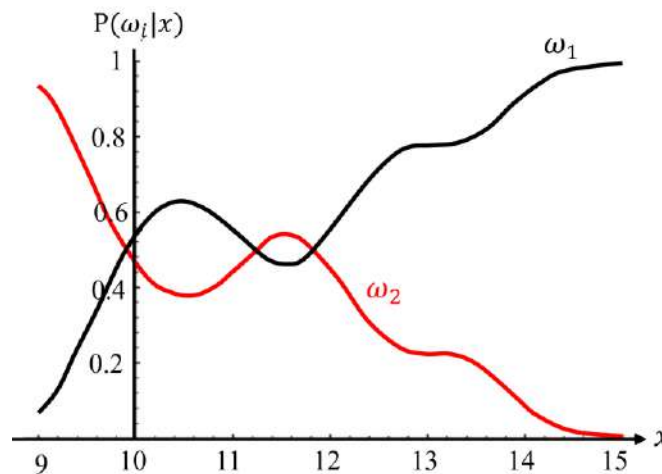


Figura 4.4. Probabilidades a posteriori calculadas para valores específicos de probabilidades a priori $P(\omega_1)$ e $P(\omega_2)$, considerando as densidades de probabilidade condicionais de classe apresentadas na Figura 3 (Duda *et al.* (2001)).

A regra de decisão Bayesiana estabelece que, para uma observação x , deve-se escolher ω_1 se $P(\omega_1|x) > P(\omega_2|x)$ e ω_2 no caso contrário. Essa decisão, embora não elimine totalmente a ocorrência de erros, minimiza a probabilidade média de erro, a qual pode ser expressa como:

$$P(\text{erro}) = \int_{-\infty}^{\infty} P(\text{erro}, x) dx = \int_{-\infty}^{\infty} P(\text{erro}|x)p(x) dx \quad (4.7)$$

De forma equivalente, a regra de decisão pode ser expressa pela comparação direta entre as verossimilhanças ponderadas pelas probabilidades *a priori*:

Decidir ω_1 se $p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2)$; caso contrário, dividir ω_2 .

Situações particulares tornam esse raciocínio ainda mais claro. Por exemplo, se $p(x|\omega_1) = p(x|\omega_2)$, a observação de x não fornece informação adicional, e a decisão depende unicamente das probabilidades *a priori*. Por outro lado, se $P(\omega_1) = P(\omega_2)$, ambas as classes são igualmente prováveis, e a decisão passa a depender exclusivamente das verossimilhanças. Em termos gerais, a classificação

Bayesiana resulta da combinação desses dois elementos, buscando sempre a minimização da probabilidade de erro (Duda *et al.*, 2001).

4.3. MÉTODO

Os métodos adotados neste estudo consistem na integração das técnicas *K-means* e Classificação Bayesiana para a definição de fácies e em seguida para um estudo de viabilidade para inversão sísmica. O fluxo de trabalho que resumo o desenvolvimento dessas duas aplicações está mostrado na Figura 4.5, e será detalhada nos subtópicos a seguir.

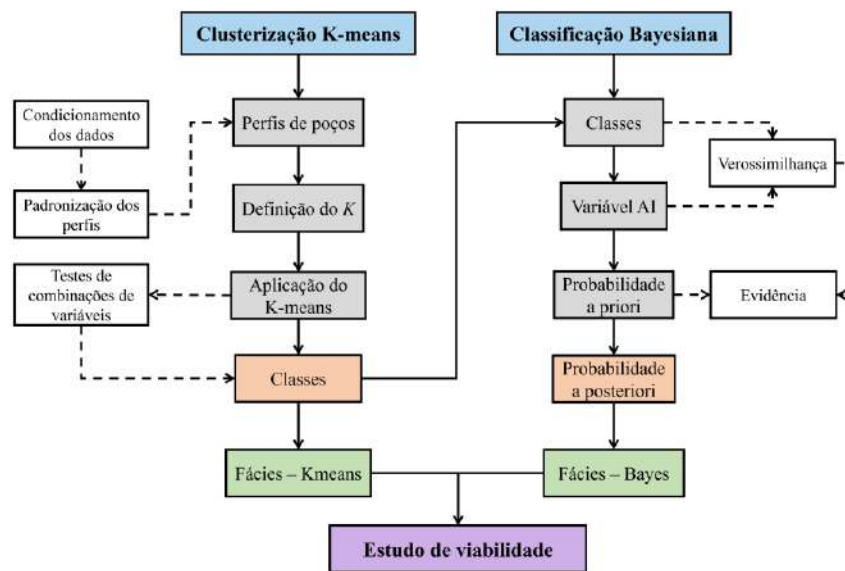


Figura 4.5. Fluxo de trabalho integrando as técnicas *K-means* e classificação Bayesiana.

4.3.1. Definição de Fácies via *K-Means*

4.3.1.1. Elaboração de hipóteses

A clusterização não supervisionada como já detalhada anteriormente, terá a função de separar em diferentes classes dados que contém características semelhantes, portanto, o significado no geral é dado, ou interpretado de acordo com o conhecimento do objeto de estudo em questão. Dessa forma, antes de aplicar o método foi elaborado algumas hipóteses e questionamentos a serem respondidos relacionadas às fácies:

i) O algoritmo deveria ser capaz de separar de forma evidente pelo menos duas classes, as quais pertenceriam as rochas folhelhos e soleira de diabásio?

ii) É observado folhelhos alterados para rochas metassedimentares descritas nos poços para as rochas muito próximas as soleiras. Será possível identificar essa terceira classe?

iii) Além dessas três fácies conhecidas, ao interpretar um conjunto de perfis, nota-se pelo menos três comportamentos diferentes associados às rochas finas. Será possível definir de maneira compreensível mais de duas classes para os intervalos de folhelho?

iv) Unificando perfis de poços e as amostras de COT, disponível em apenas um dos poços, é perceptível uma tendência de diminuição ao se aproximar das soleiras, porém longe das soleiras há variações do teor de COT. Essas variações poderiam estar relacionadas não só a geração promovida pelo calor das rochas ígneas, mas também às diferentes fácies de folhelhos mais ricas ou menos ricas em matéria orgânica. Será possível diferenciá-las de acordo com o teor de COT ou pelo menos fácies predominantes por teor de COT?

Seguindo a elaboração desses questionamentos e hipóteses de acordo com o que é interpretado dos perfis de poços e mesmo pelo que tem disponível de descrição de amostras de calha, podemos estabelecer não somente os possíveis rótulos que serão dados as classes separadas pelo K-means mas também direcionar para a definição do número de clusters que será estabelecido.

4.3.1.2. Tratamento dos perfis de poços

O método *K-means* é sensível à diferença de escala entre as variáveis, portanto, a normalização do conjunto de dados é necessária. Antes disso ainda, é importante realizar o pré-condicionamento das curvas para eliminar principalmente valores espúrios causados por um erro de medição, ou mesmo devido a zonas de desmoronamento que ocorre nas paredes dos poços. Por se tratar de um intervalo de folhelhos, essas etapas de tratamento das variáveis se tornam ainda mais cruciais, pois é justamente nessas formações que são os arrombamentos ou *washout* são mais frequentes.

As curvas selecionadas inicialmente para o estudo pertencem a suite básica de perfis de poços como: *Gamma ray* (GR), Fator fotoeletrico (PEFZ), Sônico (DTCO), Resistividade profunda (AT90), Neutrão (NPHI) e Densidade (RHOZ). O tratamento dos dados ocorreu principalmente sobre as curvas de RHOZ e DTCO que passaram pela remoção de pontos espúrios, pois foram os mais afetados pelas condições dos poços como fica evidente pela curva de Caliper (CALI e DCAL). Todos os dados passaram por uma avaliação de controle de qualidade, principalmente quanto aos intervalos de ausência de medição (*gaps*) os quais foram interpolados e junção (*merge*) de dados medidos em diferentes corridas da perfilagem. Os dados de entrada após o pré-condicionamento estão mostrados individualmente por poço nas Figura 4.6, Figura 4.7, Figura 4.8 e Figura 4.9 abaixo.

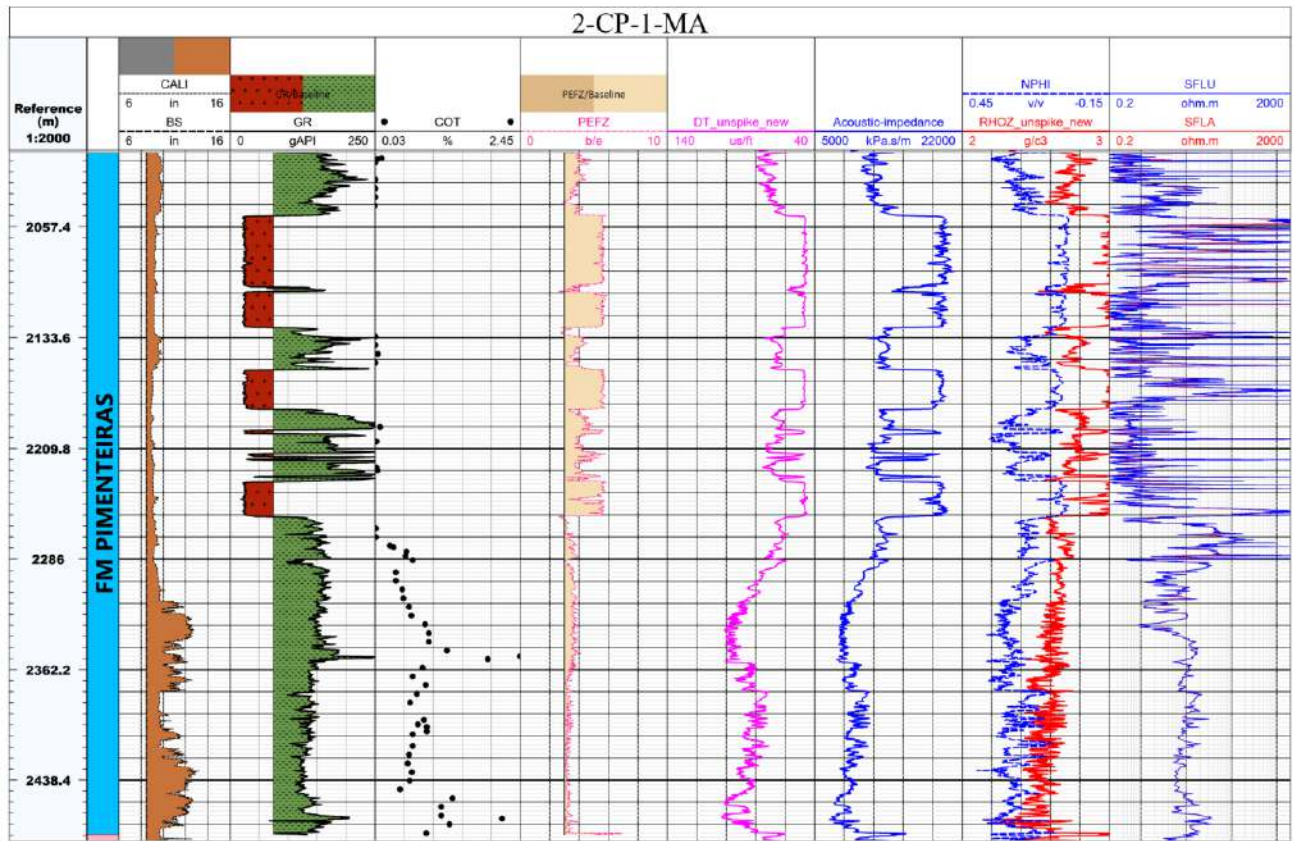


Figura 4.6. Suite básica dos perfis do poço 2-CP-1-MA após o pré-condicionamento. O comportamento das curvas sugere a existência de diferentes eletrofácies, e por isso os são candidatos a compor o conjunto de dados para a clusterização não supervisionada.

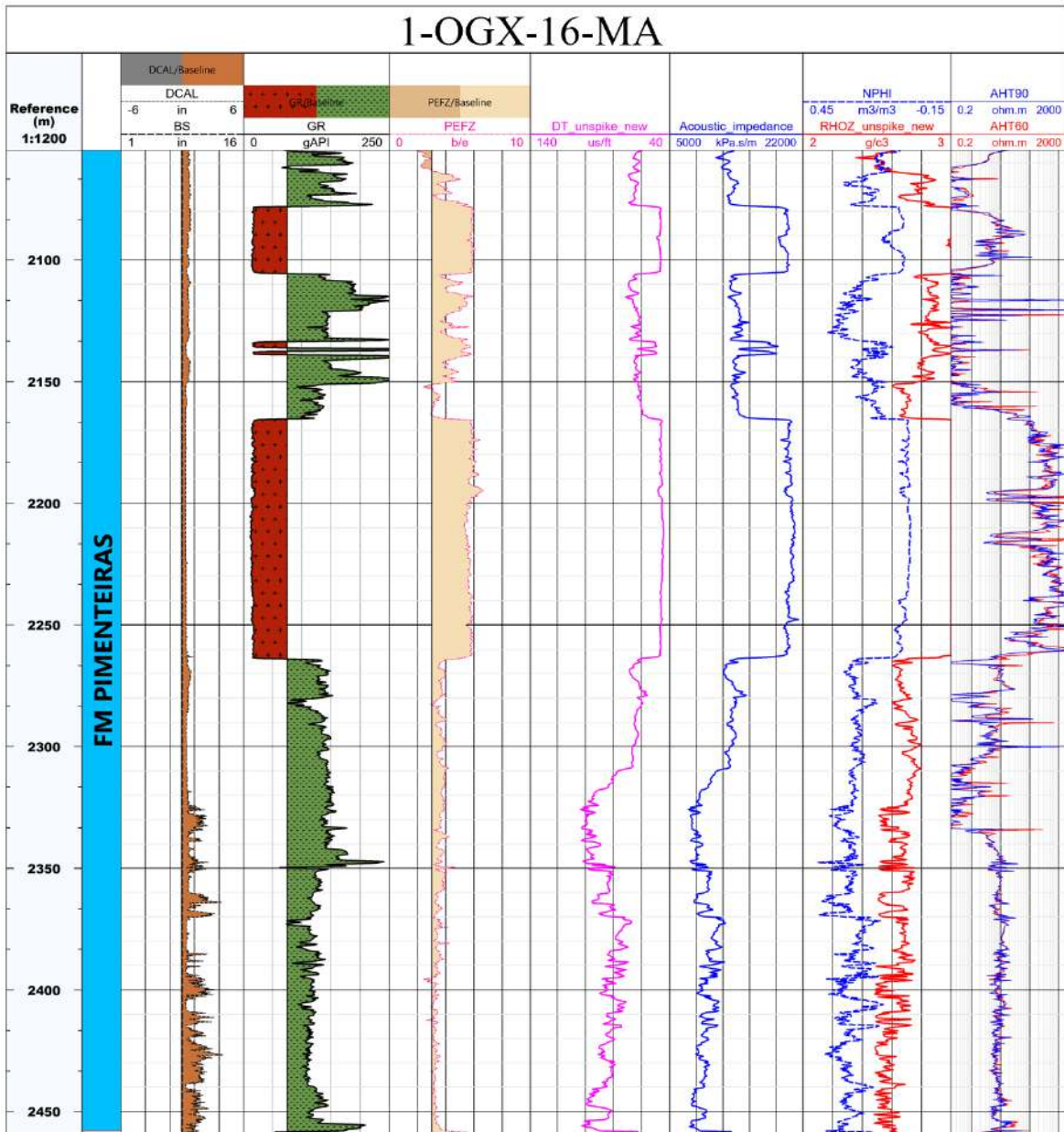


Figura 4.7. Dados disponíveis no poço 1-OGX-16-MA após o pré-condicionamento. A curva de *gamma-ray* diferencia de forma evidente os folhelhos das soleiras de diabásio, e o comportamento geral dos perfis elétricos refletem a heterogeneidade presente no intervalo da Fm. Pimenteiras.

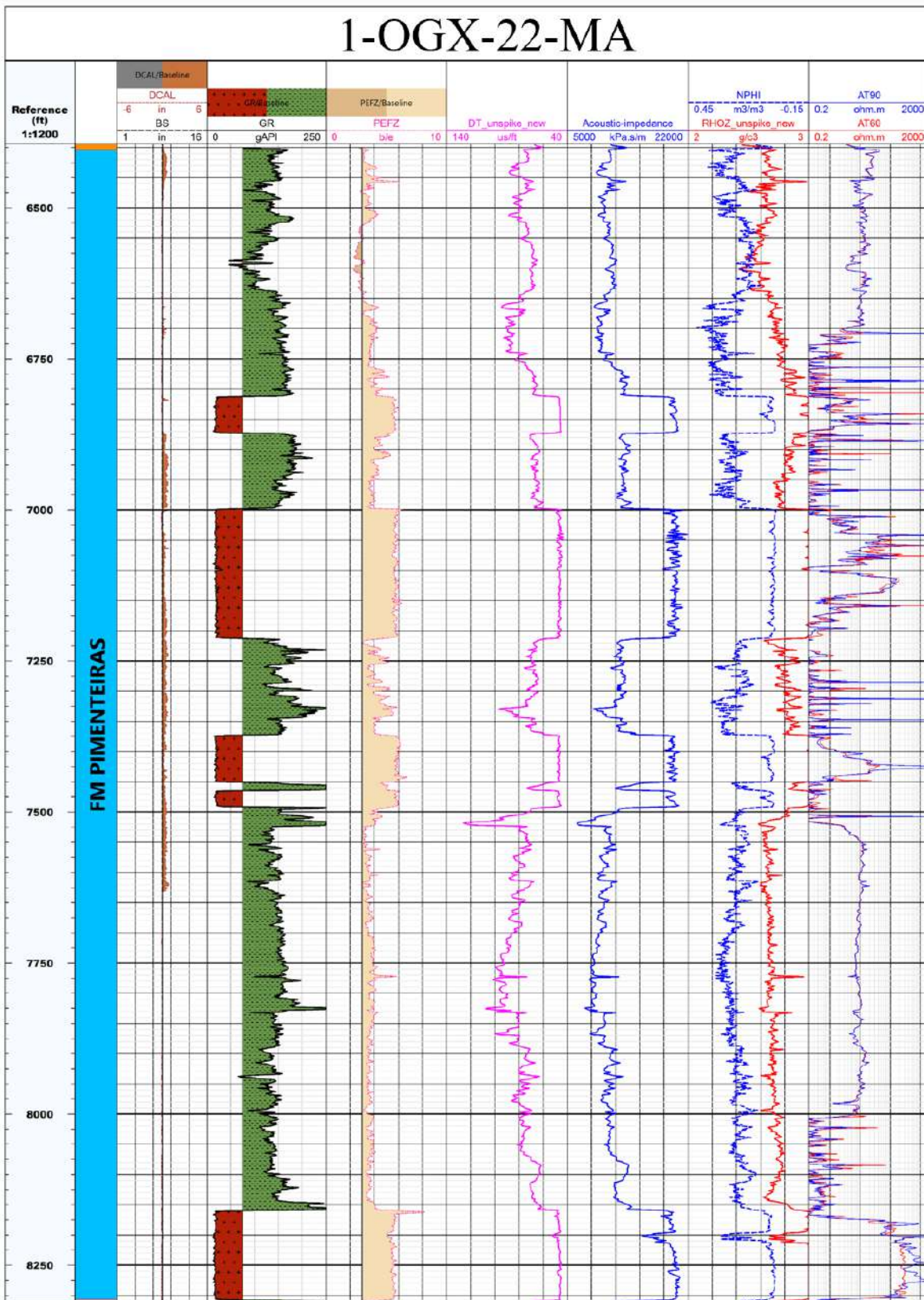


Figura 4.8. Conjunto de perfis medidos e condicionados no poço 1-OGX-22-MA, onde ocorreu intrusões não somente na porção superior da formação, mas também na base da mesma. A interpretação dos perfis mostra um comportamento semelhante as demais acusando uma possível diversidade de eletrofácies a serem investigadas.

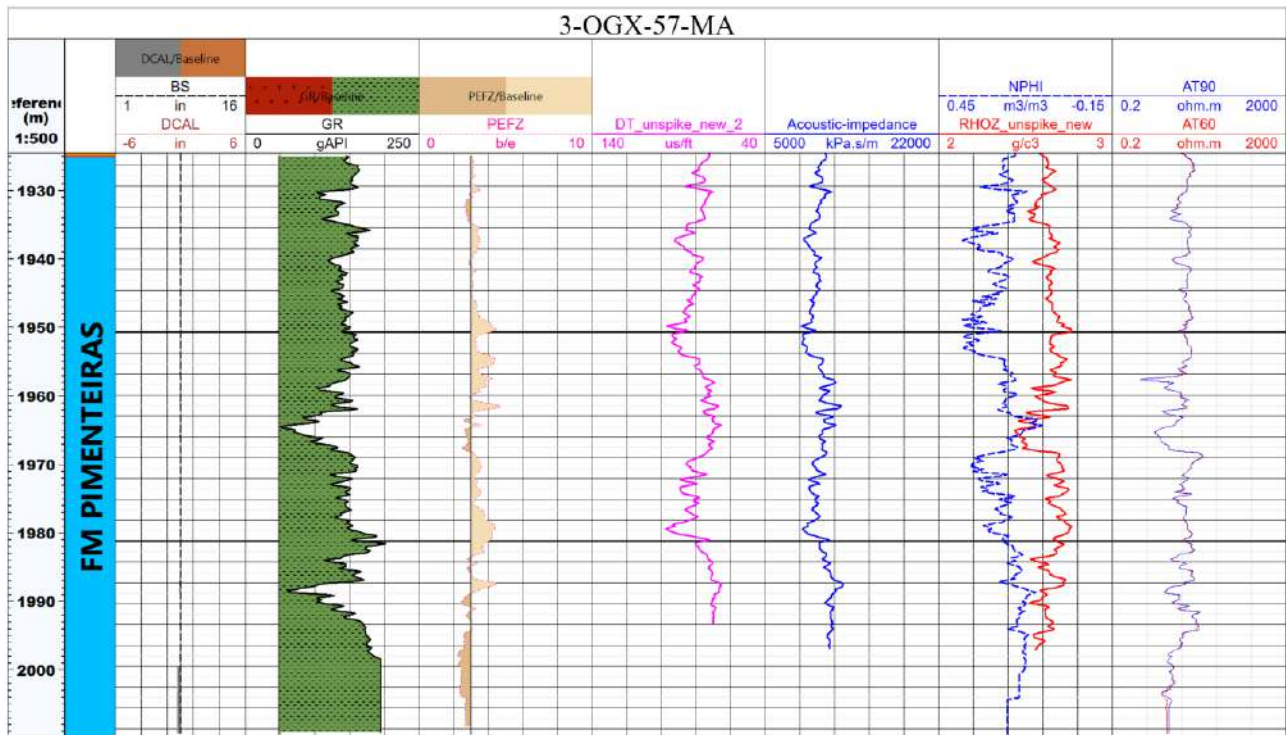


Figura 4.9. Exibição do mesmo conjunto de curvas para o poço 3-OGX-57-MA, os quais foram preconditionados e selecionados para testes de seleção de variáveis importantes para a clusterização *K-means*.

Uma vez concluído o tratamento inicial, os dados foram importados em ambiente *Python*, que serviu como base para todos os processos subsequentes. Considerando que a variável de resistividade apresenta comportamento tipicamente logarítmico, foi aplicada uma transformação logarítmica a este perfil, etapa anterior ao procedimento de padronização. A normalização foi realizada em todo o conjunto de variáveis por meio da função *StandardScaler*, disponível na biblioteca *Scikit-learn*, garantindo média nula e desvio padrão unitário para cada atributo.

Formulações matemáticas e explicações detalhadas acerca das etapas de pré-processamento de dados encontram-se descritas no Capítulo 5, dedicado aos fundamentos e metodologias de *machine learning*. As métricas estatísticas obtidas antes e após a aplicação das transformações estão apresentadas, respectivamente, na Tabela 4.1 e na Tabela 4.2, evidenciando as diferentes escalas das variáveis e mostrando os efeitos do pré-processamento sobre o conjunto de dados,

Tabela 4.1. Resumo das métricas estatísticas dos perfis candidatos a clusterização anterior a transformação dos dados.

	MD	AI	AT90	AT90log	DTCO	GR	NPHI	PEFZ	RHOZ
Unidade	m	g/cm ³ .m/s	Ohm.m	Ohm.m	µs/ft	API	v/v	b/e	g/cm ³
mean	2233.35	13386.15	46.40	1.57	68.33	112.63	0.17	4.21	2.78
std	154.18	4279.97	144.88	2.10	16.45	64.07	0.10	1.04	0.18
min	1925.00	6342.61	0.10	-2.31	43.38	10.97	0.01	2.02	2.16
0.25	2108.24	9991.32	0.81	-0.22	48.10	26.53	0.07	3.40	2.65
0.50	2235.61	12096.29	7.09	1.96	68.79	133.82	0.20	3.81	2.73
0.75	2362.86	18951.13	12.34	2.51	81.20	153.67	0.25	5.45	2.98
max	2535.56	21594.22	1556.56	7.35	125.59	368.87	0.40	8.18	3.14

Tabela 4.2. Estatísticas descritivas das variáveis de entrada após o pré-processamento (transformações logarítmicas e padronização), mostrando a homogeneização das escalas dos dados.

	MD	AI	AT90	AT90log	DTCO	GR	NPHI	PEFZ	RHOZ
Unidade	m	g/cm ³ .m/s	Ohm.m	Ohm.m	µs/ft	API	v/v	b/e	g/cm ³
mean	2233.35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
std	154.18	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
min	1925.00	-1.65	-0.32	-1.84	-1.52	-1.59	-3.41	-2.11	-1.70
0.25	2108.24	-0.79	-0.31	-0.85	-1.23	-1.34	-0.71	-0.78	-1.09
0.5	2235.61	-0.30	-0.27	0.19	0.03	0.33	-0.29	-0.39	0.24
0.75	2362.86	1.30	-0.24	0.45	0.78	0.64	1.10	1.20	0.78
max	2535.56	1.92	10.42	2.75	3.48	4.00	1.99	3.84	2.41

Após as etapas de pré-condicionamento, transformação e padronização das variáveis, realizou-se uma análise exploratória multivariada com o objetivo de avaliar o comportamento conjunto dos atributos selecionados como entrada para o algoritmo K-means. Para isso, foram gerados crossplots entre as variáveis GR, NPHI, DTCO e PEFZ, representados na forma de um pairplot, permitindo a inspeção simultânea das distribuições individuais e das relações bivariadas entre os dados.

A Figura 4.10. Pairplot das variáveis GR, NPHI, DTCO e PEFZ após o pré-processamento e padronização (z-score). As distribuições individuais são mostradas na diagonal principal, enquanto os crossplots representam as relações bivariadas entre os atributos. Os dados estão coloridos de acordo com os diferentes poços analisados. apresenta gráficos *crossplots* das variáveis GR, NPHI, DTCO e PEFZ após o processo de padronização (z-score), com os dados coloridos de acordo com os diferentes poços analisados. As distribuições apresentadas na diagonal indicam que, mesmo após a padronização, os atributos preservam características multimodais, sugerindo a presença de diferentes populações de dados possivelmente associadas a distintas eletrofácies.

Os *crossplots* evidenciam relações não lineares e sobreposição parcial entre os dados provenientes dos diferentes poços, indicando que, embora existam particularidades locais, os atributos

compartilham padrões globais comuns. Essa sobreposição reforça a adequação do uso de técnicas de aprendizado não supervisionado, uma vez que não há separação clara e prévia entre classes conhecidas.

Além disso, a análise conjunta das variáveis padronizadas demonstra que nenhuma delas domina o espaço de atributos em função de escala, condição fundamental para a aplicação do algoritmo K-means, que se baseia em medidas de distância. Dessa forma, o pairplot confirma a consistência do pré-processamento adotado e justifica a utilização dessas variáveis como entrada para a etapa de clusterização.

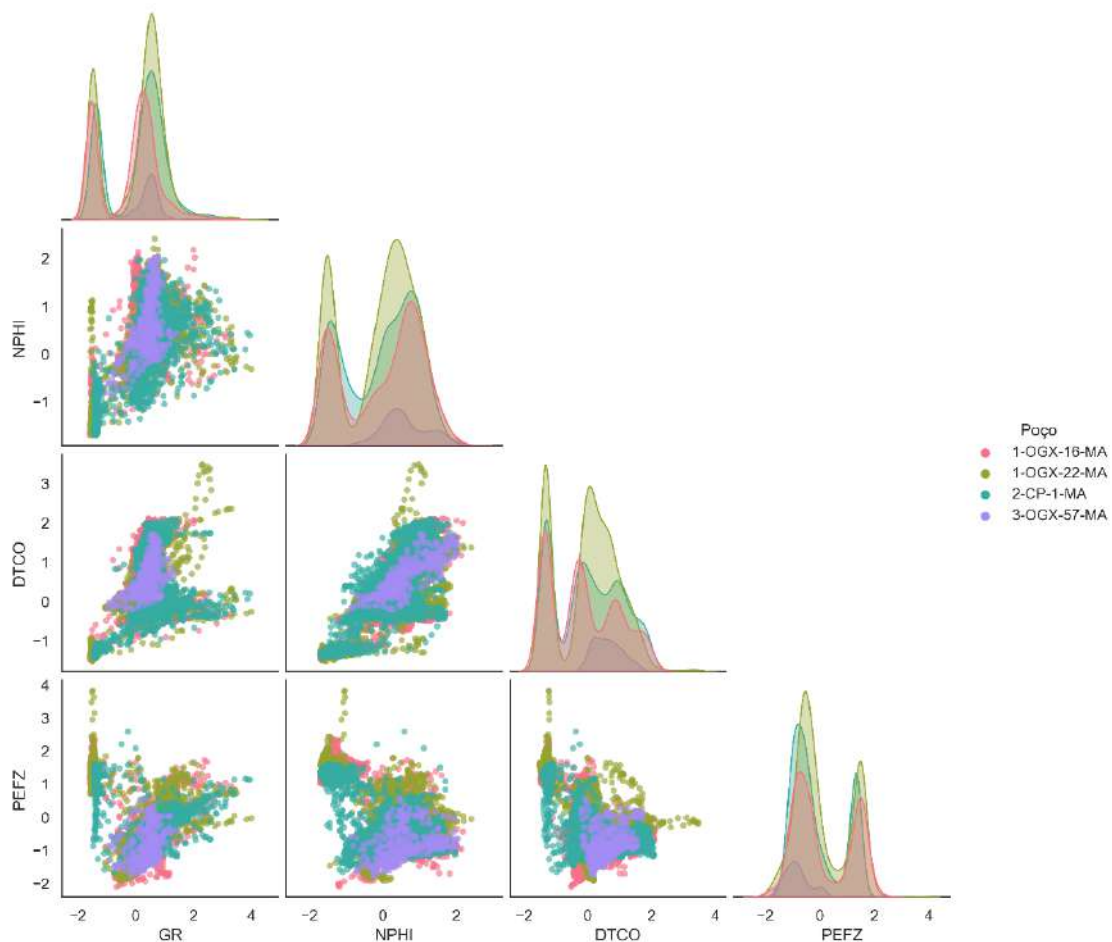


Figura 4.10. Pairplot das variáveis GR, NPHI, DTCO e PEFZ após o pré-processamento e padronização (z-score). As distribuições individuais são mostradas na diagonal principal, enquanto os crossplots representam as relações bivariadas entre os atributos. Os dados estão coloridos de acordo com os diferentes poços analisados.

4.3.1.3. Definição do número de *clusters* (K)

Mesmo com alguns critérios estabelecidos para testes das hipóteses como descrito anteriormente, foi utilizado um método de *Elbow* para determinar o número ideal de classes segundo o conjunto de dados disponível.

Esse procedimento baseia-se na análise da Soma dos Quadrados Dentro do *Cluster* (Within-*Cluster* Sum of Squares – WCSS) ou Soma dos Erros Quadrados (*Sum of Squared Errors* – SSE), que quantifica a variabilidade interna de cada agrupamento. Em termos gerais, quanto mais eficiente for a partição, menor será o valor global de WCSS. A expressão matemática que descreve esse método é dado por (Madhulatha, 2012):

$$WCSS(K) = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (4.8)$$

Onde, K é número de *clusters*, C_k é conjunto de pontos atribuídos ao *cluster* k , x_i é vetor de atributos do ponto i , e μ_k é o centróide do *cluster* k .

O método funciona pela construção de um gráfico que relaciona os valores de K com seus respectivos WCSS. Inicialmente, quando $K = 1$, o valor de WCSS é elevado, pois todos os pontos pertencem a um único *cluster*. À medida que K aumenta, o WCSS tende a diminuir, refletindo a redução da variância interna, uma vez que os pontos ficam mais próximos de seus centroides. Contudo, essa redução não ocorre de forma linear: em determinado ponto, observa-se uma queda acentuada no WCSS seguida de uma estabilização gradual. O ponto de inflexão da curva, que se assemelha ao formato de um “cotovelo”, indica o valor de K que representa o melhor compromisso entre complexidade do modelo e qualidade do agrupamento (Figura 4.11).

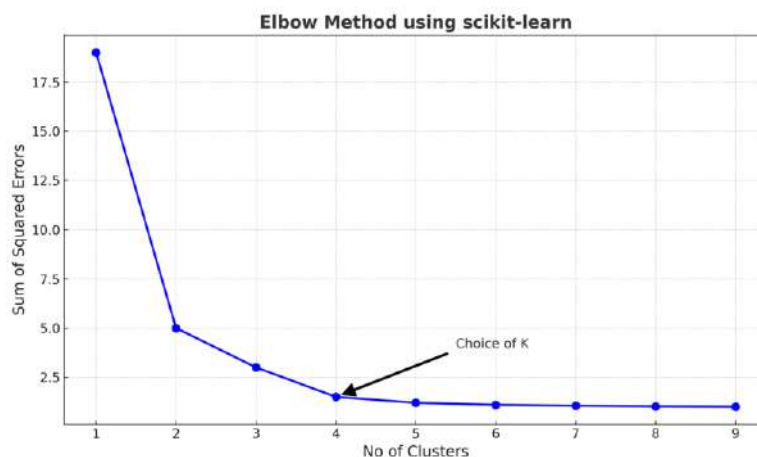


Figura 4.11. Exemplificação do método de *Elbow* e sua interpretação voltada para a escolha do número ideal de clusters. (Fonte: <https://dchandra.com/machine%20learning/2018/12/16/K-means-Clustering-Algorithm-using-scikit-learn.html>).

É importante destacar que, embora o WCSS tenda a zero quando cada ponto é considerado como um *cluster* individual ($K = n$), essa situação não possui utilidade prática, pois implica ausência de generalização. Dessa forma, o método de *Elbow* fornece uma heurística eficaz especialmente para

valores relativamente pequenos de K , permitindo identificar um número de grupos que capture a estrutura subjacente dos dados sem resultar em sobreajuste.

4.3.1.4. Aplicação do método *K-means*

A seleção das variáveis utilizadas no algoritmo *K-means* foi conduzida de forma iterativa, a partir da realização de diferentes testes. O critério inicial para a escolha do conjunto final de curvas foi a capacidade de separar adequadamente as fácies, de acordo com as hipóteses estabelecidas a partir do conhecimento prévio sobre o intervalo de interesse. Em segundo lugar, buscou-se privilegiar combinações de perfis que reproduzissem uma separação de classes consistente e visualmente interpretável em *crossplots*, especialmente na relação entre GR e AI (Impedância Acústica).

Entre os perfis considerados, o GR e o PEFZ mostraram-se candidatos adequados, pois apresentam forte relação com a litologia: o GR mede a radioatividade natural das rochas, enquanto o PEFZ reflete a influência do número atômico dos elementos constituintes dos minerais. O perfil DTCO, por sua vez, é afetado tanto pela matriz quanto pela porosidade, possibilitando diferenciar de forma qualitativa os tipos litológicos e, no caso da área de estudo, evidenciar alterações associadas ao contato com intrusões ígneas. De forma complementar, o perfil NPHI, interpretado como um indicador indireto de porosidade, também contribui para a caracterização litológica, uma vez que a perda de energia dos nêutrons emitidos pela ferramenta varia conforme os elementos presentes na rocha. Além disso, o NPHI apresenta maior profundidade de investigação, sendo menos suscetível às condições das paredes do poço.

Por outro lado, o perfil de RHOZ apresentou comportamento relativamente monótono, refletindo de forma limitada as variações litológicas, exceto nas transições mais marcantes entre folhelhos e soleiras. O perfil de resistividade, embora relevante, foi deliberadamente excluído desta etapa de análise para evitar vieses na classificação de fácies, uma vez que será empregado de forma específica e detalhada na investigação do Carbono Orgânico Total (COT), descrita no capítulo seguinte.

Em suma o algoritmo *K-means* utilizado nesse estudo, após a avaliação do número de *clusters* e dos estudos preliminares dos conjuntos de variáveis a serem utilizadas, pode ser descrito como na Figura 4.12 abaixo.

Algoritmo 1: *K-means*

Input: Dados de poços $X = \text{GR, DTCO, PEFZ, NPHI}$ e número de clusters K

Output: Atribuição de fácies (clusters) por profundidade

1. Pré-processamento:

- 1a. Remover valores nulos nas variáveis de entrada.
- 1b. Padronizar as variáveis **GR, DTCO, PEFZ, NPHI** utilizando *StandardScaler* (média = 0, desvio padrão = 1).

2. Inicialização dos centróides (*K-means++*):

- 2a. Escolher o primeiro centróide c_1 uniformemente em.
- 2b. Para cada ponto $x \in X$, calcular $D(x) = \text{distância}$ até o centróide mais próximo já escolhido.
- 2c. Selecionar o próximo centróide $c_1 = x' \in X$, com a probabilidade $\frac{D(x')^2}{\sum_{x \in X} D(x)^2}$.
- 2d. Repetir o passo 2c até obter os μ centróides iniciais.

3. Iteração do algoritmo *K-means*:

- 3a. Atribuir cada amostra ao cluster do centróide mais próximo (distância euclidiana).
- 3b. Atualizar cada centróide μ_i como a média das amostras atribuídas ao cluster i
- 3c. Repetir os passos 3a–3b até convergência (quando não há mudanças relevantes na atribuição).

4. Saída:

Atribuição final dos clusters às profundidades do poço, interpretados como **eletrofácies**.

Figura 4.12. Pseudocódigo detalhando as etapas realizadas na clusterização *K-means* executadas utilizando a linguagem *python*.

4.3.2. Estudo de Viabilidade via Classificação Bayesiana

A classificação Bayesiana de fácies foi aplicada neste trabalho, como um estudo de viabilidade para a inversão sísmica a partir das fácies definidas pelo método *K-means*. O objetivo foi compreender se as fácies definidas nos poços teriam propriedades elásticas necessárias para exercer contrastes de impedâncias suficientes de forma a diferenciá-las.

O fluxo de trabalho empregado está resumido na Figura 4.13, onde é aplicado a equação 4.4, substituindo-se os conjuntos C_i pelas fácies e a ocorrência do evento A pela impedância acústica. Produzindo, portanto, a probabilidade condicional das fácies ocorrerem dado que ocorre determinados valores e impedância acústica. De maneira simplificada, foi definido que as quatro fácies são igualmente prováveis, permitindo com que a probabilidade a posteriori dependa apenas das verossimilhanças, dado pelas funções de densidade de probabilidade calculadas utilizando-se as estimativas de densidade do *kernel* do tipo gaussiano.

Algoritmo 2: Classificação Bayesiana 1D

Input: classes estimadas pelo *K-means* f_i (fácies), probabilidade *a priori* $P(f_i)$ e a variável contínua AI (impedância acústica)

Output: Probabilidade a posteriori $P(f_i|AI)$

1. Estimação das distribuições 1D:

1a. Estimar a densidade 1D de AI condicionada à fácies f_i , para $i = 1, \dots, 4$. Via KDE (`kde = scipy.stats.gaussian_kde`)

1b. Definir as probabilidades *a priori*: $P(f_1) = P(f_2) = P(f_3) = P(f_4) = 0.25$

2. Cálculo das probabilidades a posteriori:

2a. Para cada fácies f_i : computa a verossimilhança $P(AI|f_i)$

2b. Calcula o numerador: $P(AI|f_i) \cdot P(f_i)$.

2c. Calcula o denominador: $\sum_{i=1}^4 P(AI|f_i) \cdot P(f_i)$.

2d. Retorna $P(f_i|AI) = \frac{P(AI|f_i) \cdot P(f_i)}{\sum_{i=1}^4 P(AI|f_i) \cdot P(f_i)}$, $i = 1, \dots, 4$

3. Validação:

3a. Atribuir: `Facies_Bayes[i] = argmax_f P(f_i|AI)`

3b. Calcular acurácia global: média de `Facies_Bayes == Facies_Kmeans`

3c. Matriz de confusão

Figura 4.13. Etapas realizadas na classificação Bayesiana 1D no formato pseudocódigo o qual foi utilizando a linguagem *python*.

A estimativa de densidade por *kernel* (*Kernel Density Estimation – KDE*) é uma técnica não paramétrica amplamente utilizada para inferir a função densidade de probabilidade de uma variável aleatória a partir de uma amostra finita (Silverman, 1986). Formalmente, a estimativa é expressa por:

$$\hat{f} = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (4.9)$$

Em que, n representa o número de observações da amostra, $K(\cdot)$ corresponde à função núcleo (*kernel*), que define a forma da distribuição local, h é o parâmetro de suavização conhecido como largura de banda (*bandwidth*), responsável pelo controle da amplitude da janela de suavização e X_i são as amostras observadas da distribuição de interesse.

No caso particular em que o *kernel* adotado é o gaussiano, a expressão assume a forma:

$$K(u) = \frac{1}{\sqrt{2\pi}} e\left(-\frac{u^2}{2}\right) \quad (4.10)$$

O que confere suavidade contínua à estimativa, sendo esta uma das escolhas mais recorrentes em aplicações práticas. A determinação da largura de banda é realizada por meio de um fator de

suavização (*smoothing factor*) aplicado à matriz de covariância dos dados. Por padrão, assim como aplicado neste trabalho, o *gaussian_kde* utiliza a regra de Scott, $h \propto n^{-1/(d+4)}$, (Scott, 2014).

4.4. RESULTADOS E DISCUSSÕES

4.4.1. Integração dos Métodos K-means e Classificação Bayesiana no Estudo de Viabilidade

Os resultados obtidos por meio da classificação não supervisionada utilizando o método *K-means* mostraram-se consistentes com as interpretações derivadas da suíte básica de perfis de poços, em função dos diferentes comportamentos analisados. A aplicação do método *Elbow* ao conjunto de dados indicou um número ótimo de classes compatível com as expectativas preliminares e com o contexto do alvo deste estudo. Conforme ilustrado na Figura 4.14, o gráfico da inércia resultante do método *Elbow* evidencia que o limite ideal de agrupamentos é alcançado no quarto *cluster*, ponto a partir do qual a curva apresenta tendência de estabilização. Como mencionado na seção de métodos a inércia é a soma dos quadrados das distâncias de cada ponto de dados ao centro do cluster mais próximo. Assim no método de Elbow calcula-se a distorção ou a inércia para diferentes valores de *k* e as representamos graficamente.

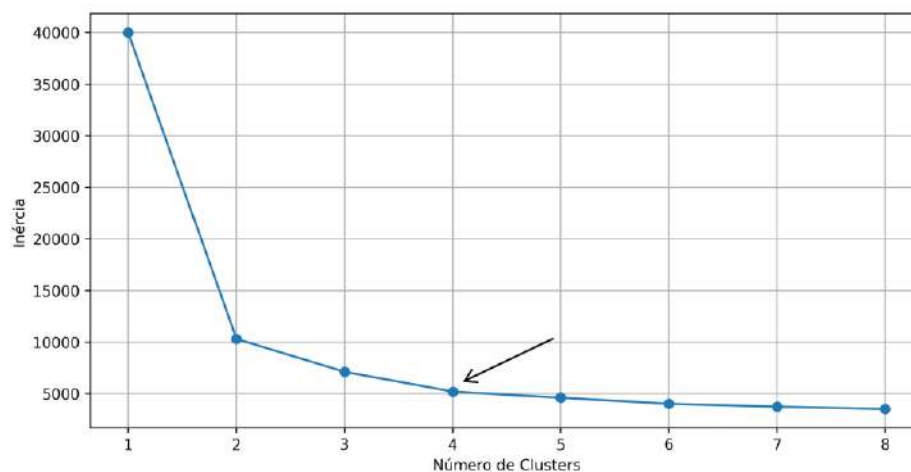


Figura 4.14. Aplicação do método de *Elbow* a partir do conjunto de dados selecionados, indicando o número ideal de *clusters*, o qual está sendo apontado pela seta preta, a partir do qual a curva começa a estabilizar.

Diante desse resultado, foram realizados testes no qual foi analisado a clusterização a partir do conjunto de curvas GR, DTCO, PEFZ e NPHI, considerando valores de *K* variando entre o mínimo de 1 e até uma classe além do número indicado pelo método *Elbow* ($K = 5$). Esses testes estão ilustrados na Figura 14, na qual são apresentados os gráficos de GR contra AI, colorido pelos diferentes *clusters*. A escolha desses parâmetros justifica-se pela necessidade de avaliar o desempenho do algoritmo *K-means*, dado que o GR é um indicador litológico robusto, enquanto a impedância acústica não apenas

reflete contrastes associados a mudanças litológicas, mas também evidencia alterações em litotipos com características muito próximas, aspecto particularmente relevante no intervalo estudado, composto predominantemente por folhelhos. Além disso, a impedância será o atributo tridimensional central para a investigação posterior, em que será avaliada sua relação com as fácies definidas neste estudo e, conseqüentemente, com o conteúdo de COT, tema que será aprofundado nos capítulos seguintes.

Nos testes apresentados na Figura 4.15, observa-se que, em uma divisão em dois agrupamentos, ocorre a separação clara entre os folhelhos e as soleiras de diabásio. Entretanto, os dados indicam fortemente que os folhelhos não apresentam um comportamento homogêneo. Esse aspecto torna-se ainda mais evidente na divisão em três *clusters*, na qual os folhelhos se distribuem predominantemente em duas faixas distintas de impedância: uma com valores abaixo de aproximadamente 10000 $\text{g/cm}^3 \cdot \text{m/s}$ e outra acima desse limite. Além disso, verifica-se uma expressiva variação nos valores de radiação gama natural, com folhelhos que alcançam valores superiores a 200 API e outros que apresentam valores consideravelmente mais baixos, como evidenciado no agrupamento de número quatro.

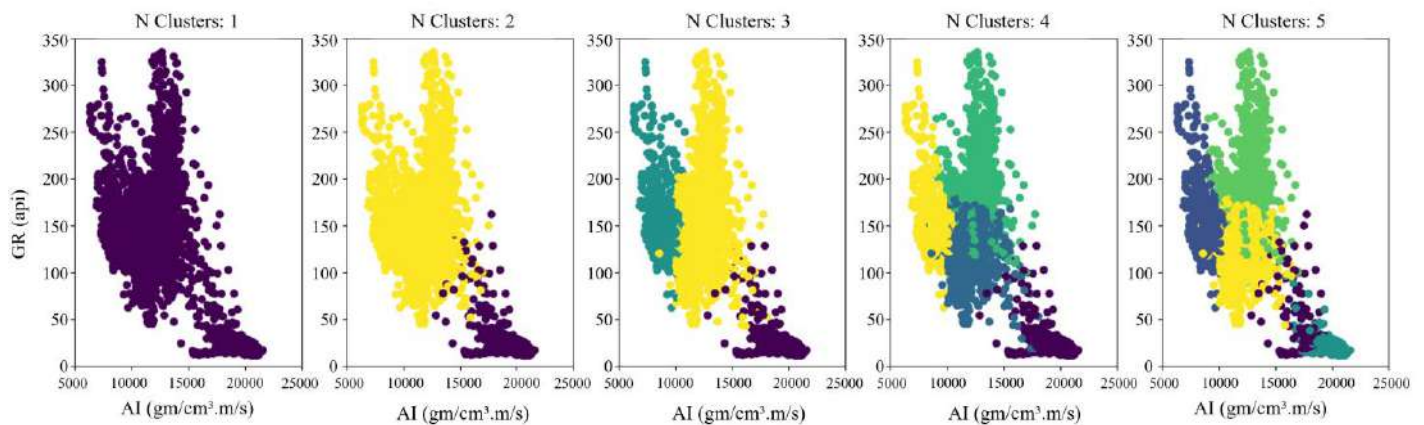


Figura 4.15. Relação entre *gamma ray* e impedância acústica para diferentes testes de número de *clusters*, ilustrando a separação progressiva dos agrupamentos representando as diferentes eletrofácies.

De fato, ao se observar o *cluster* de número cinco, percebe-se que as separações perdem parte do contexto geológico esperado para a formação analisada. Nota-se, nesse caso, uma tentativa de subdivisão da rocha ígnea, a qual já se destaca de maneira significativa por seus baixos valores de GR e elevados valores de AI. No intervalo entre 50 e 100 API de GR, alguns poucos pontos caracterizados por folhelhos podem ser confundidos com as soleiras de diabásio; contudo, a subdivisão acima de quatro *clusters* não resolve satisfatoriamente essa zona de sobreposição. Assim, definiu-se a utilização de quatro agrupamentos como representação mais adequada do conjunto de dados, os quais serão posteriormente traduzidos em eletrofácies.

A Figura 4.16 evidencia as diferentes características associadas aos quatro *clusters* definidos pelo método *K-means*, representadas pelas funções de densidade do GR e da AI. No caso da impedância acústica, observam-se picos de densidade de probabilidade bem distintos, indicando uma separação clara entre os *clusters* 0, 2 e 3, enquanto se verifica maior sobreposição entre os *clusters* 1 e 2 e, em menor intensidade, entre os *clusters* 1 e 3. Por outro lado, na análise das funções de densidade do GR, a sobreposição mais expressiva ocorre entre os *clusters* 1 e 3, refletindo a variabilidade interna desse parâmetro.

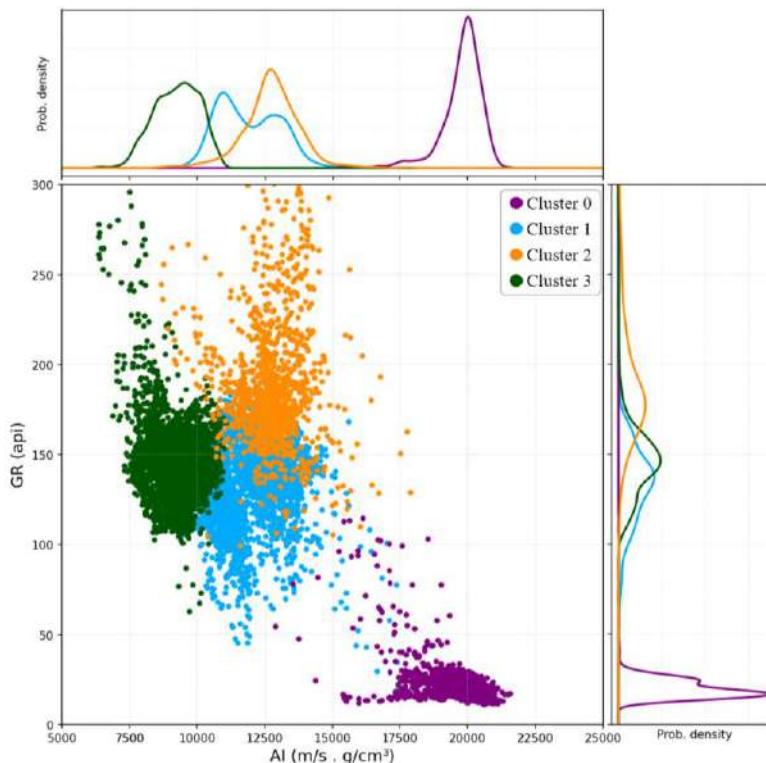


Figura 4.16. Funções de densidade de probabilidade dos *clusters* definidos pelo *K-means*, mostrando a distribuição dos valores de GR e AI para cada agrupamento.

Após os resultados obtidos na classificação de fácies, foi aplicada a classificação Bayesiana como etapa preliminar de avaliação da viabilidade para a inversão sísmica, constituindo, portanto, um passo importante para a geração dos volumes de impedância. As verossimilhanças condicionais dos *clusters* foram determinadas por meio da estimativa de densidade via *kernel*, conforme ilustrado na Figura 4.17. Observa-se, como já destacado anteriormente, uma maior sobreposição entre os *clusters* 1 e 2 no intervalo de 12000 a 14000 g/cm³·m/s. Embora os picos das funções *kernel* se apresentem bem definidos para os quatro *clusters*, espera-se maior incerteza na separação entre os *clusters* 1 e 2, uma incerteza moderada entre os *clusters* 1 e 3, e uma distinção satisfatória entre os *clusters* 0, 2 e 3.

Conforme definido na etapa metodológica apresentada anteriormente, todos os *clusters* foram considerados, *a priori*, com a mesma probabilidade de ocorrência, de modo que as probabilidades *a posteriori* são inteiramente determinadas pelas funções densidade de probabilidade (PDFs) associadas às verossimilhanças. A Figura 16 apresenta as PDFs das probabilidades *a posteriori* de cada *cluster*, demonstrando que as fácies definidas pela classificação Bayesiana ocorrem predominantemente no *cluster 3* para valores de impedância acústica inferiores a 10000 g/cm³·m/s. O *cluster 1*, por sua vez, é mais representativo no intervalo entre 10000 e 12000 g/cm³·m/s, enquanto o *cluster 2* se concentra na faixa de 12000 a 16000 g/cm³·m/s. Finalmente, valores de impedância superiores a 16000 g/cm³·m/s são atribuídos ao *cluster 0*.

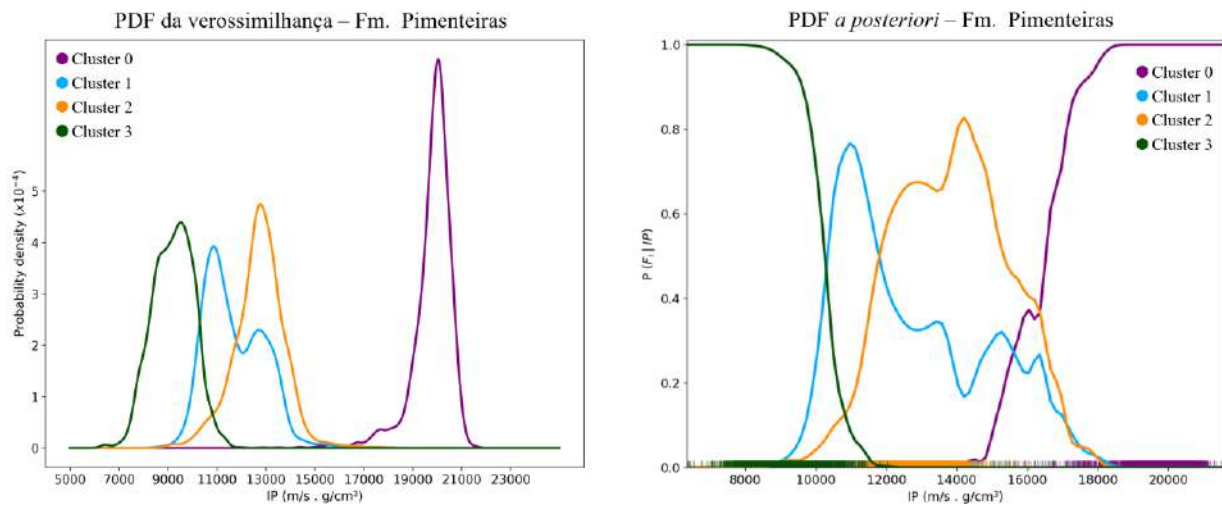


Figura 4.17. Funções de densidade e probabilidades *a posteriori* obtidas pela Classificação Bayesiana, evidenciando as diferenças e sobreposições entre os *clusters*.

Um controle de qualidade fundamental aplicado à classificação é a avaliação da matriz e da barra de confusão, que permitem não apenas quantificar o desempenho global do modelo, mas também identificar os principais pontos de ambiguidade. Por meio dessa análise, torna-se possível comparar as classes previstas com as classes reais, evidenciando tanto os acertos quanto os erros de classificação. Esse procedimento fornece uma medida direta da capacidade preditiva do classificador e revela padrões de erro recorrentes, como a confusão entre classes distintas. Tais inconsistências estão diretamente relacionadas à sobreposição das funções densidade de probabilidade condicionais, já discutida anteriormente e ilustrada na Figura 4.18.

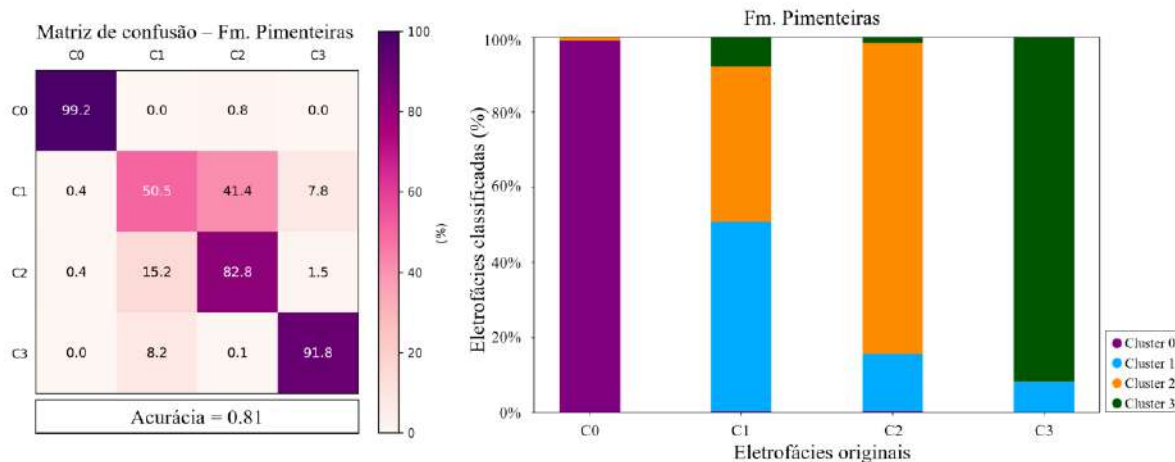


Figura 4.18. Matriz e barra de confusão da Classificação Bayesiana, apresentando os índices de acerto e os principais casos de confusão entre as eletrofácies representadas ainda pelos *clusters* obtidos do algoritmo *K-means*.

As maiores taxas de acerto, acima de 90%, como esperado, ocorrem para os *clusters* 0 e 3, em que a pequena confusão observada para esses dois *clusters* é geologicamente coerente, como será melhor detalhado a seguir na transferência de *clusters* para eletrofácies. O *cluster* 2 também apresentou desempenho muito satisfatório, com taxa de acerto de 83%, enquanto o *cluster* 1 obteve a menor previsibilidade, com 51% de acerto e uma confusão significativa de 41% em relação ao *cluster* 2. De modo geral, a classificação Bayesiana atingiu uma acurácia global de 81%, resultado considerado bastante satisfatório, indicando que a impedância acústica apresenta variações consistentes entre os agrupamentos de eletrofácies, sem grandes incertezas ou grandes dificuldades de separação entre os mesmos.

4.4.2. Interpretação dos *Clusters* em Eletrofácies

As rochas ígneas que intrudiram a Formação Pimenteiras apresentam baixos valores de GR e elevados valores de AI, o que permite distingui-las com relativa facilidade das rochas encaixantes. Essas soleiras de diabásio também exibem comportamento singular nas curvas de PEFZ e NPHI, sendo classificadas pelo algoritmo *K-means* como *cluster* 0, e aqui será reclassificado como Fácies 1 – Soleira de diabásio.

O *cluster* 2, por sua vez, foi associado à Fácies 2 – Folhelho metamorfozido, devido à sua proximidade com as soleiras e às respostas marcantes nos perfis: valores de GR acima de 180 API, PEFZ superior a 4 b/e e DTCCO relativamente baixos, variando entre 55 e 75 $\mu\text{s}/\text{ft}$. Embora essas fácies já tenham sido previamente reconhecidas nos poços, a aplicação do *K-means* foi fundamental para

refinar a diferenciação entre os tipos de eletrofácies presentes nas regiões afetadas pelas rochas intrusivas.

O *cluster* 1 foi interpretado como Fácies 3 – Folhelho razoável em COT, em concordância com os dados geoquímicos disponíveis no poço 2-CP-1-MA (tema que será aprofundado no Capítulo 5). Essa fácies ocorre em contato ou próximas às rochas intrusivas, apresentando valores de DTCO semelhantes aos da Fácies 2, porém diferenciando-se desta pelos valores mais baixos de GR e PEFZ, situados logo abaixo de 180 API e 4 b/e, respectivamente.

Por fim, o *cluster* 3 foi reclassificado como Fácies 4 – Folhelho rico em COT, caracterizado por valores elevados de GR, embora relativamente menores que os da Fácies 2, e por valores de DTCO significativamente mais altos, superiores a 80 $\mu\text{s}/\text{ft}$. A reclassificação dos *clusters* obtidos pelo *K-means* em eletrofácies está sintetizada na Tabela 4.3.

Tabela 4.3. Relação entre os *clusters* obtidos pelo algoritmo *K-means* e a reclassificação em eletrofácies, bem como a sua descrição interpretativa.

<i>K-means</i>	Códigos das Fácies	Descrição das Fácies
Cluster 0	Fácies 1	Soleira de diabásio
Cluster 2	Fácies 2	Folhelho metamorfizado
Cluster 1	Fácies 3	Folhelho razoável em COT
Cluster 3	Fácies 4	Folhelho rico em COT

As definições de fácies geradas pelo método *K-means* e pela classificação Bayesiana são apresentadas em perfil para diferentes poços: o poço 2-CP-1-MA (Figura 4.19), o poço 1-OGX-16-MA (Figura 4.20), o poço 1-OGX-22-MA (Figura 4.21) e, por fim, para o poço mais raso, 3-OGX-57-MA (Figura 4.22). Para uma compreensão mais abrangente, as figuras incluem os perfis de poço utilizados na clusterização não supervisionada (GR, DTCO, PEFZ e NPHI), bem como os dados de AI empregados na classificação Bayesiana.

A comparação entre as fácies obtidas pelo *K-means* e aquelas mais prováveis segundo a classificação Bayesiana é particularmente relevante para identificar os intervalos nos quais as confusões foram mais persistentes. Conforme já indicado pela matriz e pela barra de confusão, a maior sobreposição ocorreu entre as Fácies 2 e 3. Entretanto, observa-se que tais ambiguidades concentram-se principalmente nas porções em contato com as soleiras de diabásio. Nesses intervalos, a classificação Bayesiana tende a atribuir maior probabilidade de ocorrência à Fácies 2 (folhelho metamorfizado),

enquanto em contato direto com esta se estabelece a Fácies 3 (folhelho razoável em COT), interpretada como uma fácies transicional, menos afetada pelo metamorfismo de contato.

Os folhelhos da Formação Pimenteiras apresentam, em geral, baixos valores de impedância acústica, os quais tendem a aumentar à medida que se aproximam das intrusões ígneas. Esse comportamento foi evidenciado pelas funções de densidade de probabilidade condicional e pelas probabilidades *a posteriori*, nas quais se observam dois extremos representados pelas Fácies 1 e 4. Entre essas, destacam-se as fácies intermediárias: a Fácies 2, cuja distribuição se aproxima mais do padrão da Fácies 1, ainda que sem fortes sobreposição, e a Fácies 3, caracterizada por um comportamento intermediário entre as Fácies 2 e 4, apresentando significativa sobreposição com a Fácies 2.

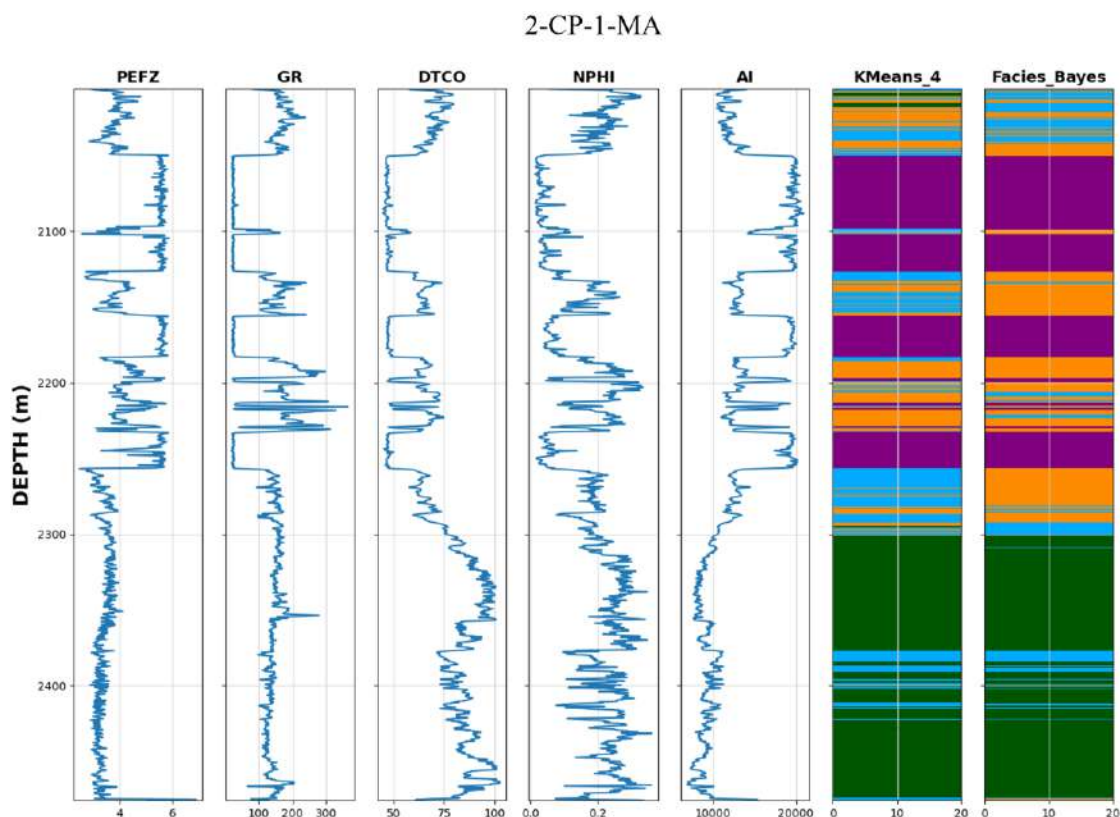


Figura 4.19. Resultado das fácies classificadas pelo *K-means* e pela classificação Bayesiana, junto aos perfis utilizados no processo.

A aplicação do método K-means na Formação Pimenteiras mostrou-se adequada para discriminar diferentes eletrofácies a partir dos perfis de poço, ainda que, em alguns intervalos, a sobreposição entre classes tenha gerado incertezas interpretativas. Esse comportamento é consistente com resultados encontrados na literatura, onde por exemplo, Glover *et al.* (2022), ao aplicar diferentes algoritmos de *clustering* em carbonatos compactos, observaram que o K-means, especialmente utilizando a distância Euclidiana, alcançou elevadas taxas de acerto na classificação de fácies,

superando 89% de eficiência em problemas com três classes, quando comparado com outros métodos de clusterização. Assim como observado neste trabalho, Glover *et al.* (2022) apontaram que a performance do K-means diminui em cenários com maior número de classes ou quando há significativa sobreposição entre atributos de diferentes fácies, situação que se assemelha à encontrada entre as Fácies 2 e 3 na Formação Pimenteiras, e principalmente para *clusters* maiores que quatro como visto nos testes realizados. Ainda assim, a simplicidade, rapidez computacional e robustez do K-means o tornam uma abordagem prática e eficiente em estudos de geociências, especialmente quando o número de *clusters* pode ser estimado a partir de dados prévios ou análises auxiliares.

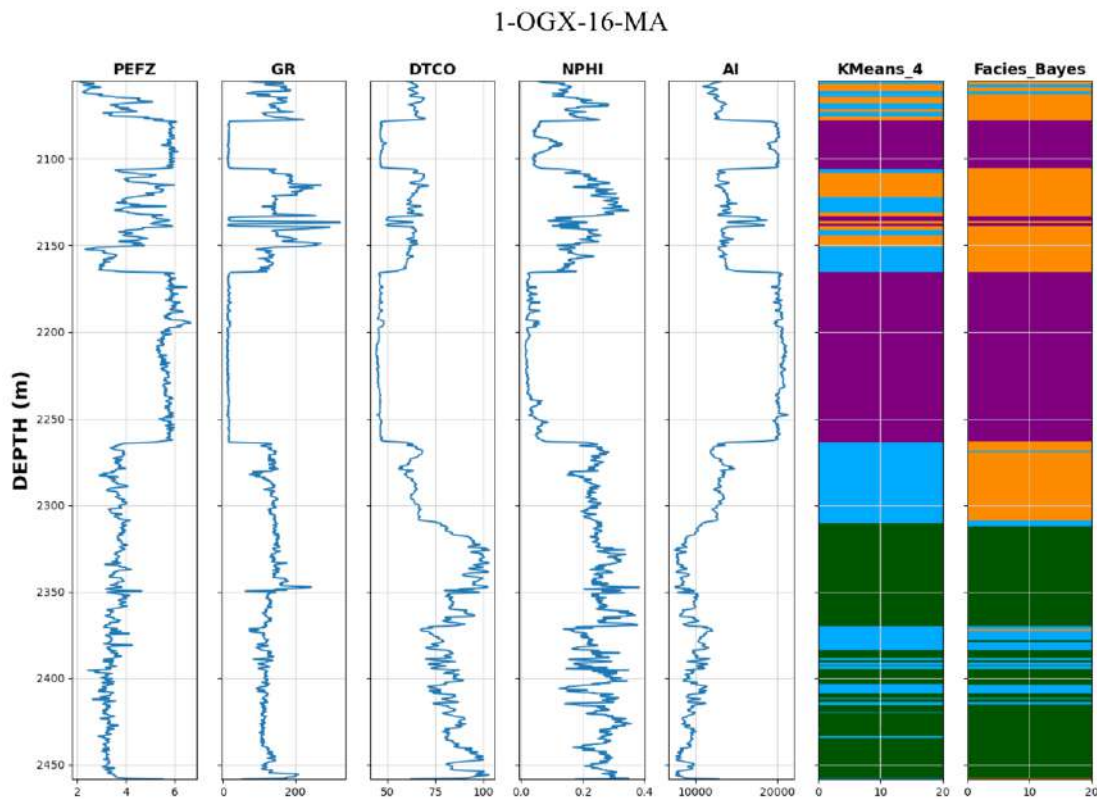


Figura 4.20. Resultado das fácies classificadas pelo *K-means* e pela classificação Bayesiana, junto aos perfis utilizados no processo.

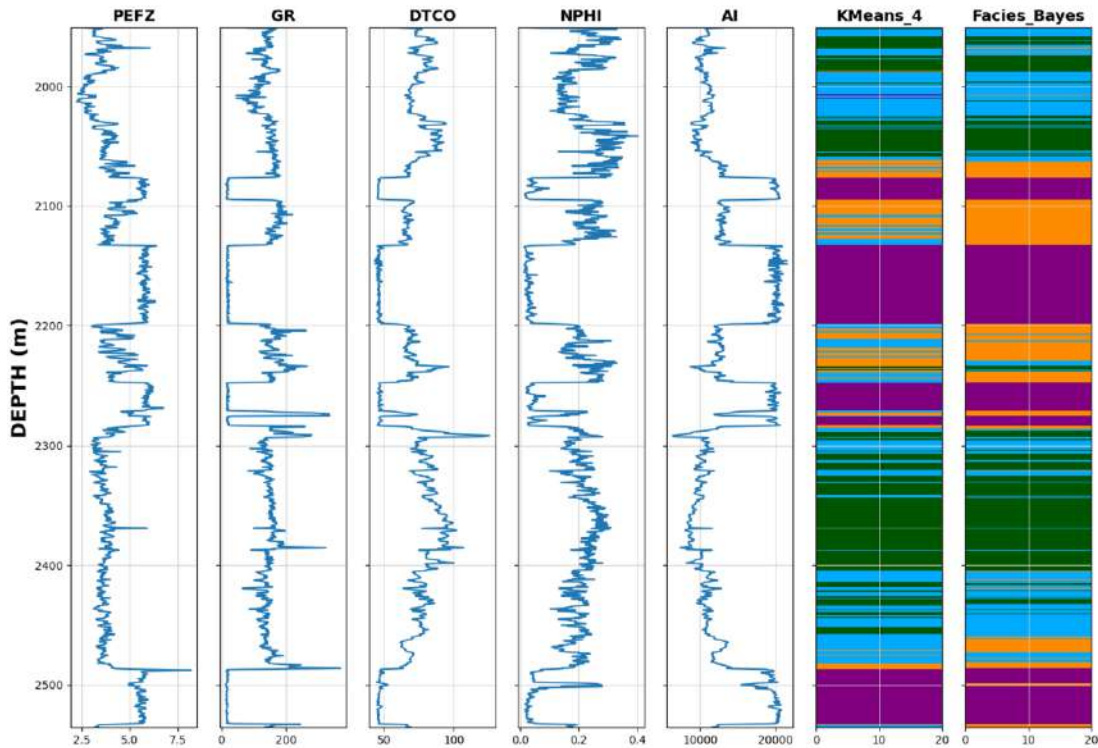


Figura 4.21. Resultado das fácies classificadas pelo *K-means* e pela classificação Bayesiana, junto aos perfis utilizados no processo.

A aplicação da classificação Bayesiana neste estudo, voltada para avaliar a viabilidade da inversão sísmica na discriminação de eletrofácies, está em consonância com resultados obtidos em outros contextos geológicos. Fernandes, Freire & Lupinacci (2022), ao investigarem reservatórios turbidíticos eocênicos da Formação Carapebus, Bacia de Campos, demonstraram que a classificação Bayesiana é uma ferramenta eficiente para integrar variáveis elásticas derivadas de inversão sísmica na caracterização litológica. Utilizando estimativas não paramétricas das PDFs de verossimilhança via *Kernel Density Estimation* (KDE), os autores observaram que a combinação dos parâmetros velocidade compressional (VP) e a razão VP/VS apresentou acurácias significativamente superiores (77,1% e 91,3% em dois poços analisados), quando comparadas ao uso isolado da impedância acústica.

No caso da Formação Pimenteiras, verificou-se que a impedância acústica foi suficiente para distinguir grupos de fácies, embora com maior sobreposição entre as Fácies 2 e 3. No presente estudo, a aplicação da classificação Bayesiana utilizando exclusivamente a impedância acústica resultou em uma acurácia global de 81%, valor considerado bastante satisfatório, sendo um desempenho que situa-se no mesmo intervalo dos resultados reportados por Fernandes, Freire & Lupinacci (2022). A comparação evidencia que, mesmo tratando apenas da impedância, os resultados obtidos para a

Formação Pimenteiras demonstram robustez equivalente à encontrada em contextos sedimentares distintos, reforçando a viabilidade da metodologia como ferramenta de caracterização preliminar.

Entretanto, é importante salientar que, assim como observado por Fernandes, Freire & Lupinacci (2022), a inclusão de variáveis elásticas adicionais, como as velocidades compressional e cisalhante ou suas razões, pode potencialmente aumentar a acurácia do modelo e reduzir ambiguidades. No caso específico da Formação Pimenteiras, essa abordagem poderia contribuir para minimizar a sobreposição observada entre as Fácies 2 e 3, proporcionando maior refinamento na distinção de fácies e ampliando a confiabilidade das previsões associadas ao conteúdo orgânico total. Dessa forma, recomenda-se que trabalhos futuros explorem a integração de múltiplos atributos elásticos, de modo a complementar os resultados aqui apresentados e avançar na caracterização das rochas geradoras da bacia.

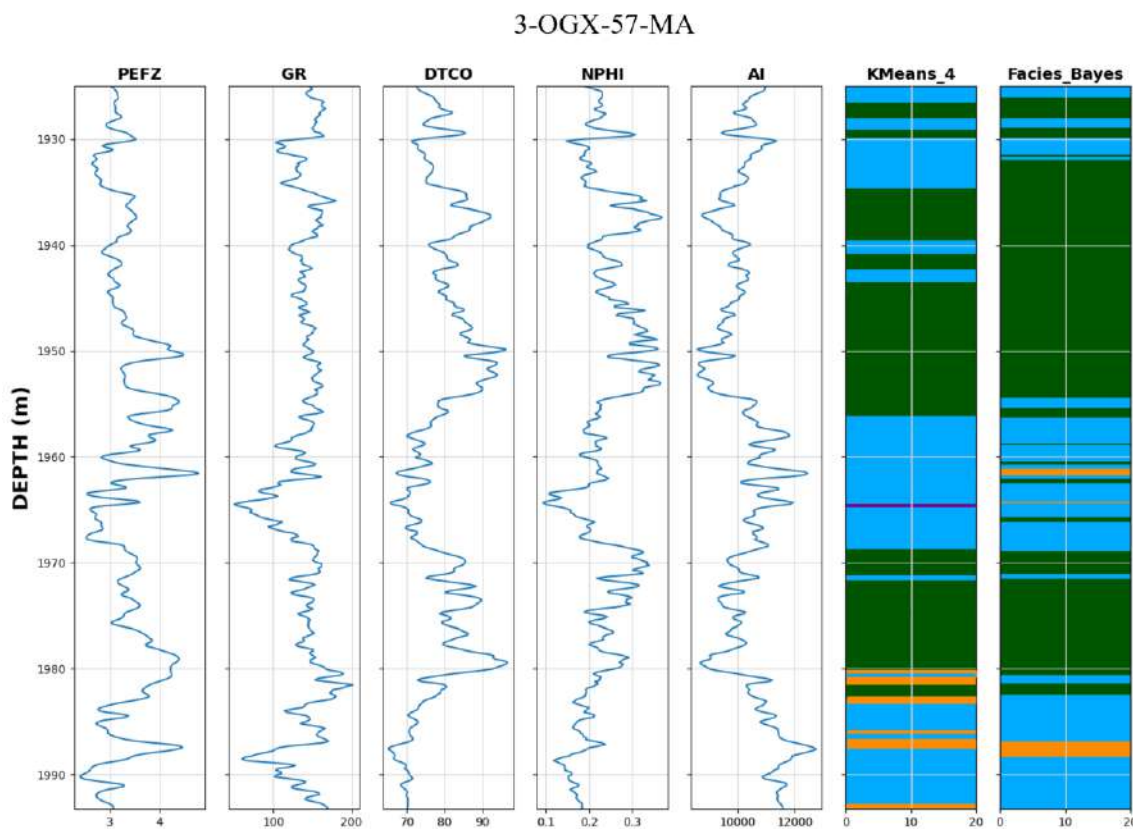


Figura 4.22. Resultado das fácies classificadas pelo *K-means* e pela classificação Bayesiana, junto aos perfis utilizados no processo.

Por fim, é importante destacar que a similaridade entre as Fácies 2 e 3, observada nos parâmetros elásticos, como velocidade compressional, densidade e, conseqüentemente, impedância acústica, pode ser associada as alterações térmicas induzidas pelo contato com as rochas ígneas intrusivas. Contudo, outras propriedades constituem melhores indicadores litológicos e permitem

diferenciá-las de forma mais precisa em diferentes tipos de folhelho, especialmente as concentrações dos elementos tório (Th), potássio (K) e urânio (U), conforme ilustrado nas Figura 4.23 e Figura 4.24. A partir dessas análises, nota-se que a Fácies 2 apresenta concentrações de U predominantemente mais elevadas (acima de 8 ppm) nos dois poços avaliados.

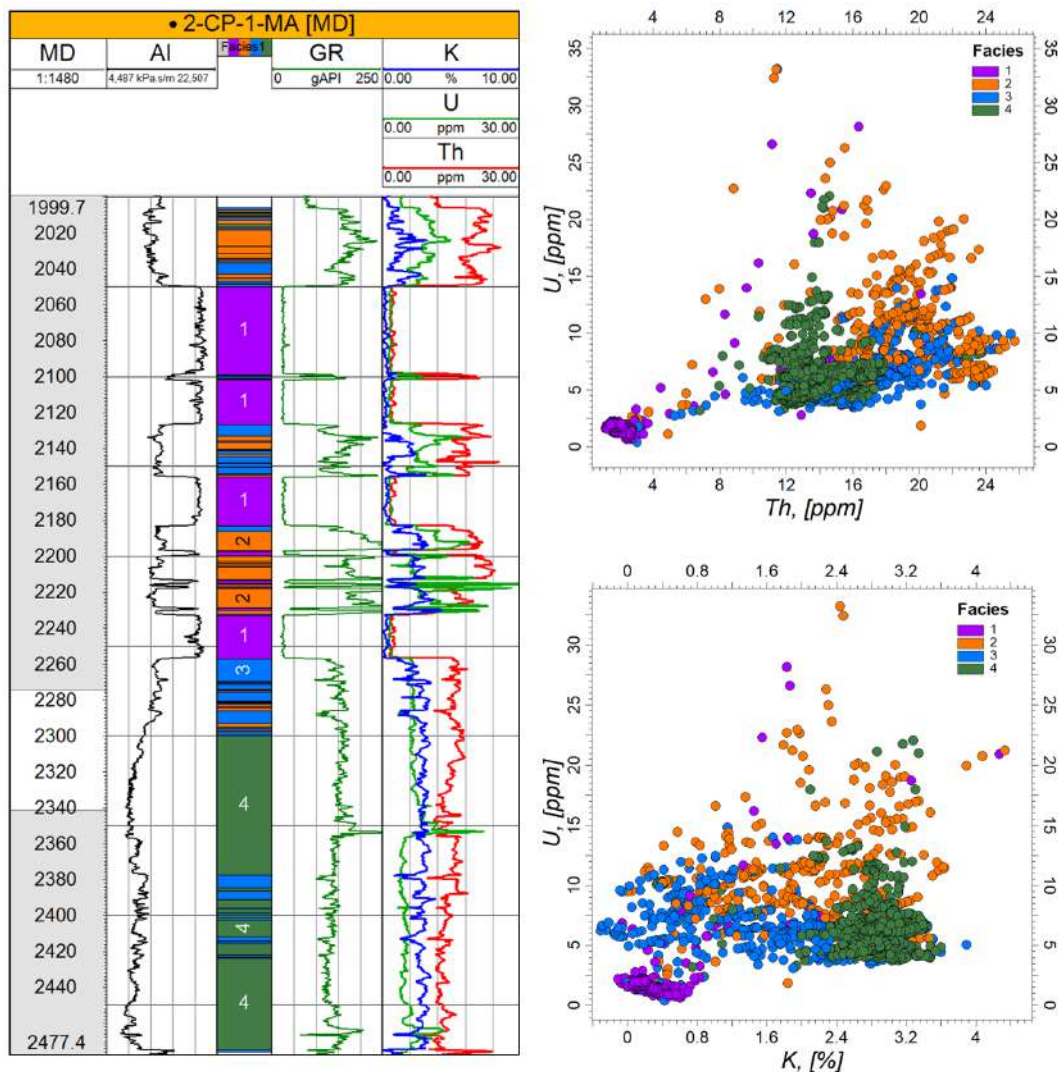


Figura 4.23. Distribuições de Tório (Th), Potássio (K) e Urânio (U) para as diferentes fácies definidas no poço 2-CP-1-MA, reforçando a interpretação dada a clusterização obtida. É possível avaliar as variações das contribuições espectrais de *gamma ray* em profundidade que destaca principalmente os altos valores de U em alguns intervalos na porção superior da formação. Os *crossplots* da relação entre U e Th, e U e K, evidenciam as diferentes concentrações desses elementos por fácies.

No poço 2-CP-1-MA, essa fácies também mostra concentrações de Th mais altas, com predomínio acima de 18 ppm, ainda que parcialmente sobrepostas aos valores de tório medidos na Fácies 3, a qual, por sua vez, exibe ampla variação nos perfis de Th e K. Já a Fácies 4, também no poço 2-CP-1-MA (Figura 4.23), ocorre de forma mais concentrada, apresentando valores relativamente mais elevados de K e menores de Th quando comparada às demais fácies de folhelho.

Um comportamento semelhante é observado no poço 1-OGX-16-MA, no qual a Fácies 2 apresenta maior enriquecimento em urânio e tório, mostrando menor sobreposição com a Fácies 3. Nesse poço, a principal sobreposição ocorre entre as Fácies 3 e 4 em relação às concentrações de tório, embora a Fácies 4 apresente uma faixa de variação mais restrita.

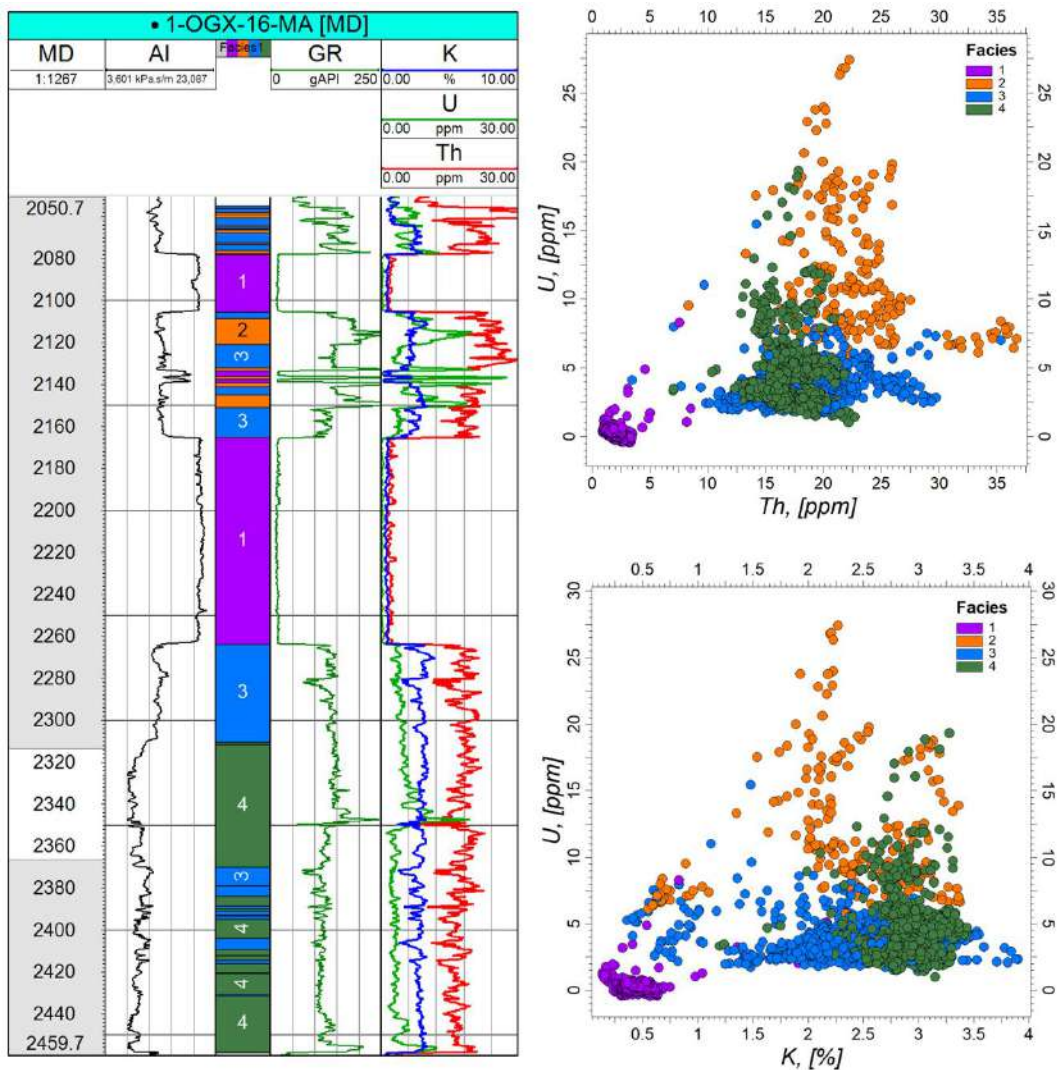


Figura 4.24. Distribuições dos elementos Th, K e U para as fácies identificadas no poço 1-OGX-16-MA, reforçando a consistência da interpretação da clusterização. A análise espectral do GR evidencia variações significativas em profundidade, com destaque novamente, para concentrações elevadas de U em intervalos específicos da porção superior da formação, associadas a Fácies 2 como pode ser melhor demonstrado pelos crossplots entre U e Th, e entre U e K, contribuindo para a diferenciação entre folhelhos.

Essa fácies se destaca, sobretudo, pelos valores mais elevados e concentrados de potássio, situados no intervalo entre 2.3% e 3.4%, enquanto as demais fácies exibem distribuições mais esparsas.

A avaliação das curvas espectrais de raios gama constitui uma análise complementar que contribui para o melhor entendimento das fácies definidas por meio da clusterização não supervisionada pelo algoritmo *K-means*. Embora essas curvas não tenham sido utilizadas como

variáveis de entrada, devido à ausência desse tipo de dado nos demais poços da área, os resultados demonstram que as fácies identificadas a partir do GR total também podem ser interpretadas e refinadas a partir das contribuições individuais de cada elemento radioativo natural emitido pelas rochas.

A definição das fácies obtidas neste trabalho para a Formação Pimenteiras corrobora com interpretações recentes da literatura que descrevem essa unidade como heterogênea, tanto do ponto de vista litológico quanto geoquímico. Mussa *et al.* (2024), ao revisar a geoquímica e a petrografia dos folhelhos devonianos da bacia, destacaram que a Formação é composta predominantemente por folhelhos ricos em matéria orgânica, com contribuição mista de querogênio tipo III (terrestre) e tipo II (marinho), refletindo variações paleoambientais entre condições oxigênicas restritas e anóxicas.

Além disso, Mussa *et al.* (2024) ressaltaram que a maturidade térmica da matéria orgânica na Pimenteiras apresenta forte variação espacial sendo mais elevada no centro, especialmente nas áreas afetadas por intrusões, onde o metamorfismo de contato promoveu maturação até estágios muito avançados. Esse comportamento está diretamente relacionado às diferenças observadas na impedância acústica das fácies aqui definidas, confirmando que o calor das rochas intrusivas é um fator-chave na diferenciação das propriedades elásticas entre folhelhos.

Portanto, as fácies definidas neste trabalho, desde o folhelho rico em COT (Fácies 4) até o folhelho metamorfozido (Fácies 2), refletem não apenas contrastes litológicos, mas também graus distintos de preservação e maturação térmica da matéria orgânica, em concordância com a síntese apresentada por Mussa *et al.* (2024). Essa convergência entre a caracterização petrofísica, geoquímica e sísmica reforça o papel da Formação Pimenteiras como um sistema petrolífero atípico, onde a interação entre sedimentação original e intrusões ígneas define a variabilidade das fácies e, conseqüentemente, o potencial gerador.

4.5. CONCLUSÃO

Os resultados apresentados neste capítulo demonstram que a integração de técnicas de *machine learning*, em especial a clusterização não supervisionada pelo algoritmo *K-means* e a posterior aplicação da classificação Bayesiana, constitui uma abordagem eficiente para a caracterização de fácies na Formação Pimenteiras. A utilização do método do *Elbow* permitiu determinar de forma objetiva o número ideal de agrupamentos, sendo quatro *clusters* definidos como os mais representativos do conjunto de dados, em consonância com o contexto geológico e com as hipóteses iniciais estabelecidas.

A análise dos agrupamentos revelou a presença de fácies distintas, englobando desde soleiras de diabásio até diferentes tipos de folhelhos, incluindo aqueles afetados pelo metamorfismo de contato.

A reclassificação dos *clusters* em eletrofácies mostrou-se coerente com as respostas dos perfis de poço e com os dados geoquímicos disponíveis, destacando-se a identificação de fácies ricas e razoáveis em COT, bem como de folhelhos metamorfizados. Essa diferenciação é relevante, pois evidencia que os folhelhos da Formação Pimenteiras não constituem um pacote homogêneo, mas sim um conjunto de rochas com variações significativas tanto em termos litológicos quanto de propriedades petrofísicas e geoquímicas.

A aplicação da classificação Bayesiana possibilitou avaliar a viabilidade da utilização da inversão sísmica como condicionantes para a distinção de fácies. Os resultados mostraram que a impedância acústica apresenta variações consistentes entre os agrupamentos, atingindo uma acurácia global de 81%. Embora tenha sido observada maior sobreposição entre as Fácies 2 e 3, especialmente em zonas de contato com as soleiras ígneas, a distinção entre os diferentes grupos mostrou-se satisfatória do ponto de vista geológico. Esses resultados reforçam que as respostas elásticas da Formação Pimenteiras são suficientemente contrastantes para que as fácies possam ser utilizadas como guias na interpretação de volumes de impedância.

Adicionalmente, a análise complementar das curvas de raios gama espectral contribuiu para refinar a diferenciação entre fácies de folhelhos, demonstrando que variações nos elementos K, Th e U podem atuar como indicadores auxiliares para uma melhor interpretação das eletrofácies clusterizadas pelo algoritmo não supervisionado. Essa abordagem evidencia o potencial da integração entre dados geofísicos e geoquímicos para aprimorar a caracterização de rochas geradoras submetidas a intrusões ígneas.

Em síntese, este capítulo estabeleceu as bases metodológicas para a definição de eletrofácies na Formação Pimenteiras, demonstrando a aplicabilidade de técnicas de *machine learning* como ferramentas de suporte à interpretação de perfis de poços. Os resultados obtidos não apenas confirmam a heterogeneidade interna da formação, mas também indicam que a combinação entre clusterização e classificação probabilística pode fornecer informações consistentes para estudos posteriores, em especial para a integração com dados de inversão sísmica e estimativas de conteúdo orgânico total, que serão explorados nos capítulos seguintes.

5. Machine Learning Aplicado a Estimativa de COT

5.1. INTRODUÇÃO

A caracterização 3D dos folhelhos, seja do ponto de vista de rochas geradoras ou alvos de reservatórios não convencionais, foi tema discutido por diversos trabalhos como de Liu *et al.*, 2019, que fizeram a predição de COT baseado em modelos petrofísicos e atributos sísmicos da inversão elástica pré-empilhamento. Publicação de Ogiesoba, O., & Hammes, U. (2014) que identificaram zonas ricas em COT a partir da análise de multiatributos e inversão sísmica pós-empilhamento. Entre outros estudos promovidos por Bredesen, K. *et al.*, 2015, Thiéblemont, A. *et al.*, 2022, Løseth, H. *et al.*, 2011 e del Monte, A. *et al.*, 2018, com foco na caracterização e identificação das rochas geradoras a partir de dados sísmicos. Todos esses trabalhos têm como dado importante o conteúdo de carbono orgânico total, que é um dado medido a partir de análises geoquímicas laboratoriais, e portanto, está disponível em apenas alguns intervalos chaves, amostrados de forma discreta e em um grupo restrito de poços. Uma vez que, a caracterização tridimensional necessita de mais informações, logo, curvas contínuas dessa propriedade, a maioria desses trabalhos baseiam-se em abordagens empíricas como as descritas por Passey *et al.*, 1990, que pode não ser a abordagem mais adequada a área de estudo deste trabalho, uma vez que, apenas um dos poços contém amostras de COT medido as quais poderiam ajustar a curva calculada pelo método de Passey e assim lidar com o fator LOM (*level of organic metamorphism*) dado este de difícil acesso e que pode variar de um poço para o outro, especialmente no contexto deste trabalho em que as intrusões ígneas são determinantes para a maturação das rochas geradoras. Além disso, os métodos empíricos propostos por Passey *et al.* (1990) e Schmoker *et al.* (1979), amplamente utilizados na indústria, podem sofrer com viés humano ao ajustar as linhas bases de folhelho a partir da curvas de resistividade e sônico, para o método de Passey, no caso de Schomoker a dependência exclusiva da relação linear do COT com o inverso da densidade, que pode nem sempre ser única a depender da complexidade e heterogeneidade dos folhelhos.

A fim de superar as limitações citadas, muitos trabalhos estão sendo publicados nos últimos anos, utilizando-se de técnicas de inteligência artificial a partir de perfis geofísicos de poços para treinar modelos a partir de diferentes técnicas de *machine learning* e assim estimar as curvas de COT. Exemplos disso podem ser vistos nos trabalhos de Zhao, Tao *et al.*, 2015, que aplicaram o método *Support Vector Machine* para estimar o COT na formação Lower Barnett Shale da Bacia Fort Worth no Texas, no artigo publicado por Emelynova, Irina *et al.*, 2016, que estimaram as curvas de COT a partir métodos como *Multi-layer Peceptron* em potenciais *shale-gas* nas bacias McArthur e Geogina, na Austrália. O método *XGBoost*, também foi utilizado para esse mesmo fim no trabalho realizado por Zhang, Yuhang, *et al.*, 2024 em potenciais *shale-oil* da Bacia Junggar na China. A Bacia de Santo, no Brasil, também foi alvo desse estudo, onde foram utilizados diferentes métodos de *machine learning* entre eles o *Random Forest* para identificar rochas potencialmente geradoras (Purificação, C. A. C. D. 2021).

Nesse contexto, é fácil notar a diversidade de métodos de inteligência artificial aplicados para o mesmo objetivo. O que não parece fácil é entender qual algoritmo específico terá o melhor desempenho em um conjunto de dados com geologia diferente de outra bacia. Por isso, esse capítulo será dedicado a estimativa das curvas de COT, a partir da combinação de três dos cinco métodos testados, sendo eles: *Random Forest*, *XGBoost*, *CatBoost*, *Multi-layer Peceptron* e *Support Vector Machine*, assim como sugerido por Mandal *et al.*, 2021 e Polikar (2012), para construir um modelo de melhor desempenho preditivo do COT de forma mais generalizada e consistente em diferentes poços (Figura 5.1). Outro fator primordial para a escolha desta abordagem, como será detalhada ao longo deste capítulo, se deve ao fato de que apenas um dos poços da área de estudo possui medições de COT. Portanto, essa abordagem parece a mais adequada pois permite que sejam utilizados poços de fora da área de estudo, que tiveram os valores de COT medidos, que sejam da mesma bacia e sequência sedimentares, as quais possuem características geológicas semelhantes. Dessa forma, esses poços externos serão utilizados para treinar diferentes modelos *machine learning*, que serão integrados para ser capaz de estimar, de forma mais precisa, os valores de COT nos poços que não possuem esse dado.

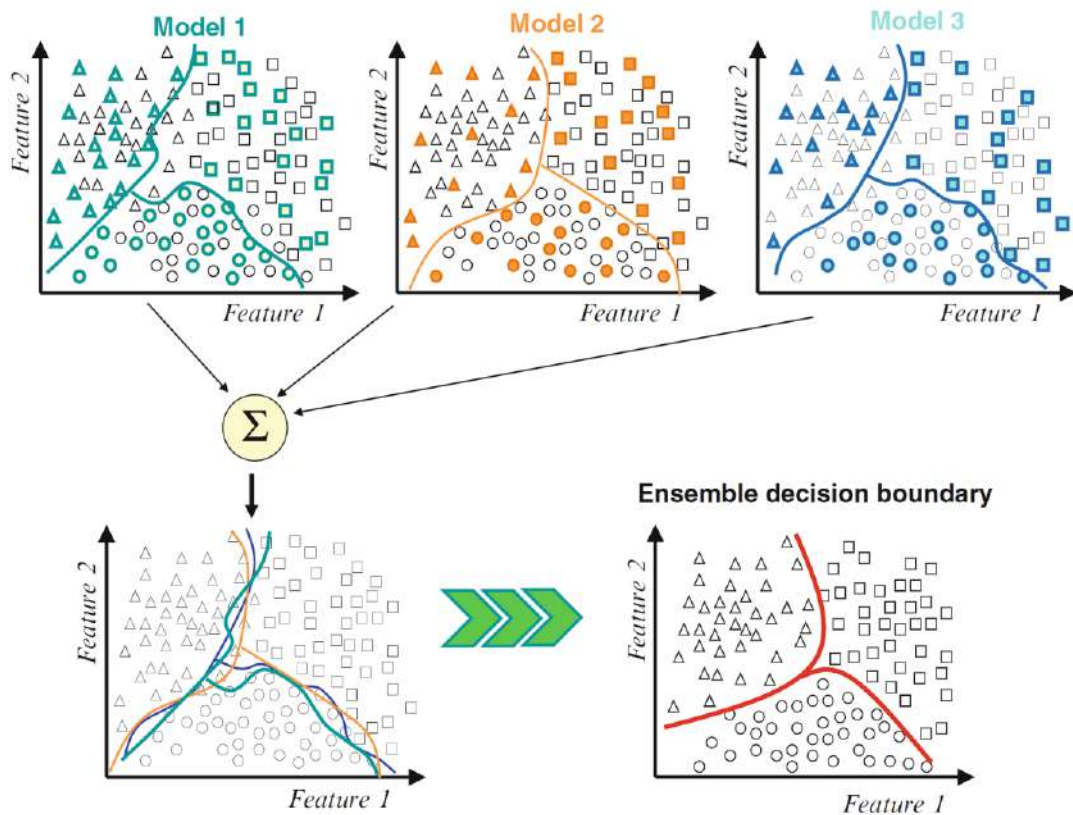


Figura 5.1. Fluxo conceitual proposto por Polikar 2012 para combinação de diferentes algoritmos de *machine learning*.

5.2. FUNDAMENTAÇÃO

5.2.1. Técnicas de *Machine Learning*

A teoria da aprendizagem estatística constitui a base teórica de grande parte dos algoritmos modernos de *machine learning* e é considerada um dos campos mais robustos dentro da inteligência artificial (Von Luxburg & Schölkopf, 2011). Seu desenvolvimento teve início na Rússia durante a década de 1960, mas foi a partir da década de 1990, com o surgimento da *Support Vector Machine* (SVM), que essa área ganhou grande destaque, tornando-se uma ferramenta padrão para tarefas de reconhecimento de padrões em diversas áreas, desde visão computacional até biologia computacional.

O *machine learning* pode ser entendido como o estudo de algoritmos e modelos computacionais que, baseados em dados, melhoram seu desempenho em uma tarefa específica ao longo do tempo, sem serem explicitamente programados para tal (Mitchell, 1997). A essência do *machine learning* reside na capacidade das máquinas de “aprender” a executar tarefas específicas a partir de exemplos, por meio de algoritmos dedicados. A ideia central é que, dada uma série de exemplos que ilustram um problema, o algoritmo possa inferir uma regra geral que não apenas explique os dados observados, mas que também seja capaz de generalizar para novos casos nunca vistos (Von Luxburg & Schölkopf, 2011).

Embora o *machine learning* tenha raízes na inteligência artificial, na estatística e na ciência da computação, ele consolidou-se como uma disciplina científica independente. Diferentemente da inteligência artificial tradicional, que busca modelar ou simular comportamentos inteligentes, o *machine learning* possui um foco mais pragmático: identificar mecanismos pelos quais computadores possam aprender tarefas específicas de forma eficiente. Muitos dos problemas abordados, como regressão, classificação e agrupamento, são familiares também para estatísticos e físicos, porém o *machine learning* aborda esses desafios sob a ótica da inferência indutiva e da capacidade de generalização dos modelos (Bishop, 2006; Von Luxburg & Schölkopf, 2011).

O *machine learning* pode ser categorizado em aprendizado supervisionado, não supervisionado e por reforço (Sutton & Barto, 2015). Entre os problemas mais estudados do aprendizado supervisionado está a classificação, na qual o algoritmo recebe um conjunto de exemplos rotulados, do tipo $(X_1, Y_1), \dots, (X_n, Y_n)$, e seu objetivo é aprender um mapeamento $f: X \rightarrow Y$ que minimize o erro das predição de novos dados o quanto possível. A classificação é um exemplo clássico de aprendizado supervisionado, no qual o algoritmo é “supervisionado” por meio dos rótulos corretos durante o treinamento, podendo assim avaliar a qualidade das suas previsões (Von Luxburg & Schölkopf, 2011; Bishop, 2006). Já o aprendizado não supervisionado trabalha com dados sem rótulos, focando na descoberta de padrões ou estruturas, como em métodos de agrupamento (*clustering*) ou redução de dimensionalidade. Por fim, o aprendizado por reforço envolve agentes que aprendem a tomar decisões sequenciais para maximizar recompensas acumuladas em um ambiente dinâmico (Sutton & Barto, 2015).

No contexto da geociência e outras áreas aplicadas, o *machine learning* tem se mostrado extremamente eficaz na modelagem de dados complexos e na predição de propriedades petrofísicas, de fácies, atributos e interpretação sísmica no geral, auxiliando na tomada de decisão e na automação de diversos processos. Técnicas de *machine learning* têm ganhado destaque nesse contexto, pois possibilitam a extração de relações complexas entre múltiplas variáveis, muitas vezes não lineares, presentes nos perfis de poços, por exemplo. Neste capítulo, serão apresentados cinco métodos de aprendizado supervisionado amplamente aplicados na predição de curvas contínuas em geociências: *Multi-Layer Perceptron* (MLP), *Random Forest* (RF), *CatBoost* (CB), *XGBoost* (XGB) e *Support Vector Machine* (SVM).

5.2.1.1. Random Forest

Algoritmos de *Random Forest* (RF) são técnicas de *machine learning* baseadas em *ensemble learning*, amplamente aplicadas para tarefas de classificação e regressão, que envolve a minimização

de impurezas, como o erro quadrático médio (MSE), para tarefas de regressão. (Breiman, 2001). O método consiste em construir múltiplas árvores de decisão durante o treinamento e combinar suas previsões para melhorar a acurácia do modelo final (Figura 5.2). Para classificação, a predição do RF é dada pela classe que obtém a maior frequência (modo) entre as árvores individuais, enquanto para regressão utiliza-se a média das previsões (Breiman, 2001; Liaw & Wiener, 2002).

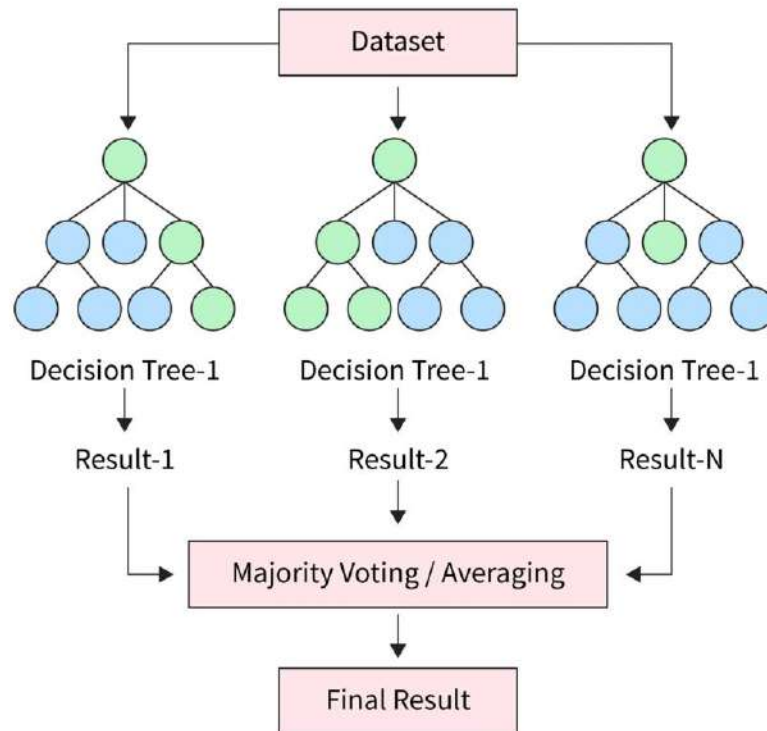


Figura 5.2. Esquema representativo do funcionamento do algoritmo Random Forest, ilustrando a construção de múltiplas árvores de decisão e a combinação dos resultados por média ou voto majoritário (Fonte: <https://www.scaler.com/topics/random-forest-in-r/>)

O conceito inicial de *Random Forest* foi desenvolvido por Ho (1995), que propôs o uso do método de subespaço aleatório para construção de múltiplas árvores, aumentando a variabilidade entre elas e diminuindo o risco de *overfitting* característico das árvores de decisão convencionais. Essa abordagem está fundamentada na ideia de discriminação estocástica (Kleinberg, 1990; Kleinberg *et al.*, 1996) e foi posteriormente aprimorada e popularizada por Breiman (2001), que consolidou o algoritmo e demonstrou sua eficácia em diversas aplicações.

No algoritmo padrão de árvore de decisão, cada nó é dividido escolhendo a melhor variável e ponto de corte dentre todas as disponíveis. Por outro lado, o algoritmo *Random Forest* introduz uma modificação crucial: para cada nó, é selecionado aleatoriamente um subconjunto de variáveis preditoras, e a melhor divisão é feita apenas entre essas variáveis (Breiman, 2001). Essa estratégia

promove diversidade entre as árvores individuais, reduzindo correlações entre elas e, conseqüentemente, melhorando a generalização do modelo.

O funcionamento do *Random Forest* pode ser descrito em etapas principais (Breiman, 2001):

1. São geradas n amostras *bootstrap* (com reposição) a partir do conjunto de dados original, formando diferentes subconjuntos de treinamento.
2. Para cada amostra *bootstrap*, é construída uma árvore de decisão sem podar, porém com a restrição de que, em cada nó, a divisão é feita entre um subconjunto aleatório de h variáveis preditoras.
3. As observações não selecionadas em cada *bootstrap*, chamadas de dados *out-of-bag* (OOB), são usadas para avaliar o desempenho do modelo sem necessidade de validação cruzada adicional.
4. A predição final para novos dados é obtida pela agregação das predições das n árvores (voto majoritário na classificação ou média na regressão).

A estimativa do erro *out-of-bag* (Erro OOB) é calculada como a média do erro nas predições feitas sobre as amostras OOB, proporcionando uma avaliação interna do modelo com baixo custo computacional:

$$ERE_{OOB} = (1/n) \sum_{i=1}^n [y_i - g_{OOB}(X_i)]^2 \quad (5.1)$$

Onde y_i é o valor real e $g_{OOB}(X_i)$ é a predição agregada das árvores para a observação i , considerando apenas aquelas árvores nas quais i foi *out-of-bag*.

Os principais hiperparâmetros do *Random Forest* incluem:

- $n_estimators$ (número de árvores): indica a quantidade de árvores que serão construídas no ensemble. Valores maiores tendem a reduzir o erro do modelo e aumentar a estabilidade da predição, porém aumentam o custo computacional. Estudos sugerem iniciar com um valor dez vezes maior que o número de variáveis preditoras e ajustar conforme a performance e recursos disponíveis (Breiman, 2001).
- $max_features$ (número de variáveis consideradas em cada nó): define quantas variáveis são aleatoriamente selecionadas para considerar a melhor divisão em cada nó da árvore. A escolha desse parâmetro é crítica para o *trade-off* entre viés e variância do modelo. Em problemas de classificação, um valor padrão comum é a raiz quadrada do número total de variáveis, enquanto

para regressão costuma-se usar um terço do total (Liaw & Wiener, 2002). Reduzir esse valor aumenta a diversidade entre as árvores, promovendo melhor generalização.

- *max_depth* (profundidade máxima das árvores): limita o número máximo de divisões feitas em cada árvore, controlando a complexidade e evitando o *overfitting*. Se não especificado, as árvores crescem até o máximo possível, podendo capturar ruídos.
- *min_samples_split* (mínimo de amostras para dividir um nó): define o número mínimo de amostras necessárias para realizar uma divisão em um nó. Valores maiores evitam divisões muito pequenas e ajudam a suavizar o modelo, reduzindo *overfitting*.
- *min_samples_leaf* (mínimo de amostras em uma folha): estipula o número mínimo de amostras que devem estar presentes em uma folha terminal. Esse parâmetro controla a granularidade do modelo e melhora sua estabilidade.
- *bootstrap*: booleano que indica se o método *bootstrap* (amostragem com reposição) será utilizado para gerar as amostras de treinamento de cada árvore. O valor padrão é True, essencial para garantir diversidade entre as árvores.
- *random_state*: controla a semente do gerador de números aleatórios para garantir resultados reproduzíveis entre execuções.

5.2.1.2. Extreme Gradient Boosting

O XGBoost (*Extreme Gradient Boosting*) é um algoritmo de *machine learning* baseado na técnica de *gradient boosting*, projetado para otimizar desempenho e eficiência em problemas de classificação e regressão (Chen & Guestrin, 2016). Trata-se de uma implementação avançada de *boosting* de árvores de decisão, que combina múltiplas árvores fracas em um modelo preditivo forte, por meio da adição sequencial de árvores que corrigem os erros residuais dos modelos anteriores (Friedman, 2001) (Figura 5.3).

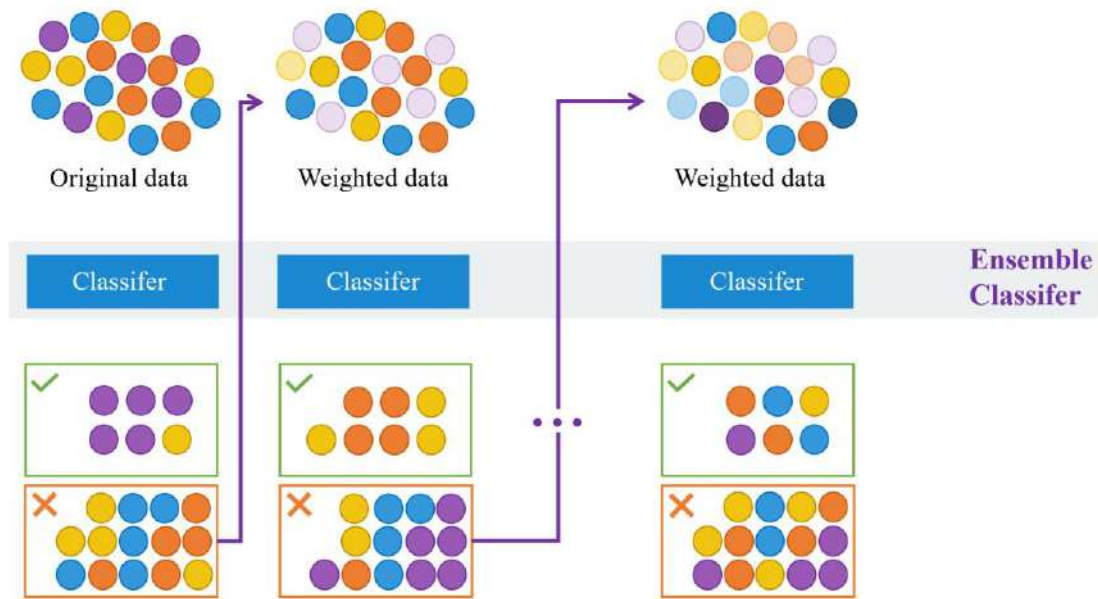


Figura 5.3. Representação do processo de *gradient boosting*, mostrando a adição sequencial de árvores para reduzir os erros residuais e aprimorar o modelo preditivo (Chen & Guestrin, 2016)

O conceito fundamental do *gradient boosting* foi inicialmente proposto por Friedman (2001), onde modelos são ajustados iterativamente minimizando uma função de perda diferenciável via descida do gradiente. O XGBoost aprimora essa abordagem com uma série de inovações computacionais e de modelagem que o tornam altamente eficiente e escalável incluindo regularização L1 e L2, suporte a dados faltantes, paralelização, uso de árvores de decisão balanceadas, em que a função objetivo a ser minimizada pode ser escrita como (Chen & Guestrin, 2016):

$$Obj^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (5.2)$$

Onde l é a função de perda, \hat{y}_i é a previsão do modelo, f_t é a nova árvore adicionada no passo t e $\Omega(f_t)$ é o termo de regularização que ajuda a prevenir *overfitting*:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (5.3)$$

Com T sendo o número de folhas da árvore, w_j o peso da folha j e γ, λ hiperparâmetros de regularização.

Diferentemente de métodos como *Random Forest*, que constroem árvores de forma independente em paralelo, o XGBoost constrói árvores sequencialmente, cada uma visando reduzir o

erro do modelo atual, através da otimização da função objetivo que inclui a função de perda e termos de regularização para controlar a complexidade do modelo.

O funcionamento do XGBoost pode ser descrito pelas seguintes etapas principais (Chen & Guestrin, 2016):

1. Inicialmente, um modelo simples é treinado, geralmente uma árvore com predição constante (como a média dos valores no caso de regressão).
2. Calculam-se os gradientes e hessianos da função de perda em relação às predições atuais para todas as amostras do conjunto de treinamento.
3. Constrói-se uma nova árvore que tenta minimizar a função objetivo baseada nessas informações de gradiente, escolhendo divisões que maximizam a redução do erro (ganho).
4. A árvore gerada é adicionada ao modelo com um peso controlado por uma taxa de aprendizado (*learning rate*).
5. O processo é repetido iterativamente até o número pré-definido de árvores ser alcançado ou até a convergência do erro.

O XGBoost incorpora várias características importantes que aumentam sua robustez e capacidade de generalização:

- Regularização: termos de regularização L1 (Lasso) e L2 (Ridge) são aplicados aos pesos das folhas das árvores, controlando a complexidade e prevenindo *overfitting* (Chen & Guestrin, 2016).
- Tratamento eficiente de dados faltantes: o algoritmo aprende automaticamente o melhor caminho para dados ausentes durante a divisão dos nós.
- Paralelização e otimização computacional: implementa algoritmos para particionamento de dados e construção de árvores, utilizando *multithreading* e *cache awareness*.
- *Pruning* (poda) de árvores: após a construção, realiza poda para eliminar ramos que não melhoram o modelo, aumentando a simplicidade e generalização.
- *Sparsity-aware split finding*: otimiza a construção das árvores quando os dados são esparsos, comum em muitos problemas reais.
- Os principais hiperparâmetros do XGBoost que impactam diretamente seu desempenho incluem:

- *n_estimators*: número total de árvores a serem construídas. Aumentar esse valor pode melhorar o ajuste, mas também aumenta o risco de *overfitting* e o tempo computacional.
- *learning_rate* (eta): taxa de aprendizado que controla o peso de cada árvore adicionada. Valores menores geralmente melhoram a generalização, exigindo mais árvores.
- *max_depth*: profundidade máxima das árvores, limitando o número de nós na árvore e impacta desempenho.
- *subsample*: proporção de amostras utilizadas para construir cada árvore, introduzindo aleatoriedade para melhorar a generalização.
- *colsample_bytree*: proporção de variáveis (*features*) amostradas para cada árvore, ajudando a controlar a correlação entre árvores.
- *gamma*: redução mínima da função objetivo necessária para realizar uma divisão em um nó. Controla a complexidade do modelo, sendo uma forma de poda antecipada.

5.2.1.3. *Categorical Boosting*

O CatBoost (*Categorical Boosting*) é, também, um algoritmo baseado na técnica de *gradient boosting* que se destaca pelo manejo eficiente de variáveis categóricas e numéricas, oferecendo desempenho robusto em tarefas de classificação e regressão (Prokhorenkova *et al.*, 2018). Embora seu principal diferencial seja o tratamento nativo de variáveis categóricas, o CatBoost também apresenta estratégias avançadas para lidar com variáveis contínuas.

O CatBoost consiste em uma implementação de algoritmo de boosting por gradiente que emprega árvores de decisão binárias como modelos preditores básicos. Uma árvore de decisão é construída por meio de uma partição recursiva do espaço de variáveis R^m em múltiplas regiões disjuntas, denominadas nós da árvore, definidas a partir dos valores de determinados atributos de divisão a . Esses atributos normalmente correspondem a variáveis binárias que indicam se uma característica x_k ultrapassa um determinado limite t (*threshold*), ou seja, $a = \mathbb{1}_{\{x_k > t\}}$, onde x_k representa uma variável numérica ou binária; no caso desta última, o valor do limite é usualmente $t = 0,5$. Cada uma das regiões finais, chamadas folhas da árvore, é associada a um valor que corresponde a uma estimativa da variável resposta y para tarefas de regressão ou ao rótulo de classe preditos em problemas de classificação. Assim, uma árvore de decisão h pode ser expressa como (Prokhorenkova *et al.*, 2018):

$$h(x) = \sum_{j=1}^J b_j \mathbb{1}_{\{x \in R_j\}} \quad (5.4)$$

Onde R_j são as regiões disjuntas correspondentes às folhas da árvore.

Além disso, exige menos ajustes finos de hiperparâmetros, gerando um resultado melhor imediatamente. Isso ocorre porque o CatBoost escolhe automaticamente valores iniciais razoáveis para os hiperparâmetros. Outra vantagem do CatBoost é que ele utiliza uma taxa de aprendizado adaptável, que pode se adaptar automaticamente durante o processo de treinamento. Isso pode reduzir a necessidade de ajuste manual da taxa de aprendizado, um parâmetro que pode afetar significativamente o desempenho de modelos com aumento de gradiente.

No que se refere à parametrização do CatBoost, destacam-se os seguintes hiperparâmetros principais (Prokhorenkova *et al.*, 2018; Dorogush *et al.*, 2018):

- *depth*: profundidade máxima das árvores
- *learning_rate*: taxa de aprendizado que controla a contribuição de cada árvore para o modelo final, ajudando a balancear precisão e risco de *overfitting*.
- *l2_leaf_reg*: regularização L2 aplicada aos pesos das folhas das árvores, importante para evitar *overfitting* na modelagem em flutuações locais dos dados contínuos.
- *random_strength*: parâmetro que adiciona uma variável aleatória com distribuição normal à pontuação da variável (*features*). Ela tem média zero e variância que é maior no início do treinamento e diminui durante o treinamento. O *random_strength* é o multiplicador da variância, o que também ajuda a prevenir *overfitting*.
- *bagging_temperature*: é responsável pelo bootstrap bayesiano, que é usado por padrão nos modos de classificação e regressão, em que cada objeto recebe um peso aleatório. Se a temperatura de bagging for igual a 1, os pesos são amostrados a partir da distribuição exponencial. Se a temperatura de ensacamento for igual a 0, todos os pesos serão iguais a 1.

Diferentemente de outras implementações de *gradient boosting*, o CatBoost utiliza o conceito de *ordered boosting*, que reduz o viés estatístico causado pelo uso das mesmas amostras para cálculo das estatísticas de variáveis categóricas e treinamento do modelo, mas esse mecanismo também beneficia o tratamento de variáveis contínuas ao melhorar a generalização do modelo (Prokhorenkova *et al.*, 2018). O que torna o CatBoost um algoritmo altamente eficaz e estável, capaz de lidar com

datasets heterogêneos, contendo tanto variáveis categóricas quanto contínuas, sem a necessidade de pré-processamento extenso (Dorogush *et al.*, 2018).

5.2.1.4. Multi-Layer Perceptron

O *Multi-Layer Perceptron* (MLP) é um modelo clássico de rede neural artificial, amplamente utilizado em tarefas de classificação e regressão (Haykin, 1999). O MLP é uma rede *feedforward* (que não contém ciclos de *feedback*) a qual garante que as saídas da rede possam ser calculadas como funções explícitas das entradas e dos pesos (Bishop, 1995). A rede neural desse tipo, é composta por múltiplas camadas de neurônios artificiais, incluindo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída (Figura 5.4), onde cada neurônio em uma camada está conectado a todos os neurônios da camada subsequente (Haykin, 1999).

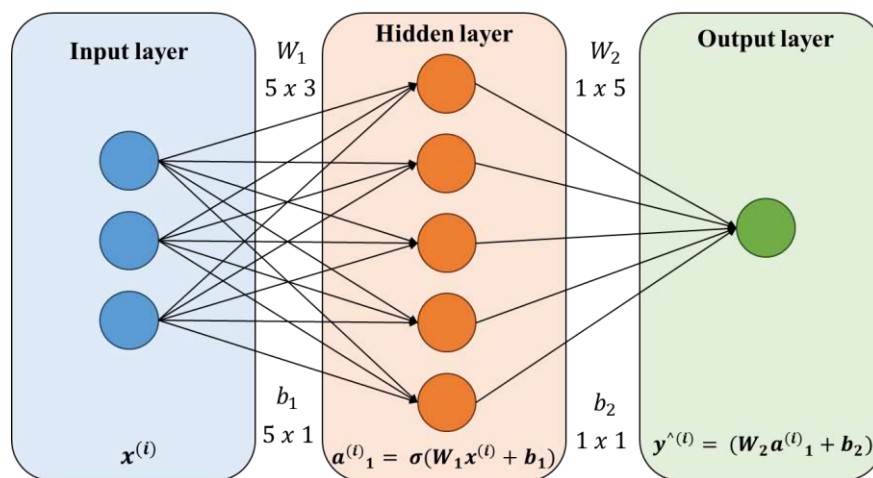


Figura 5.4. Arquitetura típica de uma rede neural Multi-Layer Perceptron (MLP), composta por camadas de entrada, ocultas e saída (modificado de Mandal *et al.*, 2021)

Cada neurônio realiza uma combinação linear ponderada das entradas recebidas e aplica uma função de ativação não linear, permitindo que o MLP realize uma propagação direta (*forward propagation*), utilizando entradas conhecidas e pesos inicializados aleatoriamente para gerar valores estimados. Em seguida, calcula-se o erro entre esses valores preditos e os valores reais (Bishop, 1995).

O treinamento do MLP é geralmente realizado por meio do algoritmo de retropropagação (*backpropagation*), que utiliza o método do gradiente descendente para minimizar uma função de custo, ajustando iterativamente os pesos das conexões da rede (Rumelhart *et al.*, 1986). Esse processo envolve a propagação do erro da saída para as camadas anteriores, atualizando os pesos de forma a reduzir o erro preditivo global. Matematicamente esse processo pode ser representado como:

$$a_i = f\left(\sum_{j=1}^n w_{ij}x_j + b_i\right) \quad (5.5)$$

Onde a_i é a saída do neurônio i , w_{ij} são os pesos, x_j são as entradas, b_i é o viés e f é a função de ativação.

De maneira geral, o algoritmo de aprendizado por retropropagação de erros compreende duas fases principais: a propagação direta (*forward*) e a retropropagação (*backward*). Na etapa de propagação direta, um vetor de entrada é aplicado aos neurônios da camada sensorial, e a ativação resultante é transmitida sucessivamente pelas camadas subsequentes da rede até a obtenção da saída, que corresponde à resposta atual do modelo. Em seguida, calcula-se o erro a partir da diferença entre essa saída e o valor alvo desejado. Esse erro é propagado no sentido inverso ao fluxo inicial, ao longo das conexões sinápticas, com o objetivo de ajustar os pesos sinápticos de maneira a reduzir estatisticamente a discrepância entre a saída real e a saída esperada (Haykin, 1999).

Antes de iniciar os experimentos, os hiperparâmetros são calibrados por meio de análises de sensibilidade aplicadas aos conjuntos de dados de treinamento e validação. O funcionamento básico do MLP pode ser descrito em etapas principais:

1. As entradas são apresentadas à camada de entrada da rede.
2. Cada neurônio da camada oculta calcula uma soma ponderada das entradas, seguida da aplicação de uma função de ativação não linear, como a sigmoide, tangente hiperbólica ou ReLU (Rectified Linear Unit).
3. O sinal processado pelas camadas ocultas é então passado para a camada de saída, que produz a predição final da rede.
4. Durante o treinamento, o erro entre a saída predita e o valor real é calculado e utilizado para ajustar os pesos através do algoritmo de retropropagação.
5. Esse ciclo se repete por diversas iterações até a convergência ou até que algum critério de parada seja satisfeito.

Os hiperparâmetros fundamentais para o MLP incluem:

- Número de camadas ocultas e neurônios por camada: influenciam diretamente a capacidade do modelo de representar padrões complexos. Redes muito pequenas podem ter baixa capacidade, enquanto redes muito grandes podem sofrer *overfitting* (Goodfellow *et al.*, 2016).
- Função de ativação: define a não linearidade introduzida em cada neurônio. A escolha da função pode afetar a convergência e a performance do modelo (Nwankpa *et al.*, 2018).

- Taxa de aprendizado (*learning rate*): controla o tamanho dos passos durante a atualização dos pesos. Taxas muito altas podem impedir a convergência, enquanto taxas muito baixas tornam o treinamento lento.
- Número de épocas (*epochs*): define quantas vezes o modelo passará por todo o conjunto de treinamento durante o treinamento.
- Algoritmo de otimização: além do gradiente descendente padrão, são utilizados métodos mais avançados como Adam, RMSProp, que aceleram e estabilizam o treinamento (Kingma & Ba, 2015).
- Regularização: técnicas como *dropout*, L2 (*weight decay*) ajudam a evitar *overfitting* e melhoram a generalização do modelo (Srivastava *et al.*, 2014).

5.2.1.5. Support Vector Machine

Support Vector Machines (SVM) são algoritmos de aprendizado supervisionado utilizados tanto para tarefas de classificação quanto de regressão, amplamente reconhecidos pela capacidade de encontrar fronteiras de decisão ótimas com alta generalização (Cortes & Vapnik, 1995). As SVMs baseiam-se no conceito de maximização da margem, onde um hiperplano ótimo pode ser compreendido como a função de decisão linear que maximiza a margem entre as duas classes, ou seja, a maior distância possível entre os vetores representativos de cada classe (Figura 5.5). Um aspecto fundamental desse conceito é que a construção desse hiperplano depende apenas de um subconjunto restrito dos dados de treinamento, conhecidos como vetores de suporte, os quais determinam a margem ótima. Além disso, foi evidenciado que, quando os vetores de treinamento são separáveis sem erro pelo hiperplano ótimo, a probabilidade esperada de erro em novos exemplos está limitada pela proporção entre o número esperado de vetores de suporte e o total de dados de treinamento utilizados (Cortes & Vapnik, 1995).

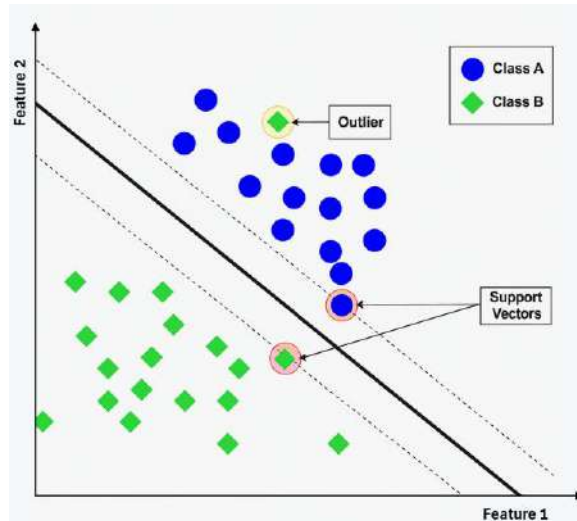


Figura 5.5. Ilustração do conceito de hiperplano ótimo em Support Vector Machine (SVM), evidenciando a maximização da margem entre classes (Fonte: <https://towardsdatascience.com/support-vector-machine-with-scikit-learn-a-friendly-introduction-a2969f2ff00d/>)

O algoritmo SVM foi originalmente proposto por Cortes e Vapnik (1995) e se fundamenta na teoria do aprendizado estatístico de Vapnik (1998). A ideia central é transformar os dados originais em um espaço de maior dimensão, onde seja possível encontrar um hiperplano linear que maximize a margem entre as classes, utilizando funções de kernel para realizar essa transformação sem necessidade de computar explicitamente as coordenadas no espaço de alta dimensão (função *kernel trick*).

O objetivo do SVR é encontrar uma função $f(\vec{x})$, com uma margem de erro caracterizada pelo intervalo $[y_i - \varepsilon, y_i + \varepsilon]$, onde, desvios são permitidos desde que não ultrapassem a margem especificada, dada por:

$$f(\vec{x}) = \langle \vec{w}, \vec{x} \rangle + b \quad (5.6)$$

A otimização (problema primal), sendo escrita como:

$$\min \frac{1}{2} \|\vec{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (5.7)$$

Sujeito a:

$$y_i = \langle \vec{w}, \vec{x} \rangle - b \leq \varepsilon + \xi_i$$

$$\langle \vec{w}, \vec{x} \rangle + b - y_i \leq \varepsilon + \xi_i^* \quad (5.8)$$

$$\xi_i, \xi_i^* \geq 0, \quad \forall i = 1, \dots, n,$$

Onde ε é a margem de tolerância, ξ_i, ξ_i^* são variáveis de folga que permite desvios maiores que ε e C é o parâmetro de penalização que controla o trade-off entre a margem e o erro de treinamento.

Para capturar a não linearidade, o SVM utiliza a função de *kernel*, substituindo o produto interno por uma função kernel $K(x_i, x_j)$, por exemplo o RBF:

$$K(x_i, x_j) = \exp \left(-\gamma \|x_i - x_j\|^2 \right) \quad (5.9)$$

Diferentemente dos classificadores lineares tradicionais, as SVMs são robustas a dados com alta dimensionalidade e conseguem lidar eficientemente com problemas não lineares por meio da utilização de diferentes funções kernel, como o kernel linear, polinomial, radial basis function (RBF) e sigmoid (Schölkopf & Smola, 2002).

O funcionamento do método SVM pode ser resumido nos seguintes passos (Cortes & Vapnik, 1995; Burges, 1998):

1. Definição do problema: dados um conjunto de treinamento com rótulos binários, o objetivo é encontrar o hiperplano que separa as duas classes com a maior margem possível.
2. Maximização da margem: a SVM resolve um problema de otimização convexa, onde maximiza a margem entre as classes sujeita a restrições que garantem a correta classificação dos dados, ou minimização do erro no caso de dados não linearmente separáveis.
3. Uso do kernel: para dados que não são linearmente separáveis, aplica-se uma função kernel para mapear os dados a um espaço de variáveis de maior dimensão, onde a separação linear seja possível.
4. Determinação dos vetores de suporte: são os pontos de dados mais próximos ao hiperplano ótimo que definem a margem; apenas esses pontos influenciam o modelo final.
5. Classificação de novos dados: a predição é feita avaliando em qual lado do hiperplano o dado de entrada se encontra.

As principais vantagens do SVM incluem sua alta capacidade de generalização, eficiência em espaços de alta dimensão e a flexibilidade proporcionada pelas funções kernel (Schölkopf & Smola, 2002). Contudo, a escolha adequada do kernel e dos parâmetros (como o parâmetro de regularização C e o parâmetro do *kernel gamma*) é fundamental para o desempenho do modelo. Os principais hiperparâmetros do SVM são:

- C (parâmetro de penalização): controla o *trade-off* entre a maximização da margem e o erro de classificação nos dados de treinamento. Valores altos de C buscam minimizar erros de treinamento, podendo causar *overfitting*; valores baixos permitem margem maior com mais erros (Cortes & Vapnik, 1995).

- Tipo de *kernel*: define a função utilizada para mapear os dados a um espaço de maior dimensão. Os *kernels* mais comuns são linear, polinomial, RBF e sigmoid (Schölkopf & Smola, 2002).
- Parâmetros do *kernel*: por exemplo, no *kernel* RBF, o parâmetro gamma controla o alcance da influência de um único exemplo de treinamento; valores baixos indicam influência ampla, enquanto valores altos limitam a influência a pontos próximos.

5.2.3. Métricas de Desempenho

A avaliação da performance de modelos preditivos é uma etapa fundamental em trabalhos que envolvem técnicas de *machine learning* e regressão, pois permite quantificar a qualidade das previsões e comparar diferentes modelos ou configurações. Para isso, diversas métricas estatísticas são empregadas, cada uma com suas particularidades e adequação dependendo do contexto e do tipo de dado analisado. Neste capítulo, são abordadas as principais métricas utilizadas para avaliar a performance de modelos de regressão, incluindo o Erro Quadrático Médio (MSE), Erro Absoluto Médio (MAE), Raiz do Erro Quadrático Médio (RMSE), Mediana do Erro Absoluto (MedAE), Coeficiente de Correlação de Pearson e o coeficiente de determinação (R^2 ou Pearson^2).

5.2.3.1. Erro Quadrático Médio (*Mean Squared Error* – MSE)

O MSE é uma das métricas mais utilizadas para avaliar a precisão de modelos de regressão. Ele calcula a média dos quadrados das diferenças entre os valores observados e os valores preditos pelo modelo. Matematicamente, é definido como:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5.10)$$

Onde y_i representa o valor observado, \hat{y}_i valor predito pelo modelo e n o número total de observações.

Valores menores de MSE indicam um ajuste mais próximo entre as previsões e os valores reais. Contudo, devido ao termo quadrático, o MSE é muito sensível a erros grandes (*outliers*), penalizando-os fortemente. Portanto, um MSE elevado pode indicar a presença de grandes discrepâncias em algumas previsões. (Hastie, Tibshirani & Friedman, 2009). Na otimização e treinamento de modelos, o MSE frequentemente serve como função de custo para ajustar os parâmetros, devido à sua propriedade diferenciável e penalização forte de erros grandes. Modelos com MSE menor em dados de teste indicam melhor capacidade de generalização. Contudo, seu alto peso para *outliers* pode afetar a análise, sendo importante interpretar esse valor junto de outras métricas.

5.2.3.2. Erro Absoluto Médio (*Mean Absolute Error* – MAE)

O MAE mede a média dos valores absolutos das diferenças entre valores observados e preditos, oferecendo uma interpretação intuitiva da magnitude do erro médio:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5.11)$$

Por utilizar o valor absoluto, o MAE é menos sensível a *outliers* comparado ao MSE, proporcionando uma avaliação mais robusta para distribuições assimétricas ou com ruídos (Willmott & Matsuura, 2005). O MAE representa o erro médio absoluto esperado nas unidades originais da variável. Valores menores indicam previsões mais precisas, sendo uma métrica facilmente compreendida por sua interpretação direta da magnitude média do erro. O MAE é útil para avaliar a performance típica de modelos quando se deseja uma métrica estável e com interpretação direta na unidade original dos dados. É especialmente valorizado em aplicações onde grandes erros são menos frequentes ou onde outliers não devem dominar a otimização do modelo. Pode ser usado para comparar diferentes algoritmos ou configurações.

5.2.3.3. Raiz do Erro Quadrático Médio (*Root Mean Squared Error* – RMSE)

O RMSE é simplesmente a raiz quadrada do MSE, trazendo a métrica para a mesma unidade dos dados originais, facilitando a interpretação:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5.12)$$

Assim como o MSE, valores menores de RMSE indicam melhor desempenho. Como penaliza erros grandes, o RMSE destaca modelos que produzem grandes erros individuais (Chai & Draxler, 2014). Por estar na mesma escala dos dados, é muitas vezes preferido para interpretação direta do erro médio típico. O RMSE é muito utilizado para comparar o desempenho entre modelos, pois mantém as propriedades do MSE, mas com unidade original dos dados, tornando-o intuitivo para avaliação prática. Modelos com RMSE menor são geralmente preferidos, especialmente em tarefas onde grandes discrepâncias são críticas, como propriedades petrofísicas.

5.2.3.4. Mediana do Erro Absoluto (*Median Absolute Error* – MedAE)

A MedAE calcula a mediana dos valores absolutos dos erros, sendo ainda mais robusta que o MAE a *outliers*, pois a mediana não é influenciada por valores extremos:

$$MedAE = median (|y_1 - \hat{y}_1|, |y_2 - \hat{y}_2|, \dots, |y_n - \hat{y}_n|) \quad (5.13)$$

A MedAE representa o erro absoluto típico central das previsões, ignorando erros extremos. Isso é útil para avaliar a performance típica do modelo quando se deseja minimizar a influência de discrepâncias esporádicas muito grandes (Wilcox, 2012). Em cenários com dados ruidosos ou com *outliers*, a MedAE oferece uma avaliação mais confiável da performance central do modelo. É especialmente útil para validar modelos quando a estabilidade frente a erros extremos é desejada e para comparar robustez entre algoritmos.

5.2.3.5. Coeficiente de Correlação de Pearson

O coeficiente de correlação de Pearson r quantifica a força e a direção da relação linear entre os valores observados e preditos. É dado por:

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \quad (5.14)$$

Onde \bar{y} e $\bar{\hat{y}}$ são as médias dos valores observados e preditos, respectivamente. O valor de r varia de -1 a 1, onde valores próximos a 1 indicam forte correlação positiva, valores próximos a -1 indicam forte correlação negativa, e valores próximos a zero indicam pouca ou nenhuma correlação linear (Benesty *et al.*, 2009). O coeficiente de Pearson é uma medida complementar que avalia a capacidade do modelo em captar a tendência linear dos dados, independentemente da escala exata do erro. Um alto valor indica que o modelo está conseguindo reproduzir a variação e a direção dos valores reais, importante para validar modelos preditivos.

5.2.3.6. Coeficiente de Determinação (R^2 ou Pearson²)

O coeficiente de determinação R^2 representa a proporção da variância dos valores observados que é explicada pelo modelo. Pode ser calculado como o quadrado do coeficiente de Pearson quando a relação é linear:

$$R^2 = r^2 \quad (5.15)$$

Um valor de R^2 próximo de 1 indica que o modelo explica grande parte da variabilidade dos dados, enquanto valores próximos a zero indicam baixo poder explicativo (Montgomery, Peck & Vining, 2012). O R^2 é amplamente utilizado para quantificar a qualidade global do ajuste do modelo, indicando quanto da variabilidade dos dados é explicada pela previsão. Modelos com valores altos de R^2 são geralmente considerados mais eficientes, embora deva ser interpretado junto com as métricas de erro para evitar ilusões em casos de *overfitting* ou dados não lineares.

5.2.4. Importância das Variáveis

A importância das variáveis, ou *feature importance*, é uma métrica que quantifica a contribuição relativa de cada atributo para o desempenho e as previsões de modelos de *machine learning*. Em algoritmos baseados em árvores, como *Random Forest*, XGBoost e CatBoost, essa avaliação é direta e amplamente utilizada, pois a importância é calculada a partir da diminuição média de impurezas *mean decrease in impurity* - MDI), também chamada de importância de Gini (Scornet, 2023) ou na função objetivo obtida pelos *splits* que envolvem as variáveis ao longo das árvores, permitindo identificar quais atributos exercem maior impacto preditivo (Breiman, 2001; Chen & Guestrin, 2016; Prokhorenkova *et al.*, 2018).

Por outro lado, em modelos como *Support Vector Machine* e *Multi-Layer Perceptron*, que possuem estruturas mais complexas e não lineares, o cálculo direto da importância das variáveis não é inerente ao algoritmo e, portanto, não é tão comum quanto em modelos baseados em árvores. Contudo, é possível estimar essa importância por meio de técnicas específicas. Para SVMs lineares, os coeficientes do hiperplano de decisão podem ser utilizados como uma medida direta da relevância das variáveis, sendo que coeficientes com maior magnitude indicam maior importância (Guyon *et al.*, 2002). Em SVMs com *kernels* não lineares, a interpretação direta dos coeficientes não é possível, e métodos como eliminação recursiva de variáveis (*Recursive Feature Elimination*) e importâncias baseadas em permutação são empregados para avaliar a influência das variáveis (Guyon *et al.*, 2002).

No caso do MLP, a avaliação da importância das variáveis é mais desafiadora devido à sua arquitetura multicamada e altamente não linear. A interpretação geralmente depende de métodos pós-treinamento, como análise de sensibilidade, relevância por propagação reversa (*Layer-wise Relevance Propagation*), análise de gradientes e técnicas de permutação de variáveis, que permitem atribuir a contribuição de cada variável para a saída do modelo (Olden & Jackson, 2002; Lundberg & Lee, 2017).

A compreensão da importância das variáveis é essencial para interpretar, validar e aprimorar modelos, além de facilitar a seleção de atributos relevantes, reduzir a dimensionalidade dos dados e a redundância dos dados. Isso é particularmente importante em aplicações críticas que demandam explicabilidade, como no caso do foco deste trabalho que é importante entender quais perfis geofísicos mais influencia ou melhor minimiza erro da variável alvo da predição. Dessa forma, é possível compreender se uma análise integrada entre dados de poços e o seu *upscale* para dados sísmicos está enviesada ou não.

5.4. MÉTODO

O fluxo de trabalho adotado para a estimativa do COT nos poços da área de estudo está resumido na Figura 5.6. A estratégia principal foi treinar modelos de *machine learning* utilizando dados de poços localizados fora da área de estudo, mas que continham medições de COT. Isso possibilitou ter uma quantidade suficiente de dados para treinar e validar os modelos.

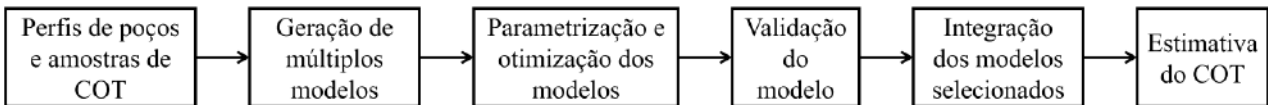


Figura 5.6. Fluxo de trabalho adotado para a estimativa do COT nos poços da Formação Pimenteiras.

Inicialmente, foram selecionados os perfis comuns a todos os poços perfilados na Formação Pimenteiras que apresentassem relação com o conteúdo de matéria orgânica. Como os poços são antigos, datados da década de 50 e 80, sendo os mais recentes perfurados em 2014, pode-se imaginar que as condições de perfuração mudaram entre esses poços e os mais antigos podem ter sofrido mais com respeito a integridade das paredes dos poços, o que influencia bastante na perfilagem de diversos ferramentas. Por isso, foi necessário realizar um pré-condicionamento das curvas para garantir que os dados estivessem uniformizados para as etapas seguintes.

Em seguida, foi feita uma análise exploratória dos dados, com gráficos, tabelas e histogramas, destacando principalmente a distribuição do COT. A partir disso, aplicaram-se transformações nos dados, como a transformação logarítmica do perfil de resistividade e a padronização (*StandardScaler*), para melhorar a performance dos modelos. Os dados foram então divididos em conjuntos de treino, validação e teste. O único poço com amostras de COT dentro da área de estudo, chamado 2-CP-1-MA, foi reservado para teste, não sendo usado, portanto, no treinamento.

Na etapa de treinamento, os modelos foram ajustados com parâmetros específicos e avaliados por métricas estatísticas gerais. A performance foi avaliada no conjunto de validação, e os melhores modelos foram selecionados com base em critérios estatísticos. Também foi realizada uma análise da importância das variáveis de entrada para entender quais curvas mais influenciaram as previsões. Os modelos selecionados foram combinados por meio de uma média simples para gerar a estimativa final do COT.

Por fim, realizou-se um pós-processamento, que consistiu em uma regressão linear para corrigir possíveis vieses nas previsões, garantindo maior precisão nas estimativas. As próximas seções

detalham cada uma dessas etapas, apresentando os procedimentos e resultados que sustentam a metodologia adotada.

5.4.1. Seleção do conjunto de dados

Os dados de perfis de poços e amostras de COT utilizados para treinar os modelos de *machine learning* foram adquiridos através do banco de dados da ANP (Agência Nacional do Petróleo) que disponibilizou todos os dados terrestres de poços e sísmicas de forma *online*. Com exceção do poço 1-OGX-110-MA que tiveram as medições de COT disponibilizadas pela Universidade Federal Fluminense através do grupo pesquisa GIECAR (Grupo de Interpretação Exploratória e Caracterização de Reservatórios). A Figura 5.7 mostra a localização dos poços utilizados, que como pode ser notados, todos eles estão fora do polígono azul que delimita a área da aquisição sísmica 3D e portanto da área de interesse deste estudo.

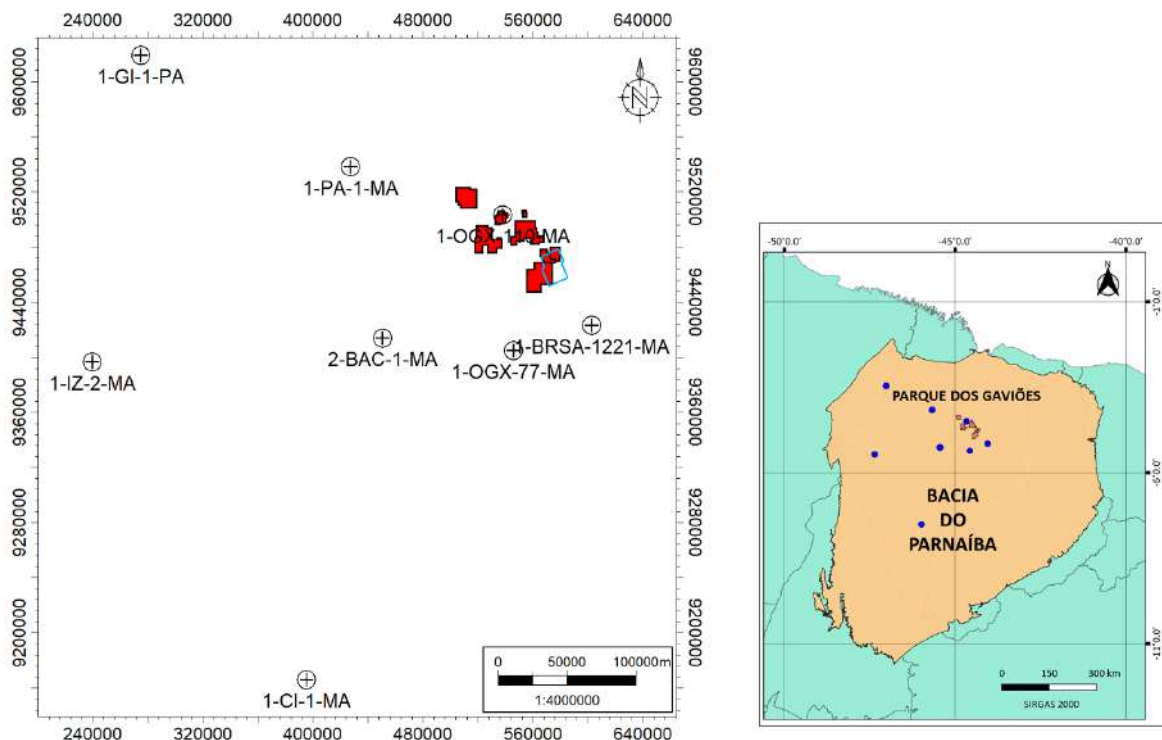


Figura 5.7. Localização dos poços utilizados neste estudo, destacando em azul a área delimitada pela aquisição sísmica 3D.

Os perfis de poços selecionados como variáveis de entrada (*features*) foram os perfis de *Gamma-Ray* (GR), Sônico (DT) e a Resistividade Profunda (RES), que neste trabalho a resistividade será abreviada como RES para homogeneizar a nomenclatura já que os poços continham nomes diferentes para este mesmo tipo de medição. Outros perfis como densidade (RHOB) ou porosidade neutrão (NPHI), *spectral gamma-ray* (K, Th e U) que poderiam beneficiar o estudo, foram

desconsiderados. Com relação ao perfil de densidade, essa escolha se deve principalmente para evitar o enviesamento de uma boa correlação entre o COT estimado e a impedância acústica, que é o objetivo específico deste trabalho e será melhor discutido no Capítulo 10. Com relação aos demais perfis citados, não foi possível utilizá-los pois suas medições foram realizadas apenas em poucos poços do conjunto de dados selecionados, por exemplo, a curva de NPHI foi perfilado em apenas três dos oito poços selecionados, as curvas de spectral gamma-ray em apenas dois poços. Dessa forma, diante das limitações encontradas, foi dado prioridade para reunir a maior quantidade de amostras de COT medido mesmo que para isso, o número de variáveis tenha que ser reduzido.

Os perfis de GR, DT e RES são comumente empregados na estimativa de curvas de COT, seja por meio de métodos empíricos tradicionais ou, mais recentemente, para o treinamento de modelos de *machine learning*, cujo uso tem crescido significativamente nos últimos anos. A curva de GR é amplamente utilizada na caracterização litológica e na identificação de intervalos ricos em matéria orgânica, pois a radioatividade natural registrada pelo GR tende a ser mais elevada em folhelhos e rochas sedimentares com maior teor de matéria orgânica, principalmente devido à presença de elementos radioativos como o urânio, detectados pela ferramenta (Hassan *et al.*, 1976).

As curvas DT e RES, por sua vez, são componentes fundamentais do método Delta Log R, técnica consolidada e amplamente aplicada na indústria para estimativa do COT (Passey *et al.*, 1990). A influência da matéria orgânica sobre essas curvas se dá pelo fato de que rochas com elevado teor de matéria orgânica geralmente apresentam velocidades de propagação do som mais baixas, associadas à baixa densidade da matéria orgânica, resultando em tempos de trânsito (DT) mais elevados quando comparadas a folhelhos pobres em componentes orgânicos (Nixon, 1973; Meissner, 1978; Passey *et al.*, 1990; Schmoker & Hester, 1983). Quanto à resistividade, esta tende a apresentar valores mais elevados em rochas geradoras que não sofreram migração primária, onde os hidrocarbonetos permanecem confinados na matriz da rocha (Myers & Jenkyns, 1992; Passey *et al.*, 1990). Dessa forma, a resposta da curva de resistividade é influenciada diretamente pelo conteúdo de matéria orgânica presente, refletindo o teor de COT.

Em resumo, após a seleção dos dados que serão destinados como variáveis de entrada para treinar os algoritmos de *machine learning*, a Tabela 5.1 ilustra a quantidade de dados de COT disponíveis nos poços selecionados de fora da área de estudo, e o mesmo para os poços da área de estudo, para os quais serão estimadas as curvas de COT conforme o principal objetivo deste estudo.

Tabela 5.1. Quantidade de amostras de COT disponíveis nos poços de fora e de dentro da área de estudo, utilizadas no treinamento e validação dos modelos

Dados de fora da área de estudo	
Poços	Nº amostras
1-BRSA-1221-MA	18
2-BAC-1-MA	34
1-CI-1-MA	27
1-IZ-2-MA	54
1-GI-1-PA	53
1-PA-1-MA	103
1-OGX-110-MA	15
1-OGX-77-MA	8
Total de amostras	312

Dados da área de estudo	
Poços	Nº amostras
3-OGX-57-MA	x
1-OGX-22-MA	x
1-OGX-16-MA	x
2-CP-1-MA	59

5.4.2. Pré-condicionamento dos dados

A conclusão bem-sucedida de um fluxo de trabalho de projeto de ML depende fortemente do número de conjuntos de dados de entrada, da qualidade dos dados, de lacunas/valores nulos limitados nos dados, do comportamento estatístico das amostras de dados e da relação entre as variáveis de entrada e saída (Lopes, R. L., & Jorge, A. 2017, Hall, B., 2016). Por isso a importância de realizar, com rigor, análises do comportamento das variáveis e o pré-condicionamento das mesmas quando necessário.

Todos os perfis utilizados como dado de entrada para este estudo passaram por etapas de pré-condicionamento, realizados de forma cuidadosa para que dados erráticos não influenciassem de forma negativa a performance dos algoritmos de *machine learning* na estimativa das curvas de interesse. Portanto, foram realizados de forma manual, a junção (*merge*) das curvas adquiridas em diferentes fases do poço, além disso, foi necessário realizar interpolações em alguns intervalos com dados faltantes. Outra etapa fundamental foi a remoção de dados espúrios, principalmente se tratando de uma formação rochosa que comumente sofre com a integridade do poço. Os folhelhos, no geral apresetam desmoronamentos mais frequentes devido a percolação do fluido de perfuração, e por isso será notado nos perfis de caliper dos poços, uma curva não tão bem comportado.

Uma forma de não descartar dados mesmo em condições de um caliper indicando arrombamento da parede poço, é realizar uma análise comparativa entre os perfis dos poços em boas condições para os poços em condições de desmoronamento. Idealmente, não é indicado considerar dados ruidosos ou com possíveis problemas de medição, para um estudo como este onde será feita a predição de um dado baseado em um outro conjunto de dados. No entanto, a quantidade de dados utilizados é igualmente importante. Ter um conjunto de dados grande o suficiente para ser representativo da variabilidade do dado permite uma melhor consistência da predição do dado alvo, e por isso, é fundamental realizar uma avaliação criteriosa antes de reduzir o conjunto de dados, somente baseado pela curva de caliper neste caso.

Os perfis de GR, DT e RES após o pré-condicionamento podem ser observados na Figura 5.8 para o conjunto de poços de fora da área de estudo e na Figura 5.9 e para os poços da área de interesse, assim como as amostras de COT disponíveis. Vale ressaltar que, ao analisar os perfis dos poços, deve-se considerar que cada um deles é apresentado com diferentes escalas verticais.

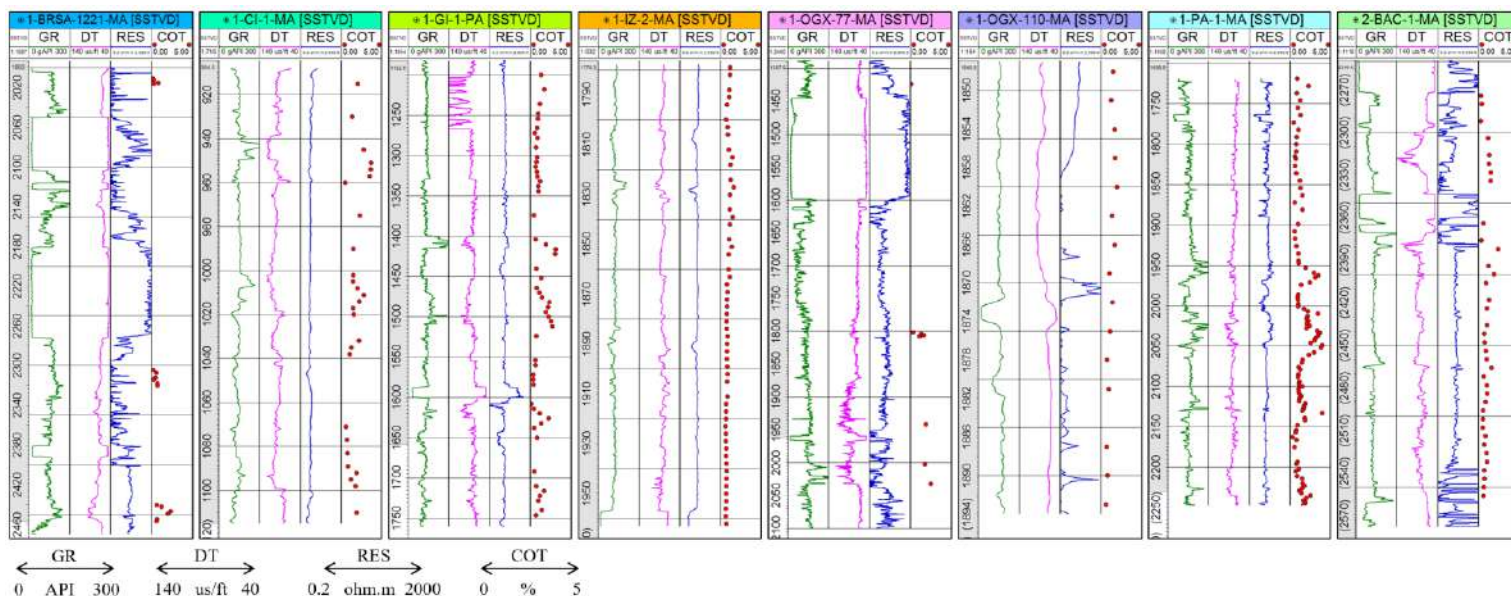


Figura 5.8. Perfis geofísicos de GR, DT e RES, após o pré-condicionamento, em poços externos à área de estudo, juntamente com as amostras de COT disponíveis.

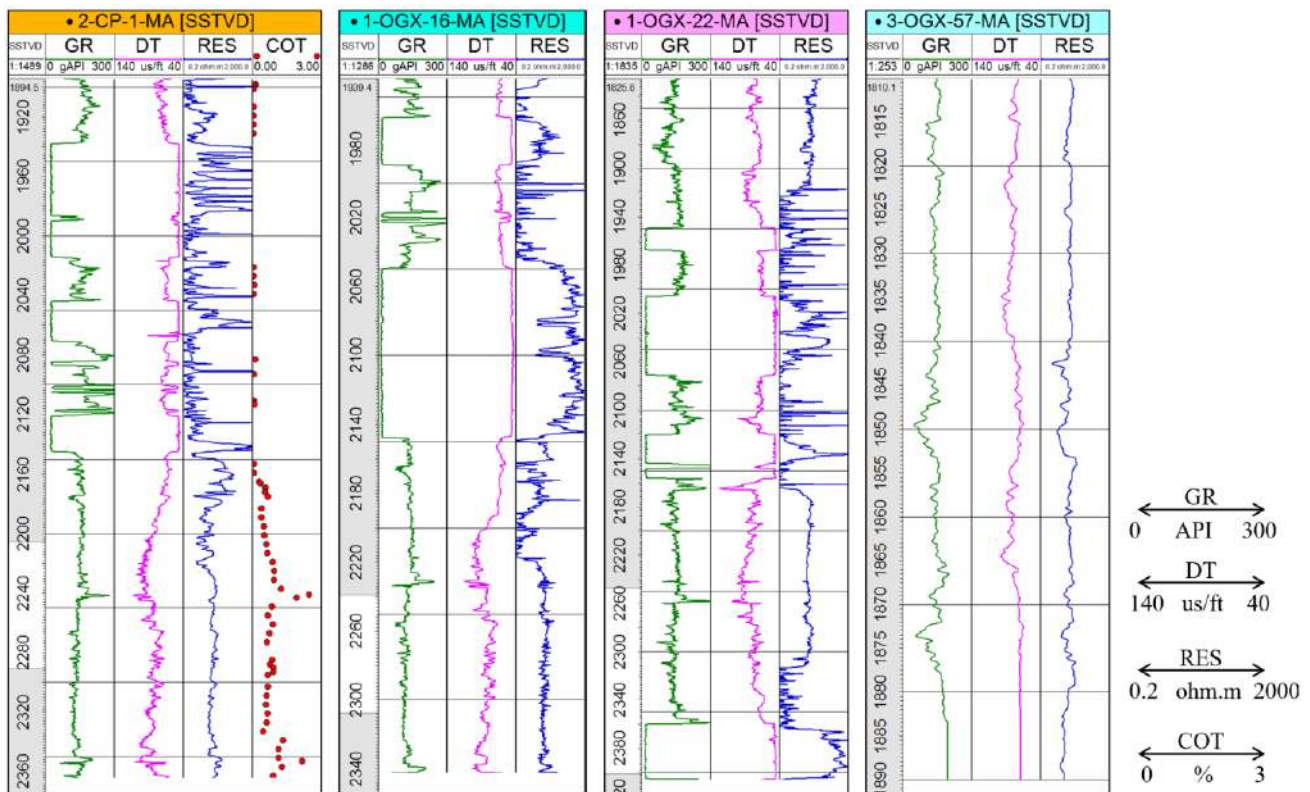


Figura 5.9. Perfis geofísicos de GR, DT e RES, após o pré-condicionamento, nos poços da área de estudo, com destaque para o poço 2-CP-1-MA, único com medições de COT.

5.4.3. Análise exploratória dos Dados

A análise exploratória dos dados também é uma etapa essencial a ser realizada antes de iniciar o treinamento dos algoritmos propriamente dito. Nesta etapa, busca-se identificar padrões, tendências e anomalias presentes nos dados selecionados, permitindo assim uma compreensão mais ampla sobre as distribuições dos dados e seus comportamentos, além de reforçar a efetividade do pré-condicionamento e transformações dos dados realizadas nas etapas anteriores ou ainda indicar possíveis melhorias a serem feitas.

Como descrito anteriormente, é fundamental ter um conjunto de dados volumoso e representativo, no entanto, considerar dados que fogem da distribuição geral do conjunto de dados pode ser bastante prejudicial para a predição do dado alvo. A análise dos histogramas dos dados de COT disponíveis para o treinamento, foi a maneira efetiva de mitigar esse problema e filtrar possíveis dados que reduziriam de forma significativa uma boa correlação entre o dado real e o estimado pelos métodos de *machine learning*.

No conjunto de dados disponibilizado pelo GIECAR, outros dois poços, 1-OGX-93-MA e o 1-OGX101-MA, também tiveram amostras de COT medidas, o que somaria 106 medições e totalizaria

477 amostras disponíveis do dado alvo. Entretanto, a análise exploratória revelou, por meio dos histogramas desses poços, que as médias, medianas e valores máximos observados estavam significativamente acima daqueles registrados nos demais poços (Figura 5.10). Essa discrepância na distribuição dos valores impactou diretamente o treinamento dos modelos e, conseqüentemente, na estimativa da curva de COT no poço de validação, como será demonstrado posteriormente no tópico dos resultados. Por isso a importância de analisar os dados não só de forma quantitativa mas também a qualidade como eles se apresentam, pois apesar de aumentar o volume de dados para o treinamento do modelo, pode-se nesse processo adicionar ruídos no lugar de informações relevantes.

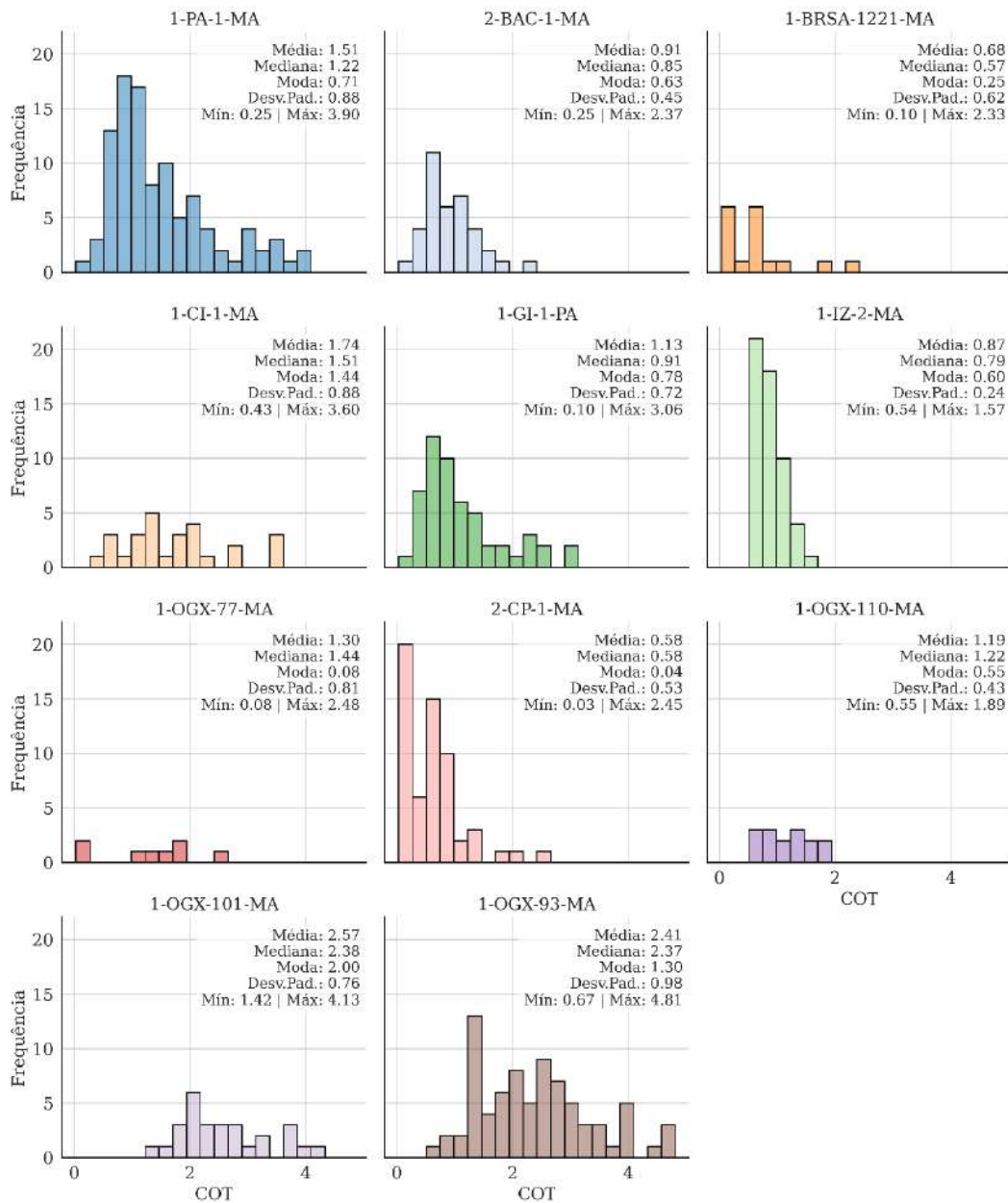


Figura 5.10. Histogramas das amostras de COT comparando a distribuição estatística para cada poço.

Diante do conjunto de dados selecionados e tratados é apresentado na Tabela 5.2 a descrição estatística básica, onde são observados valores esperados para os perfis de GR, DT e RES na área de estudo. Nota-se que os valores de COT variam dentro de um intervalo no qual a rocha geradora pode ser classificada como pobre a excelente potencial gerador (Peters & Cassa, 1994), mostrando que há uma diversidade de cenários a serem estudados pelos algoritmos de *machine learning* e uma quantidade total de 371 amostras de COT.

Tabela 5.2. Estatística descritiva dos perfis GR, DT e RES e das amostras de COT, após análise exploratória.

Unidade	COT %	DENS g/cm³	DT µs/ft	GR API	RES Ohm.m
Quantidade	371.00	30449.00	30426.00	30531.00	30531.00
Média	1.11	2.67	73.60	121.23	100.24
Desvio Padrão	0.76	0.21	17.94	57.48	327.44
Mínimo	0.03	1.70	42.98	10.97	0.07
25%	0.66	2.54	63.29	99.54	2.13
50%	0.91	2.64	74.45	133.68	7.22
75%	1.42	2.78	84.17	155.91	13.12
Máximo	3.90	3.14	183.51	409.09	2000.00

A Figura 5.11 exibe o comportamento geral dos poços destacando-se a a relação entre cada perfil e entre os perfis e as amostras de COT, onde não são observadas distribuições de dados na forma de *cluster* por poço ou mesmo seguindo alguma tendência. Também não nota-se uma alta correlação entre os perfis de poços, o que significa não haver uma redundância de informações no conjunto de dados. Com respeito a variável de saída do modeo, o COT, também não é observado uma alta correlação com as variáveis de entrada, GR, DT e RES, que é destacado pela Figura 5.12 onde observa-se uma não correlação ou uma correlação baixa tanto para uma relação linear (correlação de pearson) quanto para uma relação não linear (correlação de spearman).

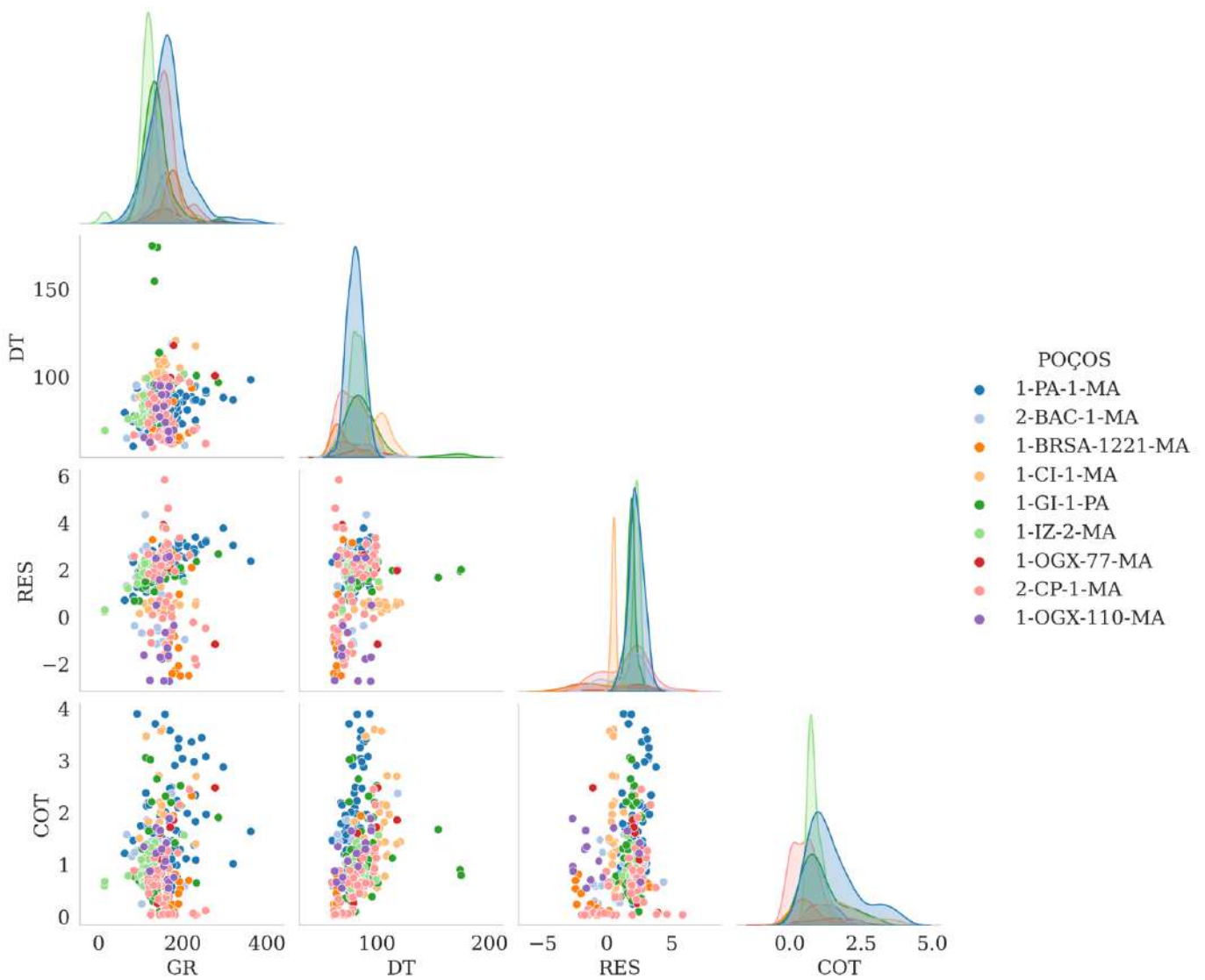


Figura 5.11. Análise conjunta de gráficos crossplot entre as variáveis e variável alvo coloridas por poços.

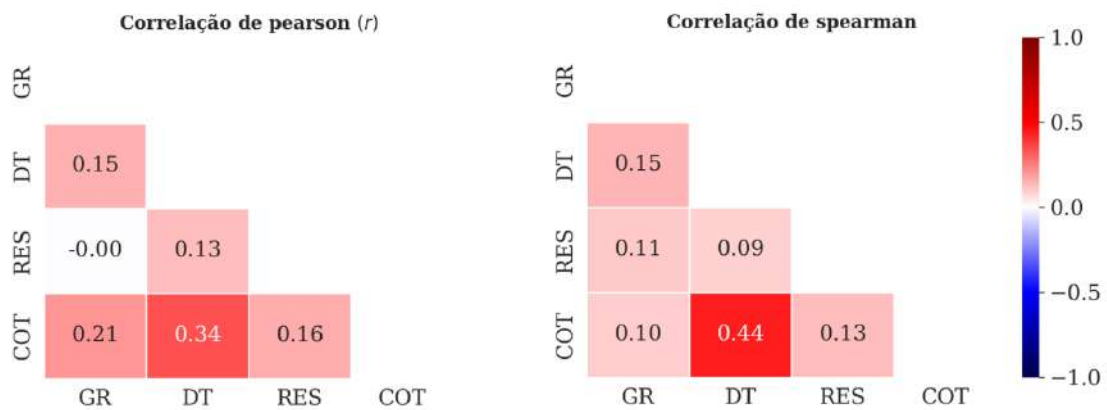


Figura 5.12. Correlação linear (Pearson) e não linear (Spearman) entre os perfis de poço (GR, DT e RES) e o COT, evidenciando baixa correlação direta.

Na Figura 5.13, é exibido gráficos com a relação entre as variáveis de entrada colorido pelo valor de COT, onde é possível evidenciar que a maior densidade de pontos com altos valores de COT, concentram-se em intervalos de altos valores de DT, RES e GR. Esse comportamento não segue uma regra direta, mas de forma global é possível observá-lo.

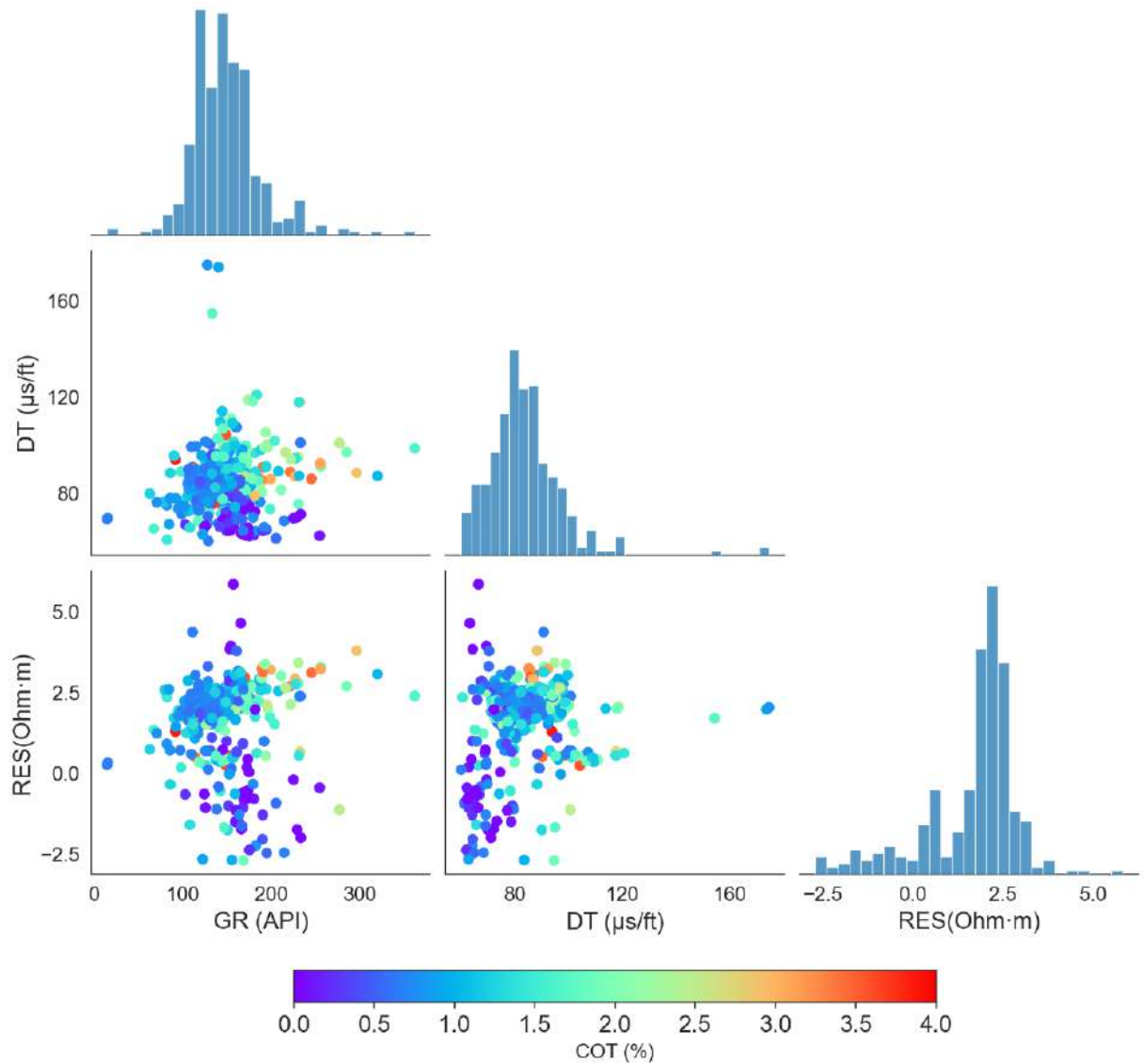


Figura 5.13. Relação entre variáveis de entrada (GR, DT e RES) coloridas de acordo com o valor de COT.

5.4.4. Transformação dos dados

Antes de qualquer transformação, o conjunto de dados analisado apresenta variáveis de entrada com escalas distintas, conforme ilustrado na Tabela 5.2. Dentre essas variáveis destacam-se os perfis de GR, DT e RES, cujos valores variam aproximadamente entre 10 e 410 API, 40 a 180 us/ft e 0,2 a 200 Ohm.m, respectivamente. Em aplicações de aprendizado de máquina, é comum a necessidade de padronização ou normalização das variáveis de entrada para garantir uma escala homogênea. Esse

escalonamento favorece uma convergência mais eficiente dos algoritmos, além de otimizar o tempo de processamento durante a modelagem (Binetti; Massarelli; Uricchio, 2024; Shirmard *et al.*, 2021).

Nesta etapa de transformação das variáveis, organiza-se os dados e aplica-se, inicialmente, transformações não lineares (logarítmica, inversa, etc) e em seguida o redimensionamentos (padronização e normalização), quando necessários. Tanto a padronização quanto a normalização possuem o mesmo objetivo: redimensionar as variáveis, colocando-as na mesma ordem de grandeza. Na padronização ocorre a transformação das variáveis originais de forma que todas tenham média zero e desvio padrão unitário, a ideia é modificar a escala de um atributo, sem contudo alterar sua distribuição. Na normalização, subtrai-se a média da variável de cada valor individual, o que centraliza os dados em torno de zero e divide-se esse resultado pela diferença entre o valor máximo e o mínimo da variável, o que garante que os valores fiquem em uma escala proporcional, confinando seus intervalos entre -1 e +1.

Ao contrário de métodos como MLP e SVM, algoritmos baseados em árvores de decisão, como Random Forest (RF), XGBoost e CatBoost, não requerem a padronização ou normalização das variáveis de entrada. Isso ocorre porque esses modelos operam por meio de divisões binárias sucessivas nas variáveis, segmentando os dados com base em limiares absolutos, em vez de cálculos que envolvam distâncias ou gradientes. Conforme destacado por Hastie *et al.* (2009) e Géron (2022), o processo de construção de árvores de decisão é invariante a transformações lineares das variáveis, de modo que alterações como padronização ou normalização não influenciam os pontos de divisão e, conseqüentemente, não afetam o resultado final do modelo.

Especificamente, no Random Forest e no XGBoost, o algoritmo seleciona, em cada nó, a variável e o ponto de corte que maximizam uma métrica de impureza (como Gini ou entropia) ou reduzem o erro quadrático médio, sem que a magnitude absoluta dos valores interfira na escolha. O CatBoost segue a mesma premissa, realizando codificações e divisões que são independentes da escala das variáveis (Prokhorenkova *et al.*, 2018). Dessa forma, esses métodos são robustos a diferentes unidades e escalas de medida, dispensando a necessidade de pré-processamentos como o *StandardScaler*.

Contudo, vale ressaltar que, em situações onde as variáveis apresentam distribuições altamente assimétricas ou com *outliers* extremos, transformações como o logaritmo podem ser aplicadas para melhorar a estabilidade do modelo e facilitar a interpretação dos resultados, mas não por exigência intrínseca do algoritmo em relação à escala dos dados. Por isso, como parte da metodologia adotada neste trabalho, transformação logarítmica foi aplicada ao perfil de resistividade em função de sua

distribuição assimétrica e do fato de seus valores variarem em múltiplas ordens de magnitude, característica típica de variáveis que apresentam comportamento aproximadamente logarítmico. Matematicamente essa transformação pode ser escrita como:

$$RES_{ln} = \ln(RES + \epsilon) \quad (5.16)$$

Onde RES_{ln} é o valor transformado da resistividade, RES é o valor original da resistividade e ϵ é uma pequena constante (por exemplo, 1×10^{-6}) usada para evitar problemas matemáticos com valores iguais a zero.

A distribuição do dado de resistividade antes e após a transformação logarítmica é mostrado na Figura 5.14.

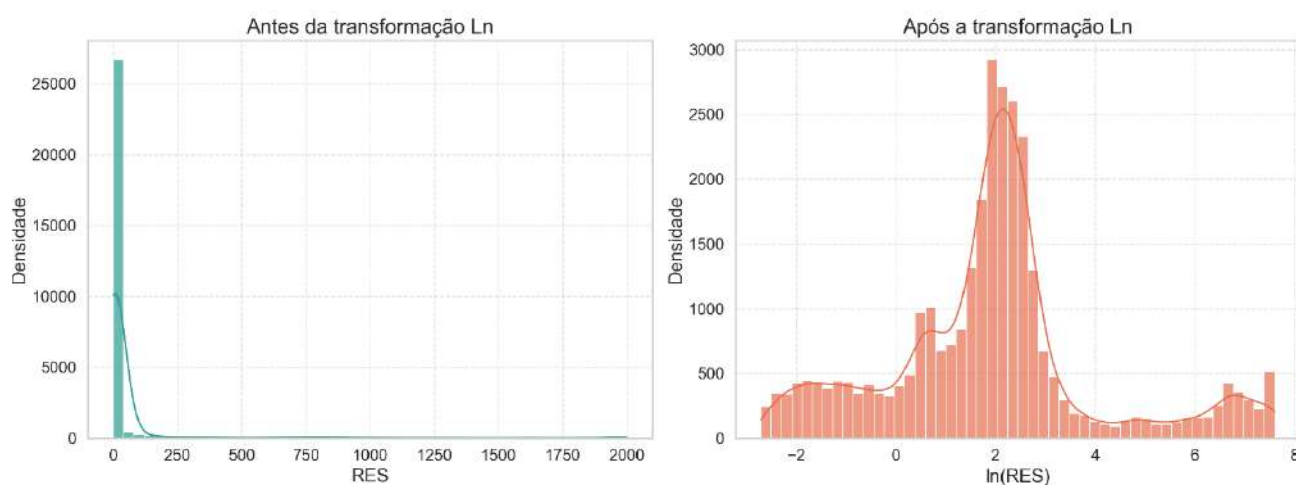


Figura 5.14. Distribuição dos valores do perfil de resistividade antes e após a transformação logarítmica aplicada.

Diversos métodos de *machine learning* são reconhecidamente sensíveis à escala das variáveis de entrada, com destaque para o Multi-Layer Perceptron e as Support Vector Machines. Segundo Hastie *et al.* (2009), algoritmos como SVM dependem significativamente das escalas das variáveis, pois as decisões sobre a margem ótima ou hiperplano decisorio são diretamente influenciadas pela magnitude das variáveis utilizadas. Da mesma forma, Haykin (2009) e Goodfellow *et al.* (2016) afirmam que redes neurais, incluindo o MLP, têm seu treinamento fortemente afetado pela variação na escala das variáveis, influenciando diretamente na velocidade e estabilidade da convergência durante o processo de aprendizado, especialmente em algoritmos baseados em gradiente, como a retropropagação de erros (*backpropagation*).

Para solucionar essa limitação, recomenda-se a aplicação da padronização das variáveis (*standardization*) (Géron, 2022; Bishop, 2006). Neste trabalho, foi utilizada o *StardScalar* da

biblioteca *Scikit-learn*, que é uma biblioteca de código aberto em Python. Matematicamente, esta transformação pode ser descrita pela seguinte equação:

$$z = \frac{x - \mu_{train}}{\sigma_{train}} \quad (5.17)$$

Onde, x representa o valor da variável original, μ_{train} é a média da variável calculada exclusivamente no conjunto de treino e σ_{train} é o desvio padrão da mesma variável, também calculado apenas com os dados de treino.

A padronização realizada nos dados podem ser avaliada através da Figura 5.15.

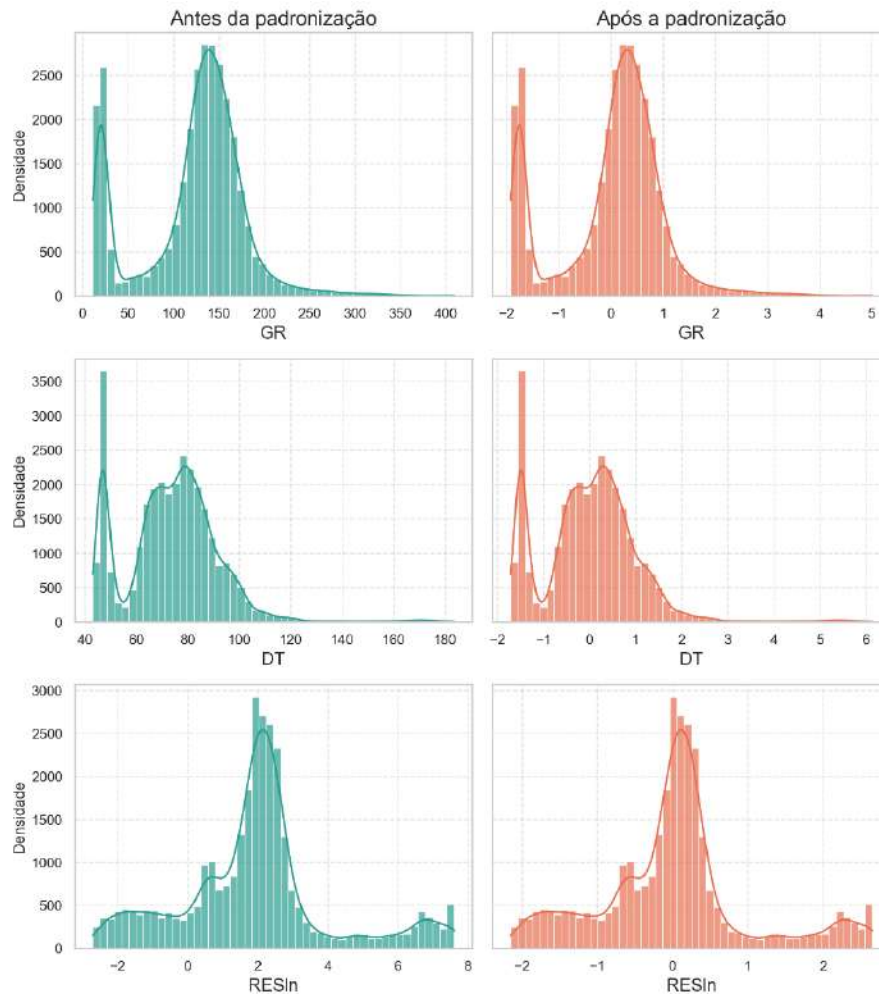


Figura 5.15. Distribuição dos dados após padronização com *StandardScaler*, demonstrando a homogeneização das escalas entre variáveis.

Esta técnica garante que todas as variáveis possuam a mesma relevância inicial no processo de aprendizado, evitando que variáveis com valores mais elevados dominem injustamente o processo de ajuste dos modelos. Como resultado, o treinamento do MLP e do SVM torna-se mais eficiente, estável

e robusto, levando à obtenção de modelos com melhor capacidade preditiva (James *et al.*, 2013; Raschka & Mirjalili, 2022).

5.4.5. Separação dos Dados

Em geral, em estudos que envolvem *machine learning*, uma parcela dos dados disponíveis é destinada ao conjunto de treinamento, no qual os modelos são ajustados, enquanto outra parcela é reservada para teste, não sendo utilizada durante o treinamento. Essa divisão permite avaliar o desempenho do modelo na capacidade de estimar a variável alvo a partir das variáveis de entrada e das parametrizações adotadas. No presente trabalho, devido à limitação do volume de dados, foi adotada a proporção de 80% das amostras para treinamento e 20% para teste dos modelos. Além disso, o poço pertencente à área de estudo foi utilizado como conjunto de dados para a validação independente, estando fora dessa divisão percentual. A Tabela 5.3 apresenta o resumo da quantidade total de amostras e a divisão exata dos conjuntos de dados utilizados.

Tabela 5.3. Distribuição das amostras nos conjuntos de treinamento, teste e validação, indicando a quantidade total e a proporção em cada conjunto.

Separação dos dados			
Classificação	Quantidade	Porcentagem treino + teste	Porcentagem Total
Treino	249	80%	67%
Teste	63	20%	17%
Validação	59	Poço: 2-CP-1-MA	16%

5.4.6. Treinamento dos modelos

Os modelos de machine learning foram implementados conforme as parametrizações específicas de cada método, conforme fundamentado previamente na seção teórica e sumarizado pela Figura 5.6. Para os algoritmos Random Forest e Support Vector Machine, foram realizados diversos experimentos visando a otimização de seus hiperparâmetros relevantes. No caso do RF, testaram-se variações no número de árvores (`n_estimators`) e no parâmetro de inicialização aleatória (`random_state`), buscando melhorar a estabilidade e o desempenho do modelo. Para o SVM, foram exploradas diferentes configurações do parâmetro de kernel, especificamente os kernels RBF e polinomial. Já para os métodos como Multi-layer Perceptron, XGBoost e CatBoost tiveram seus hiperparâmetros otimizados a partir do método *GridSearchCV*, que é uma técnica amplamente

utilizada em machine learning com o objetivo principal de otimizar os hiperparâmetros de um modelo de forma sistemática e exaustiva (Pedregosa *et al.*, 2011).

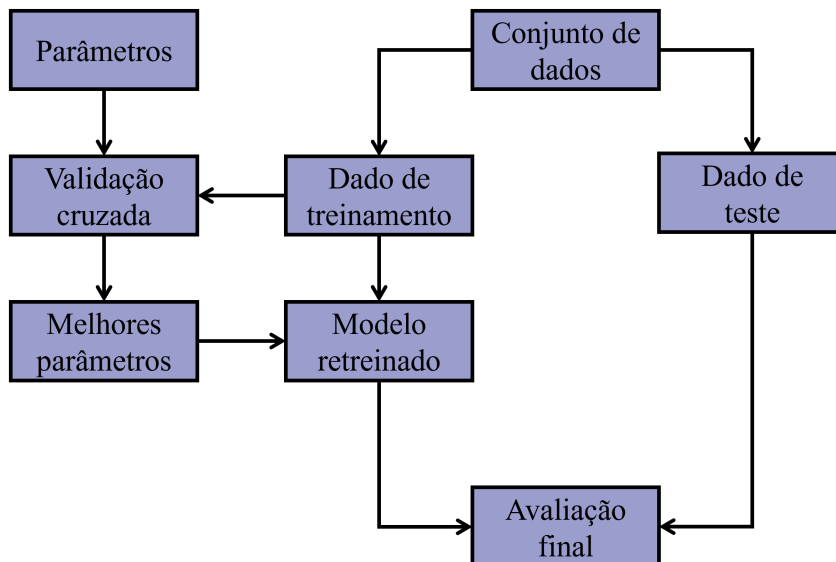


Figura 5.16. Esquema do método de validação cruzada (cross-validation) para otimização dos hiperparâmetros (Fonte: https://scikit-learn.org/stable/modules/cross_validation.html)

O *GridSearchCV* serve para encontrar a combinação ideal de hiperparâmetros para um determinado modelo de machine learning, visando maximizar sua performance (Bergstra & Bengio, 2012). Considerando que os modelos podem ter diversos parâmetros ajustáveis (como número de árvores em Random Forest, taxa de aprendizado em Gradient Boosting, profundidade da rede neural no MLP, etc.), testá-los manualmente é inviável e pode levar a escolhas subótimas. Assim, o *GridSearch* automatiza esse processo, realizando uma busca exaustiva sobre todas as combinações especificadas pelo usuário. Cada combinação de parâmetros é avaliada por meio de uma estratégia de validação cruzada (*Cross-validation*) (Figura 5.16). A combinação que apresentar o melhor desempenho, segundo uma métrica definida previamente (como RMSE, MAE ou R^2 para regressão; acurácia, F1-Score, ROC-AUC para classificação), será considerada a melhor (Hastie *et al.*, 2009).

A validação cruzada é uma técnica de avaliação amplamente usada em *machine learning*, que permite estimar a capacidade de generalização (desempenho) de um modelo sobre dados novos, ou seja, que não participaram do treinamento. O método mais comum é chamado *k-fold Cross Validation*, que segue estes passos (James *et al.*, 2013) (Figura 5.17):

1. Os dados são divididos aleatoriamente em k subconjuntos (folds) aproximadamente iguais.

2. O modelo é treinado usando k-1 folds como dados de treinamento e validado no fold restante. Esse processo se repete k vezes, de forma que cada fold seja usado exatamente uma vez como conjunto de validação.
3. Ao final, é calculada a média das métricas de desempenho obtidas em cada rodada.

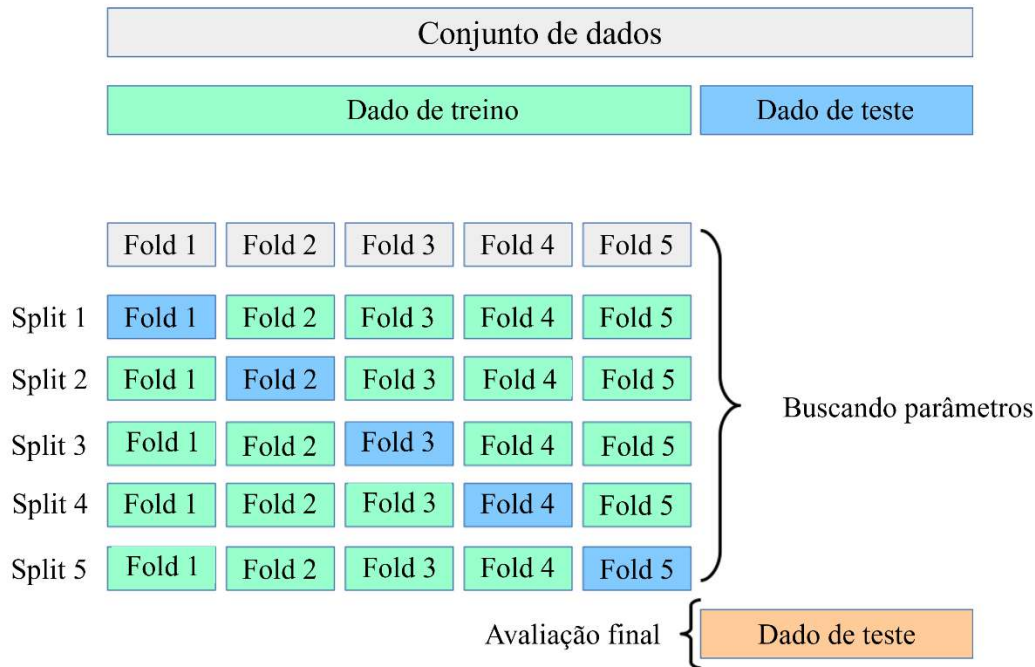


Figura 5.17. Exemplo de validação cruzada K-Fold aplicada a um conjunto de dados, no qual este é dividido em subconjuntos destinados ao treinamento, à validação e ao teste do modelo. (Fonte: https://scikit-learn.org/stable/modules/cross_validation.html)

Especificamente quanto ao MLP, dado que as redes neurais possuem uma sensibilidade intrínseca à inicialização dos pesos sinápticos, utilizou-se uma estratégia dupla visando garantir a robustez e o desempenho do modelo: a otimização sistemática dos hiperparâmetros por meio de GridSearchCV e a aplicação de técnicas de *ensemble* através do treinamento múltiplo com diferentes inicializações de pesos (Goodfellow *et al.*, 2016). Inicialmente, foi aplicado o método GridSearchCV, para determinar o valor ótimo do hiperparâmetro de regularização alfa (α). Esse hiperparâmetro controla a magnitude da penalização aplicada aos pesos da rede, influenciando diretamente na prevenção de overfitting. O conjunto de valores testados foi: $\alpha = [10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1]$. A métrica utilizada na seleção do melhor α foi o erro quadrático médio negativo (neg_mean_squared_error), obtido por meio de validação cruzada com cinco *folders* (5-fold cross-validation).

Além da otimização de hiperparâmetros, foi aplicada uma técnica de *ensemble*, conhecida como *bagging* (Breiman, 1996), para reduzir a variância associada à inicialização aleatória dos pesos na rede MLP. Nesse método, 300 redes neurais individuais foram treinadas com exatamente a mesma configuração arquitetural, porém com inicializações diferentes dos pesos sinápticos, definidas pelo parâmetro `random_state` variando de 0 até 299. Cada uma dessas redes foi treinada com *early stopping* ativado para evitar *overfitting* e garantir que cada rede individual atingisse um desempenho satisfatório. Após o treinamento individual dessas redes, suas previsões foram combinadas por meio da média aritmética simples. Este método de *ensemble* possibilita a redução significativa da instabilidade e da sensibilidade às condições iniciais da rede, proporcionando maior generalização às previsões resultantes.

Os parâmetros utilizados para cada método estão descritas nas Tabela 5.4, Tabela 5.5, Tabela 5.6, Tabela 5.7, Tabela 5.8 a seguir.

Tabela 5.4. Parâmetros utilizados para treinar o algoritmo Random Forest.

Parâmetros - Random Forest	Valores
Number of tree	100
Maximum depth	None (ilimitada)
Número de features	Padrão(auto ou todas)
Random state	45

Tabela 5.5. Otimização de hiperparâmetros utilizados para treinar o algoritmo XGBoost e a demonstração dos melhores parâmetros encontrados.

Parâmetros - XGBoost	Valores da otimização	Melhores Parâmetros
Loss function	Least square	Least square
Number of boosting stages	[50, 100, 500, 1000, 1500, 2000]	2000
<code>colsample_bytree</code> e <code>colsample_bylevel</code>	[0.6, 0.8, 1.0]	0.8
Maximum depth	None (ilimitada)	None
Learning rate	[0.0001, 0.001, 0.01, 0.1, 0.2, 0.3]	0.001
Gamma	1	1
Subsample	[0.6, 0.8, 0.85]	0.6
Lambda	[0.2, 0.4, 0.6]	0.2

Tabela 5.6. Parâmetros utilizados para a otimização e seus respectivos valores que melhor ajustaram o modelo.

Parâmetros - CatBoost	Valores	Melhores Parâmetros
Iterations	[100, 150, 200, 500, 1000]	500
Learning rate	[0.001, 0.01, 0.1]	0,01
Depth	[2, 4, 6, 8]	6
<code>l2_leaf_reg</code>	[0.2, 0.5, 1, 3]	1

Tabela 5.7. Parâmetros utilizados no ajuste do modelo Multi-Layer Perceptron.

Parâmetros - Multi-Layer Perceptron	Valores
Number of Inputs	3
Number of Outputs	1
Number of Neurons	30
Training algorithm	Adam
Alpha	1
Activation function	RELU
Maximum number of iterations	2000

Tabela 5.8. Configuração utilizadas no treinamento do modelo Support Vector Machine.

Parâmetros - Support Vector Machine	Valores
Kernel	RBF (radial basis function)
Penalty parameter C	1
Gamma	scale (1 / (n_features * X.var()))

5.4.7. Estimativa no conjunto de dados e *ensemble* dos modelos selecionados

Após o treinamento de cada um dos métodos de *machine learning* testados, o desempenho foi avaliado confrontando os dados estimados com os dados medidos e também as relações entre os dados estimados com o conjunto de dados teste, com o objetivo de ajustar os parâmetros e selecionar os modelos mais ajustados para a estimativa final nos poços da área de estudo. O critério principal de seleção consistiu na escolha de modelos que apresentassem coeficiente de correlação superior a 0,70 e coeficiente de determinação acima de 0,50 no conjunto completo de dados (treinamento mais teste). Para o conjunto de teste, foram considerados satisfatórios os modelos que atingissem coeficiente de correlação superior a 0,60 e coeficiente de determinação acima de 0,40. Além disso, foram analisadas métricas estatísticas complementares, como RMSE, MAE e MedAE. Os modelos que atenderam a todos os critérios estabelecidos foram posteriormente combinados por meio de uma média aritmética simples.

5.4.8. Pós-processamento dos dados preditos

Após a exaustiva otimização das parametrizações dos modelos de *machine learning* visando alcançar o melhor desempenho possível, foi incorporada, como etapa final da metodologia, uma correção das curvas estimadas pelo modelo selecionado. Esse pós-processamento constituiu-se em uma solução simples, direta e eficiente para corrigir a tendência de superestimativa observada nas predições.

Nesse contexto, aplicou-se o método de regressão linear como ferramenta de correção, com o intuito de tornar as curvas estimadas mais representativas não apenas em relação ao comportamento geral dos dados, como já previsto pelos modelos de *machine learning*, mas também em valores absolutos medidos. Essa correção não afeta a correlação entre variáveis, uma vez que a dispersão relativa dos dados é mantida, todos os pontos são deslocados de forma uniforme, preservando sua disposição relativa. Como a covariância e os desvios-padrão permanecem proporcionais, coeficientes como o de correlação de Pearson permanecem inalterados após a correção.

Erros sistemáticos desse tipo podem manifestar-se como viés (super ou subestimação) e/ou distorções de escala nas predições, resultantes de limitações do modelo, presença de ruído nos dados ou diferenças estatísticas entre o conjunto de treinamento e as condições reais de aplicação (James *et al.*, 2013). Para corrigir esses desvios, o ajuste por regressão linear é amplamente empregado, pois utiliza as observações reais disponíveis para calibrar as estimativas produzidas pelo modelo, assegurando maior correspondência aos valores medidos e, conseqüentemente, maior confiabilidade.

Matematicamente, o método estabelece uma relação linear entre as predições do modelo, \hat{y} , e as observações reais, y , considerando as amostras nas quais ambas as medidas estão disponíveis:

$$y = a \cdot \hat{y} + b + \varepsilon \quad (5.18)$$

Em que y representa o valor real medido da propriedade (neste estudo, as amostras de COT), \hat{y} é o valor predito pelo modelo ML, a é o coeficiente angular (inclinação) da regressão, representando ajuste de escala, b é o coeficiente linear (intercepto), representando o viés ou deslocamento constante e ε é o termo residual ou erro aleatório, assumido com média zero.

Os coeficientes a e b são estimados por meio do método dos mínimos quadrados, que minimiza a soma dos quadrados das diferenças entre os valores reais e os valores ajustados pela regressão:

$$\min_{a,b} \sum_{i=1}^n (y_i - a\hat{y}_i - b)^2 \quad (5.19)$$

A aplicação prática do ajuste por regressão linear envolve (i) selecionar os pontos onde as predições do modelo ML e as observações reais coexistem, (ii) calcular os coeficientes a e b que melhor ajustam a relação linear entre as duas variáveis, (iii) aplicar a equação calibrada para ajustar todas as predições do modelo, incluindo aquelas em pontos sem observações reais, por meio da transformação:

$$\hat{y}_{\text{corrigido}} = a \cdot \hat{y} + b \quad (5.20)$$

Após a correção, recomenda-se realizar uma nova avaliação quantitativa por meio de métricas como MAE, RMSE, MedAE e MSE, a fim de verificar a redução desses valores e confirmar a efetividade do procedimento.

5.5. RESULTADOS

5.5.1 Treinamento dos modelos *machine learning*

Os resultados que serão apresentados a seguir, foram baseados em um conjunto de oito poços perfurados fora da área de interesse, distribuídos pela Bacia do Parnaíba, dos quais, foram considerados os perfis de *gamma ray*, sônico e resistividade profunda, além de 371 amostras com medições de COT (alvo a ser estimado). Do total de 371 amostras de COT disponíveis, 80% foram empregadas no treinamento dos modelos de *machine learning*, a partir de cinco técnicas de regressão supervisionada: três baseadas em árvores de decisão (Random Forest, XGBoost e CatBoost), uma rede neural do tipo *Multi-Layer Perceptron* e a última na técnica de classificador de margem máxima, o *Support Vector Machine* configurado para regressão.

De acordo com a abordagem adotada para a definição dos hiperparâmetros de cada método, conforme detalhado nas tabelas apresentadas no tópico anterior, foram obtidos os coeficientes de correlação (R) e de determinação (R²) entre os valores estimados pelos modelos e os valores medidos de COT. Esses resultados estão sintetizados na Figura 5.18, que mostra os coeficientes individualmente por poço para cada método utilizado, assim como média dos poços por método aplicado.

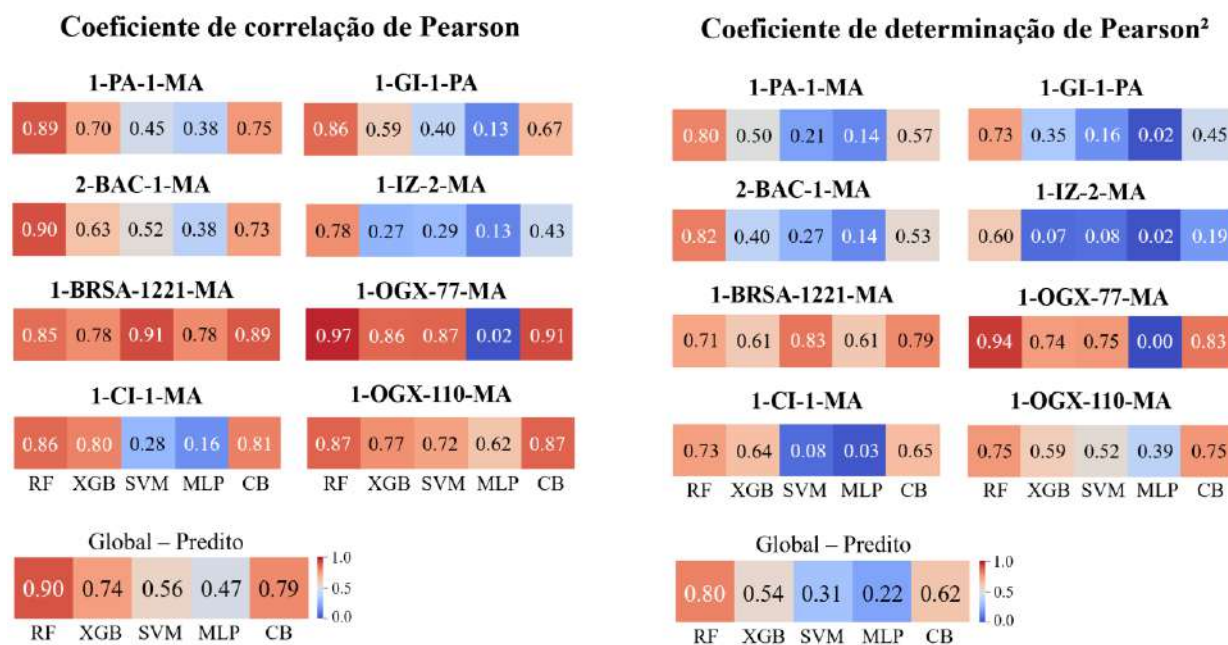


Figura 5.18. Coeficientes de correlação e determinação por poço e por método.

Essas duas métricas, funcionaram neste estudo, como ferramentas principais para a determinação dos modelos que foram selecionados para definir um regressor agregado que combina os modelos de melhor desempenho para o *ensemble learning* que retorna a estimativa média. Os modelos que resultaram um coeficiente de correlação global acima de 0.7 e um coeficiente de determinação global acima de 0.5 foram selecionado, sendo eles o Random Forest, XGBoost e o Catboost. Os gráficos mostrando a relação entre os dados medidos e estimados de COT por método são exibidos na Figura 5.19, onde fica evidente uma dispersão maior dos dados para os métodos SVM e para o MLP, e boas correlações para o *ensemble* dos três métodos selecionados. Apesar dos valores de correlação para o *ensemble* ser um pouco menor do que o método RF, o objetivo desse estudo é buscar generalizar um método que terá uma boa performance de forma consistente, um vez que, os poços alvo da estimativa não possuem amostras medidas de COT para confrontar os dados estimados com dados reais, com exceção do poço 2-CP-1-MS, que tem a função de validar o método adotado.

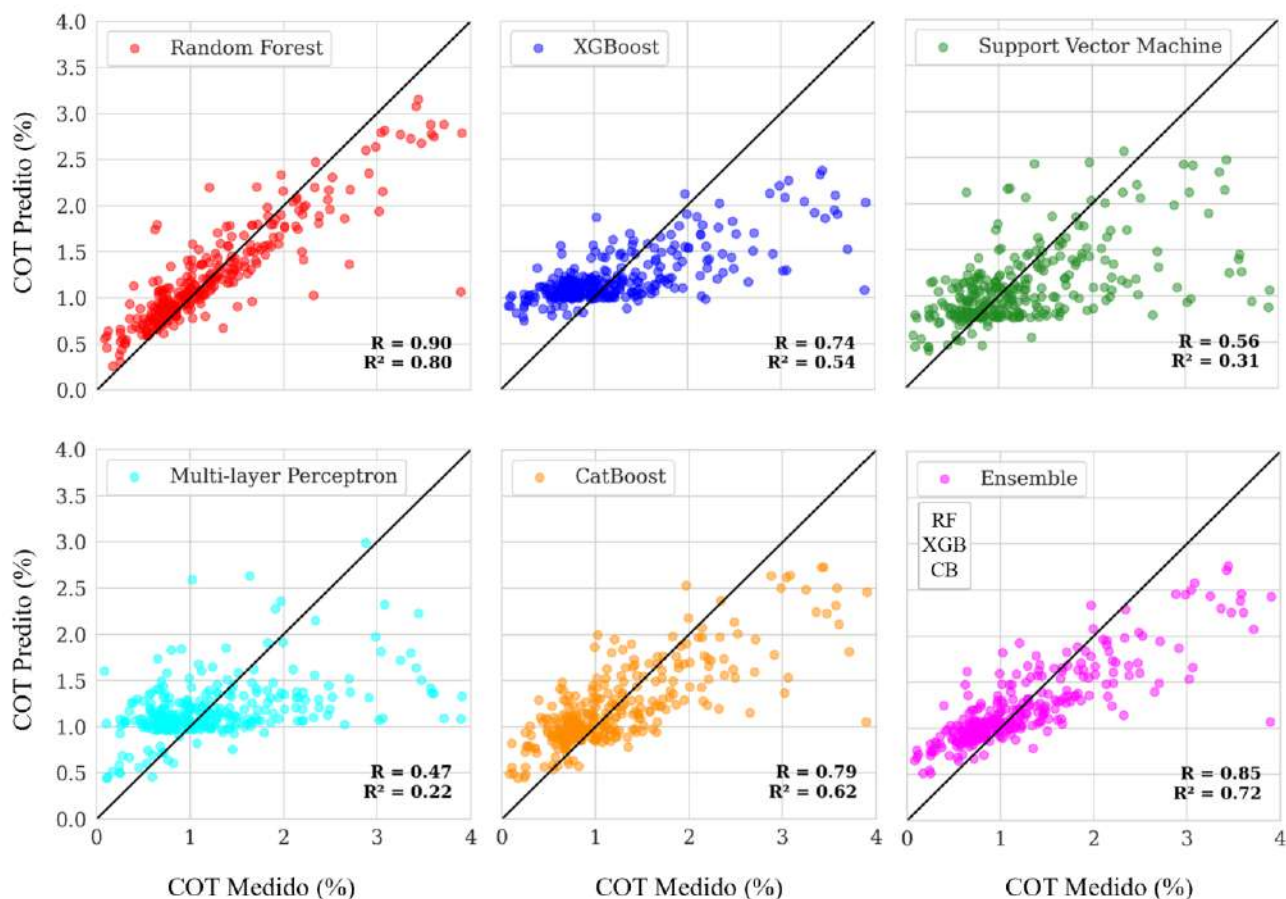


Figura 5.19. Crossplot correspondente as correlações entre o COT predito e medido por cada método de *machine learning* e pelo *ensemble* dos métodos de melhor performance. Os gráficos representam o conjunto de dados completo (teste e treino).

A Figura 5.20 mostra dois poços utilizados no treinamento dos modelos, onde são exibidas as curvas de COT estimadas por cada método. É possível analisar por meio desse exemplo, que nenhum dos métodos foram capazes de estimar valores extremos de COT, sejam eles valores baixos, próximo de 0% ou valores muito alto próximo de 4%. Além disso, é notável que os métodos como RF, XGB e CB de maneira geral são os que melhor acompanham a tendência dos dados medidos, enquanto que o MLP mostra comportamento inversos em alguns intervalos, mesmo se tratando de dois poços que obtiveram-se os valores de correlação mais razoáveis de todos os métodos.

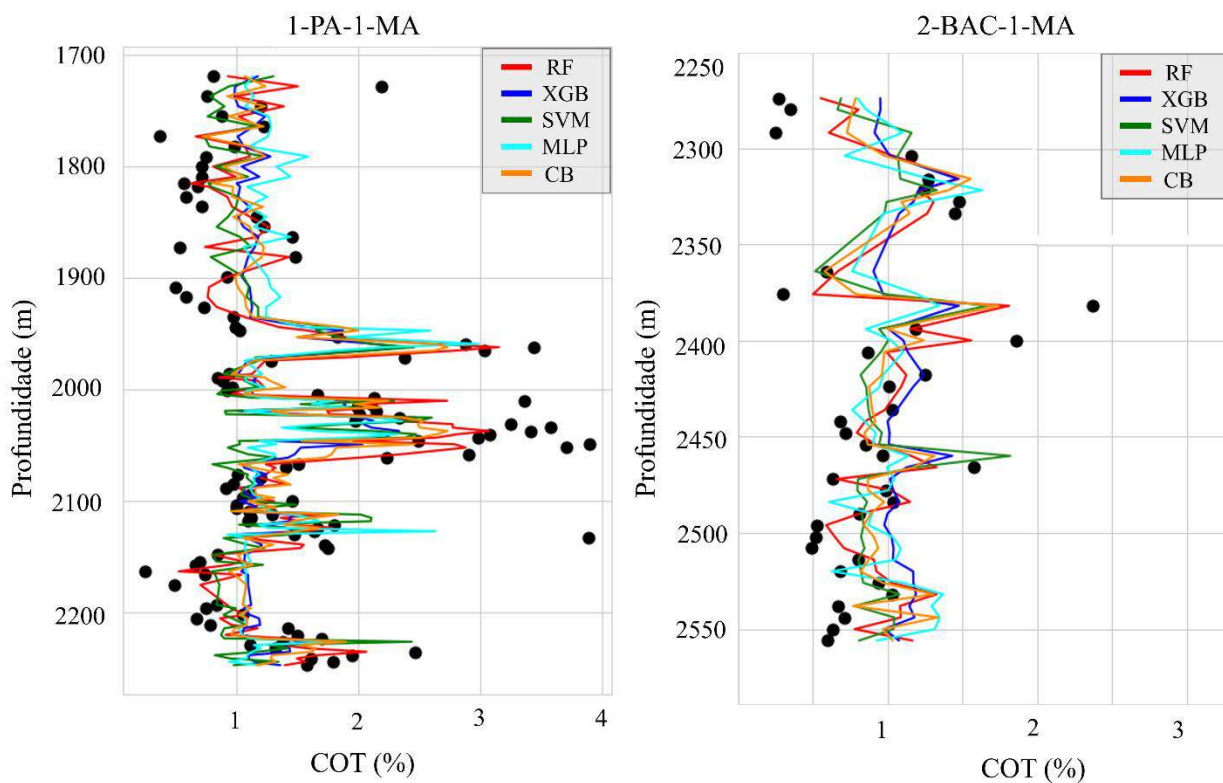


Figura 5.20. Exemplo das curvas estimadas por métodos em dois poços do conjunto de dados de treinamento.

Outras métricas estatísticas podem ser analisadas na Figura 5.21, considerando apenas os 20% dos dados que foram utilizados como dado teste. Como pode-se observar os métodos de RF, XGB e CB resultaram nos maiores valores de R e R^2 , dentro do critério de seleção estabelecido, acima de 0,4 para o R^2 e acima de 0.6 para o R. Com relação aos erros, sejam eles o MAE, MedAE e o RMSE, nota-se valores muito próximo entre os métodos, dos quais os RF e CB foram os que apresentaram menores erros de forma geral.

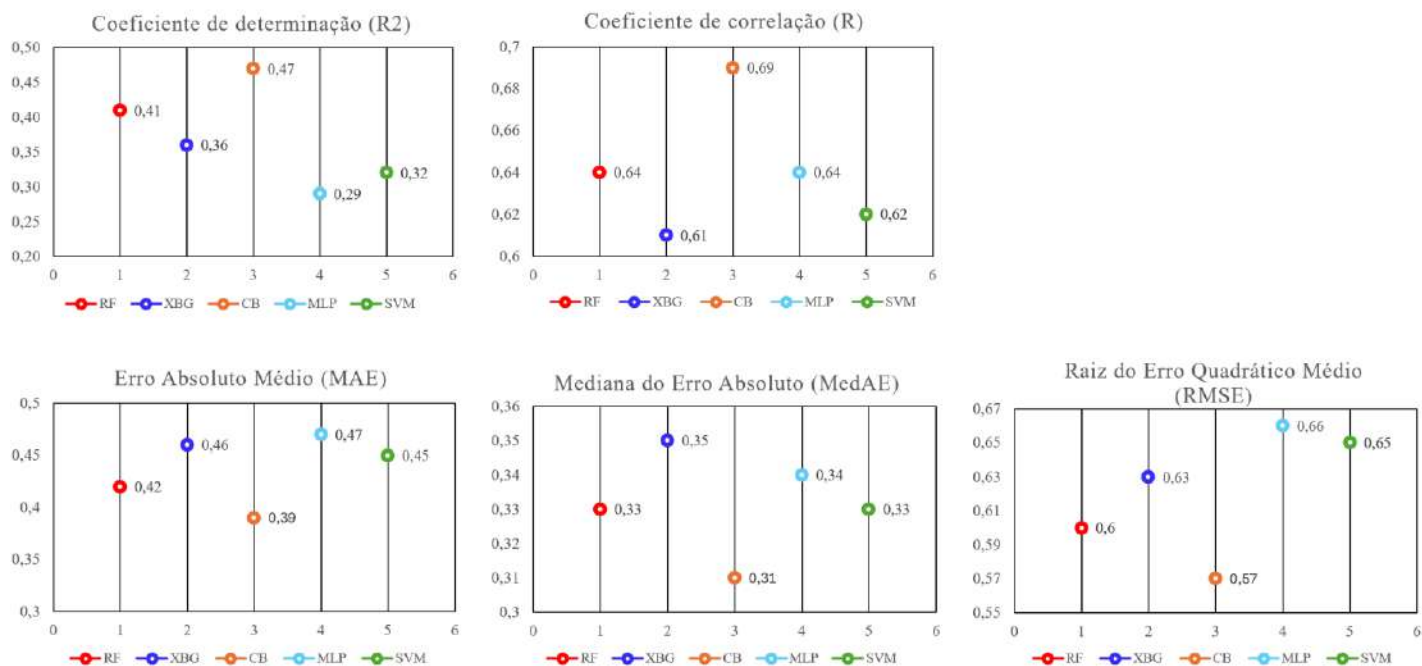


Figura 5.21. Métricas estatísticas correspondente ao conjunto de dados de teste para avaliação da performance de cada método utilizado.

Como etapa complementar da avaliação dos modelos selecionados, foi realizada uma análise gráfica aplicada ao conjunto de teste, conforme apresentado nas Figura 5.22, Figura 5.23 e Figura 5.24, correspondentes aos métodos Random Forest, XGBoost e CatBoost, respectivamente. Para cada modelo, foram construídos três tipos de gráficos:

- (i) crossplot entre os valores medidos e os valores preditos de COT;
- (ii) gráfico dos resíduos em função dos valores preditos, com o objetivo de identificar padrões sistemáticos de erro ao longo da faixa de estimativas;
- (iii) distribuição dos resíduos representada por meio de uma curva de densidade ajustada a uma distribuição gaussiana, com marcações das posições da média e da mediana.

O gráfico de densidade (iii) permite avaliar a simetria, a dispersão e a tendência central dos resíduos, sendo desejável uma distribuição aproximadamente normal centrada em zero, o que indica ausência de viés sistemático nas predições.

Em todos os casos, observou-se a presença de um ponto extremo, com valor medido próximo de 4% de COT, que foi sistematicamente subestimado pelos três modelos, constituindo-se no maior erro individual identificado.

Os modelos RF e CB apresentaram melhor desempenho, com resíduos concentrados principalmente na faixa de -0.5 a 0.5, e distribuições residuais simétricas, cujas médias e medianas se

encontram muito próximas de zero. Esse comportamento sugere boa acurácia e precisão, com leve predominância de superestimações.

Por outro lado, o modelo XGBoost apresentou maior dispersão dos resíduos, com erros superiores a 0.5 em diversos pontos. Embora a média e a mediana da distribuição residual também estejam próximas de zero, observou-se um afastamento ligeiramente maior entre elas, indicando uma assimetria mais acentuada e, conseqüentemente, maior incerteza associada às suas predições.

Random Forest

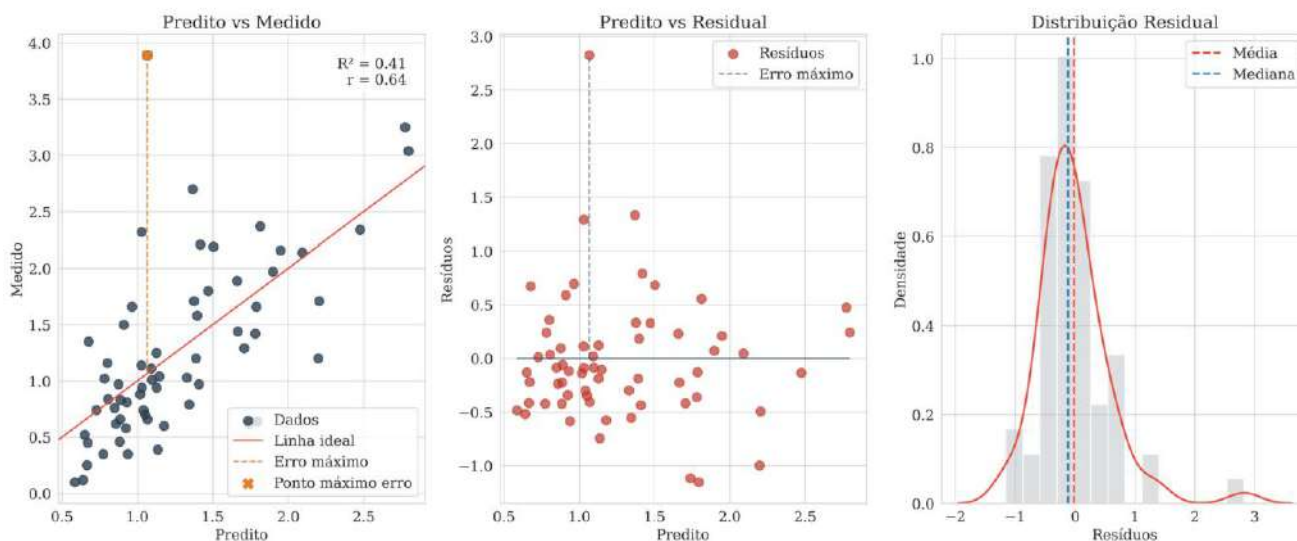


Figura 5.22. Avaliação da performance para o método *Random Forest* para o conjunto de dados teste, onde é observado os coeficientes de correlação evidenciando o ponto de maior erro, os valores residuais por amostra e a distribuição de densidade do residual.

XGBoost

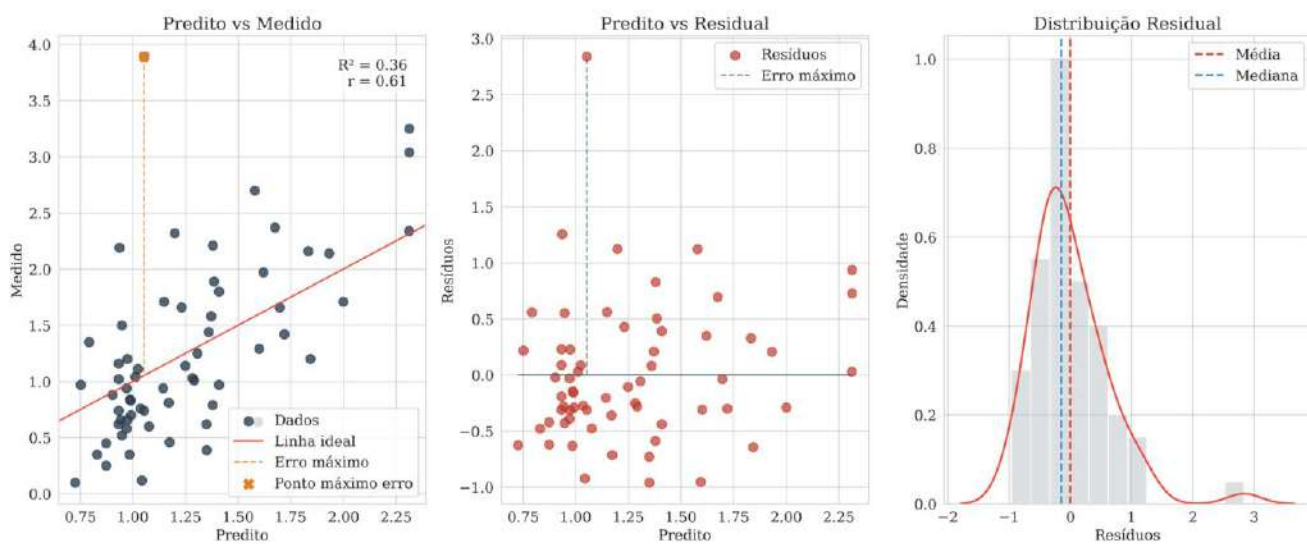


Figura 5.23. Avaliação da performance para o método *XGBoost* para o conjunto de dados teste, onde é observado os coeficientes de correlação evidenciando o ponto de maior erro, os valores residuais por amostra e a distribuição de densidade do residual.

CatBoost

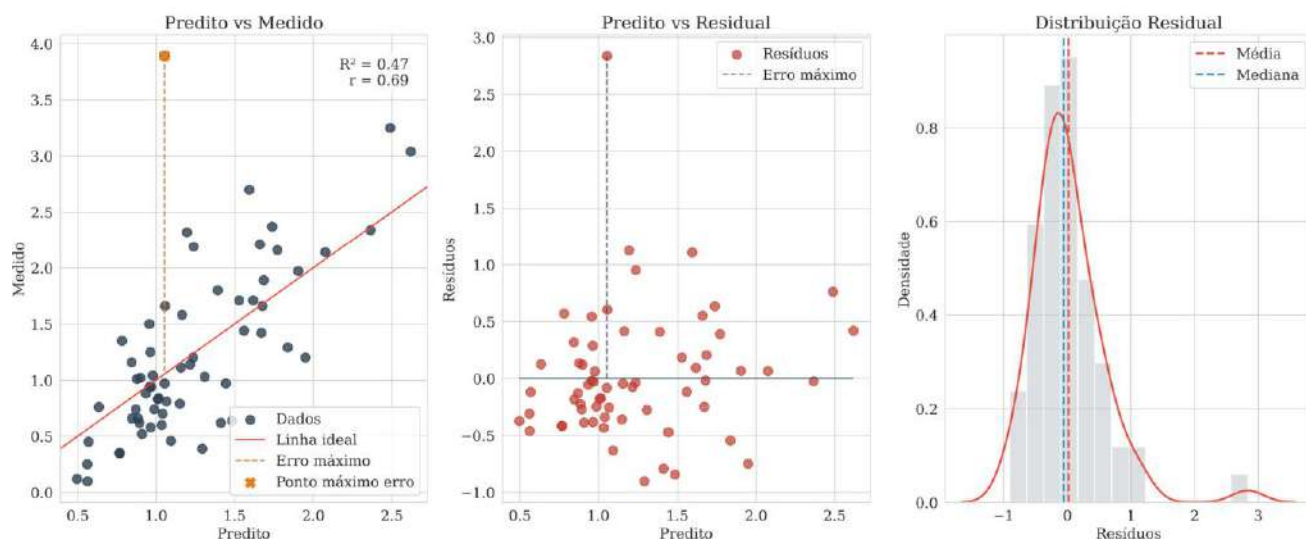


Figura 5.24. Avaliação da performance para o método *Catboost* para o conjunto de dados teste, onde é observado os coeficientes de correlação evidenciando o ponto de maior erro, os valores residuais por amostra e a distribuição de densidade do residual.

A importância das variáveis, estimada a partir do conjunto de dados pelos três métodos selecionados (RF, XGB e CB), está apresentada na Tabela 5.9. Observa-se que todas as variáveis possuem relevância semelhante, indicando consistência entre os algoritmos. Em todos os casos, a curva GR destacou-se como a de maior impacto, o que está em concordância com a literatura, que associa a radioatividade natural registrada nesse perfil, especialmente pelo urânio em folhelhos, à maior presença de matéria orgânica (Hassan *et al.*, 1976). Nota-se ainda que, no método XGB, a variável DT apresentou importância ligeiramente superior à RES, embora essa diferença seja pouco significativa.

Tabela 5.9. Importância das variáveis nos métodos RF, XGB e CB.

Importância das Variáveis		
	Variável	Importância
RF	GR	0.35
	RES	0.35
	DT	0.30
XGB	GR	0.37
	DT	0.32
	RES	0.31
CB	GR	0.36
	RES	0.35
	DT	0.29

5.5.3. Estimativa de COT no poço de validação

Como já mencionado, o único poço localizado na área de estudo com dados de COT disponíveis é o poço 2-CP-1-MA, o qual possui um total de 59 medições. Esse poço foi reservado exclusivamente

para validação dos modelos treinados, sendo tratado como poço cego (*blind well*), ou seja, sem qualquer influência no processo de treinamento.

O desempenho das estimativas de COT nesse poço, geradas pelos métodos selecionados, Random Forest, XGBoost, CatBoost, e também pelo modelo final, construído a partir da combinação dos três algoritmos (ensemble), pode ser avaliado por meio dos coeficientes de correlação e dos *crossplots* apresentados na Figura 5.25.

É importante destacar que, embora o método Random Forest tenha apresentado melhor desempenho em termos de correlação durante as previsões realizadas nos poços utilizados no treinamento (conjuntos de treino e teste), o melhor resultado no poço cego foi obtido com o método CatBoost. Apesar de o CatBoost apresentar um coeficiente de correlação ligeiramente superior ao do modelo combinado, optou-se por adotar o ensemble como modelo final para estimar as curvas de COT nos demais poços da área de interesse.

Essa escolha fundamenta-se no fato de que os poços da área de interesse não possuem medições de COT disponíveis, o que inviabiliza a verificação direta de qual método isolado ofereceria a melhor correlação. Nesse contexto, entende-se que a combinação entre os algoritmos com melhor desempenho individual tende a produzir resultados mais robustos e consistentes. Dessa forma, o ensemble foi definido como o modelo a ser adotado na estimativa final de COT.

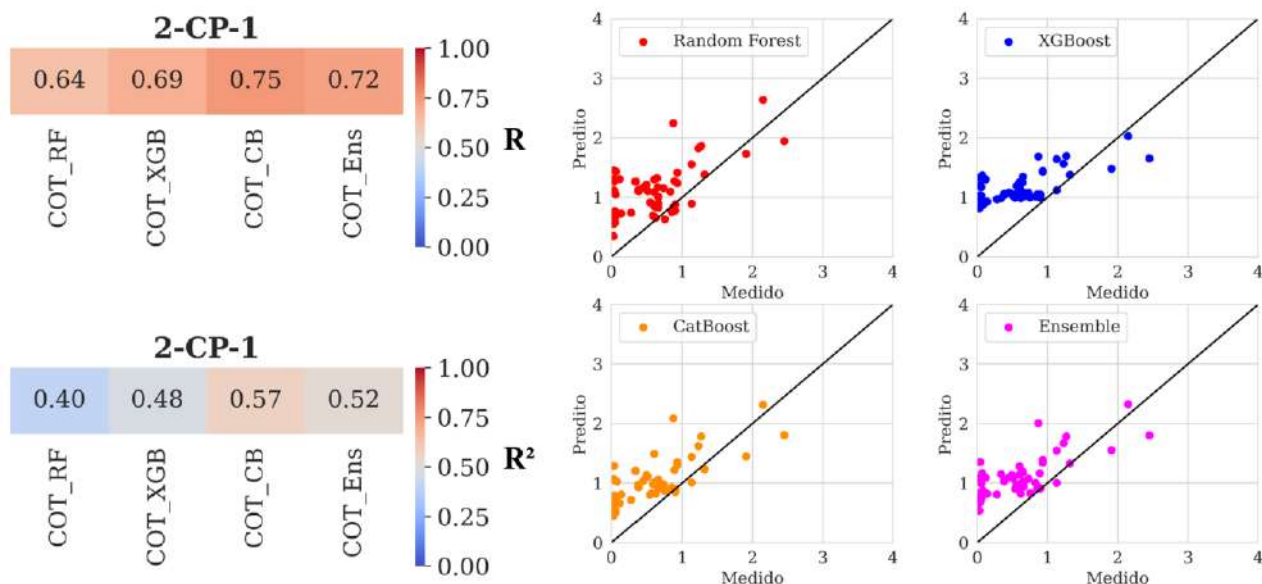


Figura 5.25. Visão geral sobre a performance dos métodos selecionados (RF, XGB e CB) e a combinação dos mesmos. É exibido os valores de coeficientes de correlação e determinação por método, bem como, os gráficos *crossplots* individualizados e o combinado.

Durante a etapa de seleção dos dados, conforme descrito no tópico anterior, verificou-se que outros dois poços adicionariam aproximadamente 100 amostras de COT. No entanto, a análise exploratória revelou que esses poços apresentavam comportamento estatístico significativamente distinto em relação ao restante do conjunto de dados, destoando das tendências gerais observadas.

Diante dessa divergência, optou-se por excluir esses dois poços do conjunto de dados utilizado no treinamento e teste do modelo final. Ainda assim, realizou-se posteriormente um teste adicional considerando um conjunto alternativo que incluía os dados desses poços, submetendo-os aos mesmos procedimentos de pré-processamento e parametrização aplicados anteriormente. Os resultados confirmaram que a inclusão dessas amostras impactava negativamente a qualidade das previsões no poço de validação.

A Figura 5.26 apresenta uma análise comparativa baseada nesse conjunto com dez poços, evidenciando os efeitos da inclusão dos dados adicionais. Quando comparada à Figura 5.25, que se refere ao modelo final treinado com oito poços, observa-se que o *crossplot* entre os valores medidos e estimados torna-se mais disperso, com os pontos significativamente afastados da linha de máxima correlação. Além disso, os coeficientes de desempenho apresentaram queda acentuada: o coeficiente de correlação foi reduzido de 0.72 para 0.36, enquanto o coeficiente de determinação passou de 0.52 para apenas 0.13.

Esses resultados destacam a importância de uma análise criteriosa dos dados de entrada em estudos de *machine learning*, reforçando a necessidade de equilíbrio entre a quantidade e a qualidade das amostras utilizadas para o treinamento dos modelos.

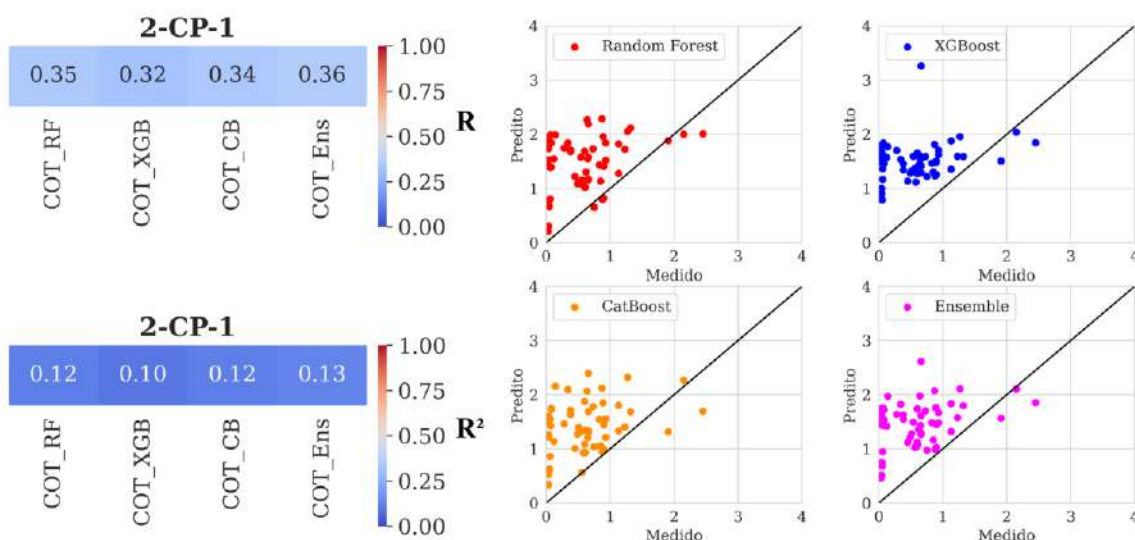


Figura 5.26. Visão geral sobre a performance dos métodos selecionados (RF, XGB e CB) e a combinação dos mesmos considerando os poços 1-OGX-101-MA e 1-OGX-93-MA os quais foram retirados do conjunto de dados de treinamento na etapa de análise exploratória.

É exibido os valores de coeficientes de correlação e determinação por método, bem como, os gráficos crossplots individualizados e o combinado evidenciando o impacto gerado ao inserir dados não coerentes comparados o conjunto geral.

As curvas de COT estimadas através dos métodos selecionados e pelo modelo final estão apresentados na Figura 5.27 junto as amostras medidas no poço 2-CP-1-MA, onde observa-se que o modelo foi eficiente em prever principalmente as tendências de COT associadas as rochas abaixo da soleira de diabásio representada pela fácies 1 em roxo. Também fica evidente que cada método individualmente tiveram previsões mais precisas em diferentes intervalos. No entanto, como já avaliado em gráficos crossplots da Figura 5.25, que houve uma maior dificuldade na previsão de valores muito baixo de COT entre 0 e 0.5%, medições em rochas intercaladas com as intrusões ígneas representadas pelas fácies 2 e 3 (laranja e azul respectivamente).

Tendo em vista que o dado estimado no geral apresentou superestimativas com relação ao dado medido no poço de validação, foi realizado uma correção para deslocar a curva estimada para se aproximar das magnitudes do dado real.

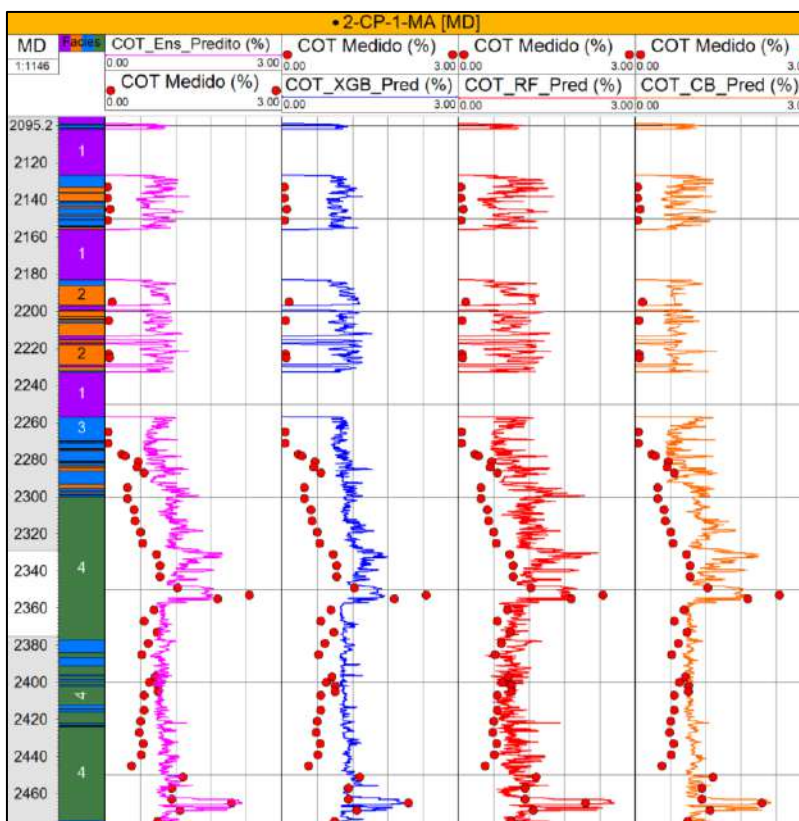


Figura 5.27. Curvas de COT estimadas pelos métodos selecionados durante a análise de performance e o resultado da combinação dos três métodos.

A Figura 5.28a apresenta a comparação entre as curvas de COT interpoladas apenas nas profundidades correspondentes às amostras medidas. Nessa figura, a curva em preto representa os

valores reais interpolados, a curva em vermelho corresponde às estimativas geradas pelo modelo de machine learning, e a curva em verde mostra os valores corrigidos por meio do ajuste linear, conforme detalhado na Seção 5.4.8. Os valores de coeficientes angular e linear estimados nessa aplicação de correção por regressão linear estão descritos na Tabela 5.10 abaixo.

Tabela 5.10. Coeficientes angular e linear obtidos na regressão e aplicada na correção da curva estimada de COT no poço de validação.

Parâmetros de correção	
Coeficiente angular (<i>a</i>)	Coeficiente linear (<i>b</i>)
1.457	-0.6433

Complementarmente, a Figura 5.28b exibe os gráficos dos resíduos ao longo da profundidade, calculados entre os valores medidos e os valores estimados, antes e após a correção. A partir desses gráficos, observa-se que os resíduos da curva corrigida encontram-se mais concentrados em torno de zero, especialmente entre os intervalos de -0.5 e 0.5. Por outro lado, os resíduos da curva estimada antes da correção apresentam maior dispersão, com predominância de valores negativos e várias observações próximas de -0.75, evidenciando uma tendência de superestimação.

Além da análise dos resíduos, os resultados também foram avaliados por meio de métricas estatísticas de erro, MAE, MSE, RMSE e MedAE apresentadas na Figura 5.28c. Todas essas métricas indicaram redução significativa dos erros após a aplicação da correção por regressão linear, reforçando a efetividade do procedimento de calibração adotado.

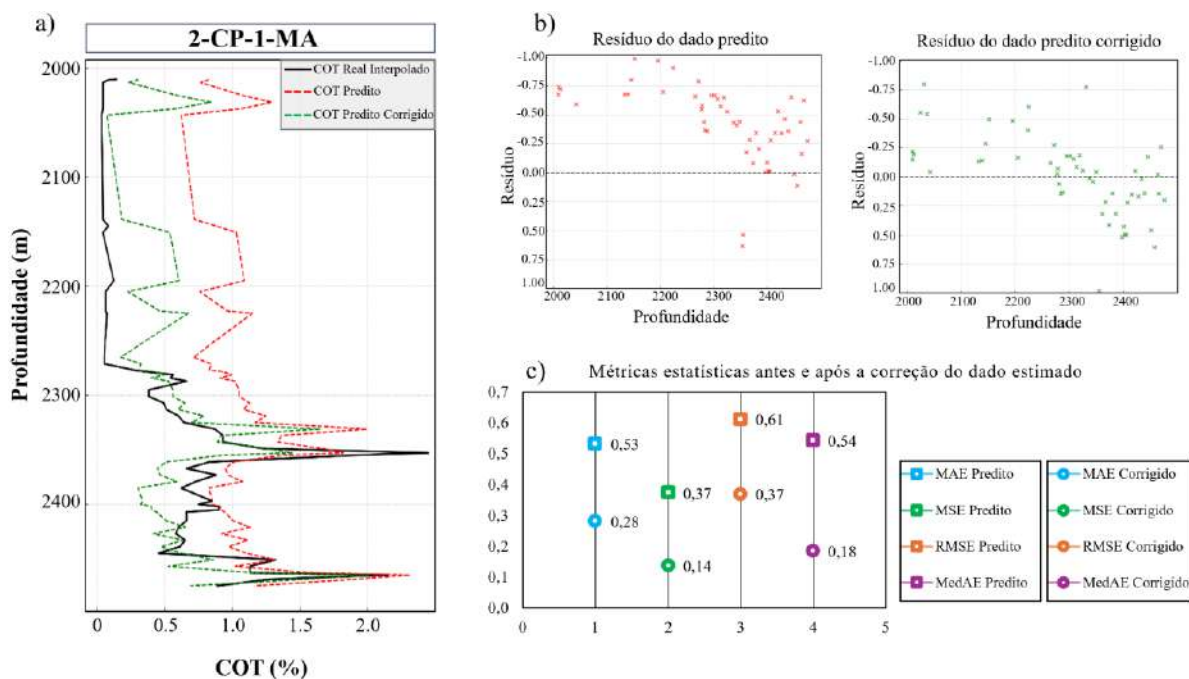


Figura 5.28. Painel de controle da correção da curva de COT estimada. Em a) curvas interpoladas nos intervalos das amostras de COT medidos no poço de validação (2-CP-1-MA), entre o dado real, predito e o predito corrigido. Em b) os valores residuais por amostra para o dados estimados antes e após a correção.

Comparando as curvas estimadas pelo modelo *ensemble* antes e após a correção através da Figura 5.29 é possível observar um melhor ajuste geral, estimando bem não só a tendência principal do dado medido mas também caindo sobre a maioria dos dados reais do poço. Os gráficos *crossplot* da Figura 5.29 mostram que os dados com maior erro, que são referentes aos dados de menor COT (menos que 0.5%), após a correção são melhores resolvidos. Ao realizar máscaras com polígonos como mostrado nos *crossplots* (retângulos em vermelho) para isolar as amostras que sofreram mais com a superestimativa, é possível observar em quais intervalos elas ocorrem, como já mencionado, próximo as facies correspondente as soleiras, e destaca também uma menor quantidade de amostras que tiveram seus valores superestimados pelo modelo machine learning final.

O resultado da correção da curva estimada de COT é apresentada na Figura 5.29 onde é mostrado a comparação entre a curva de COT estimada pelo modelo final (*ensemble*) e a mesma curva após correção por regressão linear, aplicada com base nas amostras reais disponíveis no poço 2-CP-1-MA. Essa correção tem como objetivo ajustar sistematicamente a escala e o viés da estimativa original, promovendo um deslocamento (*shift*) da curva estimada para torná-la mais compatíveis aos dados medidos.

Visualmente, observa-se que, após a aplicação da correção, a curva estimada apresenta um alinhamento mais consistente com os dados medidos de COT. Além de seguir adequadamente a tendência geral do perfil, a curva corrigida também passa a refletir, de forma mais fiel, os valores quantitativos observados nas amostras reais. Isso indica uma melhora substancial não apenas na representação do comportamento global do dado, mas também na sua precisão ponto a ponto ao longo da profundidade. Esse ajuste é especialmente relevante nos intervalos onde a estimativa original apresentava superestimativas sistemáticas, sobretudo em regiões de baixo teor de COT (inferiores a 0,5%).

Além das curvas, a Figura 5.29 também exibe dois gráficos *crossplot*: o primeiro relaciona os valores estimados antes da correção com os valores medidos, enquanto o segundo mostra a relação entre os dados medidos e os valores estimados após a correção. Em ambos os gráficos, foi aplicada uma máscara poligonal (em destaque por retângulos vermelhos) para isolar as amostras com os maiores desvios residuais. Essa seleção permite identificar os pontos com maior erro de predição, enfatizando que, após a correção, esses desvios são significativamente atenuados.

A aplicação dessa máscara também possibilitou mapear, ao longo da profundidade do poço, os intervalos onde essas amostras com maior erro estão localizadas. A análise integrada com o perfil de fácies revelou que essas amostras se concentram predominantemente em zonas associadas à ocorrência de soleiras ígneas. Tal observação sugere uma possível relação entre as características litológicas dessas fácies e a dificuldade do modelo em realizar previsões precisas nesses contextos, mesmo após o ajuste.

Por fim, é importante destacar que a quantidade de amostras com superestimação relevante foi reduzida após a correção, indicando que o procedimento de calibração contribuiu de forma eficaz para melhorar a acurácia local da curva estimada, especialmente em regiões críticas do perfil.

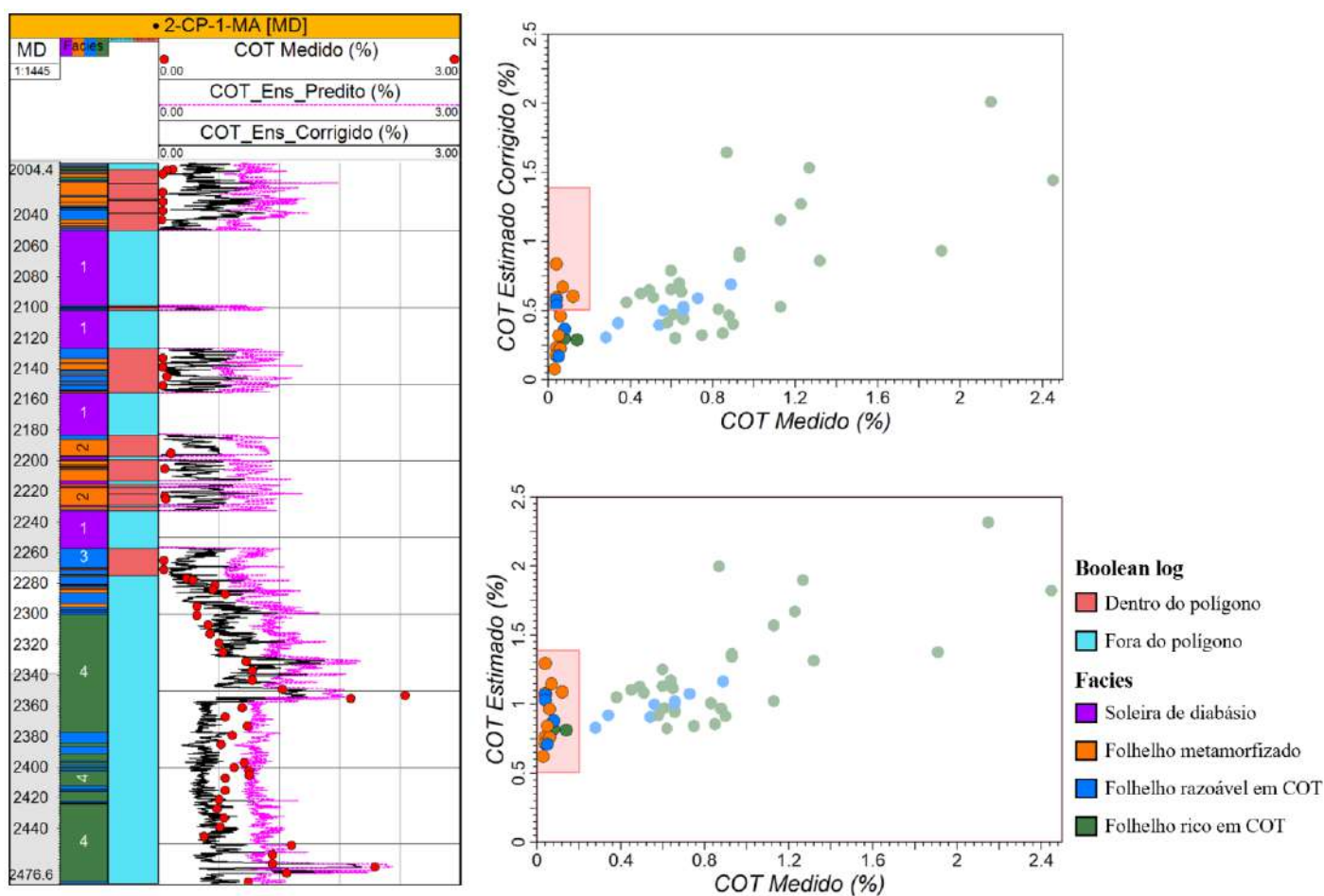


Figura 5.29. Curva de COT estimado após a correção (preto) e antes da correção (rosa) mostrando um melhor ajuste. O perfil de fácies é exibido como uma avaliação complementar para evidenciar a variação de COT por fácies. O perfil booleano corresponde a máscara feita pelo polígono vermelho no crossplot entre o COT estimado e medido. A máscara foi criada contemplando as amostras com valores superestimados pelo *machine learning*.

Tendo em vista, o melhor ajuste obtido após a correção da curva estimada no poço de validação, 2-CP-1-MA, foram aplicadas correções nas curvas estimadas para os poços 1-OGX-16-MA, 3-OGX-

57-MA e 1-OGX-22-MA, utilizando os mesmos coeficientes linear e angular mostrados na Tabela 5.10. Dessa forma, o resultado final das curvas de COT para os poços da área de estudo é estimado e mostrado através da Figura 5.30.

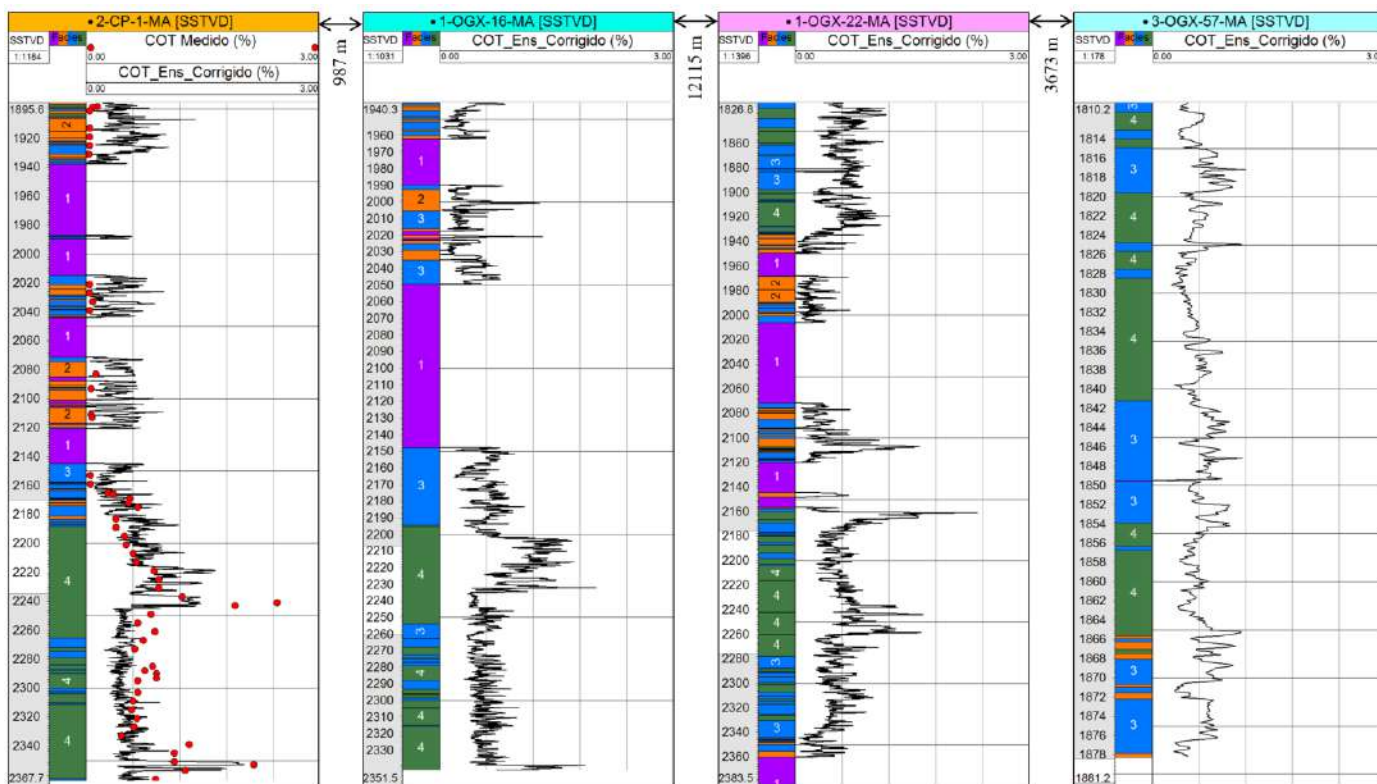


Figura 5.30. Curvas de COT previstas nos poços da área de estudo.

Conforme já indicado na análise desenvolvida no Capítulo 4 quanto a classificação das fácies, os resultados obtidos neste estudo, agora avaliados em maior nível de detalhe, confirmam que as fácies rotuladas em verde correspondem aos folhelhos com os maiores teores de COT, podendo corresponder até mesmo em rochas imaturas. Por outro lado, os folhelhos representados pelos rótulos azul e laranja apresentam distribuições de COT bastante semelhantes (Figura 5.31), caracterizadas por valores mais baixos. Essa redução pode ser atribuída, principalmente, à maior proximidade dessas rochas com soleiras de diabásio, cuja intrusão promove uma maturação térmica quase instantânea da matéria orgânica presente, resultando, portanto, em menores teores de COT nesses intervalos.

Apesar da semelhança nos valores de COT, a Figura 5.32 evidencia diferenças claras entre essas duas fácies, principalmente associadas às propriedades do perfil de raios gama. A fácies de folhelho metamorfizado, identificada pelo rótulo laranja, apresenta valores de GR consideravelmente elevados, o que reforça sua distinção em relação às demais. Ainda na Figura 5.32, observa-se a contribuição

relativa de cada variável (GR, RES e DT) no processo de estimativa do COT. Destaca-se que os altos valores de GR registrados nessas zonas podem ter contribuído para a superestimativa do COT pelo modelo, uma vez que, em termos gerais, elevados níveis de GR estão associados à presença de matéria orgânica.

Essa inferência é corroborada pelos resultados mostrados na Tabela 5.9, a qual aponta o perfil GR como a variável de maior importância para todos os algoritmos de *machine learning* aplicados no estudo.

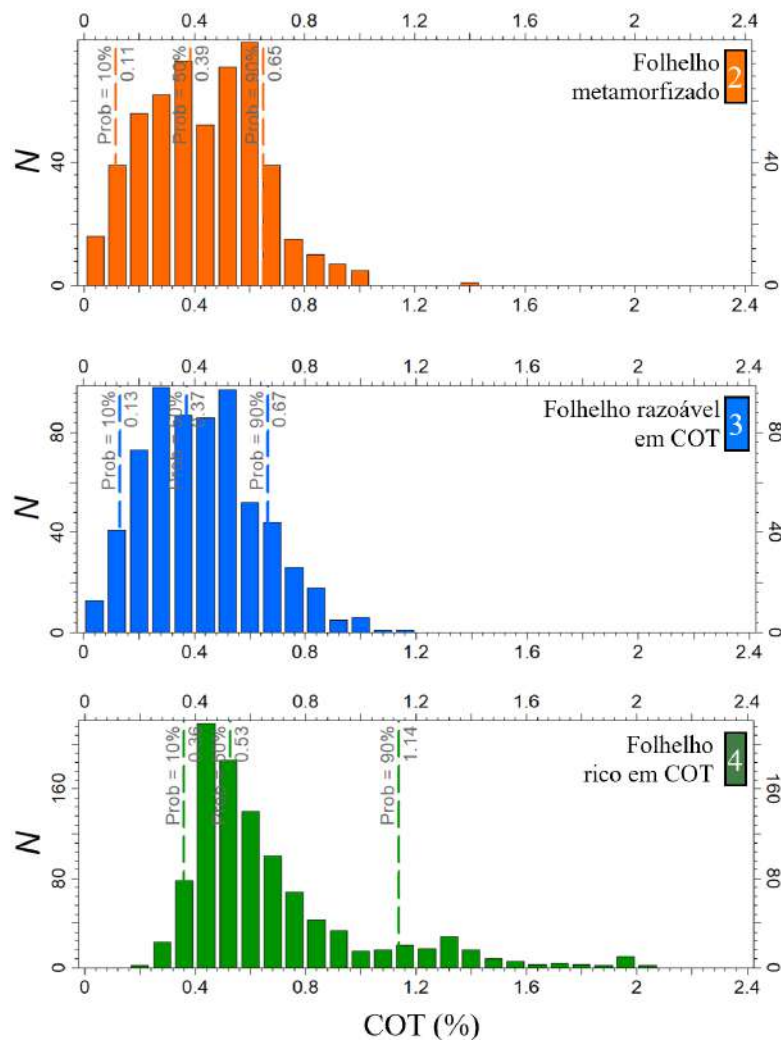


Figura 5.31. Histogramas da distribuição de COT por fácies, destacando as suas diferentes concentrações, onde nota-se que ocorre uma maior sobreposição nas fácies 2 e 3, enquanto que para as fácies 4 há um predomínio de valores relativamente mais altos.

Além disso, a Figura 5.32 também evidencia que, de modo geral, os maiores valores de COT estão associados à ocorrência simultânea de altos valores nos perfis de GR, DT e RES, porém não de forma linear. Essa associação reforça a ideia de que a combinação dessas variáveis bem distribuídas é um indicativo importante para a predição de teores elevados de matéria orgânica.

Entretanto, em rochas localizadas próximas às soleiras ígneas, observa-se um comportamento distinto, especialmente no perfil sônico (DT). Nesses intervalos, devido ao metamorfismo de contato, as rochas metassedimentares tendem a se comportar de maneira mais compacta e rígida, resultando em tempos de trânsito do som mais baixos. Esse comportamento afeta diretamente a resposta do DT, contribuindo para a estimativa de valores mais baixos de COT pelos modelos de *machine learning*, mesmo em presença de valores elevados de GR e variações bruscas de resistividade, que podem alternar entre mínimos muito baixos e picos acentuados.

Adicionalmente, um fator que contribui para o aumento das incertezas na estimativa de COT nesses intervalos é a baixa disponibilidade de medições reais de COT em rochas afetadas por soleiras ígneas. Essa escassez de amostras no processo de treinamento limita a capacidade dos algoritmos de *machine learning* em capturar corretamente os padrões associados a esse tipo de fácies, o que pode levar a previsões menos confiáveis e com maior dispersão residual.

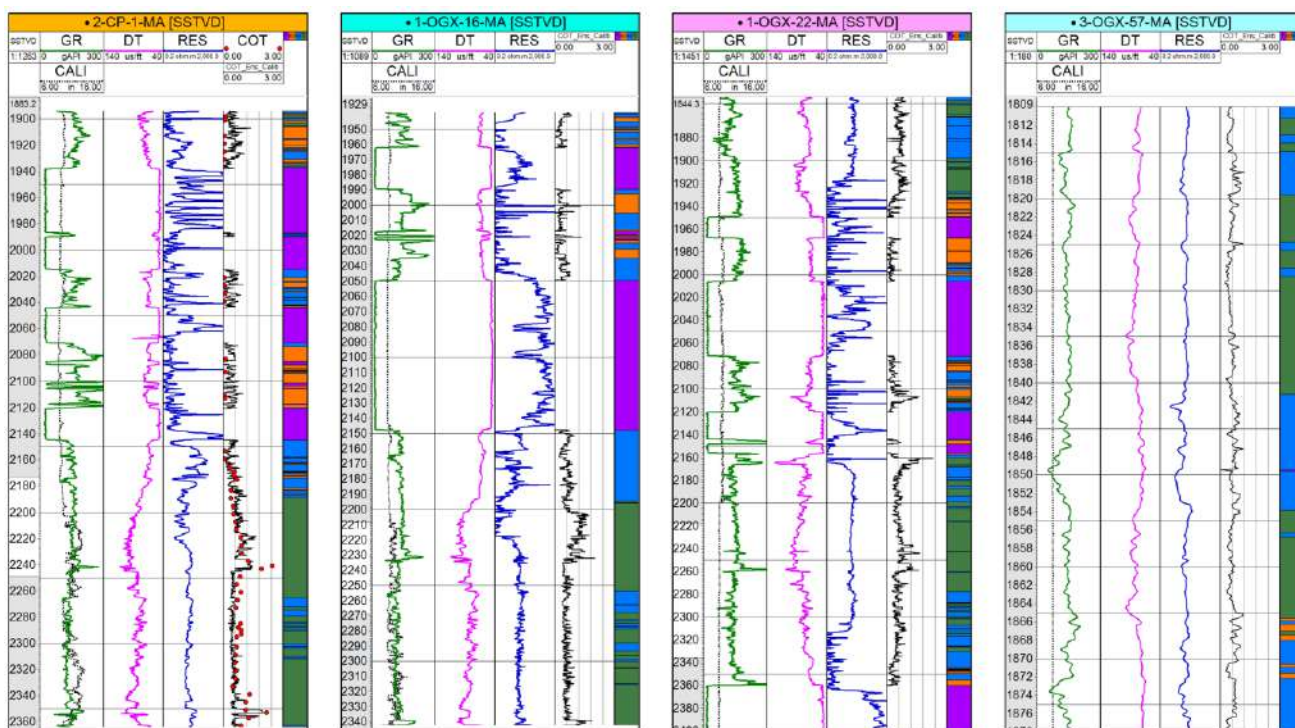


Figura 5.32. Exibição das variáveis utilizadas para estimar a curva de COT a partir do modelo *ensemble*, bem como do resultado obtido para o COT estimado junto a discretização de diferentes fácies definidas para o intervalo de estudo. Além disso, também é exibido o perfil de Caliper reforçando a qualidade dos poços, mostrando que para alguns intervalos é observado desmoronamento não tão importantes.

5.6. DISCUSSÕES

A estimativa indireta do carbono orgânico total em rochas geradoras é alvo de estudo para o desenvolvimento de metodologias desde a década de 70, sendo os métodos de Passey (1990) e de

Schmoker (1979) os mais conhecido e utilizados por muitas anos como método padrão. Devido a crescente dos usos de métodos de *machine learning* na área da geociências, muitos pesquisadores mudaram a abordagem na estimativa de curvas de COT e passaram a elaborar metodologias para adquirir esse dados a partir de diferentes algoritmos de *machine learning* seja por meio de métodos de árvore de decisão, como o *Random Forest*, *Extreme Gradient Boosting*, como a partir de método de *Artificial Neural Network*, *Multi-layer Perceptron*, *Support Vector Machine* entre muitos outros.

Não foram encontrados na literatura estudos que abordem um cenário geológico semelhante ao apresentado neste capítulo, no qual intervalos de rochas geradoras são intrudidos por soleiras ígneas. Nesse contexto, as intrusões não apenas forneceram o calor necessário para que as rochas atingissem a maturação térmica, e portanto entrasse na janela de geração de gás, mas também promoveram alterações mineralógicas e texturais decorrentes do metamorfismo de contato, impactando diretamente as respostas dos perfis elétricos. Nos intervalos mais afetados pelas intrusões ígneas, onde são observados os maiores erros de estimativa, fica compreendido que mesmo se tratando de uma metodologia robusta gerando bons resultados, o volume de informações trabalhadas podem não ter sido suficientes para prever esse comportamento particular.

Diante das limitações de dados disponíveis na área de estudo e da complexidade geológica da Bacia do Parnaíba, observa-se que, mesmo com resultados considerados satisfatórios, ainda que não excelentes, a metodologia aplicada se mostra a mais adequada para este contexto. Esse cenário contrasta com os trabalhos de Liu *et al.* (2021) e Khan *et al.* (2023), que, embora também tenham trabalhado com conjuntos de dados de COT relativamente limitados, alcançaram coeficientes de correlação superiores a 0,7 por meio do uso de algoritmos como RF, XGB, SVM e MLP. A diferença fundamental reside na maior diversidade de perfis incorporados nesses estudos, incluindo densidade, porosidade neutrão e perfis espectrais de raios gama (como tório e potássio), o que ampliou a capacidade de captura de variações associadas ao teor orgânico.

Além disso, a estratégia adotada neste trabalho, com a validação independente no poço 2-CP-1-MA, representou uma abordagem metodológica mais rigorosa. Embora essa opção tenha reduzido a quantidade de dados disponível para o treinamento dos modelos, ela aumentou a robustez da generalização, garantindo maior confiabilidade aos resultados. Essa escolha metodológica, contudo, também pode contribuir para a diminuição dos valores de R^2 quando comparados àqueles obtidos em validações menos restritivas.

Em contrapartida, Shalaby *et al.* (2019) demonstraram que os modelos matemáticos clássicos, como os de Schmoker e Hester (1983), Passey *et al.* (1990) e Zhao *et al.* (2016), alcançaram elevada

correlação (R^2 superior a 0,9) nas formações analisadas, quando comparadas as estimativas de COT com resultados de análises geoquímicas. Por outro lado, as previsões obtidas por meio de redes neurais artificiais (ANN) apresentaram desempenho significativamente inferior, com valores máximos de R^2 em torno de 0,4. Apesar desse contraste, os autores ressaltam o caráter promissor dos métodos de *machine learning*, destacando, entretanto, a necessidade de bases de dados mais extensas para validar sua aplicabilidade e aprimorar a precisão das estimativas, uma vez que, o estudo contou com apenas 95 amostras de COT, número considerado limitante. Além disso, recomendam a experimentação com diferentes algoritmos e técnicas de regressão, bem como a aplicação de métodos de transformação de dados, como estratégias para otimizar a performance preditiva dos modelos.

Outros trabalhos mais recentes, como os de Sun *et al.* (2023) e Bione *et al.* (2024), evidenciaram resultados opostos. Nessas pesquisas, a estimativa de COT por meio de algoritmos de *machine learning*, como *Random Forest* e *XGBoost*, apresentou correlações superiores a 0,7 entre os valores medidos e estimados. Já o método empírico do $\Delta\log R$ demonstrou desempenho significativamente inferior, com correlações próximas de 0,1 em Sun *et al.* (2023) e até mesmo negativas em Bione *et al.* (2024). Este último estudo, em particular, destacou-se pelo grande volume de dados analisados, reunindo 6.353 observações de COT em 130 poços da Bacia do Espírito Santo, o que reforça a robustez estatística dos resultados obtidos pelos modelos de *machine learning*.

Com o objetivo de construir um modelo o mais generativo possível, no presente trabalho foi definido a curva estimada de COT a partir da combinação dos métodos de melhor desempenho, considerando que a acurácia pode variar entre os diferentes poços, como evidenciado pelos coeficientes de correlação. A adoção dessa estratégia busca garantir que o modelo seja capaz de representar a heterogeneidade observada no poço da área de estudo, mesmo sem que este tenha sido incluído no treinamento dos algoritmos, assegurando, assim, estimativas mais consistentes em poços sem dados geoquímicos disponíveis para validação. Metodologias semelhantes foram descritas em trabalhos recentes. Goliatt *et al.* (2023) propuseram o método denominado *Super Learner*, no qual algoritmos como MLP, SVR, *K-Nearest Neighbors* (KNN) e *Linear Regression* (LR) foram empilhados em etapas sucessivas, resultando em um modelo robusto que atingiu coeficiente de correlação de $R=0,80$. De forma análoga, Mandal *et al.* (2021) desenvolveram uma abordagem de *Ensemble Learning*, combinando algoritmos como RF, MLP, SVM e *Gradient Boosting Regressor* (GBR), alcançando igualmente um coeficiente de correlação de R^2 aproximadamente de 0,80. A principal diferença entre os estudos citados e o presente trabalho refere-se ao número de variáveis utilizadas no treinamento e ao contexto geológico investigado. Enquanto Goliatt *et al.* (2023) e Mandal *et al.* (2021) aplicaram

suas metodologias em reservatórios não convencionais, este estudo concentra-se na estimativa de COT em rochas geradoras intrudidas por soleiras ígneas, sem qualquer interpretação voltada para reservatórios.

Com o intuito de avaliar de forma abrangente os trabalhos mais recentes publicados sobre a estimativa de COT, a Tabela 5.11 apresenta um resumo das principais informações desses estudos, contemplando os métodos empregados, a máxima performance alcançada, as variáveis de entrada utilizadas nos modelos de *machine learning* e a quantidade de amostras da variável alvo. Esse panorama permite uma análise comparativa mais ampla, evidenciando os avanços metodológicos observados na literatura. De modo geral, constata-se que os estudos com maior disponibilidade de dados alcançaram correlações mais elevadas. Entretanto, não apenas o volume de dados se mostra determinante: a integração de diferentes algoritmos e metodologias também desempenha papel central. Destacam-se, por exemplo, os trabalhos de Goliatt *et al.* (2023) e Mandal *et al.* (2021), que propuseram modelos de *Stacking* e *Ensemble Learning*, respectivamente, combinando múltiplos algoritmos para aumentar a robustez preditiva. Além disso, outras estratégias inovadoras incluem o uso de algoritmos genéticos para otimização de hiperparâmetros (Cabral *et al.*, 2025) e a aplicação de Análise de Componentes Principais (PCA) para redução de dimensionalidade e melhoria do desempenho dos modelos (Zhang *et al.*, 2024). Em contraste com os trabalhos reunidos na tabela, o presente estudo introduz a aplicação do algoritmo CatBoost, que apresentou a melhor performance no poço de validação, com coeficiente de correlação de 0,75 e coeficiente de determinação de 0,57. Outro diferencial relevante desta pesquisa foi a utilização de dados de poços externos à área de estudo para o treinamento dos modelos, posteriormente aplicados aos poços locais, estratégia que reforça a robustez e a capacidade de generalização da metodologia proposta.

Tabela 5.11. Resumo comparativo dos trabalhos recentes na estimativa de COT por machine learning, detalhando métodos utilizados, variáveis de entrada, tamanho das amostras e desempenho.

Referências	Métodos	Nº de amostras	Perfis utilizados	R ² Treino	R ² Teste	
Elkhatny, 2018	ANN + SaDE	489	(82% treino, 18 % teste)	GR, DT, RD, RHOB	0.97	0.97
Shalaby et al., 2019	ANN	95	(75% treino, 25% teste)	GR, RHOB, NPHI, LLD	0.96	0.42
Liu et al., 2021	XGB, RF, KNN, SVM, MLR	104	(80% treino, 20% teste)	DT, DEN, CNL, K, TH, RD	0.91	-
Mandal et al., 2021*	Ensemble: RF, MLP, SVM, GBR	283	(60% treino, 20% teste, 20% validação)	RHOB, DT, LLD, NPHI	0.86	0.86
Sun et al., 2023	RF, SVR, XGB	816	(70% treino, 30% teste)	GR, DT, DEN, RD, RS, CAL, PE, TH, K, U, CNL, SP, MSFL, GRSL	0.91	0.87
Khan et al., 2023	RF, SVR, XGB	282	(70% treino, 30 % teste)	GR, DT, RT, DEN	0.95	0.85
Goliatt et al., 2023*	Stacking ML: MLP, SVR, KNN, RF, LR	462	(70% treino, 30% teste)	GR, DT, RT, DEN	0.80	-
Bione et al., 2024	XGB (PySpark)	6353	90% treino, 10% teste)	GR, DT, RESD, RHOB, NPHI, CALI	0.93	0.71
Zhang et al., 2024	XGB + PCA	82	(70% treino, 30% teste)	DT, GR, DEN, RT, CNL	0.68	0.57
Cabral et al., 2025	XGB, SVM, KNN, ELM + Algoritmo Genético	309	(80% treino, 20% teste)	GR, DT, RT, DEN, NPHI	0.75	0.42
Este estudo (Nobre, 2025)*	RF, XGB, CB, MLP, SVM	371	(67% treino, 17% teste, 16% validação)	GR, DT, RES	0.72	0.52

* Correlações considerando a combinação dos métodos como proposto pelos autores

Pode-se afirmar que a utilização de dados de poços externos à área de estudo não constitui um fator limitante, uma vez que, o contexto geológico é o mesmo, por se tratar da mesma bacia sedimentar. Esta é caracterizada por baixa complexidade estrutural, pela presença de camadas com ampla continuidade lateral e, sobretudo, por corresponder à mesma unidade estratigráfica, a Formação Pimenteiras. Nessa perspectiva, espera-se que a influência das intrusões ígneas, principal desafio para a estimativa de COT, apresente comportamento semelhante em toda a bacia. Dessa forma, a metodologia adotada pode ser aplicada com relativa confiança, sem a introdução de incertezas significativas nesse aspecto.

Em relação à baixa correlação observada entre as variáveis de treinamento (GR, DT e RES) e os valores de COT, destaca-se que os métodos de *machine learning*, conforme avaliado neste estudo, apresentam capacidade de lidar adequadamente com esse tipo de comportamento. Esse resultado está em consonância com o reportado em outros trabalhos da literatura (Zhang *et al.*, 2024; Liu *et al.*, 2021; Mandal *et al.*, 2021; Bione *et al.*, 2024), nos quais os algoritmos foram capazes de identificar padrões heterogêneos mesmo diante de relações fracas ou não lineares entre as variáveis de entrada e o COT.

As baixas correlações observadas neste estudo, associadas ao desempenho dos métodos MLP e SVM, podem estar relacionadas ao número reduzido de variáveis utilizadas no treinamento, bem como

à parametrização dos modelos, ainda que tenham sido realizadas otimizações e múltiplos testes. Em trabalhos comparativos nota-se que, quando aplicados esses mesmos métodos, foram utilizadas pelo menos quatro variáveis de entrada, e não apenas três como no presente estudo (Liu *et al.*, 2021; Mandal *et al.*, 2021; Sun *et al.*, 2023; Khan *et al.*, 2023).

Como esperado, o SVM tende a lidar melhor com conjuntos de dados limitados em comparação ao MLP, fato também evidenciado pelas métricas de desempenho obtidas neste estudo. Outro aspecto relevante refere-se ao impacto do ruído presente em alguns perfis, especialmente na resistividade, que pode ter influenciado negativamente o desempenho. Algoritmos como MLP e SVM são mais sensíveis a variáveis com comportamento ruidoso, ao contrário de outros métodos mais robustos, o que ajuda a explicar a diferença em relação aos resultados reportados em trabalhos anteriores, nos quais os perfis de resistividade apresentaram comportamento mais estável e menos heterogêneo.

Por fim, os valores estimados de COT para os poços da área de estudo podem ser classificados, segundo Peters & Cassa (1994), como rochas de potencial gerador variando de pobre a bom, uma vez que, os teores se situam entre 0% e 2%. As fácies discretizadas no Capítulo 4 estão diretamente correlacionadas à variação do COT estimado, de modo que a fácies denominada Folhelho rico em COT corresponde predominantemente a intervalos de bom potencial gerador; a fácies Folhelho razoável em COT apresenta, em sua maioria, teores razoáveis, mas inclui também trechos classificados como pobres em COT; já a fácies Folhelho metamorfozado é caracterizada, de forma predominante, por baixos teores, sendo, portanto, considerada pobre em potencial gerador.

Cabe destacar que, embora os folhelhos da Formação Pimenteiras tenham sido discretizados em diferentes fácies no capítulo anterior, tais classes inicialmente não receberam uma qualificação concreta, uma vez que, em termos gerais, tratam-se de rochas folhelhosas. Entretanto, já havia indícios de variações significativas ao longo da formação, com a presença de rochas metassedimentares descritas em poços da área e variações no teor de COT atribuídas a diferentes períodos de afogamento da bacia. Esses aspectos já haviam sido apontados nos levantamentos de Rodrigues (1995), que identificou intervalos de folhelhos radioativos com potencial para concentrar maiores teores de COT. Assim, estabeleceu-se uma associação preliminar entre as fácies descritas nos poços e o potencial gerador correspondente (pobre, razoável ou rico em COT).

5.7. CONCLUSÃO

O desenvolvimento deste capítulo permitiu demonstrar, de forma abrangente, o potencial das técnicas de *machine learning* aplicadas à estimativa de Carbono Orgânico Total (COT) em contextos

geológicos, como a Formação Pimenteiras, marcada pela heterogeneidade litológica e pela influência de intrusões ígneas sobre a maturação da matéria orgânica. A pesquisa partiu de um cenário desafiador: a disponibilidade restrita de amostras medidas de COT, limitadas a apenas um poço da área de estudo (2-CP-1-MA), o que inviabilizaria abordagens puramente empíricas. Nesse sentido, a utilização de algoritmos supervisionados treinados com dados de poços externos mostrou-se uma solução metodológica robusta, garantindo a geração de curvas contínuas de COT e permitindo caracterização mais detalhada do potencial gerador da unidade.

A comparação entre diferentes algoritmos, Random Forest, XGBoost, CatBoost, Multi-Layer Perceptron e Support Vector Machine, evidenciou que cada método apresenta vantagens e limitações específicas, relacionadas à sua capacidade de generalização, sensibilidade à escala dos dados e robustez frente a baixa relação linear entre as variáveis. A aplicação de pré-condicionamento rigoroso (interpolação, remoção de *outliers* e transformação logarítmica da resistividade), associada à padronização das variáveis, foi fundamental para assegurar que o treinamento ocorresse de maneira consistente e que os modelos alcançassem níveis satisfatórios de desempenho.

Os resultados demonstraram que os algoritmos baseados em árvores (RF, XGB e CB) apresentam maior estabilidade em dados heterogêneos, enquanto os métodos baseados em neural network e redes de maximização da margem (MLP e SVM) se mostraram mais sensíveis à escolha de hiperparâmetros e à padronização. Ainda assim, a estratégia de combinação de modelos (*ensemble* simples por média), aliada ao pós-processamento por regressão linear, permitiu reduzir vieses individuais, aumentando a confiabilidade das curvas finais. Essa abordagem reforça a importância da integração metodológica como caminho para superar as limitações de cada algoritmo isolado.

Outro ponto relevante refere-se à análise da importância das variáveis. As curvas de Gamma Ray (GR), Sônico (DT) e Resistividade (RES) mostraram-se consistentes com a literatura quanto à sua relação com a presença de matéria orgânica. Entretanto, observou-se baixa correlação linear direta com o COT, reforçando que métodos de aprendizado de máquina, capazes de captar relações não lineares complexas, são mais adequados do que abordagens empíricas tradicionais. Em termos de performance, as métricas estatísticas utilizadas (MSE, RMSE, MAE, MedAE, Pearson e R^2) confirmaram que os modelos alcançaram resultados robustos, ainda que com limitações decorrentes do volume de dados disponível.

Os resultados indicam que, nos poços da área de estudo, os teores estimados de COT se concentram majoritariamente entre 0% e 2%, o que enquadra a maior parte dos intervalos nas classes de pobre a boa qualidade de rocha geradora segundo a convenção de Peters & Cassa (1994). Em

especial, os trechos com $COT < 0,5\%$ caracterizam-se como pobres, e foram justamente esses os intervalos onde o modelo apresentou superestimação antes da calibração final, comportamento mitigado após a correção por regressão, que reduziu os desvios nos baixos teores ($\leq 0,5\%$). Quando se observa o conjunto de treinamento e validação (poços externos), a distribuição estatística dos dados abrange cenários de pobre a excelente potencial gerador, evidenciando diversidade composicional e de maturidade suficiente para treinar modelos generalizáveis, embora, na área 3D, prevaleçam teores até $\sim 2\%$. Esse contraste reforça que a transferência de aprendizado para a área alvo herda a competência para diferenciar as qualificações de potencial gerador, mas está condicionada ao intervalo de COT efetivamente presente na área (predominantemente pobre–boa).

Os rótulos de fácies por cor utilizados nas figuras reforçam esse quadro: os rótulos verdes correspondem aos folhelhos com maiores teores de COT, enquanto azul e laranja apresentam distribuições mais baixas de COT, relacionadas à maior proximidade de soleiras diabásicas, o que induz maturação térmica rápida e redução do COT remanescente. Ainda, a fácies folhelho metamorfozido (laranja) tende a exibir GR mais elevado, distinguindo-se no comportamento pretrofísico; por sua vez, essa elevação de GR pode ajudar a explicar superestimativas locais do modelo, já que, em termos médios, altos GR associam-se à matéria orgânica, padrão que o algoritmo aprende e extrapola nessas zonas. Do ponto de vista multivariado, observou-se que as ocorrências de COT mais alto tendem a coincidir, em escala global, com valores simultaneamente altos de GR, DT e RES—sem que isso implique correlação linear forte ponto a ponto, o que justifica o uso de ML para capturar relações não lineares. Nas proximidades de soleiras, entretanto, o DT diminui (rocha mais rígida/compacta), o que diminui a predição de COT mesmo quando GR é alto e a RES oscila fortemente, um padrão que aumenta as incertezas justamente nas fácies afetadas por metamorfismo de contato (e nas quais há pouca medição direta de COT para treino).

Portanto, conclui-se que a metodologia proposta contribui significativamente para a predição indireta do COT em folhelhos radioativos, oferecendo um caminho promissor para ampliar o entendimento sobre o potencial gerador de unidades pouco amostradas. Além disso, o trabalho reforça a importância da combinação de múltiplos algoritmos, do tratamento criterioso dos dados de entrada e da integração com conhecimentos geológicos prévios. A partir dos resultados aqui alcançados, abre-se a possibilidade de aplicações futuras voltadas à integração com atributos sísmicos e à expansão do escopo para caracterização tridimensional do potencial gerador, alinhando-se às práticas modernas de geociência quantitativa como será estabelecida nos próximos capítulos da tese.

6. Precondicionamento do Dado Sísmico

6.1. INTRODUÇÃO

Dados sísmicos terrestres são conhecidos por serem mais ruidosos quando comparados com os dados marinhos, como resultado da natureza da aquisição. Algumas fraquezas incluem o desacoplamento da fonte e dos receptores, a não uniformidade topográfica da área pesquisada e a presença de vários obstáculos, como rios, lagos e cidades. Outra característica comum observada em dados sísmicos terrestres é o ruído produzido pelo arranjo de aquisição, chamado de *footprint*. A Bacia do Parnaíba, área de estudo deste trabalho, é formada por sucessões de rochas sedimentares e soleiras espessas de diabásio, que podem facilmente exceder 150 m de espessura. Os diques e soleiras intrudem em diferentes níveis da bacia, o que também interfere na qualidade dos dados, refletindo muita energia das ondas sísmicas nessas rochas ígneas de alta impedância acústica. Portanto, este capítulo será dedicado ao fluxo de pré-condicionamento sísmico, baseado na combinação dos métodos Dip Steering Median Filter (DSMF) e Spectral Blueing (SSB), aplicados ao dado sísmico 3D da Bacia do Parnaíba, na área de produção de gás do Parque dos Gaviões.

O fluxograma de trabalho, como apresentado na Figura 6.1, é subdividido em duas etapas, a primeira com a aplicação da técnica do Dip Steering Median Filter, o qual tem a função de filtrar ruídos aleatórios e melhorar a relação sinal-ruído, e na etapa seguinte é empregada a técnica de Spectral Blueing que possui a finalidade de equilibrar o espectro sísmico, recuperando a energia das altas frequências, enfraquecidas após a primeira técnica, embora, devido ao efeito de propagação e atenuação das ondas sísmicas, elas já fossem relativamente mais fracas do que as baixas frequências no dado original.

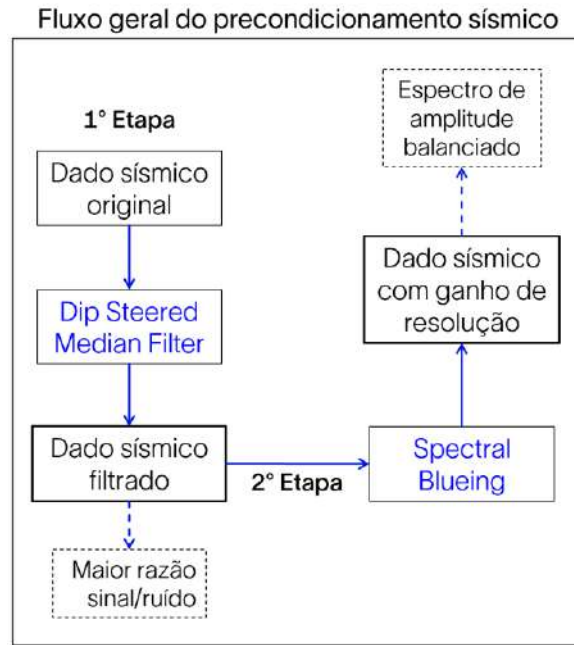


Figura 6.1. Visão geral do fluxo de trabalho de condicionamento aplicado aos dados sísmicos.

6.2. FUNDAMENTAÇÃO

6.2.1. DIP STEERED MEDIAN FILTER

O Dip Steering Median Filter é uma técnica de filtro de mediana orientado para estrutura. A orientação do filtro é dada por um cubo de mergulho de eventos sísmicos, estimado traço a traço. Assim, essa técnica de filtro é adequada para remover ruídos de alta energia, respeitando a inclinação predominante dos eventos, além de não causar uma suavização significativa dos dados, como pode ocorrer em um filtro de média. Também será observado uma melhora na continuidade lateral dos eventos sísmicos promovido por esta técnica (Tingdahl, Kristofer & de Groot, Paul., 2003).

O filtro de mediana é uma técnica amplamente reconhecida no processamento de imagens e tem sido eficaz na remoção de ruídos. Em comparação com o filtro de média, o filtro de mediana tende a manter as bordas da imagem, enquanto o filtro de média tende a suavizá-las (Gonzalez, 1987). No processamento sísmico, tanto a filtragem mediana quanto o de média são utilizadas. Normalmente, o valor médio ou mediano dentro de uma janela de pesquisa horizontal é atribuído à posição da amostra de saída. No entanto, esse tipo de filtro favorece eventos horizontais em detrimento dos mergulhos, tornando-o inadequado para muitas aplicações sísmicas, onde os eventos geológicos tendem a ocorrer com um determinado mergulho.

O processo para filtrar o dado sísmico a partir desta técnica, se inicia com a estimativa de um volume que contenha o mergulho dos eventos sísmicos, também chamado de *steering cube* (Figura

6.2), onde essas informações são calculadas traço a traço do dado e são utilizadas para guiar o filtro de mediana.

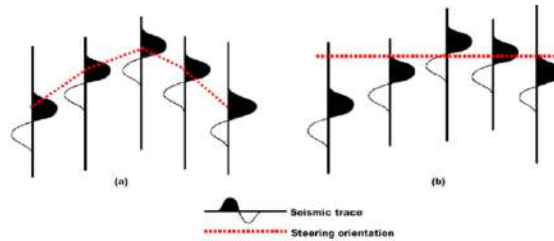


Figura 6.2. Comparação entre os dados com e sem *steering cube*, representando os traços sísmicos com informações de inclinação e azimute (linha tracejada) (Tingdahl, Kristofer & de Groot, Paul., 2003)..

Os dados de saída do método da estimativa do *steering cube* foram utilizados inicialmente como insumos para o cálculo de atributos sísmicos, e assim, realizar interpretações voltados principalmente na identificação de discontinuidades, que pudessem revelar feições estruturais e estratigráficas como demonstrado no trabalho de Marfurt et al. 1998, onde o cálculo do mergulho e azimute dos eventos sísmicos, incorporou maior robustez para uma análise multiatributo, ao se tratar de atributos sísmicos complexos como o de Coerência. Pouco tempo depois, foram apresentadas aplicações do *steering cube* como recurso para guiar filtros de mediana, com o intuito de remover ruídos não coerentes e melhorar a continuidade lateral dos eventos, preservando também bordas em dados sísmicos empilhados (Kristofer M. Tingdahl 1999, Kristofer M. Tingdahl e Paul F.M. De Groot 2003).

Existem diferentes metodologias para a estimativa de mergulho a partir da amplitude sísmica tais como *Fourier transformed* e *Phase-gradient*, e nesse trabalho foi utilizado a técnica de PCA (*Principal Component Analysis*) por ser entre os três citados, de mais rápida performance, o que permitiu gerar diferentes resultados em curto tempo e com menor exigência computacional. A técnica de Análise de Componentes Principais (PCA) empregada neste estudo utiliza a Decomposição de Autovalores (EVD - *Eigen Value Decomposition*) para identificar a direção das mudanças de amplitude máximas, representada como um vetor de três componentes, sendo o maior autovalor indicativo dessa direção (Wu, S., & Wang, Y., 2022).

A direção associada ao maior autovalor corresponde à normal à frente de onda em qualquer posição ao longo do refletor sísmico, sendo, portanto, perpendicular ao dip. Os segundo e terceiro autovalores situam-se no plano tangente ao refletor sísmico, fornecendo informações adicionais sobre a geometria local. O EVD é aplicado em uma base de amostra por amostra, utilizando seis produtos de gradiente de amplitude que foram previamente calculados. Esses gradientes são suavizados

volumetricamente usando filtros gaussianos recursivos, o que assegura uma representação mais suave e contínua dos dips.

6.2.2. SPECTRAL BLUEING

O Spectral Blueing é um método que combina o espectro sísmico médio com a forma do espectro da série de refletividade de perfis de poços, que é rica em altas frequências (Blache-Fraser, G., & Neep, J., 2004). Esse processo é usado para modelar operadores a fim de aumentar a quantidade de altas frequências nos dados sísmicos, melhorando sua resolução. Neste estudo, cinco poços distribuídos irregularmente foram utilizados e os operadores estimados a partir do Spectral Blueing foram definidos para variar no tempo, visto que a variação do espectro de amplitude com o tempo era evidente.

Um dado sísmico de alta resolução é fundamental para estudos qualitativos e quantitativos mais detalhados da subsuperfície, especialmente em formações delgadas, como o alvo deste trabalho. Conforme citado, o foco do estudo recairá sobre níveis estratigráficos ricos em carbono orgânico total, que podem ser encontrados em camadas de poucos metros de espessura. Nessas formações, os intervalos de folhelhos com maior radioatividade (e, portanto, maior potencial gerador) variam entre 20 e 40 metros de espessura, conforme mapeado em poços distribuídos pela bacia no trabalho de doutorado de Rodrigues (1995).

Portanto, para obter dados sísmicos de maior resolução, foi aplicada a técnica de Spectral Blueing para melhorar a resolução sem aumentar o ruído a níveis inaceitáveis. Essa técnica fundamenta-se na premissa de que os dados de refletividade obtidos em poços exibem, de maneira geral, um comportamento no qual amplitudes aumentam com frequências mais elevadas (conhecido como "*blue spectrum*"). No entanto, durante o processamento sísmico, as amplitudes costumam ser *whitened* (Blache-Fraser, G., & Neep, J. (2004); Walden, A. T., & Hosken, J. W. J., 1985).

Conforme detalhado no estudo de Kazemeini et al. (2010), os espectros de potência são frequentemente categorizados em cores como *white*, *blue* e *red*, de acordo com as propriedades dos sinais, que são proporcionais a f^β . As características específicas de cada tipo de sinal são ilustradas na Figura 6.3. O espectro de potência de um sinal *white* é plano, resultando em $\beta = 0$, o que indica ausência de correlação entre valores vizinhos e distribuição uniforme de energia em qualquer largura de banda de frequência. Em contraste, um sinal *red* apresenta correlação entre valores vizinhos e sua energia tende a se concentrar em frequências mais baixas, com $\beta < 0$. Por último, o sinal *blue* exibe correlação negativa entre valores vizinhos e concentra sua energia em frequências mais altas, com $\beta > 0$.

No estudo de Walden e Hosken (1985), de acordo com diversas séries de coeficientes de reflexão analisados, os resultados indicaram que a refletividade da Terra não é *white*, mas sim representada por uma função de f^β , onde $\beta > 0$ o que define um sinal *blue*. Esse tipo de refletividade sugere uma seção geológica composta por poucas camadas espessas e numerosas camadas finas em comparação com o comprimento de onda sísmico. Esse fenômeno é atribuído aos processos de deposição e erosão na Terra o que é diferente da atenuação da alta frequência durante a propagação da onda. Devido à natureza periódica dos processos deposicionais, é mais provável que os valores vizinhos na série de coeficientes de reflexão apresentem correlação negativa, resultando em um espectro *blue*. Considerando que uma série de refletividade tende a exibir um espectro *blue*, os métodos convencionais de *spectral whitening*, como a deconvolução de pico, que buscam alcançar um espectro plano *white*, acabam amplificando excessivamente as frequências baixas, comprometendo assim a resolução sísmica (Haffner e Cheadle, 1999; Saggaf e Robinson, 2000).

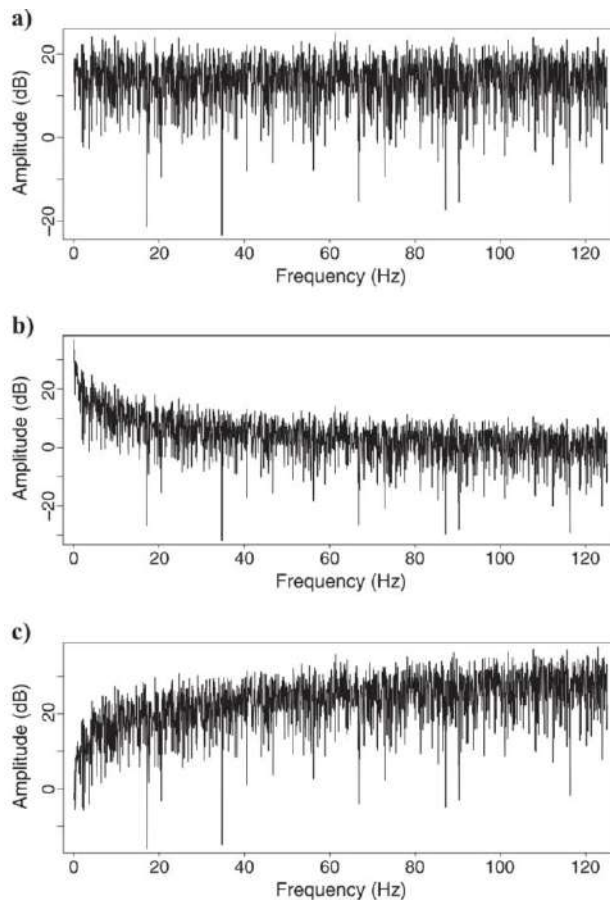


Figura 6.3. (a) Espectro do sinal branco, (b) espectro do sinal vermelho e (c) espectro do sinal azul proporcionais a f^β . O espectro plano do sinal branco demonstra que a energia é distribuída igualmente entre todas as frequências. Em contraste, para os sinais vermelho e azul, a energia é predominantemente concentrada em frequências baixas e altas, respectivamente (Walden e Hosken, 1985).

6.3. MÉTODO

6.3.1. DIP STEERED MEDIAN FILTER

O fluxo de trabalho para a primeira etapa de condicionamento do dado está detalhada na Figura 6.4, que inicia-se com a estimativa do steering cube que será utilizado para guiar o filtro de mediana para produzir dados condicionados menos ruidosos.

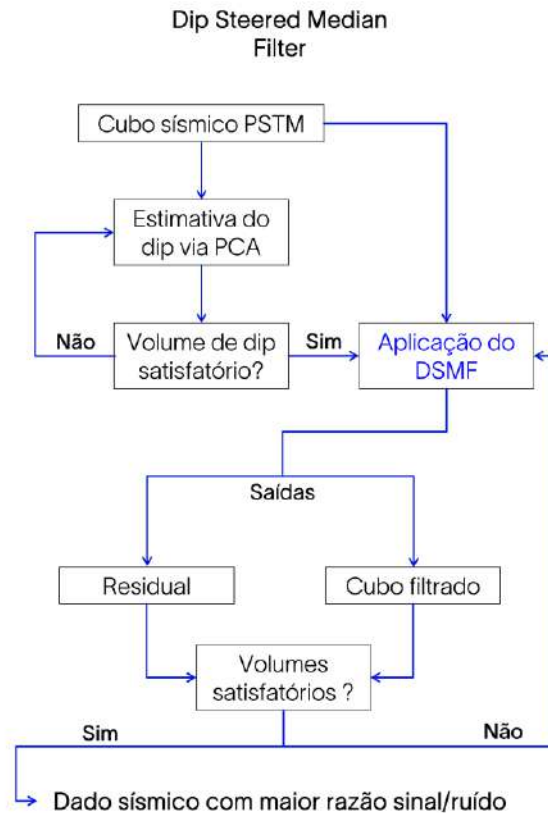


Figura 6.4. Fluxo de trabalho da técnica Dip Steered Median Filter.

6.3.1.1. Estimativa do steering cube

É observado no dado sísmico que os refletores, no geral, seguem um padrão plano paralelo, exceto em áreas onde ocorrem saltos de soleiras. Estes saltos não apenas formam altos estruturais, mas também podem gerar falhas de baixo rejeito. Por isso, ao parametrizar o *steering cube*, foi estabelecido como critério, um tamanho de janela capaz de identificar mudanças sutis no mergulho dos refletores. Isso é crucial para obter o melhor resultado na etapa seguinte de filtragem, garantindo que apenas o ruído seja observado no dado residual desta análise e evidenciando o que é interpretado como sinal sísmico relevante. A parametrização que atingiu os critérios estabelecidos foi a estimativa do mergulho a partir do método PCA, realizada a cada uma inline e uma crossline e a cada oito amostras em Z

(1x1x8). Diferentes parametrizações foram testadas, porém quanto maior o espaçamento entre a inline e a crossline mais suavizado era o resultado do volume de dip, e mudanças de mergulhos mais sutis e de baixo ângulo poderiam não ser bem representados.

Na Figura 6.5 pode ser observado o resultado do volume de dip sobre uma inline, crossline e uma time slice, apresentado em transparência sobre a amplitude sísmica. Neste atributo, é possível notar que as variações laterais do mergulho dos refletores ocorrem de forma muito suave e com baixos valores, sua maioria no entorno de 20° , onde as mudanças abruptas são observadas principalmente nas regiões de saltos de soleiras. Este comportamento era o esperado, visto o padrão sísmico predominantemente plano paralelo, e, portanto, considerou-se, este como a parametrização mais adequada para a obtenção do melhor resultado a ser utilizado como dado de entrada para o filtro posteriormente aplicado.

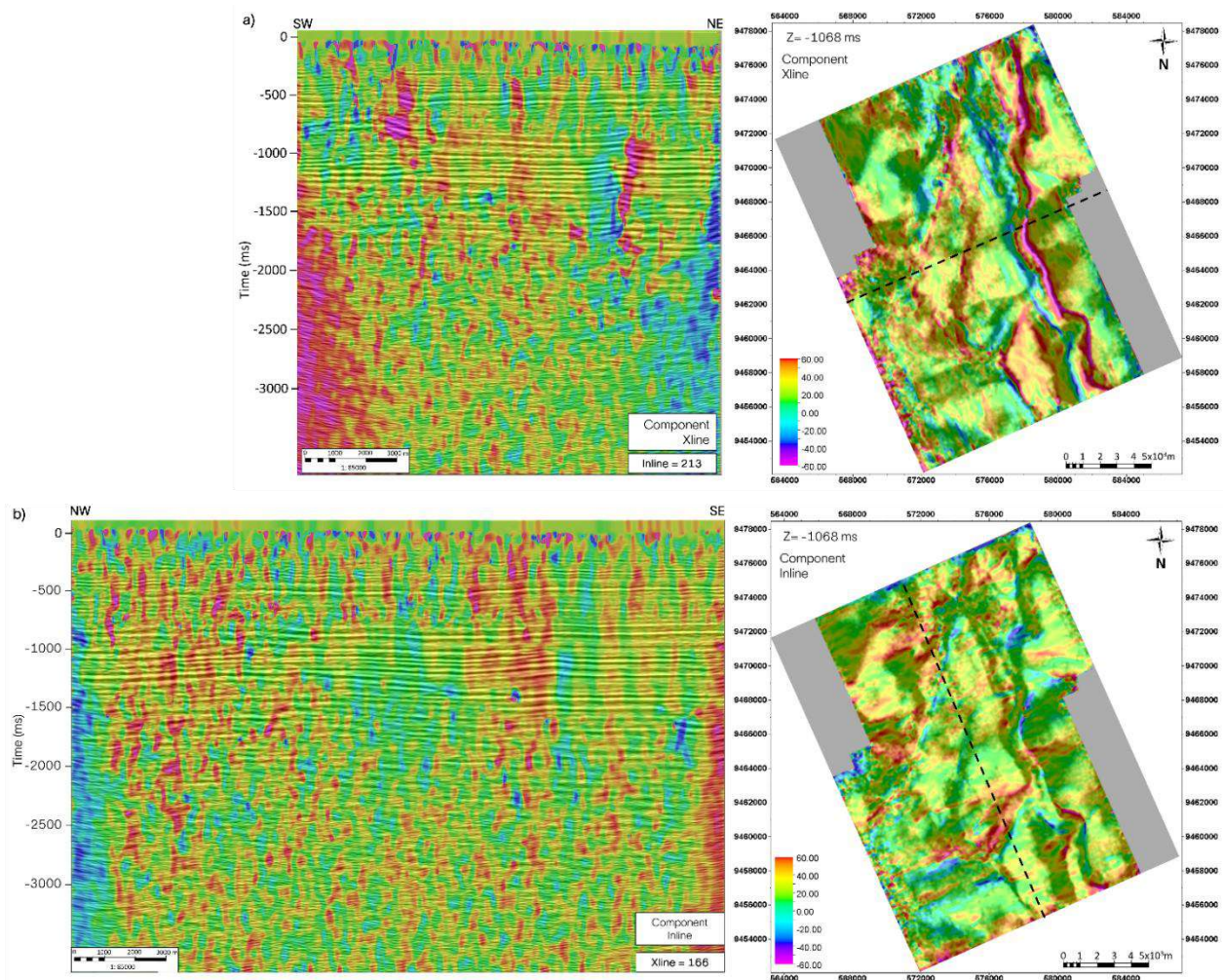


Figura 6.5. Atributo de volume de *dip* sobre a amplitude sísmica com transparência, destacando as maiores variações de mergulho dos refletores sísmicos. Em a) o componente de mergulho na direção da *crossline* é exibido sobre uma *inline* e em b) o componente de mergulho na direção da *inline* é exibido sobre uma *crossline*. O mesmo é demonstrado em uma *timeslice*.

6.3.1.2. Aplicação do filtro

O filtro de mediana atribuirá o valor mediano das amostras de entrada à posição central da amostra, tendo como guia o volume de mergulho dos eventos sísmicos calculados anteriormente. Geralmente este filtro é aplicado utilizando a uma mesma configuração de janela a qual foi adotada para calcular o steering cube. No entanto, o melhor desempenho alcançado neste estudo, foi aplicando-se o filtro a partir de uma janela com cinco amostras a cada duas inlines e duas crosslines (2x2x5). Além disso, foi observado melhores resultados, quando aplicado novamente o filtro sobre o dado já filtrado com essa mesma configuração de janela do DSMF. Em resumo, a técnica DSMF foi aplicado em duas interações com arranjo 2x2x5 tendo como dado de entrada, o steering cube estimado com o arranjo 1x1x8. Diferentes filtros foram testados no dado sísmico, alterando-se o tamanho das janelas do filtro, assim como a utilização de diversos volumes de mergulho como o dado de entrada. Porém, eles não foram tão eficientes em reduzir o máximo dos ruídos sem eliminar sinais importantes. Usualmente, aplica-se o mesmo tamanho de janelas nessas duas etapas, no entanto, a partir dos diversos testes e combinações de parâmetros, os mesmos não produziram resultados tão eficientes quanto a combinação escolhida neste trabalho.

É fundamental que se faça uma análise criteriosa, após a aplicação do filtro principalmente com a extração do cubo residual entre os dados antes e após a filtragem. Uma vez que, a área coberta pelo dado sísmico 3D era relativamente pequena (menor que 1GB de dado), foi possível gerar múltiplas realizações modificando os parâmetros de janelas para o dado completo, à medida que se analisava os resultados obtidos. Mesmo utilizando-se de técnicas mais otimizadas para a extração do volume de mergulho dos eventos sísmicos, para volumes de dados sísmicos mais densos cobrindo áreas muito extensas, esse processo pode ser bastante custoso computacionalmente, sendo necessário aplicar o filtro em subvolumes para identificar os parâmetros mais adequados de acordo com a natureza dos ruídos e a complexidade geológica.

O dado com menor conteúdo de ruídos, obtido após essa primeira etapa de condicionamento, foi então utilizado como dado de entrada para a segunda etapa do fluxo de trabalho proposto, conforme será detalhado a seguir.

6.3.2. SPECTRAL BLUEING

A proposta da técnica spectral blueing é combinar o espectro sísmico médio com o espectro de refletividade dos poços para projetar e aplicar um ou mais operadores (wavelets) ao dado pós-empilhamento, como mostrado no fluxo de trabalho da Figura 6.6. Dessa forma, é possível aumentar a

quantidade de altas frequências no dado sísmico, melhorando a resolução (Blache-Fraser e Neep, 2004). Vale destacar que essa técnica aprimora a resolução apenas dentro da banda de frequência sísmica, sem ultrapassá-la.

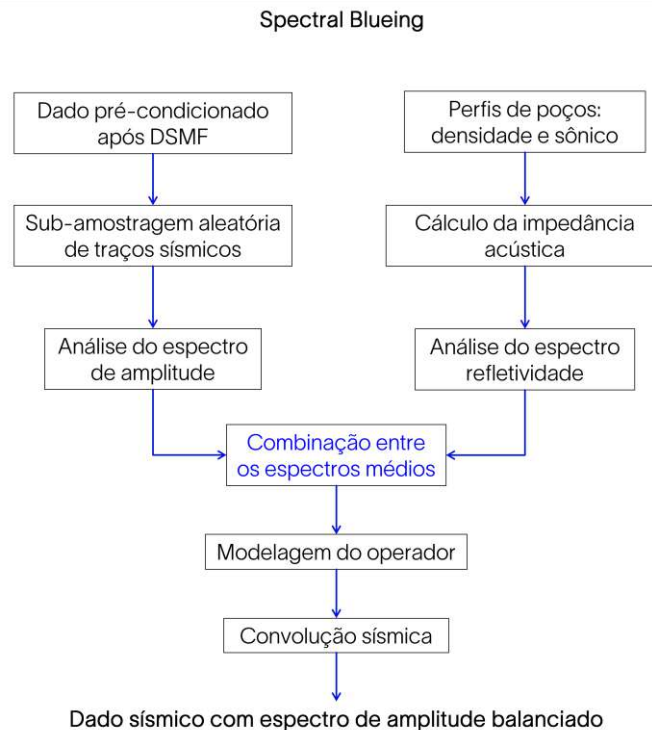


Figura 6.6. Fluxo de trabalho da técnica Spectral Blueing (Blache-Fraser e Neep, 2004).

No entanto, ao projetar o operador com base em um espectro sísmico médio dentro de uma janela de tempo ampla, pode ocorrer uma superamplificação das frequências altas nas regiões mais rasas e das frequências baixas nas regiões mais profundas da seção sísmica (Neep, 2007).

Isso resulta na formação de lóbulos laterais nas áreas mais rasas e na diminuição da resolução nas áreas mais profundas. Para solucionar esse problema, Neep (2007) propõe a técnica de "time variant spectral blueing", na qual múltiplos operadores são criados para diferentes janelas de tempo. Uma vez que foi observada uma variação no espectro de frequência do dado para diferentes intervalos de tempo, essa técnica foi aplicada seguindo as seguintes etapas:

- 1) Definição dos zonas de intervalo de tempo para os quais seria modelado de forma individual cada operador (Figura 6.7). Essa etapa foi baseada nos intervalos observados para a geração das wavelets estatísticas, as quais foram extraídas para realizar o processo de amarração sísmica-poço no dado original (antes do pré-condicionamento). Nesse trabalho, também se optou por realizar a amarração sísmica-poço utilizando três wavelets estatísticas extraídas em três diferentes intervalos. O processo de amarração será descrito mais adiante pois este será um dos controle de qualidade utilizado e por ter sido realizado em três momentos

diferentes (1) no dado original, (2) no dado após a primeira etapa do condicionamento de atenuação dos ruídos via Dip steered median filter e por fim (3) no dado após o fluxo completo de condicionamento resultado do spectral blueing.

- 2) Foram selecionados aleatoriamente quarenta traços sísmicos para gerar a média do espectro de amplitude do dado sísmico (Figura 6.8a e b).
- 3) O perfil de impedância acústica dos poços utilizados no estudo é transformado em séries de refletividade, em seguida é calculado o espectro de amplitude tanto individualmente por poço quanto o espectro médio (Figura 6.8c e d)
- 4) Inicialmente um operador é modelado automaticamente ao combinar e ajustar os dados de entrada pois essa técnica foi desenvolvida para ser orientado ao dado. No entanto é importante que se faça ajustes nos parâmetros de corte de frequência alta e baixa, assim como no slope do operador (Figura 6.9a). Alguns recursos para o controle de qualidade podem ser utilizados durante o processo da modelagem dos operadores, como a correlação entre o traço sísmico e do poço (Tabela 6.1), além de uma análise qualitativa do resultado, uma vez que o operador modelado é convolvido com o dado sísmico em tempo real.
- 5) Por fim, é realizado a convolução dos três operadores, modelados para três diferentes intervalos de tempo, com o dado sísmico.

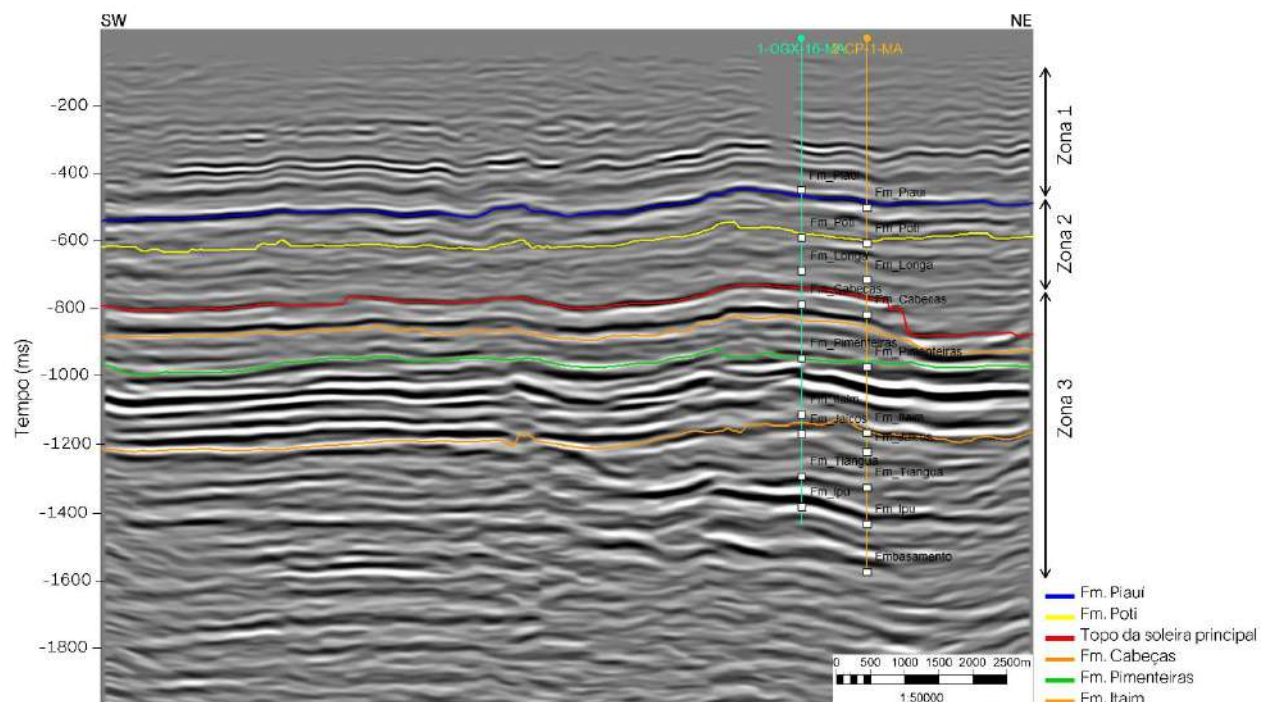


Figura 6.7. A definição de intervalos de tempo para a modelagem individual de cada operador, considerando a variação vertical da frequência do dado sísmico. A interpretação das principais seqüências dentro da bacia serve como referência eficaz para a delimitação de cada zona.

A primeira zona foi definida no intervalo entre 140 ms a 500 ms, que compreende formações mais recentes à Formação Piauí, compostas por rochas clásticas, carbonáticas e evaporíticas como as camadas de anidrita encontradas na Formação Pedra de Fogo. Foi estabelecido para a segunda zona o intervalo entre 500 ms e 750 ms, portanto, o intervalo constituído de rochas clásticas (arenitos e folhelhos) das Formações Piauí e Poti. A última zona definida (entre a Formação Cabeças e o topo do embasamento), contém o intervalo de interesse deste trabalho de forma geral, entre 750 ms a 1600 ms aproximadamente, e foi assim separada das demais, por compor sequências estratigráficas que foram cortadas por intrusões ígneas. Essas mudanças litoestratigráficas ficam muito bem refletidas na variação do conteúdo de frequência do dado sísmico, que varia não somente pela atenuação natural do sinal com a profundidade, mas também por conter rochas de maior absorção de energia da onda acústica, como é o caso das espessas camadas de soleiras de diabásio que ocorrem na terceira zona.

Vale ressaltar que as zonas definidas são cortes de tempo e por isso, os intervalos estratigráficos citados serviram como uma referência para o melhor entendimento da variação do espectro de frequência do dado analisado com a geologia e Formação predominante de cada intervalo (Figura 6.7).

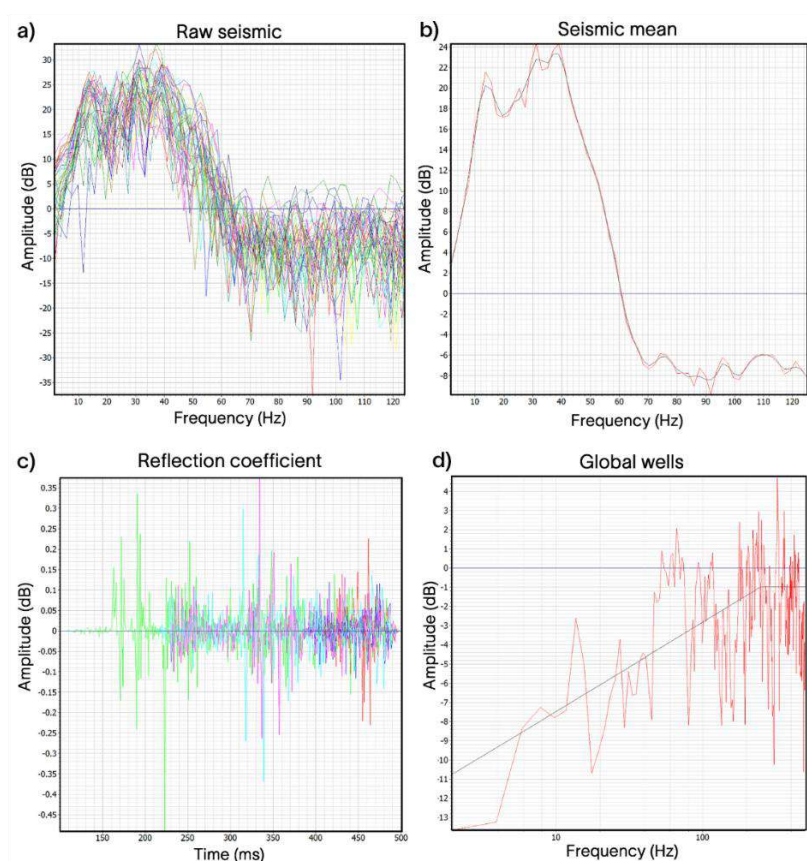


Figura 6.8. Dados de entrada utilizados para modelar o operador da primeira zona, onde em (a) pode ser observado o espectro de frequência dos quarenta traços sísmicos selecionados aleatoriamente, em (b) o espectro sísmico médio e o mesmo suavizado como

mostrado pelo traço em preto, em (c) as séries de refletividade derivadas de cada poço, e por último, em (d) é exibido o espectro de frequência médio dos poços, o qual foi filtrado em 250 Hz, e também a curva de ajuste do espectro (em preto).

Na Figura 6.9a é exibido o espectro médio do coeficiente de reflexão limitado por banda (verde), o espectro médio sísmico suavizado (vermelho) e o espectro do spectral blueing derivado (azul). O operador Spectral Blueing molda o espectro médio sísmico suavizado para o ajuste da curva do coeficiente de reflexão limitado por banda em cada frequência. Ainda na Figura 6.9b é mostrado o resultado do operador que será posteriormente convolvido com o dado sísmico no intervalo da zona 1 e na Figura 6.9c o seu espectro de frequência.

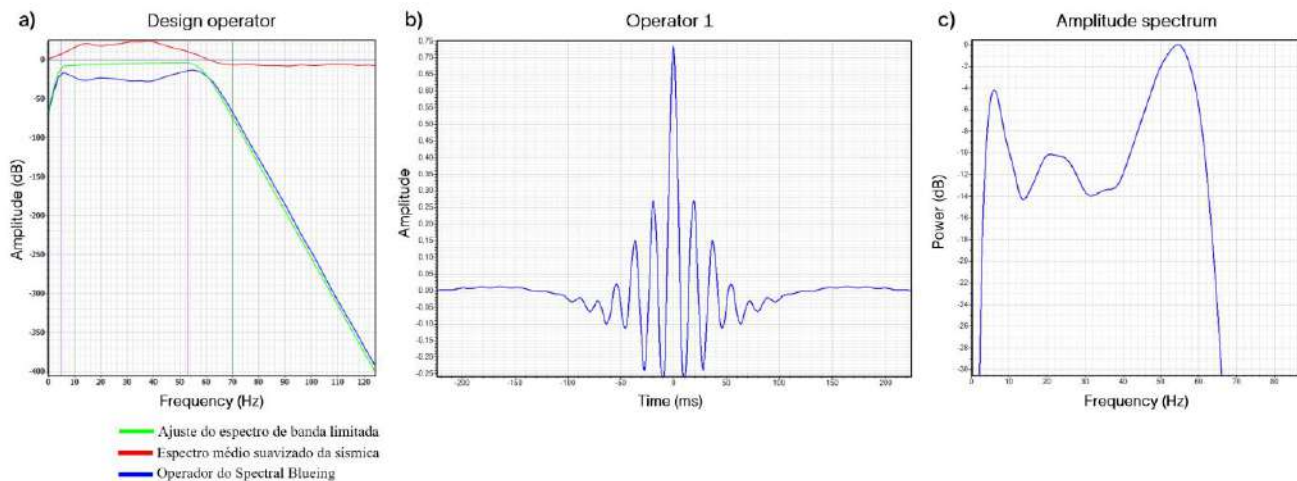


Figura 6.9. Construção do operador para a primeira zona, em a) os cortes de frequência baseado nos espectros do dado de entrada e saída, em b) o resultado do operador e em c) o espectro de frequência do sinal modelado.

As parametrizações utilizadas para os cortes de alta e baixa frequência que modularam o operador estão expressas na Tabela 1a, assim como os valores do coeficiente Beta (f^β) em (a) ajustado por poço e o valor da média utilizada na operação. Além disso, ainda na Tabela 1b tem se exibido o valor da correlação entre o traço sísmico e o traço extraído nos poços. Esses valores de correlação podem ser utilizados para calibrar o operador modelado, porém é muito importante ressaltar que esses valores estarão relacionados diretamente a um bom ajuste entre poço e a sísmica após a amarração. Dessa forma, é essencial que se faça uma avaliação quantitativa associada a uma avaliação qualitativa, observando o resultado preliminar da convolução entre o operador que está sendo modelado com a sísmica, assim ficará evidente quais são as possíveis causas de uma baixa correlação.

Tabela 6.1. Parâmetros utilizados para modelar o operador da primeira zona. Em a) aos valores dos cortes de frequências e em b) é mostrado as correlações com os poços após a aplicação do operador com a sísmica na zona 1, e os valores dos coeficientes beta.

a)

Low Cut		High Cut	
Full (F2):	5 Hz	Full (F3):	53 Hz
-60 dB (F1):	0 Hz	-60 dB (F4):	70 Hz
Slope:	26 dB/oct.	Slope:	-150 dB/oct.
Attenuation:	-60.4 dB	Attenuation:	-60.2 dB

b)

Well Name	Well Log	Well to Seismic Corr. Coef.	Beta
3-OGX-51DP-MA	AI	0.62	0.340
3-OGX-57-MA	AI	0.57	0.359
2-CP-1-MA	AI	0.43	0.090
1-OGX-16-MA	AI	0.08	0.249
1-OGX-22-MA	AI	0.06	0.095
Average Correlation Coefficient	0.35	Average Beta	0.233

Por exemplo, observando a Figura 6.10 onde é feita a avaliação do dado antes e após o spectral blueing, o poço 1-OGX-16-MA, na zona 1, além de estar localizado em uma região de sobre na sísmica (ausência completa de refletores sísmicos) também é possível notar um desajuste entre os refletores do poço e da sísmica, assim como no caso do poço 1-OGX-22-MA. No entanto, como este não é o intervalo de interesse do estudo de maneira geral, espera-se obter valores de correlação mais baixas, uma vez que, a amarração poço/sísmica realizado teve como foco a zona 3. Logo, será observado que para a maioria dos poços o a amarração realizada na zona 3 foi o suficiente para ajustar consideravelmente bem todo o poço.

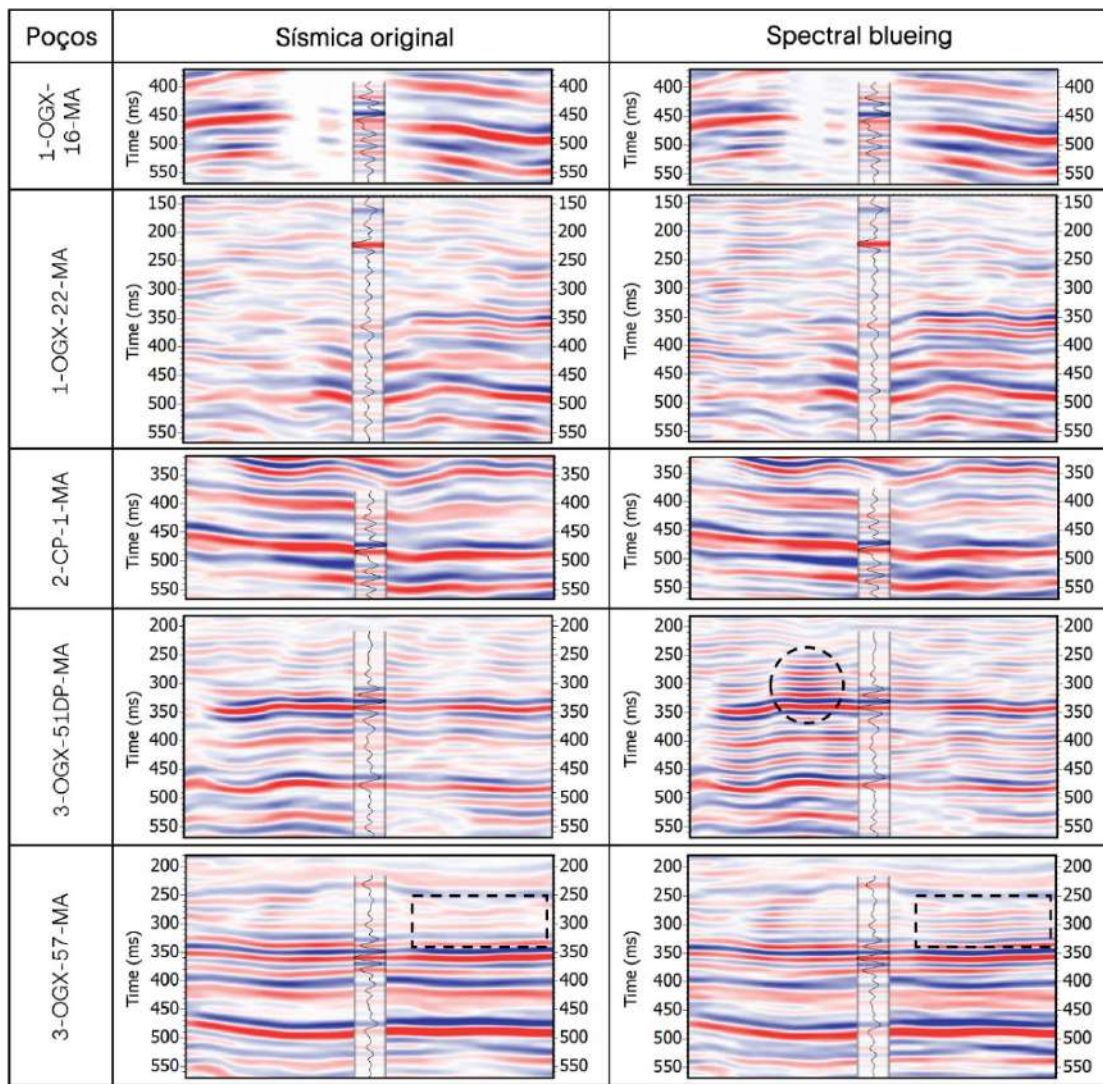


Figura 6.10. Comparação entre o dado sísmico original e após a aplicação do spectral blueing. É destacado no retângulo tracejado preto como é melhorado a continuação lateral dos refletores, bem como a resolução após a aplicação do fluxo do SBB. Também é evidenciado pelo círculo preto tracejado um efeito chamado de ringing (reverberação) que pode ocorrer nesse tipo de técnica.

Uma das razões de aplicar essa técnica variando com o tempo é para evitar a ocorrência do efeito *ringing* como destacado na Figura 10 pelo círculo preto tracejado, que neste estudo teve esse efeito minimizado quando se avalia o resultado no volume sísmico 3D pré-condicionado. O estudo publicado por Kazemeini, S. H. *et al.* 2010, mostrou que a aplicação desta técnica em dados pré-empilhamento e em dados pós-empilhamento aumenta a resolução sísmica da mesma maneira, no entanto, em dados pré-empilhamento o artefato *ringing* é minimizado quando comparados com o resultado do spectral blueing no dado empilhado.

Assim como mostrado para a primeira zona, na Figura 6.11 está exibido os dados de entrada que foram utilizados para modelar o operador que será convolvido com esse segundo intervalo de

tempo. Aqui vale ressaltar a variação com o tempo do conteúdo de frequência em distribuição e potência.

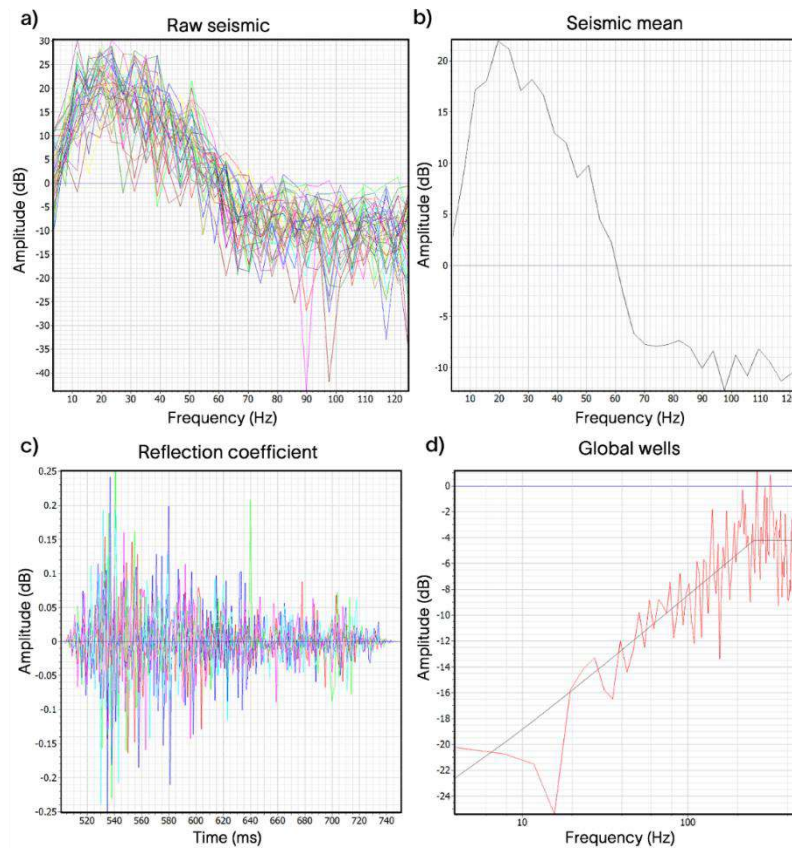


Figura 6.11. Modelagem do operador para a segunda zona. Em a) o espectro de frequência dos quarenta traços sísmicos em b) o espectro sísmico médio, em c) as séries de refletividade derivadas de cada poço, e em d) o espectro de frequência médio dos poços com a curva de ajuste do espectro (em preto).

Após os ajustes realizados nos cortes de frequência como exibidos numericamente na Tabela 6.2a e visualmente na Figura 6.12a e o valor médio de beta (Tabela 6.2b) ajustado ao valor médio do espectro de frequência do coeficiente de reflexão dos poços foi obtido o operador 2 como mostrado na Figura 6.12b.

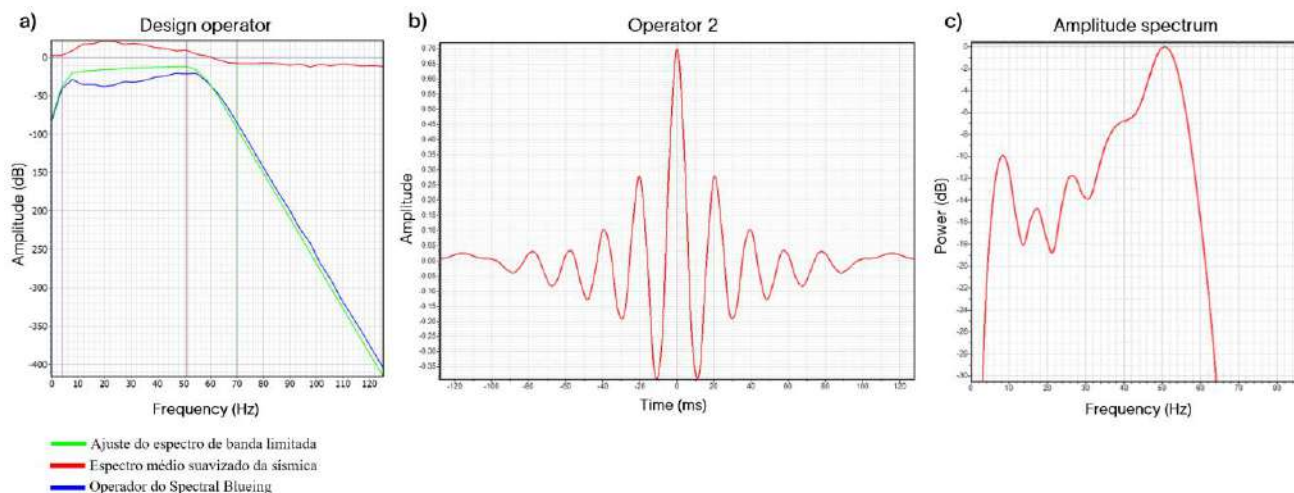


Figura 6.12. Construção do operador para a segunda zona, em a) os cortes de frequência baseado nos espectros do dado de entrada e saída, em b) o resultado do operador e em c) o espectro de frequência do sinal modelado.

A correlação média entre o traço sísmico e do poço para esta segunda zona é melhor quando comparado com o da primeira zona (Tabela 6.2b). No resultado preliminar (Figura 6.13), pode-se observar uma melhora significativa da continuidade lateral dos refletores assim como a resolução, e que neste intervalo a correlação também está sendo afetada por um pequeno deslocamento vertical entre os traços.

Tabela 6.2. Parâmetros utilizados para modelar o operador da segunda zona. Em a) aos valores dos cortes de frequências e em b) é mostrado as correlações com os poços após a aplicação do operador com a sísmica na zona 1, e os valores dos coeficientes beta.

a)

Low Cut		High Cut	
Full (F2):	4 Hz	Full (F3):	51 Hz
-60 dB (F1):	0 Hz	-60 dB (F4):	70 Hz
Slope:	30 dB/oct.	Slope:	-131 dB/oct.
Attenuation:	-60.0 dB	Attenuation:	-59.8 dB

b)

Well Name	Well Log	Well to Seismic Corr. Coef.	Beta
3-OGX-57-MA	AI	0.64	0.528
2-CP-1-MA	AI	0.49	0.692
3-OGX-51DP-MA	AI	0.40	0.659
1-OGX-22-MA	AI	0.39	0.197
1-OGX-16-MA	AI	0.35	0.455
Average Correlation Coefficient	0.45	Average Beta	0.538

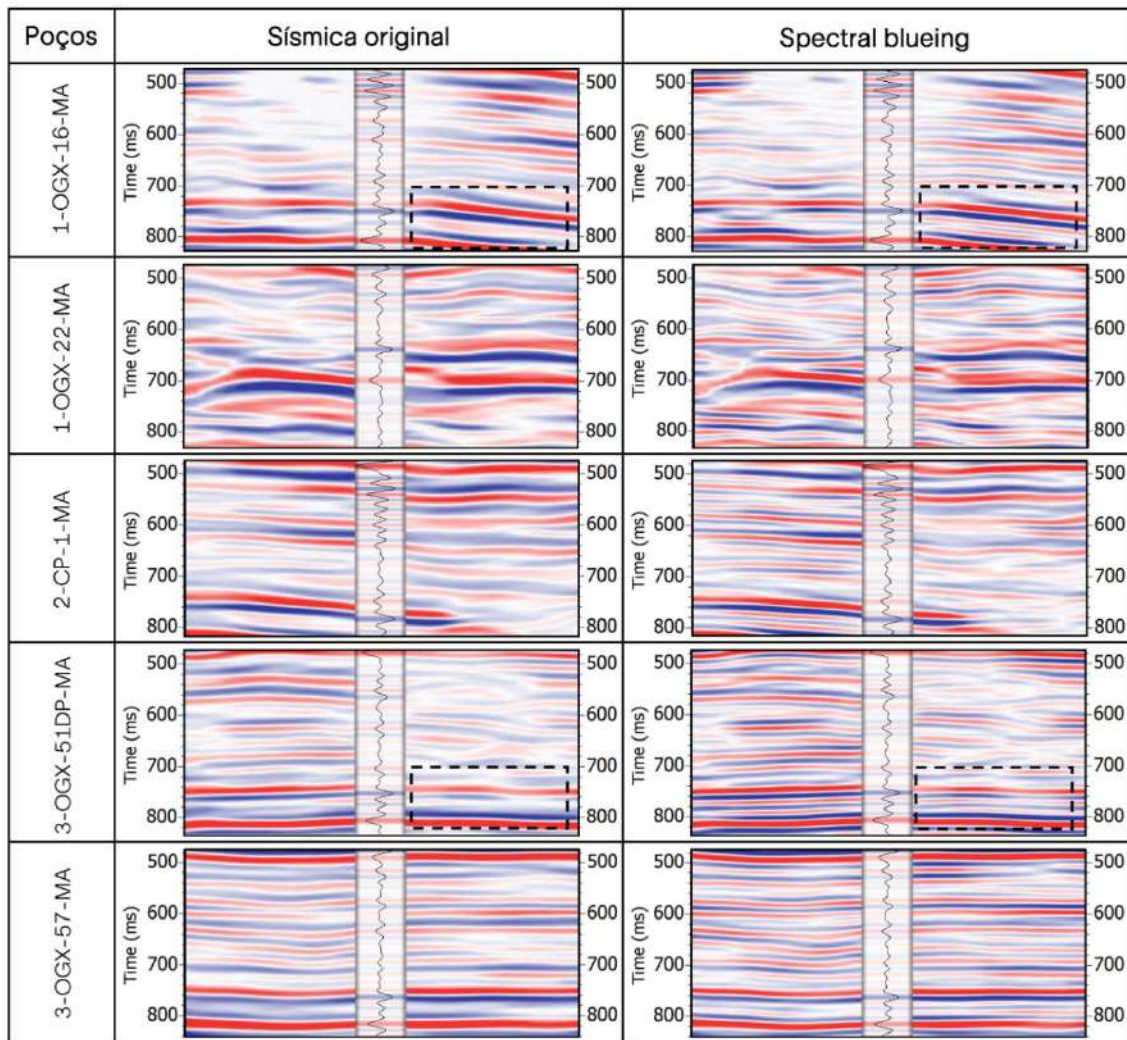


Figura 6.13. Os resultados preliminares da convolução utilizando o operador modelado para a segunda zona indicam seções sísmicas que passam pelos poços. Isso permite uma avaliação da correlação entre os traços sísmicos extraídos dos poços e os refletos sísmicos. Além disso, permite uma comparação dos dados antes e depois do SSB, destacando as melhorias dentro do retângulo preto tracejado.

Para a última zona, a intervalo de interesse deste trabalho, também é demonstrado todos os dados de entrada para a modelagem do operador que aumentará o conteúdo de altas frequência para o dado sísmico final. Na Figura 6.14, pode ser observado os quarenta traços sísmicos em a), a média do espectro de frequência dos quarenta traços sísmicos em b), os coeficientes de reflexão extraídos em cada poço em c) e por fim em e) temos a média do espectro de frequência dos poços que definirá o valor médio dos coeficientes beta que está expresso na Tabela 6.3b.

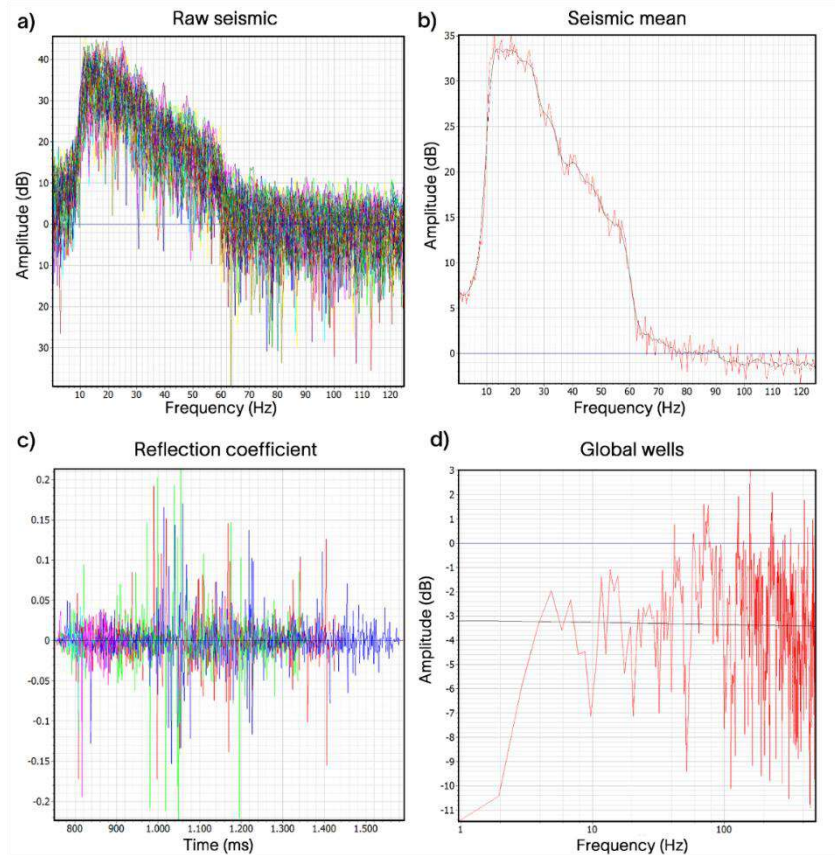


Figura 6.14. Modelagem do operador para a terceira zona. Em a) o espectro de frequência dos quarenta traços sísmicos em b) o espectro sísmico médio, em c) as séries de refletividade derivadas de cada poço, e em (d) o espectro de frequência médio dos poços com a curva de ajuste do espectro (em preto).

A forma como o operador foi projetado pode ser avaliada pela Figura 6.15a junto com a Tabela 6.3a contendo os valores das frequências de corte. Ainda na Figura 6.15b e Figura 6.15c temos o resultado do operador 3 e o espectro de frequência dele respectivamente.

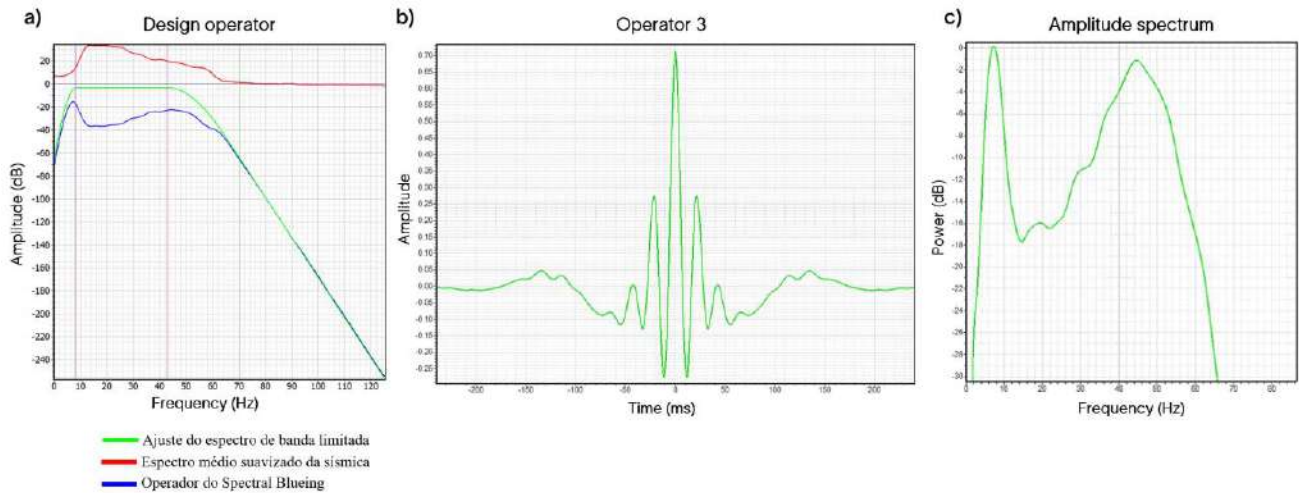


Figura 6.15. Construção do operador para a terceira zona, em a) os cortes de frequência baseado nos espectros do dado de entrada e saída, em b) o resultado do operador e em c) o espectro de frequência do sinal modelado.

Como esperado, o coeficiente de correlação entre o traço extraído nos poços e os traços sísmicos na zona 3 é maior comparado com as demais zonas, devido a amarração poço/sísmica (Tabela 6.3b).

Tabela 6.3. Parâmetros utilizados para modelar o operador da terceira zona. Em a) aos valores dos cortes de frequências e em b) é mostrado as correlações com os poços após a aplicação do operador com a sísmica na zona 1, e os valores dos coeficientes beta.

a)

Low Cut		High Cut	
Full (F2):	8 Hz	Full (F3):	43 Hz
-60 dB (F1):	0 Hz	-60 dB (F4):	70 Hz
Slope:	20 dB/oct.	Slope:	-85 dB/oct.
Attenuation:	-60.0 dB	Attenuation:	-59.8 dB

b)

Well Name	Well Log	Well to Seismic Corr. Coef.	Beta
3-OGX-57-MA	AI	0.84	0.211
3-OGX-51DP-MA	AI	0.79	0.186
1-OGX-16-MA	AI	0.61	0.146
2-CP-1-MA	AI	0.61	0.173
1-OGX-22-MA	AI	0.49	0.182
Average Correlation Coefficient	0.67	Average Beta	0.180

As seções sísmicas passando pelos poços para a zona 3, podem ser observadas na Figura 6.16, onde foram destacadas diversas regiões de melhoria do dado após a aplicação do spectral blueing. O resultado no geral mostra um ganho qualitativo de resolução onde os refletores estão mais bem definidos, com menor espessura temporal, melhor continuidade lateral dos refletores sísmicos. Além disso, a comparação entre o traço sísmico e o traço calculado nos poços mostram uma alta correlação dos eventos, o que permitiu ajustar de maneira mais consistente que as demais zonas as parametrizações realizadas durante a modelagem dos operadores os quais geraram o resultado deste estudo de condicionamento.

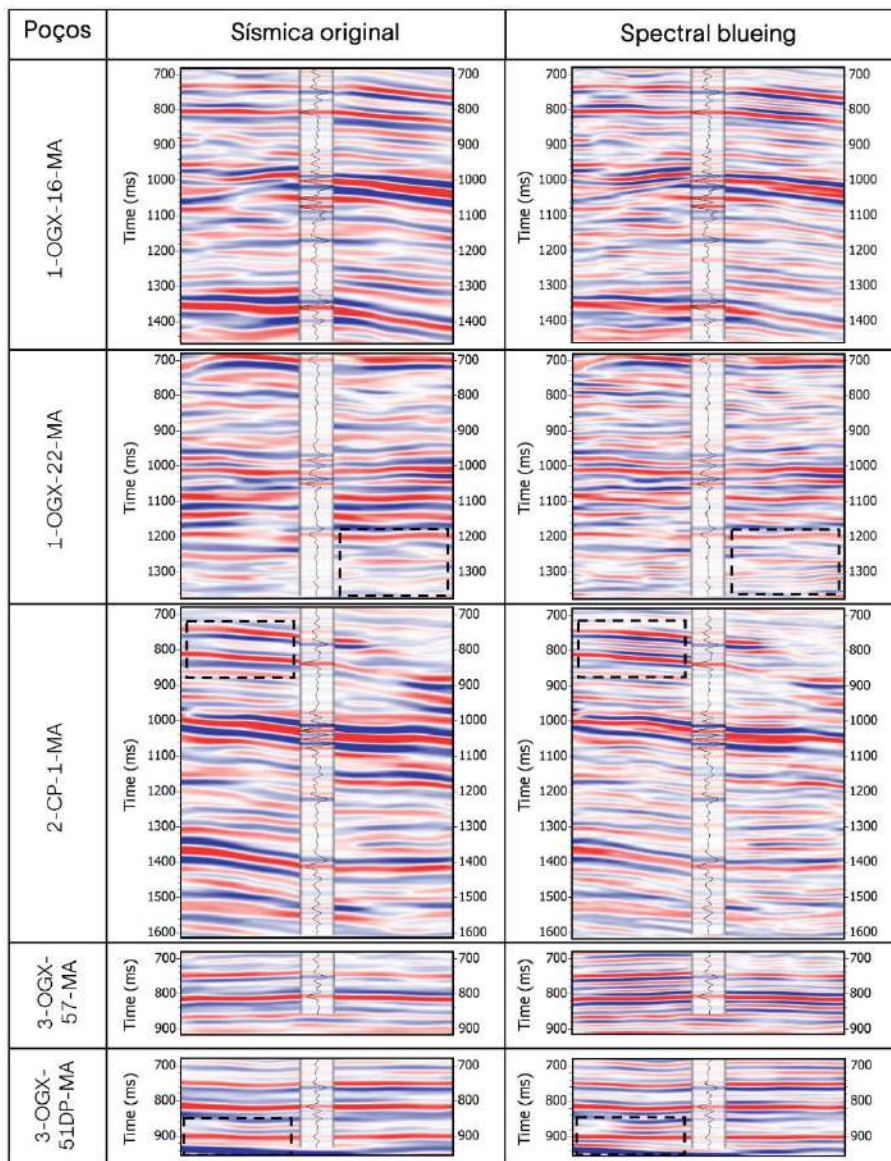


Figura 6.16. Os resultados preliminares da convolução utilizando o operador modelado para a terceira zona são apresentados, mostrando seções sísmicas que passam pelos poços. Isso nos permite avaliar a correlação entre os traços sísmicos extraídos dos poços e os refletores sísmicos. Além disso, podemos comparar os dados antes e depois do SSB, destacando as melhorias indicadas pelo retângulo preto tracejado.

6.5. RESULTADOS E DISCUSSÕES

O fluxo de condicionamento sísmico adotado foi fundamental para a atenuação de ruídos, especialmente o ruído de "footprint" e ruídos aleatórios como pode ser observado na Figura 6.17. É notavelmente evidente que o DSMF foi muito eficiente quanto a suavização do dado sem remover estruturas e sinais sísmicos de interesse, resultando em um dado sísmico com melhor continuidade lateral, tanto em porções mais rasas quanto em regiões mais profundas, onde comumente é observado uma relação sinal/ruído baixa à medida que se aproxima do embasamento. Mesmo sendo um dado

sísmico terrestre que pode ser considerado de boa qualidade quando comparados com dados sísmicos de outras bacias terrestres brasileiras, como a Bacia do Paraná, também conhecida por ter espessas camadas de soleiras de diabásio que atenuam fortemente a propagação das ondas sísmicas, Bacia do Reconcavo, Acre, Amazonas, entre outras, para estudos de caracterização sísmica com foco em inversão e modelagem de intervalos delgados e afetados pelas intrusões ígneas torna-se fundamental a obtenção de dados melhorados, menos ruidosos para que as feições estratigráficas e estruturais sejam ressaltadas. No resultado mostrado em corte no tempo e em seção na Figura 6.17 é possível observar expressões sísmicas mais bem definidas na zona de interesse (folhelhos da Fm. Pimenteiras abaixo de 900 ms), mas também fácies sísmicas nos arenitos da Fm. Poti que antes do condicionamento predominava um padrão caótico e refletores enfraquecidos.

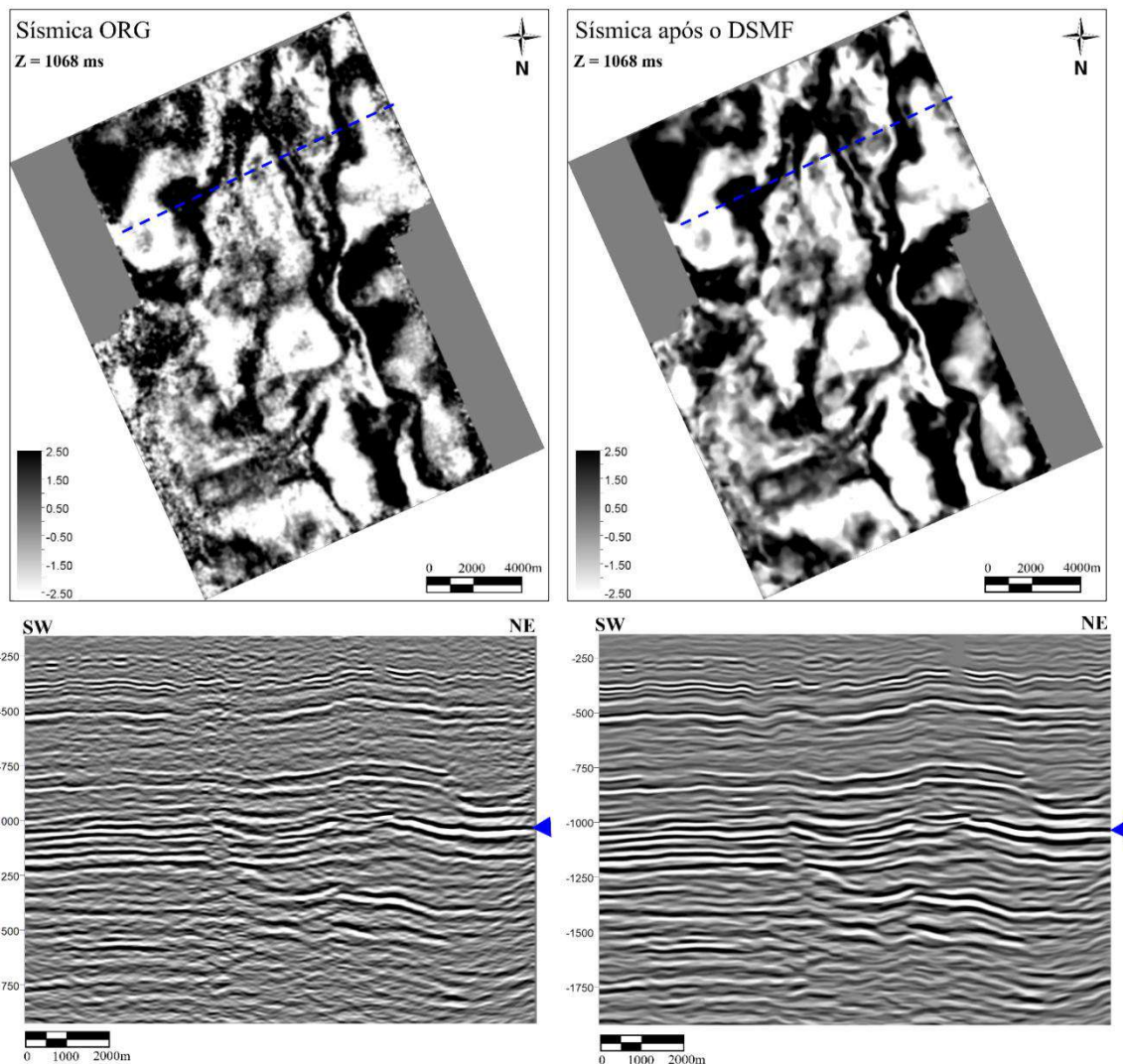


Figura 6.17. Resultado da aplicação da primeira etapa do condicionamento. É possível notar a atenuação significativa do ruído no dado após o DSMF em comparação com o dado sísmico original. A linha azul tracejada sobre o *time slice* identifica a localização da *inline* mostrada, e o triângulo azul identifica por onde passa a *time slice* exibida como referencial.

No dado residual obtido pela diferença entre o dado original e o dado condicionado pelo DSMF (Figura 6.18), reforça que grande parte do ruído de alta energia foi filtrado do dado sísmico. Também é possível observar que parte do sinal sísmico que representa alguma informação importante foi razoavelmente removido, o qual podemos associar principalmente aos refletores fechamento estrutural das intrusões ígneas que ganharam ângulos de mergulho quando se movimentaram verticalmente para cima, movimentos conhecidos como saltos de soleiras. Mesmo que esses sinais tenham sido atenuados pelo filtro, este foi considerado o melhor resultado possível, uma vez que, esse padrão não foi observado de forma persistente por todo o dado sísmico, logo, a suavização não se deu de forma agressiva nas bordas dos refletores de maior mergulho. Este é um importante estudo de controle de qualidade para o método aplicado, o qual auxiliará na parametrização tanto do *steering cube* quanto do filtro de mediana.

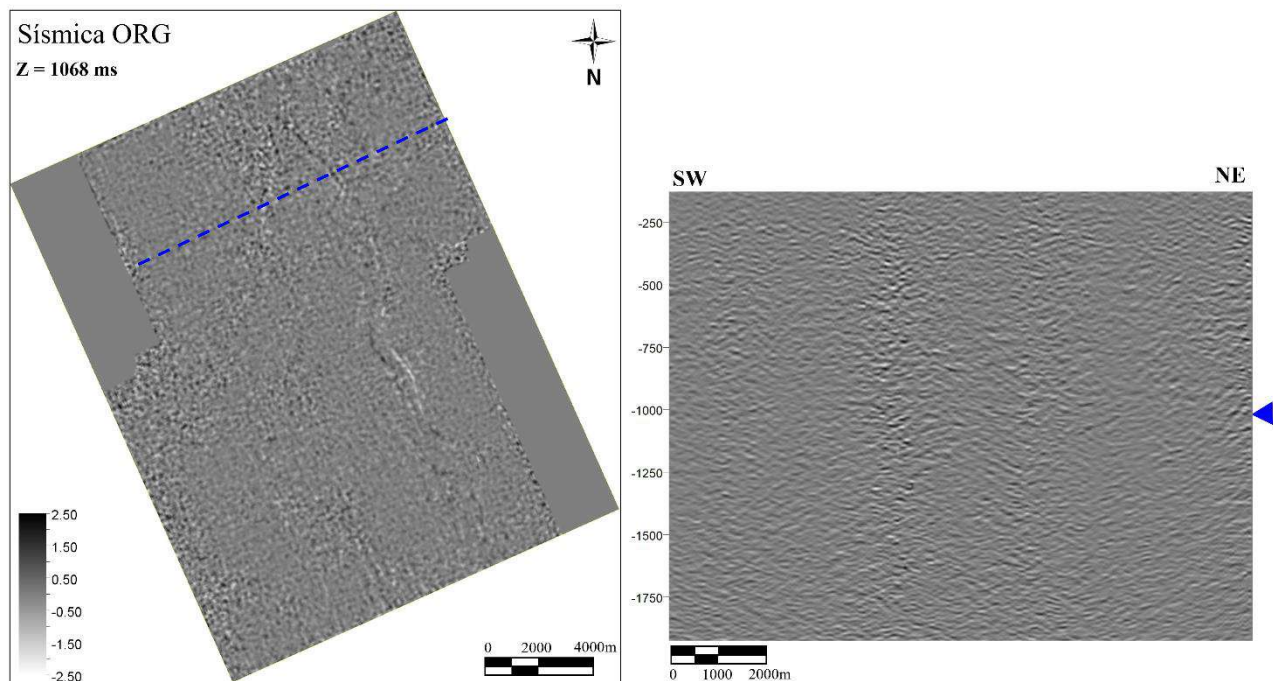


Figura 6.18. O volume de amplitude residual foi extraído da diferença entre os dados sísmicos original e o dado filtrado pelo DSMF, o que revelou principalmente uma redução de ruído e, em menor grau, a filtragem de alguns eventos sísmicos na direção NW-SE.

O resultado da segunda etapa do condicionamento, com a aplicação do *Spectral Blueing* sobre o dado filtrado, pode ser avaliado nas seções sísmica (*inline*) e nos cortes em tempo (*time slice*) apresentada na Figura 6.19. É notável o ganho de resolução e energia de alta frequência do dado após o SSB, assim como a melhor definição e continuidade lateral dos refletores. A quantificação do ganho de resolução sísmica obtido após essa abordagem de condicionamento será detalhada no Capítulo

7, porém de forma qualitativa podemos observar o nível de detalhe dos eventos sísmicos alcançado a partir deste fluxo de trabalho, sem que possíveis ruídos fossem ressaltados.

Além dos ganhos obtidos com esta técnica, é importante notar que o método SSB, mesmo quando aplicado variando no tempo, pode produzir efeitos de *ringing* devido aos lóbulos laterais do operador. Embora esses efeitos sejam diminuídos, eles ainda podem ser detectados nos dados. O estudo publicado por Kazemeini, S. H. et al. (2010) mostrou que a aplicação desta técnica em dados pré-empilhamento e pós-empilhamento melhora a resolução sísmica de maneira semelhante. No entanto, em dados pré-empilhamento, os artefatos como os de *ringing* podem ser minimizados de forma mais significativa em comparação com os resultados do *Spectral Blueing* aplicado em dados pós-empilhamento. Apesar deste efeito, os resultados do estudo atual produziram dados com resolução melhorada e definição mais clara de eventos sísmicos. Este aprimoramento é crucial para identificar estruturas geológicas e características estratigráficas, bem como para interpretar quantitativamente do intervalo de interesse. Isso é particularmente relevante em áreas como a Bacia do Parnaíba, onde intrusões ígneas afetam a qualidade dos dados. Além disso, após o aprimoramento, torna-se possível interpretar camadas mais finas nos dados sísmicos, facilitando estudos mais detalhados e precisos relacionados a amarração sísmico, poço, inversão sísmica e modelagem dos folhelhos geradores da área de estudo.

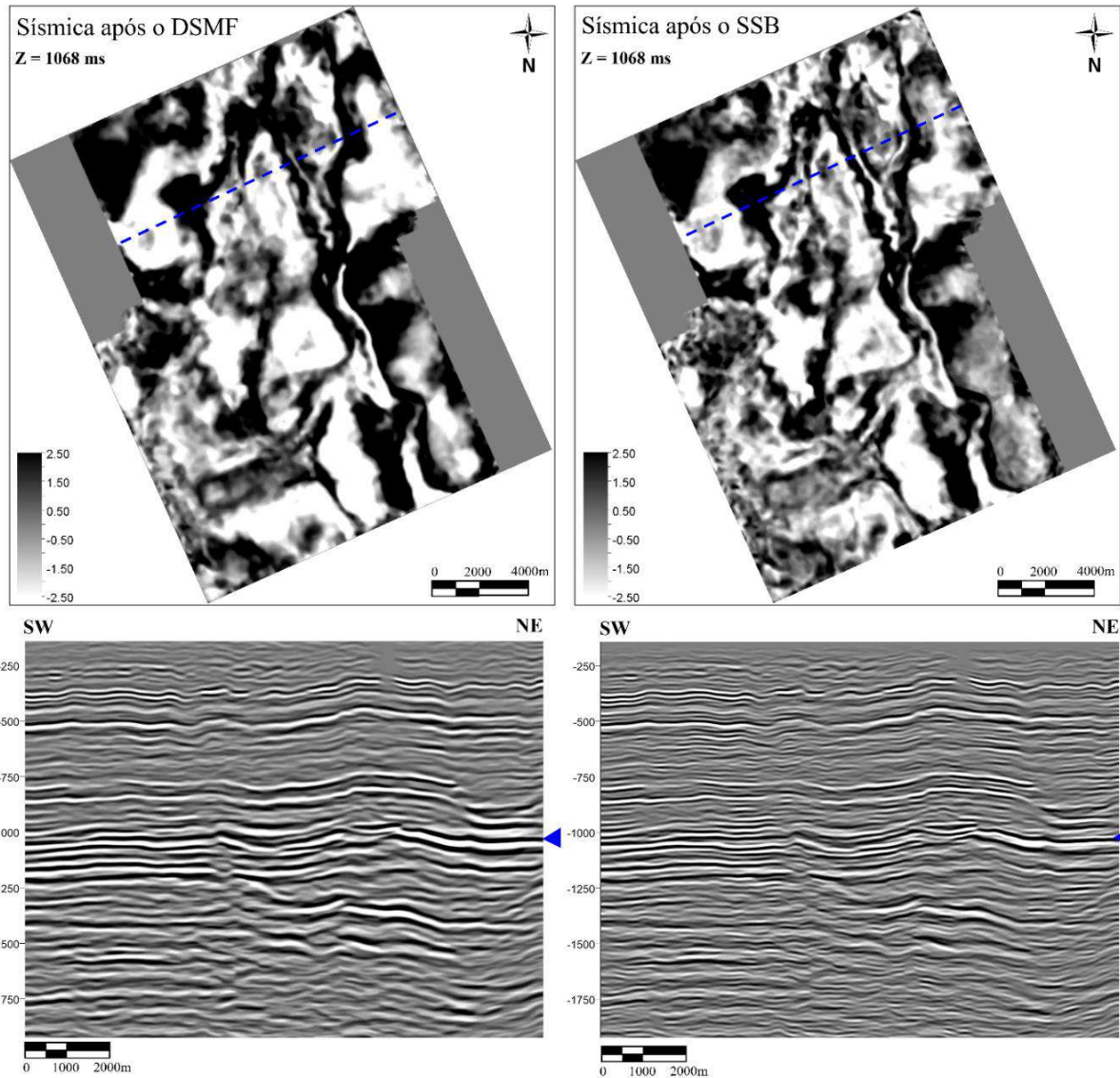


Figura 6.19. Comparação de dados sísmicos após o DSMF e o SSB em time slices e em seções sísmicas, ilustrando melhorias na resolução.

Vale também ressaltar, que ainda que o posicionamento dos poços estejam mais restritos as bordas dos dados, ou seja, não tem-se poços bem distribuídos pela área do levantamento sísmico, considerou-se que os operadores estimados a partir da combinação dos espectros dos poços com o espectro da sísmica são representativos da área de forma geral (ver distribuição de poços no Capítulo 1). Essa premissa foi considerada principalmente pela gênese geológica da formação de interesse desse estudo (Capítulo 2), sendo está composta por camadas de folhelhos plano-paralelo, sem evidências de estruturas muito complexas causadas por falhamentos de grande rejeito ou dobramentos. Se tratando de uma formação geológica com extensa continuidade lateral, além do fato do estudo ser restrito a um

dado sísmico que cobre uma área relativamente pequena de apenas 300 km², considerou-se razoável aplicar o operador extraídos nessas condições com uma confiabilidade alta.

Outro ponto que também merece atenção é o condicionamento dos perfis de poços, como essa técnica se baseia nesses dados, é importante que seja realizado um controle de qualidade dos perfis, como foi realizado neste caso sobre os perfis de densidade e sônico, evitando, dessa forma, o uso dados espúrios, ou ausentes de informações em determinados intervalos do poço, perfis muito afetados pela integridade das paredes dos poços, entre outros fatores.

A fim de destacar alguns ganhos obtidos na imagem sísmica após o condicionamento do dado, é ilustrado na Figura 6.20 algumas regiões da *inline* apresentada na Figura 6.19, para exemplificar melhorias específicas com o uso das duas técnicas combinadas. Regiões com sinais enfraquecidos apresentaram um ganho significativo de energia, resultando em uma melhor definição dos refletores. Camadas finas foram mais bem resolvidas devido ao aumento na resolução. Áreas anteriormente caracterizadas por sinais caóticos exibiram uma melhor continuidade dos refletores após a aplicação do Spectral Blueing. Finalmente, regiões afetadas pela atenuação do sinal sísmico devido a intrusões de soleiras tiveram seu conteúdo de alta frequência recuperado, conforme demonstrado no último exemplo da Figura 6.20.

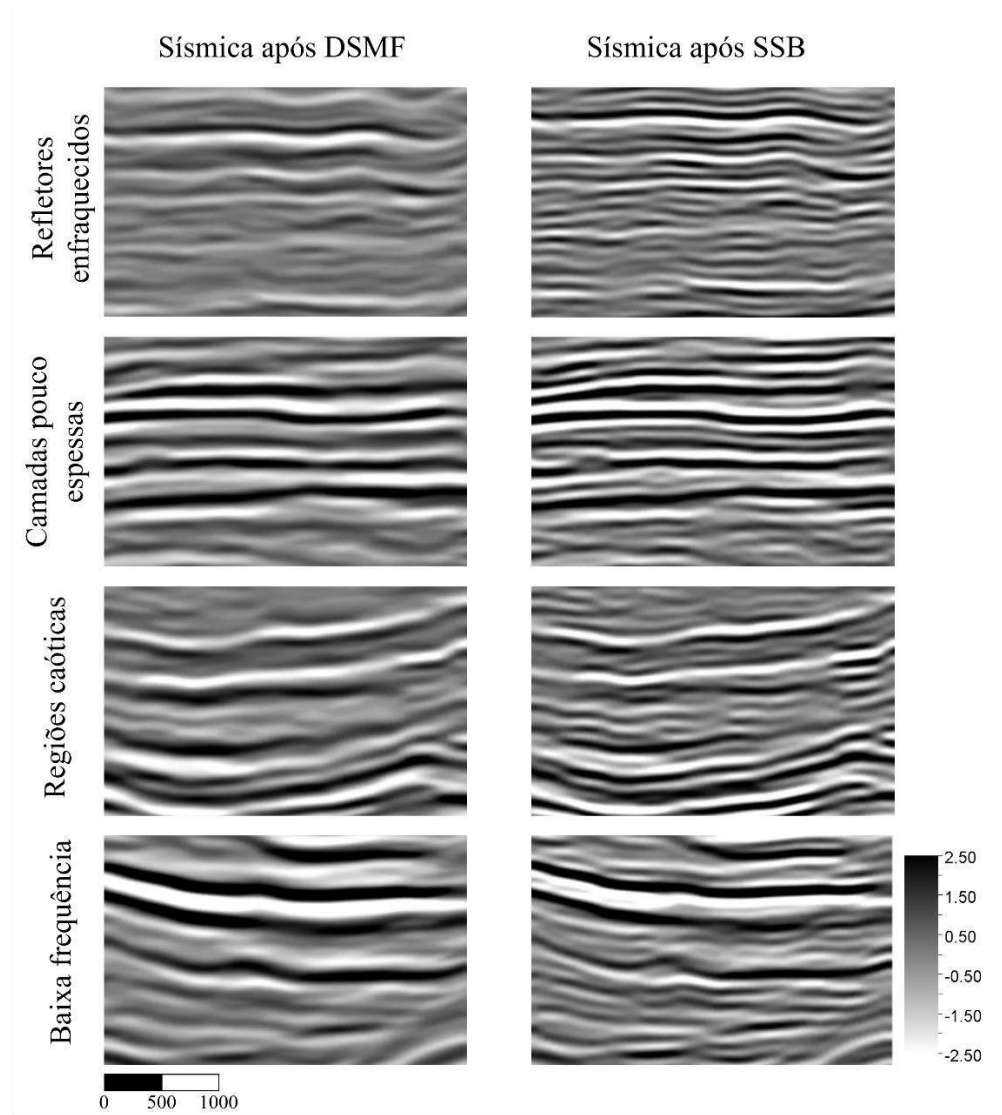


Figura 6.20: Interpretação das melhorias promovidas pela técnica *Spectral Blueing*.

Outra análise que pode ser realizada para compreender a relevância do estudo de condicionamento do dado sísmico, é confrontando-o com os dados de fácies classificadas nos poços por exemplo (classificação de fácies mostrada no Capítulo 4). Na Figura 6.21, é mostrado as inlines que passam verticalmente pelos poços. No primeiro *track* está exibido o dado sísmico condicionado com até a cinco xlines a esquerda do poço e a direita do poço tem-se o dado sísmico original limitado até cinco xlines a partir do poço no último *track*. A partir dessa análise ao confrontar os dados do poço com os dados sísmicos é observado que as camadas parecem ser mais bem definidas pelo dado de melhor resolução adquirido após os fluxos de condicionamento do dado, principalmente, mas não apenas, na interface com as soleiras de diabásio.

Além disso, devido ao ganho de resolução e conseqüentemente a melhor definição dos eventos sísmicos, o dado condicionado parece refletir melhor a heterogeneidade das fácies classificadas nos

poços ao longo da Fm. Pimenteiras, sendo possível a partir deste dado realizar estudos e caracterização deste intervalo de forma mais refinada, contemplando os contrastes encontrados nos folhelhos gerados pela variação do tipo e do conteúdo de matéria orgânica e pelo contato com as rochas ígneas como discutido no Capítulo 3.

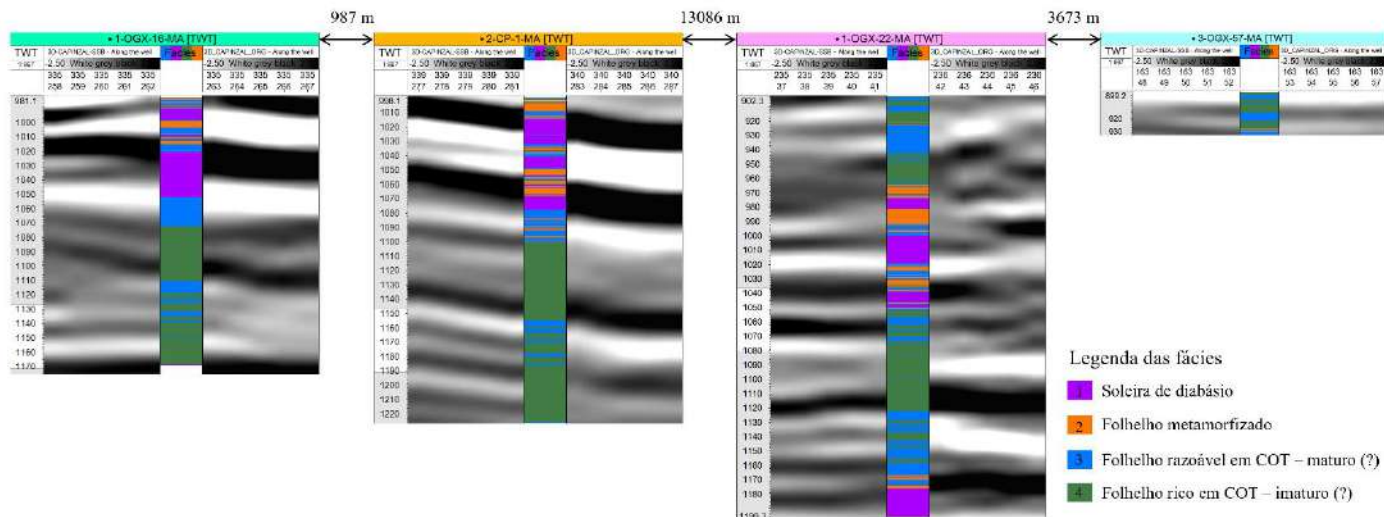


Figura 6.21. Análise dos dados sísmicos pré-condicionado (track 1) e original (track 3) juntos as fácies da Fm. Pimenteiras (track 2).

6.7. CONCLUSÃO

O fluxo de trabalho de pré-condicionamento, combinando as técnicas de *Dip Steered Median Filter* (DSMF) e *Spectral Blueing* (SSB), demonstrou ser altamente eficaz no aprimoramento da resolução de dados sísmicos terrestres da Bacia do Parnaíba. O DSMF atenuou significativamente os ruídos de alta energia, como o ruído de *footprint*, preservando e aumentando a continuidade lateral dos eventos sísmicos. Posteriormente, a aplicação do SSB recuperou expressivamente o conteúdo de alta frequência, que havia sido atenuado durante a etapa inicial de redução de ruído, resultando em um espectro de frequências mais equilibrado aumentando a resolução sísmica, que neste capítulo foi demonstrado de forma qualitativa. Ainda com as limitações dos métodos que são conhecidos e demonstrado neste capítulo, ao filtrar alguns sinais importantes (fechamento estrutural das soleiras) e gerar alguns artefatos (*ringing*), foi observado que eles não afetam de forma significativa o resultado. As etapas de condicionamento dos perfis de poços utilizados, geração de diferentes dados de entrada (*steering cube*) para o DSMF e a geração de diferentes operadores por zona no SSB, foram fundamentais para a obtenção de um dado com melhor resolução e definição de eventos sísmicos. Fácies sísmicas que antes não era possível de serem identificadas, ficaram mais claras após este estudo

mesmo para as áreas fora do intervalo de interesse que também foram cuidadosamente modificados pelo fluxo de o condicionamento proposto. A baixa correlação na etapa do SSB como mostrada principalmente na zona 1, pode estar relacionada primeiro devido à ausência de dado sísmico exatamente na posição do poço e pelo não ajuste do poço com a sísmica neste intervalo que foi ainda mais afetado por conta da caocidade dos refletores. Por fim, confrontando as fácies dos poços com os dados sísmicos antes e após o fluxo de condicionamento do dado, pode-se concluir que o resultado obtido facilitará o processo de interpretação sísmica e proporcionará melhores resultados na inversão e caracterização sísmica do intervalo de interesse.

7. Análise do Espectro de Frequência e Estudo de Resolução Sísmica

7.1. INTRODUÇÃO

Neste capítulo é apresentada uma análise detalhada do espectro de frequência dos dados sísmicos originais e dos dados resultantes após cada etapa do pré-condicionamento. O objetivo é corroborar a efetividade do fluxo de trabalho adotado e quantificar os ganhos em resolução sísmica. A avaliação integra duas frentes complementares: (i) análise espectral (curvas de espectro de frequência) e (ii) espectrogramas obtidos por Transformada Wavelet Contínua (Continuous Wavelet Transform – CWT). Essa combinação permite identificar aumento de energia em altas frequências e a obtenção de um espectro mais equilibrado (broadband), associado a melhorias na resolução temporal e frequencial. Além disso, é desenvolvido um estudo de resolução fundamentado no modelo de cunha e nos conceitos clássicos de Kallweit e Wood (1982) e Widess (1973), com ênfase na quantificação do ganho de resolução vertical decorrente do pré-condicionamento por meio da comparação das espessuras mínimas detectáveis em cada conjunto de dados.

7.2. METODOLOGIA

7.2.1 Espectro de frequência

O espectro de frequência é uma representação gráfica que mostra como a energia de um sinal sísmico se distribui entre suas diferentes componentes de frequência. Ao decompor o sinal no domínio da frequência, o espectro expressa a amplitude ou a potência associada a cada frequência predominante (Yilmaz *et al.* 2001). Essa análise é fundamental para compreender como essas frequências se distribuem ao longo do tempo, fornecendo informações importantes quanto a relação sinal/ruído, a modificação do espectro após processos de filtragem e ganho do dado sísmico.

A seguir serão demonstradas as análises e comparação entre os espectros de frequência dos dados sísmicos resultantes de cada etapa do fluxo de trabalho, extraídos do intervalo de interesse (750

– 1600 ms). É importante destacar que o condicionamento, realizado por meio da combinação dos métodos *Dip Steered Median Filter* (DSMF) e *Spectral Blueing* (SSB), está detalhado no Capítulo 6. Tal abordagem permitiu obter dados sísmicos com uma razão sinal/ruído melhorada e um conteúdo de altas frequências aumentado, e será discutido através da análise dos gráficos a seguir, as modificações no conteúdo de frequência dos dados, provocadas pelo fluxo de trabalho proposto.

7.2.2. Continuous Wavelet Transform (CWT)

O espectrograma obtido por meio da técnica CWT é um método de análise bastante útil aplicado a análise de dados sísmicos (sinais não estacionários) antes e após o condicionamento, pois essa abordagem permite realizar interpretações detalhadas ao considerar tanto a resolução temporal quanto a resolução de frequência do sinal (Addison, 2002). Como detalhado no trabalho de Paul S. Addison em 2002, a CWT de um sinal é obtida envolvendo o sinal com uma função *wavelet* mãe que é escalada e transladada. Dessa forma, a *wavelet* sofre dilatação (escala) e deslocamento (translação) ao longo do sinal permitindo o ajuste da função para capturar características de diferentes tamanhos ao longo das variações temporais, ao contrário da Transformada de Fourier, que fornece apenas informações de frequência.

Uma das características mais poderosas das *wavelets* é a sua capacidade de realizar uma análise multiresolução, o que significa que ela pode analisar simultaneamente componentes de alta e baixa frequência. Isso é especialmente aplicável para dados sísmicos, onde diferentes frequências podem corresponder a diferentes respostas geológicas.

Segundo Addison, 2002, a Transformada *Wavelet* $T(a,b)$ pode ser expressa pela equação:

$$T(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \bar{\psi} \left(\frac{t-b}{a} \right) dt \quad (7.1)$$

onde $x(t)$ representa o sinal de entrada, $\bar{\psi}$ é o conjugado da *wavelet*, a corresponde ao parâmetro de escala e b o parâmetro de translação. A análise de cada ponto do sinal avalia sua contribuição em relação à *wavelet*, conforme a regra de multiplicação de sinais. A contribuição é negativa, ou destrutiva, quando a *wavelet* e o sinal têm polaridades opostas. Por outro lado, a contribuição é positiva, ou construtiva, quando a *wavelet* e o sinal compartilham a mesma polaridade. Isso permite uma análise detalhada do comportamento do sinal em diferentes escalas e posições.

Neste estudo, a CWT será utilizada para gerar espectrogramas, que são representações visuais da variação da amplitude das diferentes frequências ao longo do tempo. Essa técnica é essencial para identificar eventos sísmicos e analisar a presença de ruídos, proporcionando uma visão clara das

características temporais e espectrais do sinal. Embora várias *wavelets* possam funcionar como "wavelet mãe", este trabalho emprega a *wavelet* de Morlet, definida para valores de $f_0 \gg 0$ como (Addison, 2002):

$$\psi(t) = \frac{1}{\pi^{1/4}} e^{i2\pi f_0 t} e^{-t^2/2} \quad (7.2)$$

7.2.3. Estudo de resolução sísmica

A interpretação sísmica é intrinsecamente limitada pela natureza da banda de frequência dos dados sísmicos e pela forma da *wavelet* envolvida no processo de aquisição e processamento. Um dos desafios centrais reside na distinção entre camadas espessas, cuja separação entre interfaces gera eventos sísmicos individualmente resolvíveis, e camadas finas, cuja resposta sísmica resulta de interferências entre os refletores superior e inferior. Essa distinção está no centro do conceito de resolução sísmica vertical, sendo essencial para a correta caracterização de feições geológicas e estratigrafias complexas.

Segundo Simm & Bacon (2014), existe um limite inferior de espessura, controlado pela largura de banda e pelo formato da *wavelet*, abaixo do qual duas interfaces não podem mais ser resolvidas de maneira única. Essa limitação introduz incertezas significativas, especialmente em interpretações quantitativas, como aquelas baseadas em amplitude 3D e os derivados atributos. Ao mesmo tempo, os autores destacam que o conceito de resolução não deve ser confundido com detectabilidade: mesmo quando duas interfaces não são resolvidas como eventos distintos, sua presença pode ainda ser detectada por meio da variação de amplitude gerada pela interferência construtiva ou destrutiva dos sinais.

Essa diferenciação entre resolução e detecção é também fortemente enfatizada por Kallweit & Wood (1982), que definem resolução como uma limitação associada à banda de frequência, enquanto a detecção está mais relacionada a condições de aquisição e à relação sinal-ruído (S/R). Segundo os autores, um evento detectável pode ou não ser resolvível, e a separação entre esses conceitos é essencial para interpretar corretamente a resposta sísmica de camadas finas.

7.2.3.1. Critérios de Resolução Sísmica

O principal objetivo do trabalho desenvolvido por Kallweit e Wood (1982) foi descrever conceitos da resolução sísmica unificando os diferentes pontos de vista de Rayleigh, Ricker e Widess, removendo as considerações de polaridade da resolubilidade. A convergência para o reconhecimento

de que existe um limite crítico de separação abaixo do qual dois refletores deixam de ser resolvidos como eventos distintos é baseado nos critérios estabelecidos por Jenkins *et al.* (1958), Ricker (1953) e Widess (1953) e resumidos por Kallweit e Wood (1982), como mostra a Figura 7.1.

O critério de Rayleigh originalmente descrito no contexto da óptica, estabelece um limite para a capacidade de resolução de sistemas ópticos. Em termos ópticos, uma fonte pontual pode ser comparada a um pico de reflexão, enquanto o instrumento óptico é análogo à Terra, e o padrão de difração desempenha um papel equivalente ao de uma *wavelet* de banda limitada. A resolução de um instrumento óptico está relacionada à sua habilidade de distinguir imagens separadas de objetos próximos, sendo limitada pelas características do padrão de difração, de modo similar ao comportamento de uma *wavelet* sísmica. De acordo com o critério de Rayleigh, conforme descrito por Jenkins e White (1957), dois objetos pontuais são considerados resolvidos quando a separação entre eles é igual ou superior à distância entre o máximo central de um padrão de difração e o seu primeiro mínimo adjacente, definida como a separação pico-vale ($b/2$). Quando a separação entre os objetos é menor do que essa distância, as imagens se fundem, impossibilitando a resolução individual dos objetos. Essa ideia é diretamente comparável à interação entre a *wavelet* sísmica e as reflexões associadas a interfaces próximas. Kallweit e Wood (1982) destacaram que, embora a amplitude da *wavelet* esteja intrinsecamente relacionada à resolução sísmica, a largura da *wavelet* não apresenta essa relação direta no caso de *wavelets* com espectros brancos. Isso implica que o limite de resolubilidade descrito por Rayleigh está mais diretamente associado à primeira derivada da *wavelet* sísmica, relacionada ao tempo de vale.

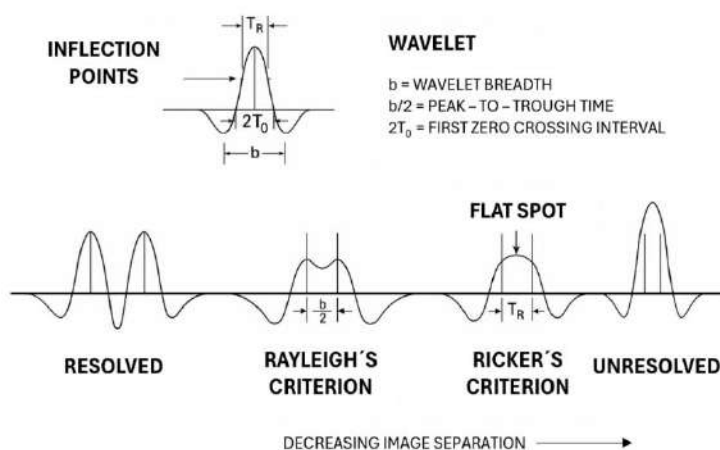


Figura 7.1. O limite de resolução segundo Rayleigh é atingido quando dois eventos sísmicos estão separados por um intervalo de tempo correspondente à distância entre o pico e o vale. Já o critério de Ricker define o limite de resolução como o intervalo de tempo igual a separação entre os pontos de inflexão da *wavelet* (Kallweit e Wood, 1982).

O critério de Ricker (1953) por sua vez, avalia a forma de onda composta resultante da superposição de dois eventos sísmicos. Ele define o limite de resolução como a separação em que o máximo central da forma de onda apresenta curvatura zero, ou seja, um "ponto plano". Essa condição ocorre quando a distância entre os dois eventos equivale à separação entre os pontos de inflexão do lóbulo principal da wavelet convolucional. Ricker também enfatizou a relação entre a resolução e a diferenciação da wavelet, observando que essa abordagem é válida para picos de polaridades iguais e amplitudes equivalentes.

Widess (1973), por outro lado, focou sua análise em eventos de polaridades opostas e amplitudes iguais. Ele observou que, à medida que a separação entre os eventos diminui, a forma de onda composta tende a se comportar como a derivada da *wavelet* original (Figura 7.2). Widess identificou que o limite de resolução para camadas finas ocorre quando a separação entre os eventos é igual a 1/8 do comprimento de onda da frequência dominante da *wavelet*. Para separações menores, as alterações no tempo entre os picos tornam-se imperceptíveis, embora, em teoria, seja possível distinguir eventos baseando-se nas variações de amplitude. No entanto, ele ressalta que essa abordagem apresenta desafios práticos, como a necessidade de calibração precisa para garantir resultados confiáveis.

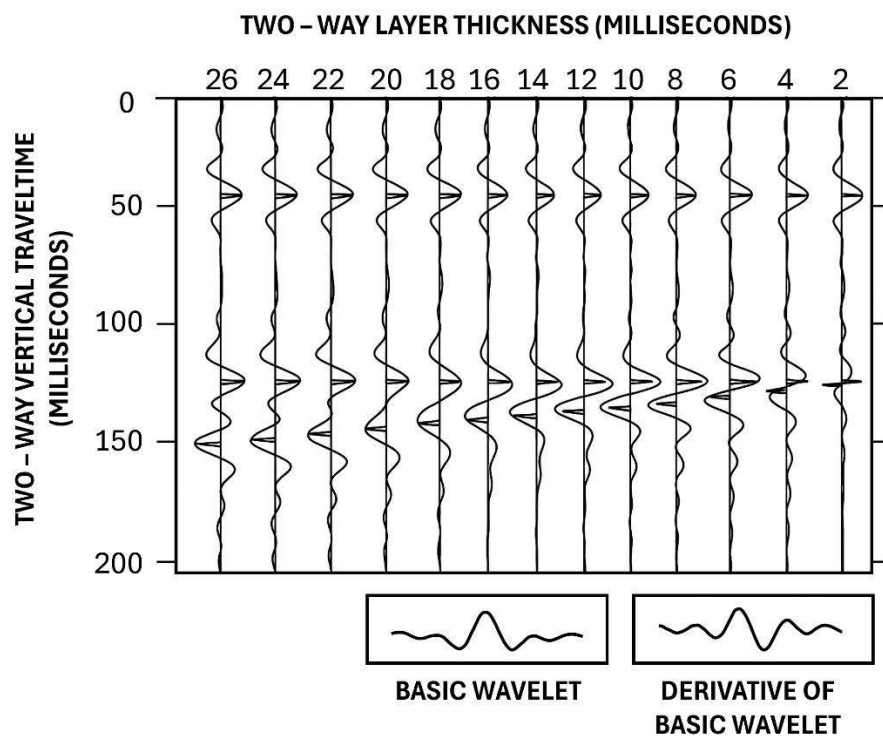


Figura 7.2. Como observado por Widess (1973) a medida que a separação entre os picos diminui uma wavelet convolvida com dois picos de igual amplitude e polaridade oposta converge para a derivada da wavelet (Kallweit e Woods, 1982).

Dessa forma, os autores fazem uma análise de estudos sobre resolução sísmica vertical e definem a resolução vertical em termos práticos, onde os seus resultados são comparados com estudos semelhantes de Rayleigh, Ricker (1953) e Widess (1973, 1980), em que a resolução temporal de uma *wavelet broadband* é controlada por sua frequência terminal mais alta, f_u , desde que a razão de banda da *wavelet* exceda duas oitavas, como serão mais bem descritas pelas equações matemáticas a seguir.

7.2.3.2. Espessura de *Tuning*

Como visto anteriormente o limite de Rayleigh é o tempo do pico ao vale de uma *wavelet*, que corresponde à espessura de tuning, enquanto o limite de Ricker, que é a separação temporal entre os pontos de inflexão da *wavelet*, corresponde à resolução temporal. Kallweit e Wood (1982), então derivam fórmulas simples que relacionam a espessura de tuning e a resolução temporal com a banda de frequência espectral para *wavelets* de Ricker e *wavelet* senoidais.

Para as *wavelets* de Ricker, Kallweit e Wood (1982) definiram a resolução temporal e espessura de tuning em termos de frequência de pico (f_p), que se refere ao componente de frequência que exhibe o valor mais alto no espectro de amplitude, que também pode ser deduzida em termos de frequência dominante (f_d) que é o recíproco do tempo de vale a vale ou da amplitude em torno do lobo central ($1/b$) como definido por Widess (1973). Portanto a resolução temporal para a *wavelet* de Ricker pode ser descrita como:

$$T_R = 1/3.0f_p, T_R = 1/2.31f_d \quad (7.3)$$

E a espessura de tuning como:

$$b/2 = 1/2.6f_p, b/2 = 1/2.0f_d \quad (7.4)$$

As *wavelets* senoidais, por outro lado, são definidas pelas frequências terminais inferior (f_l frequência mínima), e superior (f_u frequência máxima), onde, portanto, a banda de frequência da *wavelet* apresenta todas as frequências presentes com amplitudes iguais. Nesse contexto, Kallweit e Wood (1982) demonstraram que a resolução temporal de um *wavelet broadband* com um espectro plano é governada por sua frequência terminal mais alta, f_u e que relações similares para a frequência dominante definida por Widess (1973) podem ser também estabelecidas para este caso. Assim a resolução temporal para *wavelets* senoidais aproxima-se de:

$$T_R = 1/1.5f_u, T_R = 1/2.14f_d \quad (7.5)$$

E a espessura de tuning de:

$$b/2 = 1/1.4f_u, b/2 = 1/2.0f_d \quad (7.6)$$

Com isso, Kallweit e Wood (1982) concluem que no caso de a espessura de *tuning* onde $\lambda = V/f_d$ é o comprimento de onda dominante através de uma camada de velocidade de intervalar V, o limite resolvível de uma camada é dado por:

$$\Delta z_b = \frac{\lambda}{4} \quad (7.7)$$

7.2.3.3. Modelo de Interferência

O modelo de cunha, originalmente proposto por Widess (1973) e amplamente discutido por Simm e Bacon (2014), é uma abordagem eficaz e prática para ilustrar a interferência de *wavelets* sísmicas e calcular limites de resolução vertical. O modelo proposto por Simm e Bacon (2014) para demonstrar esse estudo consiste em uma camada de areia em forma de cunha inserida entre camadas de folhelho, onde os coeficientes de reflexão no topo e na base da cunha possuem a mesma magnitude, mas polaridades opostas (Figura 7.3a). Uma das principais aplicações do modelo de cunha é a análise da relação entre a espessura da camada e a resposta sísmica. Devido à largura de banda limitada das *wavelets* sísmicas, existe uma espessura mínima abaixo da qual os loops sísmicos de pico e vale apresentam separação constante, independentemente da espessura real da camada. Essa espessura crítica, comumente denominada "espessura de tuning", representa o ponto em que a interferência destrutiva é máxima, resultando em uma amplitude mínima (Figura 7.3c). Abaixo da espessura de tuning, a amplitude sísmica aumenta à medida que a camada se torna mais fina, enquanto acima desse ponto a amplitude e a separação temporal entre os loops podem ser usadas para estimar a espessura da camada, embora com algumas limitações (Simm e Bacon, 2014).

Simm e Bacon (2014) destacam que, quando a espessura da camada está abaixo da espessura de tuning, a separação entre os loops de vale e pico pode levar a uma superestimativa significativa da espessura real (Figura 7.3b). Por outro lado, para espessuras ligeiramente superiores à espessura de tuning, a estimativa com base nessa separação pode resultar em uma subestimativa modesta. Além disso, o comportamento da curva de tuning é fortemente influenciado pelo formato da *wavelet* utilizada na modelagem (Figura 3d).

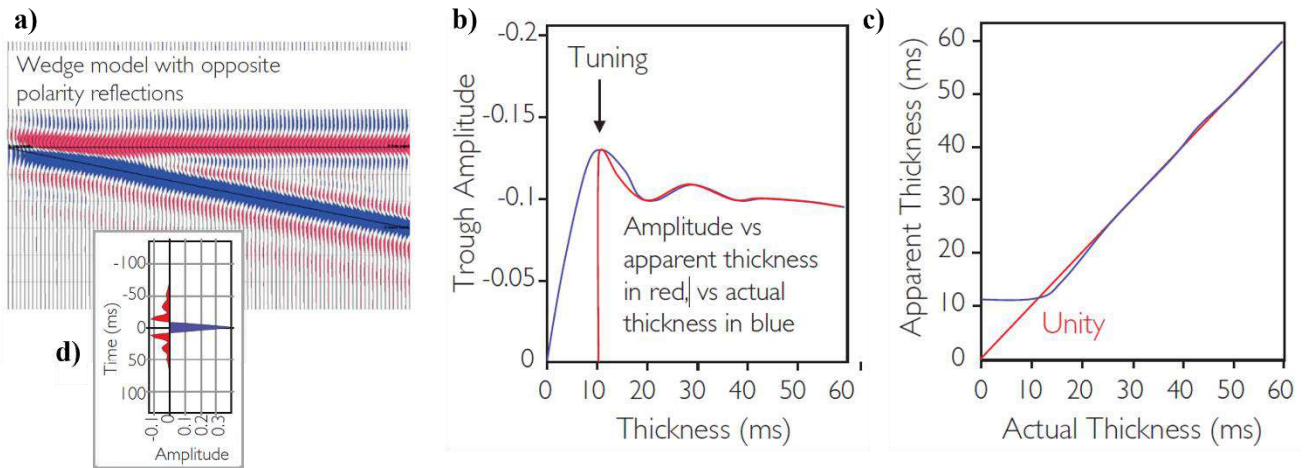


Figura 7.3. Modelo simples de interferência, onde é apresentado em a) o modelo de cunha com reflexões de polaridades opostas, em b) o *crossplot* entre a amplitude do topo contra a espessura aparente, em c) o *crossplot* entre a espessura real contra a espessura aparente. Por fim em c) é exibido a *wavelet* utilizada para gerar os traços sísmicos (Simm e Bacon, 2014).

Portanto, o modelo de cunha é uma ferramenta essencial para compreender os efeitos da interferência de *wavelets* sísmicas e os limites de resolução em camadas delgadas, permitindo análises detalhadas que combinam amplitude sísmica e espessura aparente.

7.3. RESULTADOS

7.3.1. Análise do espectro de frequência

Na Figura 7.4, os espectros de frequência dos dados sísmicos em diferentes estágios do fluxo de trabalho são apresentados. O espectro do dado sísmico original é representado pela curva verde, enquanto a curva azul corresponde ao espectro resultante após a aplicação DSMF. Já a curva vermelha representa o espectro de frequência do dado sísmico final, ou seja, após a aplicação SSB.

Conforme observado na Figura 7.4a, a comparação entre os espectros do dado original e do dado filtrado pelo DSMF evidencia uma redução da energia nas altas frequências, especialmente no intervalo de 30 Hz a 60 Hz. Quantitativamente, essa redução é caracterizada por uma diminuição média de 19,5% na energia, com valores mínimos de 8% e máximos de 30,6%. Essa perda de energia está associada ao processo de atenuação de ruído de alta frequência promovido pelo DSMF, como discutido no Capítulo 6. Após a aplicação do método, as seções sísmicas indicam que uma parcela significativa do ruído de alta energia foi efetivamente atenuada, o que é corroborado pela análise do espectro de frequência. Apesar da eficácia do DSMF na mitigação do ruído, foi identificado, no dado residual, a presença de alguns eventos sísmicos dentro do intervalo de frequência mencionado. Embora esses

eventos não sejam predominantes, sua atenuação parcial sugere que o filtro afetou não apenas o ruído, mas também, em menor escala, componentes importantes do sinal sísmico.

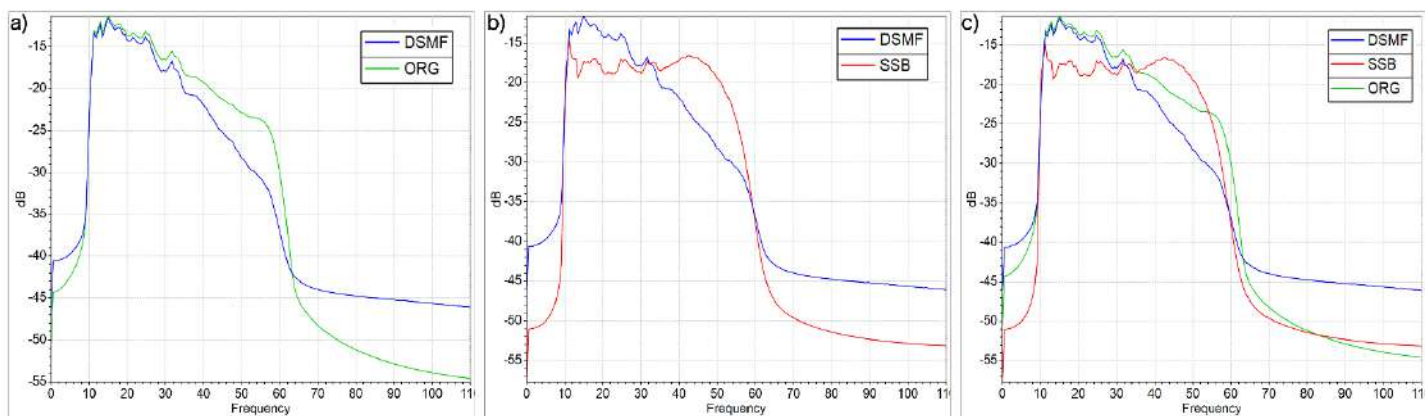


Figura 7.4. Comparação dos espectros de frequência dos dados de entrada e saída em cada etapa de pré-condicionamento. No painel a), os espectros de frequência dos dados originais são exibidos em verde, enquanto os dados pós-DSMF são mostrados em azul, ilustrando o efeito de filtragem em altas frequências. No painel b), a comparação o espectro dos dados após DSMF com os dados após SSB, representados em vermelho, o que demonstra a recuperação de energia em altas frequências. Por fim, o painel c) fornece uma comparação abrangente do espectro de frequência para todos os três conjuntos de dados sísmicos.

Ao comparar os espectros de frequência do dado sísmico após a aplicação do DSMF e do SSB, conforme ilustrado na Figura 7.4b, observa-se uma recuperação significativa da energia no intervalo de frequências previamente atenuado. Especificamente, houve um ganho médio de 29% na energia, com um ganho máximo de até 49%. Importante destacar que essa recuperação não foi acompanhada por um aumento de ruídos, como evidenciado nas seções sísmicas analisadas no Capítulo 6.

Por fim, ao comparar os três gráficos dos espectros de frequência do dado original e após cada etapa do condicionamento (Figura 7.4c), constata-se que, além da recuperação da energia que havia sido atenuada, também foi possível obter um ganho de energia nas altas frequências e um balanço no espectro de forma geral. Em outras palavras, a alta energia observada no intervalo entre 10 Hz e 30 Hz nos dados originais e após o DSMF teve sua energia reduzida após a aplicação do SSB, resultando em um espectro balanceado no intervalo de frequência até quase 60 Hz.

7.3.2. Espectrograma CWT

O espectrograma foi gerado a partir das seções (*inline* e *xline*) que cruzam a posição do poço 2-CP-1-MA. A Figura 7.5 apresenta os resultados dos espectrogramas obtidos utilizando uma *wavelet* de Morlet, abrangendo o intervalo de 140 a 2000 ms. Diferentemente do intervalo de análise do espectro de frequência apresentado anteriormente, o intervalo analisado no espectrograma CWT extrapola o intervalo de interesse principal deste trabalho. Essa escolha metodológica foi motivada pelo fato de

que o fluxo de condicionamento foi projetado considerando todas as seqüências sedimentares da bacia representadas no dado sísmico disponível, com o intuito de capturar suas particularidades e, assim, fornecer subsídios para análises e aplicações em estudos futuros. Portanto, a análise mais ampla do espectrograma busca avaliar de maneira abrangente o impacto do fluxo de condicionamento adotado no dado sísmico.

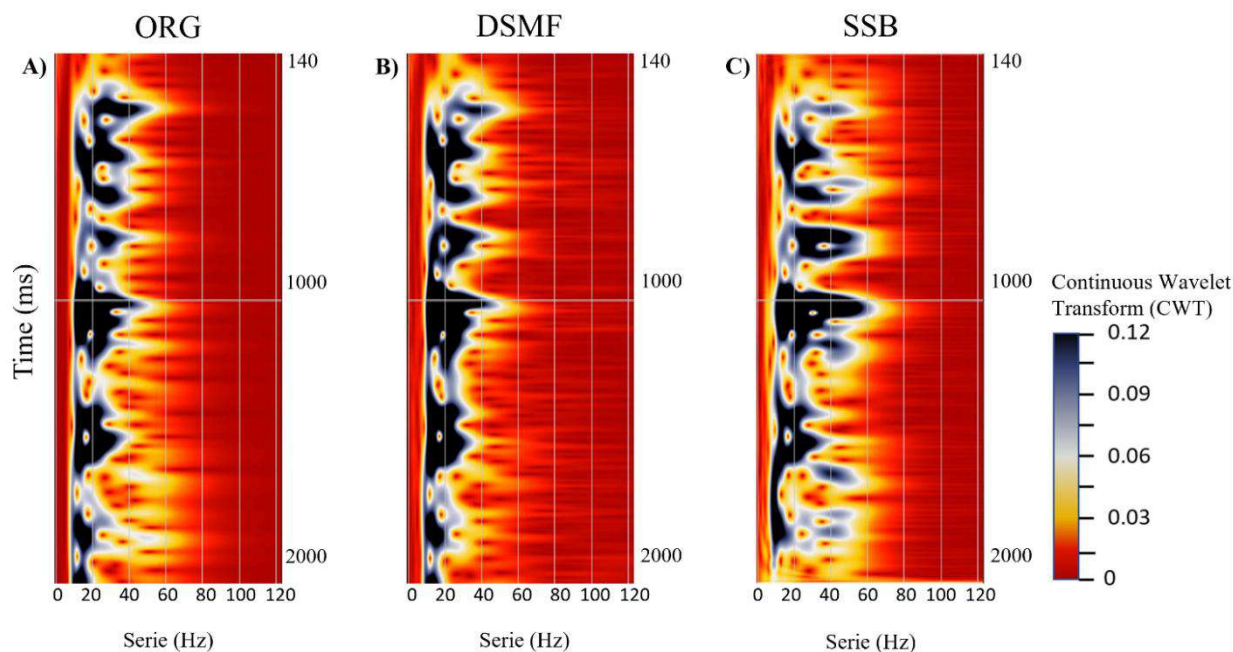


Figura 7.5. Espectrograma estimado pela técnica da Transformada *Wavelet* Contínua (CWT), em a) para o dado original em b) para o dado após o DSMF e em c) para o dado após o SSB.

As alterações no espectro de frequência ao longo das etapas de melhoria do dado sísmico são claramente evidenciadas pela análise da Figura 7.5. Ao comparar o espectrograma do dado original (Figura 7.5a) com aquele obtido após a aplicação do filtro na primeira etapa do condicionamento (Figura 7.5b), observa-se uma atenuação significativa da energia em altas frequências em diversos intervalos de tempo. Essa redução de alta energia pode ser atribuída, principalmente, à remoção do ruído de alta frequência associado ao *footprint*, que foi efetivamente filtrado, conforme comprovado pela análise visual das seções sísmicas e pelo dado residual gerado.

Entretanto, além da eliminação do ruído, também foi observada uma suavização e atenuação das altas frequências dentro do intervalo de interesse principal. Esse efeito, embora esperado em decorrência do processo de filtragem, ressalta a necessidade de um balanço cuidadoso entre a supressão de ruídos e a preservação de componentes do sinal sísmico, especialmente em frequências mais altas, que estão diretamente relacionadas à resolução temporal.

Por outro lado, o espectrograma do dado sísmico processado com o SSB, apresentado na Figura 7.5c, ressalta, assim como observado nos espectros de frequência da Figura 7.4, um ganho significativo de energia, acompanhado de um aumento na resolução em altas frequências. Esse ganho é perceptível em todo o dado sísmico, porém, mais importante, é particularmente notável dentro do intervalo de interesse, onde está localizada a Formação Pimenteiras, alvo principal deste estudo.

Observa-se, especificamente, um aumento expressivo de energia para frequências superiores a 30 Hz. Esse aprimoramento é de grande relevância, pois impacta positivamente a resolução do intervalo de interesse. O ganho em resolução temporal e espectral proporcionado pelo fluxo de condicionamento adotado contribui diretamente para uma melhor definição do sinal sísmico nessa formação, conforme mostrados pelos resultados apresentados neste estudo. Essa melhoria representa um avanço significativo na qualidade dos dados sísmicos analisados, reforçando a eficácia do fluxo de trabalho implementado.

7.3.3. Estimativa do ganho de resolução no dado condicionado

É importante notar que o fluxo de trabalho aplicado ao ganho de alta frequência resultou em dados sísmicos com um espectro de frequência plano, referido como *broadband*. Isso significa que o espectro de frequência demonstra uma distribuição ampla e relativamente uniforme de energia em uma grande faixa de frequência. E por isso, as descobertas de Kallweit e Wood (1982) serão cruciais para este estudo na estimativa precisa da resolução vertical e da espessura de *tuning* do dado pré-condicionado.

Seguindo os critérios previamente estabelecidos para a realização de estudos de resolução sísmica, conforme detalhado nas seções anteriores, nesta seção será apresentado a análise do ganho de resolução obtido após o pré-condicionamento dos dados sísmicos. Para isso, dois estudos complementares foram conduzidos, com o objetivo de mensurar as espessuras mínimas detectáveis em cada conjunto de dados. O primeiro estudo se baseia na extração de sismogramas sintéticos a partir do modelo de cunha, para analisar a interferência das *wavelets* extraídas dos dados sísmicos original e pré-condicionado. E o segundo será baseado nas equações de aproximação da espessura de *tuning* como demonstradas por Kallweit e Wood (1982), com o objetivo de estimar as espessuras resolvidas dentro da resolução sísmica de cada dado, para litologias comumente encontradas na Fm. Pimenteira.

Para a geração dos sismogramas sintéticos, foram calculadas *wavelets* estatísticas a partir dos dados sísmicos original e condicionado, utilizando as posições dos poços como referência. As *wavelets* foram extraídas dentro de uma janela de tempo compreendida entre 750 ms e 1600 ms. A

wavelet utilizada em cada cenário foi obtida como a média de quatro wavelets extraídas em cada posição de poço, sendo essas apresentadas na Figura 7.6.

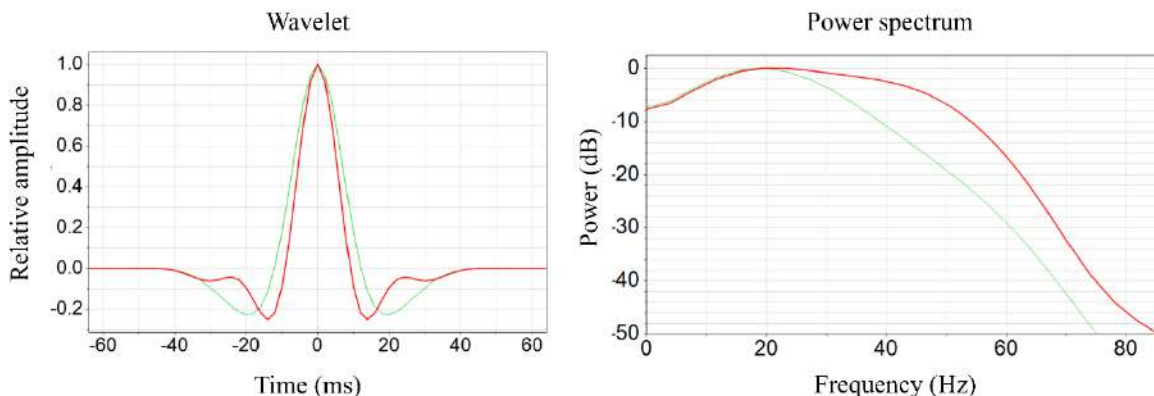


Figura 7.6. *wavelets* estatísticas médias extraídas nas posições dos poços, em verde do dado original e em vermelho do dado após o SSB, juntamente com seus respectivos espectros de frequência.

O modelo de cunha foi construído com base nos dados do poço 1-OGX-16-MA, abrangendo um intervalo que, do topo para a base, inclui: (i) uma camada de soleira de diabásio com 97 m de espessura e uma impedância média de $20.300 \text{ m/s} \cdot \text{g/cm}^3$; (ii) uma camada espessa de folhelho com 203 m, caracterizada por variações de impedância com uma média de $10.000 \text{ m/s} \cdot \text{g/cm}^3$; e (iii) uma camada de arenito pertencente à Formação Itaim, com uma impedância média de $12.300 \text{ m/s} \cdot \text{g/cm}^3$, superior à dos folhelhos subjacentes, e com 137 m de espessura. Essas camadas e suas respectivas litologias estão representadas na Figura 7.7, no *track "Facies Wedge"*, delimitadas entre os marcadores Horizon 2 e Horizon 5.

É importante ressaltar que, para este estudo inicial, as fácies foram simplificadas em três categorias principais: soleira, folhelho e arenito. Dessa forma, não foram consideradas as intercalações detalhadas via classificação não supervisionada, conforme descrito no Capítulo 4. O estudo de resolução individualizado para essas fácies, incluindo as intercalações, será apresentado nas seções subsequentes, com base nas equações de Kallweit e Wood (1982).

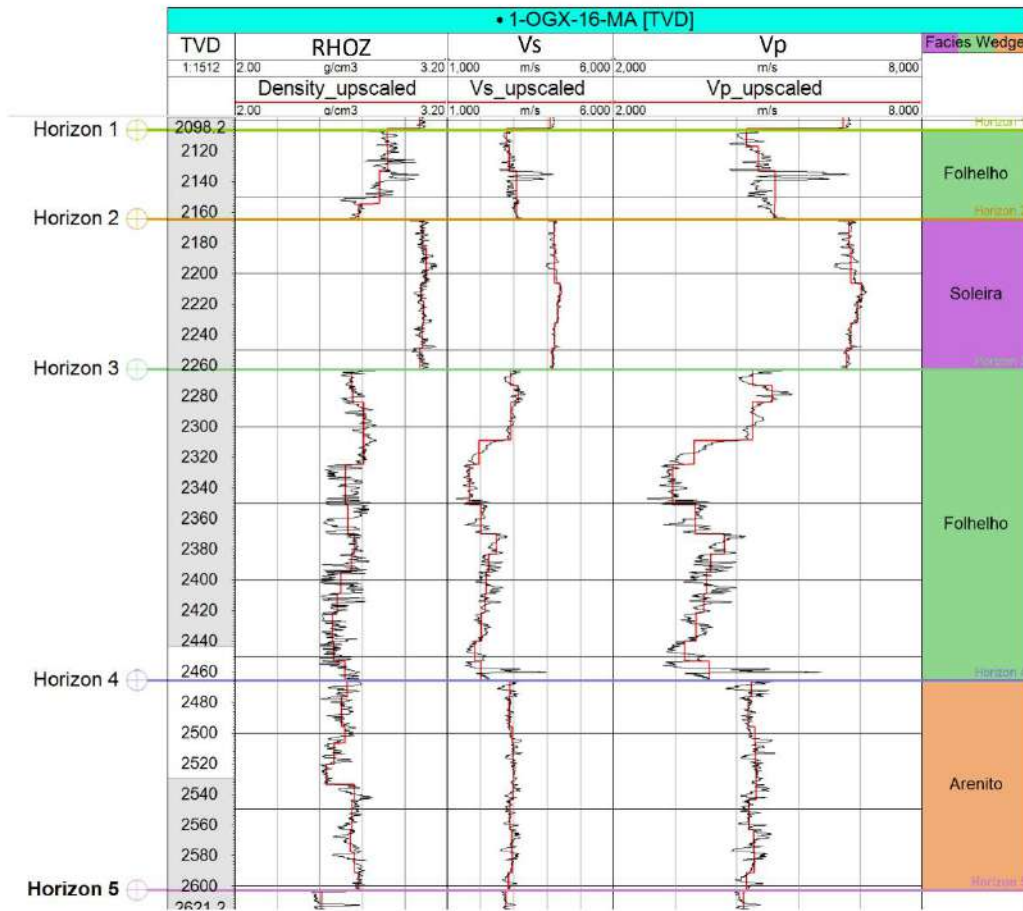


Figura 7.7. Camadas utilizadas no modelo de cunha e suas propriedades elásticas, incluindo valores de Vp, Vs e densidade.

Os modelos de parâmetros elásticos foram construídos a partir das curvas de velocidade compressional, velocidade cisalhante e densidade do poço 1-OGX-16-MA. Esses dados foram reescalados utilizando a média de Backus, com uma janela de 10 m, para representar a espessura mínima. Os perfis reescalados estão destacados pelas curvas em vermelho na Figura 7.7, enquanto os modelos gerados a partir desses perfis são apresentados na Figura 7.8.

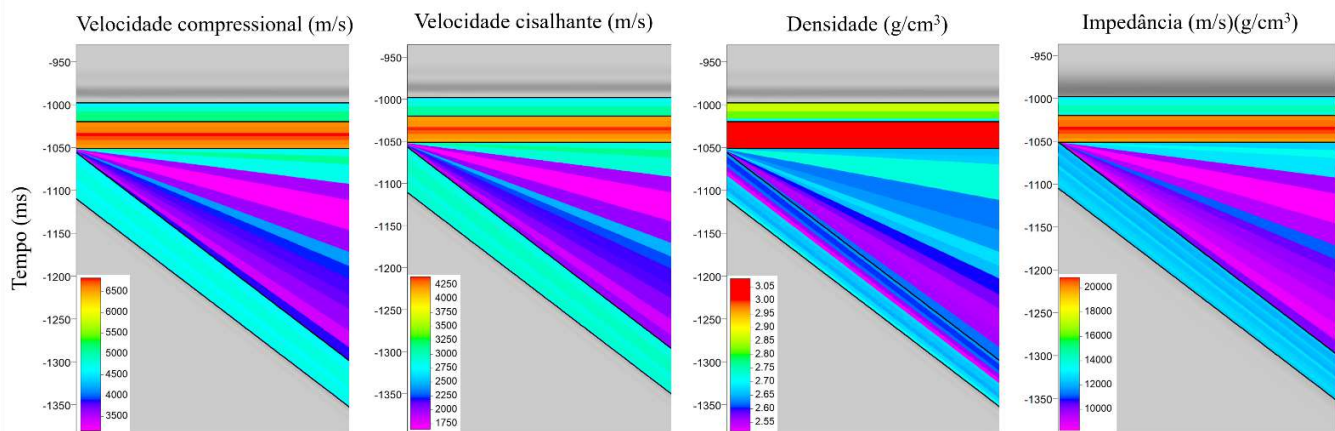


Figura 7.8. Modelos de cunha com suas propriedades elásticas, que servirão de base para a geração do sismograma sintético.

A modelagem do sismograma sintético foi realizada por meio da convolução das wavelets extraídas de cada dado sísmico com as séries de coeficientes de reflexão calculadas a partir das propriedades elásticas do modelo de cunha, utilizando a equação de Zoeppritz (Figura 7.9a). Esse procedimento permitiu a geração dos traços sintéticos, representando a resposta sísmica associada às condições do modelo geológico. Os resultados obtidos estão apresentados na Figura 7.9, que exibe uma seção passando pelo poço utilizado como referência. Nessa seção, são destacados os marcadores que delimitam os topos de cada camada litológica do modelo de cunha, onde pode-se observar o intervalo correspondente à cunha, reproduzido no intervalo de folhelhos, localizado entre os marcadores Horizon 3 e Horizon 4.

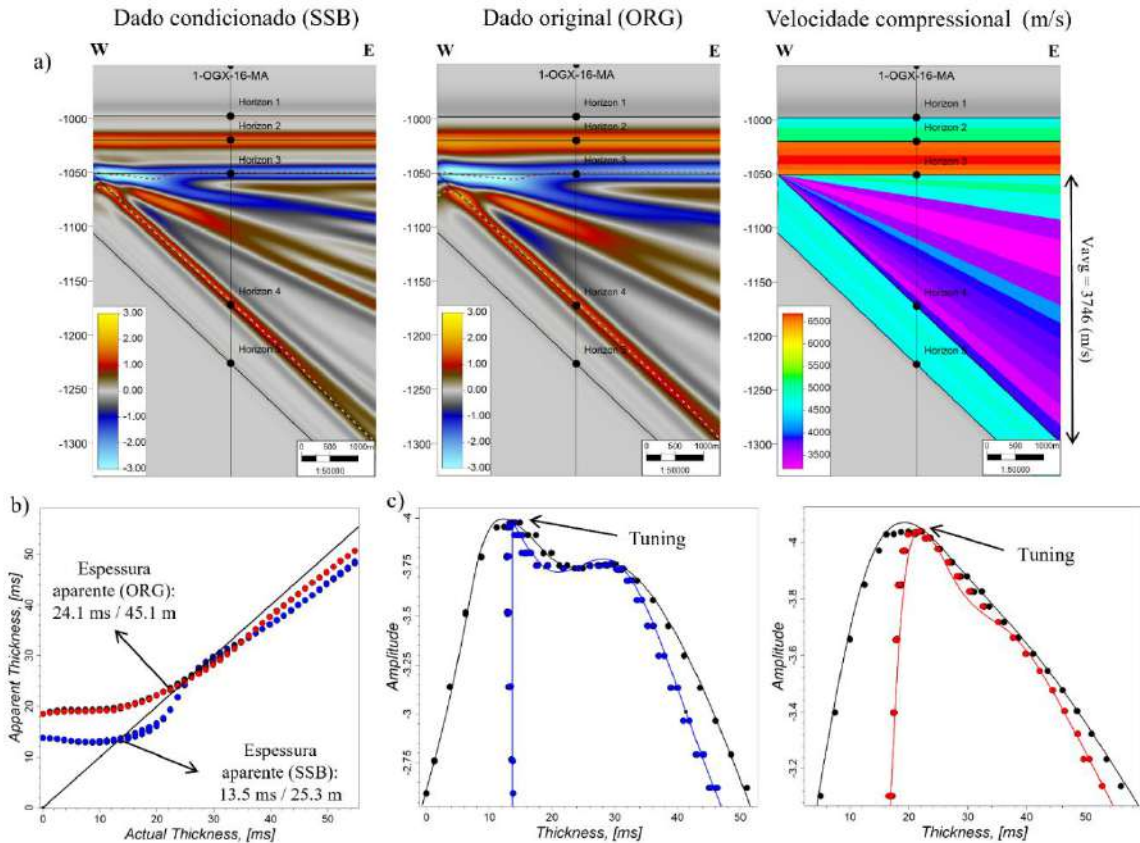


Figura 7.9. Os resultados do sismograma sintético gerado para o modelo de cunha são apresentados com base nas frequências dos dados sísmicos antes e depois do fluxo de trabalho de pré-condicionamento na parte a). Na parte b), são apresentados gráficos que comparam as espessuras real e aparente em milissegundos e metros para os dados originais e aprimorados. Por fim, na parte c), a amplitude é exibida em relação às espessuras real e aparente, com ênfase específica na amplitude de *tuning*.

A partir dos sismogramas sintéticos gerados, foi possível interpretar os horizontes correspondentes ao topo e à base da camada modelada, permitindo a construção de um gráfico comparando a espessura real com a espessura aparente (Figura 7.9b). Os resultados revelaram que a espessura aparente estimada para o dado original foi de 24,1 ms, enquanto para o dado pré-condicionado esse valor diminuiu consideravelmente para 13,5 ms. Convertendo essas medidas para unidades de metros, considerando a velocidade média da zona modelada de 3746 m/s (Figura 7.9a), obteve-se uma espessura mínima de 44,1 m para o dado original e de 25,3 m para o dado pré-condicionado. Na Figura 7.9c, são apresentados os gráficos de amplitude em função da espessura aparente, extraídos do topo da formação modelada. Esses gráficos mostram que a máxima amplitude ocorre em 13,9 ms para o dado pré-condicionado e em 23,4 ms para o dado original, correspondendo assim, as espessuras de *tuning* para cada conjunto de dados.

A fim de compreender o limite de resolução de acordo com as fácies encontradas na Formação Pimenteiras, associados aos folhelhos e soleiras de diabásio (ver Capítulo 4), foram calculadas a espessura de *tuning* a partir das relações estabelecidas por Kallwet e Wood (1982), considerando os critérios da *wavelet* Ricker e *wavelet* senoidal, para o dado original e pré-condicionado respectivamente. As propriedades de densidade e velocidade compressional de cada uma dessas fácies, estão mostradas nos gráficos de densidade abaixo (Figura 7.10) para os quatro poços da área de estudo, de onde foram consultados os valores de maior ocorrência para serem utilizados nos cálculos de estimativa das espessuras. O gráfico de densidade da Figura 7.10, chama atenção para o comportamento característico de cada fácies, mesmo se tratando de folhelhos para as fácies 2, 3 e 4. É possível notar uma sobreposição dos valores de velocidade entre a fácies 2 e 3, porém o gráfico de densidade separa de forma clara os diferentes valores de velocidade que mais ocorrem em cada uma delas.

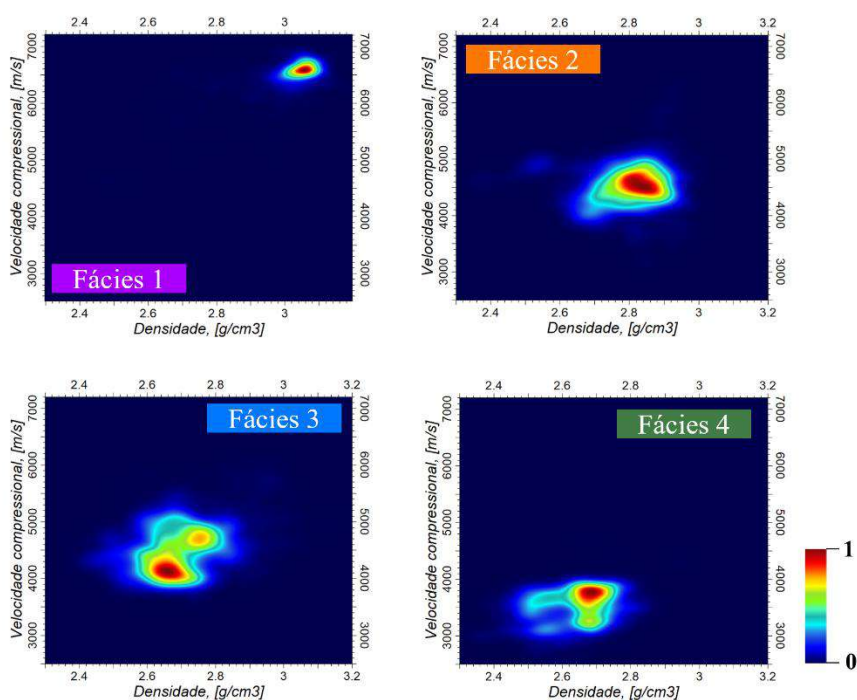


Figura 7.10. Gráficos de densidade 2D dado pela distribuição de velocidade compressional contra a densidade por fácies, onde os maiores valores (cores quentes) representam a região de maior concentração dos dados.

Portanto a frequência dominante e o comprimento de onda do pulso sísmico para o dado sísmico original foram calculados a partir do desenvolvimento das equações (7.4) e (7.7) e para o dado pré-condicionado segundo as equações a partir das equações (7.6) e (7.7), sendo possível estimar a espessura de *tuning* para as quatro fácies como descritos pela Tabela 7.1. Foi estimado a partir desse estudo um ganho de resolução de 38%.

Tabela 7.1. Avaliação da resolução sísmica em dados antes e depois do fluxo de trabalho de condicionamento para as litologias típicas da Formação Pimenteiras na Bacia do Parnaíba.

Fácies	Velocidade compressional	Dado original	Dado condicionado
	(m/s)	(m)	(m)
1 Soleira de diabásio	6570	84,23	52,14
2 Folhelho metamorfizado	4500	57,69	35,71
3 Folhelho razoável em COT	4100	52,56	32,54
4 Folhelho rico em COT	3700	47,44	29,36

Ganho de resolução: 38%

A partir da análise das espessuras das fácies identificadas nos poços, observa-se, com base na Figura 7.11, que é possível encontrar espessuras dentro dos limites de resolução estimados para o dado pré-condicionado, conforme mostrado na Tabela 7.1. Uma exceção a essa observação é o poço 3-OGX-57-MA, que não atravessou completamente a Formação Pimenteiras, resultando, de maneira geral, em espessuras reduzidas. Para a Fácies 2, foi possível identificar uma espessura acumulada que excede o valor de 35 m estimado como limite da espessura de tuning no poço 2-CP-1-MA. Esse resultado leva em consideração a presença de soleiras de diabásio de baixíssima espessura que intrudiram nesta camada, conforme observado no intervalo entre 2186 m e 2232 m.

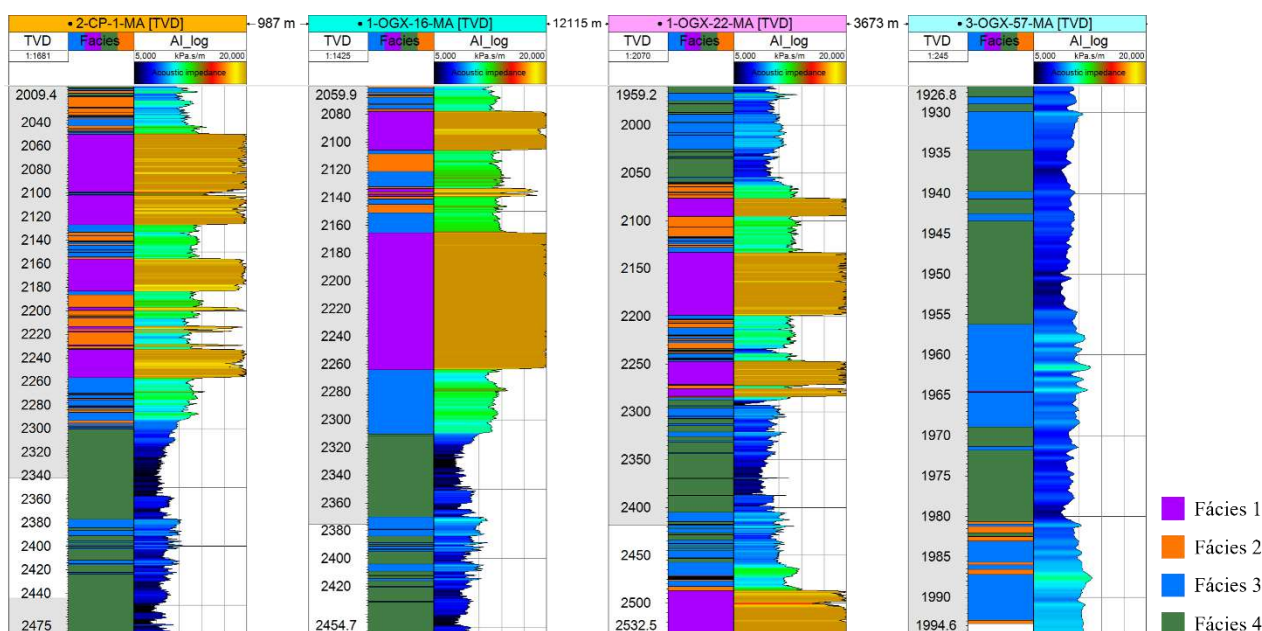


Figura 7.11. Perfil de fácies e impedância acústica dos poços, a fim de demonstrar principalmente as espessuras de cada fácies encontradas.

7.5. CONCLUSÃO

A análise integrada dos espectros de frequência, espectrogramas e sismogramas sintéticos demonstrou de forma quantitativa a eficácia do fluxo de condicionamento adotado. Observou-se

uma significativa atenuação do ruído de alta frequência após a aplicação do DSMF, seguida por uma recuperação eficiente do conteúdo de altas frequências com o uso do SSB, resultando em um espectro mais balanceado e com características *broadband*. A utilização do espectrograma gerado pela Transformada Wavelet Contínua reforçou esses resultados, evidenciando a presença de maiores de energia em faixas superiores a 30 Hz após o condicionamento indica um ganho real em resolução sísmica. Esse aprimoramento foi quantificado por meio da modelagem de sismogramas sintéticos com base em um modelo de cunha representativo da Formação Pimenteiras, em que os resultados revelaram uma redução significativa na espessura mínima resolvível, com uma diminuição de 44,1 m para 25,3 m, o que corresponde a um ganho de resolução de aproximadamente 38%. A estimativa das espessuras de *tuning* para diferentes fácies, baseada nos critérios de Kallweit e Wood (1982), corrobora esses achados e destaca a capacidade do dado pré-condicionado em distinguir feições geológicas sutilmente espessas, antes indetectáveis no dado original nos poços da área de estudo, como por exemplo as fácies 2 e 3.

8. Amarração Sísmica-Poço e a Construção do Grid 3D

8.1. INTRODUÇÃO

Este capítulo apresenta a amarração sísmica-poço, a interpretação dos horizontes que definem o topo e a base da Formação Pimenteiras e a construção do grid estratigráfico 3D que servirá de suporte à inversão sísmica e modelagem do carbono orgânico total. A amarração sísmica-poço é etapa essencial para integrar, de forma consistente, as escalas de medida dos perfis geofísicos de poço e do volume sísmico, assegurando a calibração temporal-estratigráfica necessária à interpretação confiável de limites de formações e ao condicionamento de modelos numéricos posteriores. No contexto desta tese, a amarração permitirá ainda quantificar o impacto das etapas de pré-condicionamento sísmico, por meio da comparação do desempenho do vínculo poço-traço antes e após cada tratamento aplicado ao dado. Outra informação importante resultado deste estudo será a geração da wavelet representativa do intervalo de interesse para a inversão sísmica tema do próximo.

8.2. FUNDAMENTAÇÃO

A amarração sísmica-poço (*seismic-well tie*) constitui uma etapa essencial na integração entre dados sísmicos e de poço, possibilitando estabelecer a correspondência entre os eventos observados nos registros sísmicos e as interfaces geológicas identificadas nos perfis geofísicos. Segundo Simm e Bacon (2014), o processo de amarração possui múltiplos objetivos, que incluem: (i) verificar se os dados sísmicos apresentam fase zero e, se necessário, ajustá-la; (ii) realizar a identificação estratigráfica de horizontes, correlacionando marcadores geológicos do poço com reflexões sísmicas; (iii) extrair a wavelet para fins de modelagem ou inversão sísmica; e (iv) avaliar se os dados sísmicos foram processados em amplitude verdadeira e, portanto, exibem comportamento AVO adequado, permitindo ajustes de amplitude quando necessário (*offset scaling*).

O sucesso desses objetivos depende de uma integração criteriosa entre a experiência interpretativa regional e a análise detalhada do poço, unindo os dados de sísmica de superfície, perfis sintéticos e perfis sísmicos verticais (VSPs). Conforme enfatizam Simm e Bacon (2014), o processo

de amarração pode ser visto como um “experimento” do intérprete, em que se testa a coerência entre a geologia e a resposta sísmica. A precisão da amarração e da wavelet extraída fornece o arcabouço para avaliar a qualidade da calibração e a confiabilidade das inversões sísmicas subsequentes.

8.2.1. Escalas e *Upscaling* de Velocidades

A calibração entre as escalas de poço e sísmica requer a consideração cuidadosa do *datum* de referência. Em dados terrestres, compreender o datum é fundamental, enquanto dados marinhos são usualmente referenciados ao nível médio do mar. Dessa forma, as velocidades derivadas de perfis de poço devem ser upscaladas para a escala sísmica antes da amarração, de modo que a resposta sísmica simulada reflita corretamente a resolução do dado.

De acordo com Simm e Bacon (2014), quando a razão entre o comprimento de onda (λ) e o espaçamento entre camadas (d) é menor ou igual a 1, condição típica das frequências de perfis sônicos, o método de média temporal (*time average*) é apropriado. Esse método consiste em somar os tempos de trânsito através de cada camada, o que, em perfis sônicos regularmente espaçados, equivale à média aritmética do tempo de trânsito (*slowness*) sobre uma janela móvel.

8.2.2. Calibração de Perfis de Velocidade

O processo de calibração dos perfis busca ajustar discrepâncias entre os tempos derivados do perfil sônico e os tempos observados nos *checkshots* (Figura 8.1). O fluxo de trabalho típico, conforme descrito por Simm e Bacon (2014), envolve:

1. Integração do perfil de velocidade a partir do ponto mais superior ou inferior do *checkshot* que o conecta;
2. Cálculo do *drift* (diferença entre o tempo sísmico e o tempo integrado do perfil);
3. Avaliação da qualidade dos *checkshots* com base nas curvas de *drift* e nas velocidades medidas;
4. Ajuste da função de *drift* até que as diferenças entre a curva tempo-profundidade e o perfil calibrado sejam inferiores a 2 ms;
5. Aplicação da correção de *drift* ao perfil integrado, resultando em uma curva tempo-profundidade calibrada e um perfil de velocidade corrigido.

O *drift* é interpretado como um efeito físico do meio, geralmente relacionado à dispersão de velocidades entre as frequências sísmicas e do perfil (Stewart et al., 1984). Normalmente, as velocidades sônicas são mais altas que as sísmicas, resultando em *drift* positivo, isto é, tempos

integrados mais curtos no perfil em relação à sísmica. O tipo de ajuste utilizado (linear com pontos de inflexão, polinomial ou spline) deve refletir as principais mudanças litológicas ou discontinuidades estratigráficas, evitando a introdução de reflexões artificiais no perfil sintético.

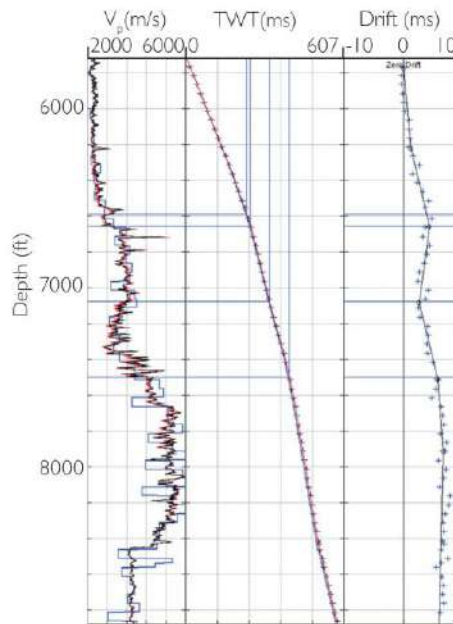


Figura 8.1. Esquema ilustrativo do processo de calibração entre o perfil sônico e o checkshot, destacando o cálculo do drift e o ajuste tempo–profundidade conforme o método descrito por Simm & Bacon (2014).

8.2.3. Extração e Ajuste da *Wavelet*

A estimativa da *wavelet* sísmica é uma das etapas mais críticas na calibração entre o dado sísmico e o poço, sendo determinante para a qualidade das inversões e para a interpretação quantitativa das amplitudes. Em termos conceituais, a *wavelet* representa a forma do pulso refletido por um refletor unitário e sintetiza os efeitos do sistema de aquisição, da resposta de propagação na subsuperfície e do processamento aplicado.

De modo geral, a extração de *wavelets* pode ser realizada por dois enfoques principais:

- (1) o método determinístico, fundamentado em dados de poço e no ajuste direto entre sismograma sintético e traço sísmico, e
- (2) o método estatístico, que estima a *wavelet* diretamente a partir dos próprios dados sísmicos, sem controle de poço.

O método determinístico é amplamente descrito por Simm & Bacon (2014) e White & Simm (2003) como o padrão de boas práticas em *well tie*. Nesse procedimento, a *wavelet* é estimada a partir da correspondência entre o sismograma sintético, obtido por convolução da série de refletividade

(calculada a partir da velocidade compressional e densidade calibrados), e o traço sísmico. A técnica proposta por White (1980), e posteriormente refinada por Walden & White (1998), trata o problema como uma relação de entrada e saída ruidosas, ajustando a wavelet por meio de mínimos quadrados e medindo a qualidade do ajuste pela Proporção de Energia Prevista (PEP):

$$PEP = 1 - \frac{E_{res}}{E_{trace}} \quad (8.1)$$

em que E_{res} representa a energia dos resíduos entre o traço sísmico e o sintético, e E_{trace} é a energia total do traço. Esse processo fornece não apenas a wavelet estimada, mas também medidas quantitativas de acurácia e aderência, expressas, por exemplo, pela raiz do erro quadrático normalizado (NMSE), que reflete a incerteza da fase e amplitude da wavelet.

Na prática, a wavelet é obtida sobre uma janela temporal, assumindo-se que dentro desse intervalo o caráter espectral e a fase permanecem aproximadamente estacionários. Segundo White & Simm (2003), wavelets muito curtas tendem a distorcer o sinal (bias), enquanto wavelets excessivamente longas incorporam ruído, sendo preferível adotar comprimentos intermediários que maximizem o *goodness-of-fit* sem superajuste.

Além do método determinístico, Simm e Bacon (2014) também descrevem a extração estatística da wavelet, uma técnica útil em situações nas quais não há poços disponíveis ou quando se busca uma estimativa inicial da wavelet representativa da área de estudo. Essa abordagem parte da suposição de que a série de refletividade é aleatória e não correlacionada, permitindo considerar o traço sísmico como a convolução de uma wavelet estacionária com essa série.

Os métodos estatísticos de extração de wavelet permitem estimar a forma e a fase do pulso sísmico diretamente a partir dos dados sísmicos, sem a necessidade de controle de poço. Além disso, os métodos estatísticos possibilitam a estimativa de wavelets com fase variável no tempo, superando a limitação dos métodos determinísticos, que dependem da curta extensão e da qualidade dos perfis de poços. Essa característica os torna particularmente úteis em áreas sem um bom controle de poço, como ferramentas de controle de qualidade, extrapolação de fase ou mesmo como método independente.

Esse método fornece uma estimativa da amplitude espectral da wavelet, e a fase pode ser posteriormente ajustada, por exemplo, para fase zero ou fase média, conforme o tipo de processamento e o objetivo da interpretação. Por não depender de dados de poço, trata-se de uma técnica de caráter global, frequentemente empregada para fins de zero-phasing, balanceamento de amplitude ou como referência inicial para comparação com wavelets extraídas de poços calibrados.

Contudo, conforme alertam Simm e Bacon (2014), a principal limitação do método estatístico é a sua dependência das suposições de aleatoriedade e estacionaridade da refletividade, o que pode comprometer a fidelidade da fase e a representatividade espectral da wavelet em áreas geologicamente complexas.

8.2.4. Construção do Sismograma Sintético

A construção do sismograma sintético é o passo subsequente à estimativa da wavelet e constitui o núcleo da amarração sísmica. De acordo com White & Simm (2003), o processo parte da geração da série de refletividade em tempo duplo (*two-way time*), obtida a partir dos perfis calibrados de velocidade e densidade.

O modelo sísmico é então construído pelo modelo convolucional, representando o sinal sísmico como a soma interferente de pulsos de reflexão de mesma forma, porém com amplitudes e polaridades distintas. O traço sintético resultante é dado por:

$$S(t) = r(t) * \omega(t) \quad (8.2)$$

onde $S(t)$ é o traço sintético, $r(t)$ é o coeficiente de reflexão em cada interface, $\omega(t)$ é a wavelet sísmica e t é o tempo de dupla viagem.

O objetivo é obter um traço sintético cuja forma e fase se ajustem ao traço sísmico real na vizinhança do poço. A qualidade do ajuste é avaliada por medidas quantitativas, como o coeficiente de correlação (R) ou a proporção de energia prevista (P)

Além disso, o método de coherency matching (White, 1980) pode ser aplicado para estimar simultaneamente o posicionamento ótimo do amarre, a forma da wavelet e a acurácia da fase, evitando práticas não recomendadas como o “*stretch and squeeze*” de sintéticos.

A qualidade do *well tie* depende fortemente da largura de banda e relação sinal-ruído dos dados sísmicos. White & Simm (2003) demonstram que, quando a largura de banda é limitada (<20 Hz), a ambiguidade entre tempo e fase aumenta, o que pode levar a erros de até 90° na interpretação da polaridade das reflexões. Assim, dados com larga faixa de frequência (>2 octavas) e S/N elevado são essenciais para obtenção de *ties* confiáveis.

Após a geração do sismograma sintético, é necessário avaliar a qualidade da amarração sísmica (*well tie quality*), o que se faz por meio da análise de correlação cruzada (cross-correlation) entre o traço sísmico observado e o sintético. Conforme discutido por Simm & Bacon (2014), a correlação

cruzada fornece uma métrica quantitativa e objetiva do grau de semelhança entre os dois sinais, complementando a análise visual tradicional (Figura 8.2).

A função de correlação cruzada mede a similaridade entre o traço sísmico e o sintético em função do deslocamento temporal. O pico máximo da função de correlação indica o grau de similaridade entre os sinais, enquanto a posição desse pico fornece o deslocamento temporal ótimo que maximiza o alinhamento entre o sintético e o dado real. Em amarrações bem calibradas, os valores de correlação são tipicamente superiores a 0,7 ou 0,8, o que indica boa correspondência de forma de onda e fase.

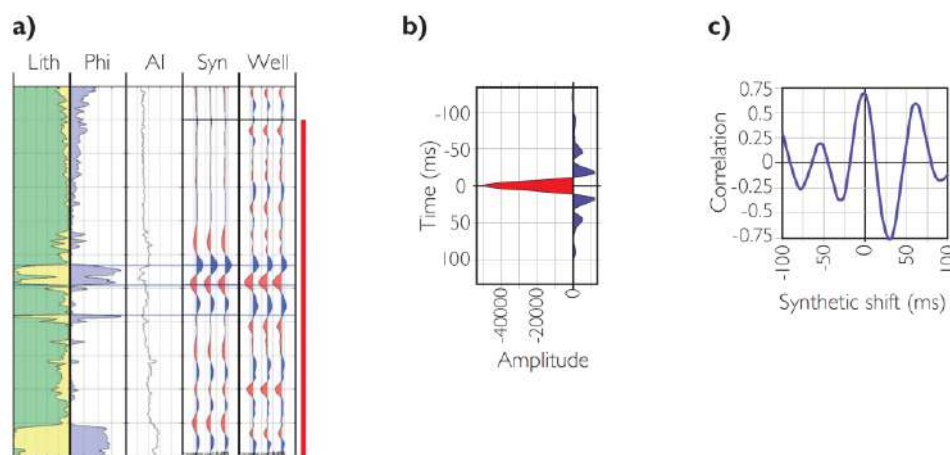


Figura 8.2. Representação esquemática da construção do sismograma sintético a partir da série de refletividade e da wavelet estimada, evidenciando o modelo convolucional clássico e a função de correlação cruzada utilizada para avaliação da qualidade da amarração (Simm & Bacon, 2014).

Simm & Bacon (2014) destacam que a análise deve ser realizada dentro de uma janela temporal centrada no intervalo de interesse, evitando a influência de ruídos ou reflexões fora da zona correlacionada. Um pico de correlação centrado em zero lag indica alinhamento adequado e fase consistente; já um pico deslocado revela a necessidade de ajuste temporal (*stretch* ou *squeeze*) ou de revisão da calibração tempo–profundidade. Correlações baixas podem indicar discrepâncias na fase, wavelets incorretas ou ruído no dado sísmico.

8.3. MÉTODO

A etapa de amarração sísmica-poço foi conduzida com base no fluxo de trabalho adaptado de Simm e Bacon (2014), conforme ilustrado na Figura 8.3. Esse processo tem como objetivo estabelecer a correspondência entre os eventos sísmicos e as interfaces geológicas observadas nos perfis de poço,

assegurando a coerência entre os domínios de tempo e profundidade e fornecendo a base para a extração de wavelets representativas e calibração de amplitudes.



Figura 8.3. Fluxograma do processo de amarração sísmica-poço adotado neste estudo, adaptado de Simm & Bacon (2014).

Inicialmente, foram realizados procedimentos de condicionamento dos perfis sônico (DT) e de densidade (RHOZ), com o objetivo de eliminar pontos espúrios (*spikes*) e inconsistências associadas à aquisição. Esse tratamento foi conduzido com o apoio da análise dos perfis *Gamma Ray* (GR) e Caliper (CALI), utilizados como referência para identificar intervalos afetados por instabilidades na parede do poço e possíveis arrombamentos.

O condicionamento dos perfis foi mais intensivo nos poços 1-OGX-16-MA e 2-CP-1-MA, nos quais se observou uma maior ocorrência de anomalias, correlacionadas a valores elevados no perfil CALI, conforme apresentado na Figura 8.4. Para correção dos dados espúrios, foram aplicados filtros de que identifica estatisticamente pontos espúrios dado uma janela de investigação, para remoção e interpolação local.

Na sequência, foi aplicada uma filtragem em frequência sobre o perfil sônico (DT), de modo a compatibilizar seu conteúdo espectral com o da sísmica. Por natureza de aquisição, os perfis de poço apresentam componentes de alta frequência que não estão presentes na banda sísmica. Assim, o filtro de frequência foi projetado buscando alinhar as bandas dominantes de ambos os domínios. Esse procedimento, ilustrado nas Figura 8.4 e Figura 8.5, resultou em curvas sônicas suavizadas e compatíveis com a faixa de frequência sísmica. Ainda na Figura 8.5 é mostrado o espectro do dado sísmico e do DT após a filtragem.

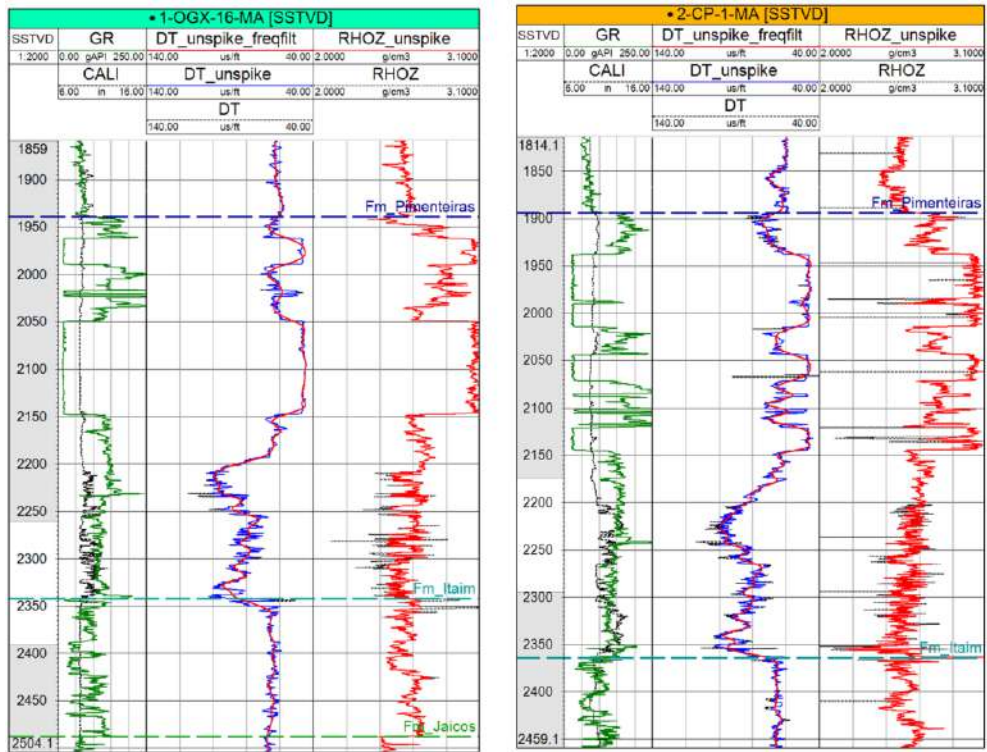


Figura 8.4. Perfis sônico (DT), densidade (RHOZ), gamma ray (GR) e caliper (CALI) após o condicionamento, destacando a remoção de pontos espúrios e a correção de intervalos afetados por instabilidade de parede nos poços 1-OGX-16-MA e 2-CP-1-MA.

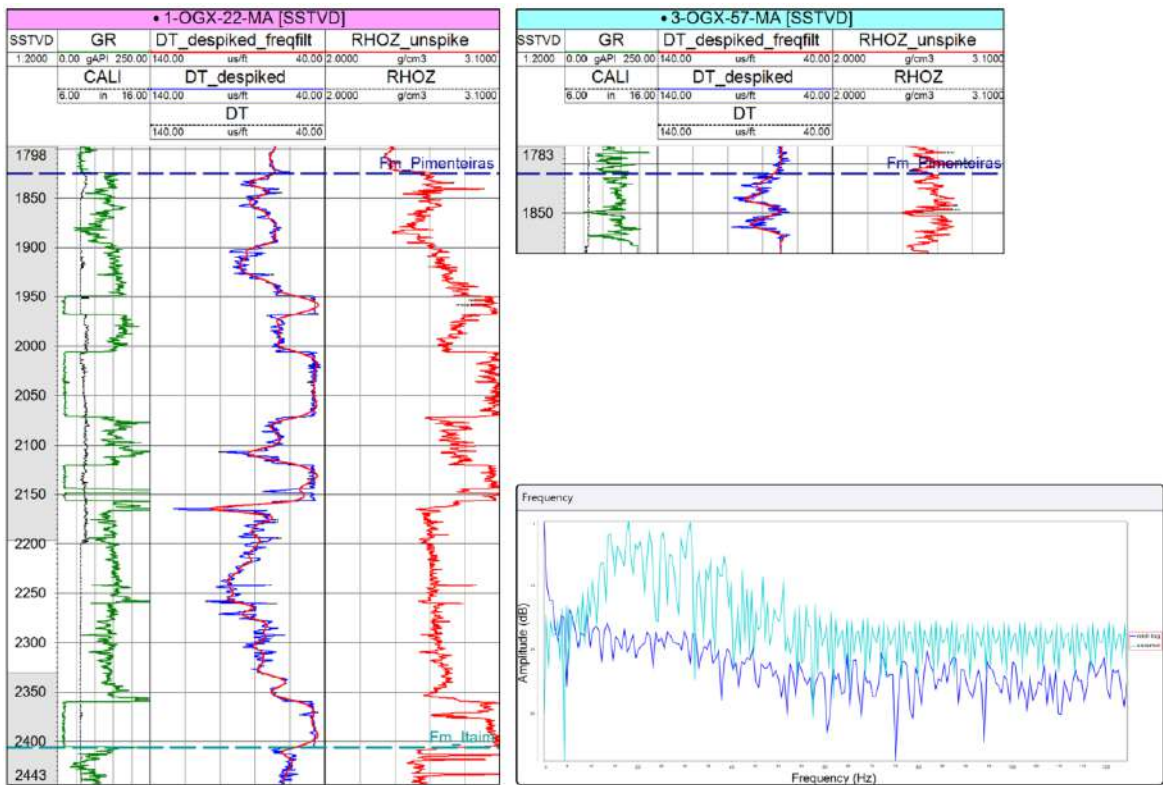


Figura 8.5. Perfis sônico (DT), densidade (RHOZ), gama natural (GR) e caliper (CALI) após o condicionamento para os poços 1-OGX-22-MA e 3-OGX-57-MA, além de demonstrar a filtragem em frequência aplicada a todos os poços.

Com os perfis condicionados, realizou-se a calibração do perfil sônico a partir dos dados de checkshot disponíveis. Na área de estudo, apenas os poços 2-CP-1-MA e 1-OGX-16-MA dispõem de medições diretas de checkshot, o que permitiu gerar curvas tempo–profundidade (TDR) mais precisas para esses casos.

Para os poços 3-OGX-57-MA e 1-OGX-22-MA, foi adotada uma calibração indireta, utilizando como referência os checkshots dos poços vizinhos, de modo a estabelecer uma relação inicial tempo–profundidade aproximada. Ressalta-se que a limitação de dados de checkshot representa uma fonte de incerteza local, mitigada por meio da correlação estratigráfica e do controle regional. As medições disponíveis foram interpoladas linearmente, priorizando o intervalo de interesse e ajustando-se aos marcadores estratigráficos das Formações Pimenteiras e Itaim, como ilustrado na Figura 8.6.

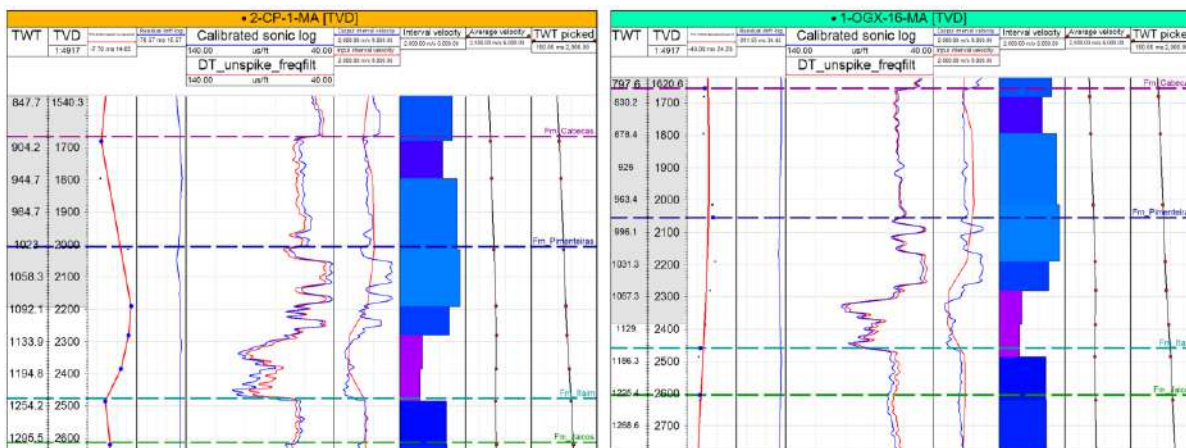


Figura 8.6. Calibração do perfil sônico a partir dos dados de checkshot medidos nos poços 1-OGX-16-MA e 2-CP-1-MA.

A extração das wavelets sísmicas foi realizada pelo método estatístico, aplicando o *taper* do tipo Papoulis nas janelas temporais centradas na posição dos poços e restritas ao intervalo de interesse. O processo buscou obter wavelets representativas em forma de de fase zero, coerentes com a resposta do dado sísmico na região. Durante a análise espectral, observou-se que a frequência dominante dos dados sísmicos varia com o tempo, refletindo diferenças nas propriedades de atenuação e no conteúdo de ruído. Por essa razão, as wavelets foram extraídas em intervalos estratigráficos específicos, correspondentes às formações principais, permitindo incorporar variações de amplitude ao longo do tempo.

Por fim, considerando que um dos objetivos deste estudo é avaliar o impacto do pré-condicionamento sísmico na amarração, as wavelets foram analisadas de forma independente para três versões do dado sísmico: original, pós-*Dip Steered Median Filter* e pós-*Spectral Blueing*. Dessa forma,

os resultados apresentados a seguir concentram-se no intervalo de interesse, permitindo comparar a influência de cada pré-condicionamento na qualidade da amarração sísmica-poço.

8.3.1. Dado Sísmico Original

O processo de amarração sísmica dos poços iniciou-se com a geração dos sismogramas sintéticos, utilizando como dados de entrada os perfis DT e RHOZ, previamente condicionados e calibrados, além da curva TDR obtida a partir da calibração entre o perfil sônico e os dados de *checkshot*.

A partir desses perfis, foi calculada a impedância acústica, definida pelo produto entre a velocidade ($1/\Delta t$) e a densidade (ρ), da qual deriva a série de refletividade do poço. Essa série foi então convolvida com a wavelet estatística extraída, resultando na geração do sismograma sintético, que representa a resposta sísmica teórica no domínio do tempo para a posição do poço. O sismograma sintético foi posteriormente ajustado ao dado sísmico real, extraído da seção que intercepta a localização do poço.

O objetivo principal dessa etapa foi alinhar o sismograma sintético ao dado sísmico com o mínimo de distorção temporal, evitando a aplicação excessiva de transformações de *stretch* e *squeeze*, que podem comprometer a fidelidade geológica do ajuste. Para isso, adotou-se uma abordagem baseada na definição de âncoras de correlação, pontos de controle selecionados em intervalos de maior confiabilidade geológica, determinados a partir do conhecimento regional da bacia e da análise comparativa entre o traço sintético e o traço sísmico real.

Após a definição das âncoras, pequenas correções de *stretch* e *squeeze* foram aplicadas apenas em intervalos específicos, visando otimizar o ajuste sem alterar significativamente a relação tempo–profundidade. Essa etapa foi conduzida com atenção especial à consistência das velocidades derivadas.

Durante o processo de amarração, informações quantitativas como o coeficiente de correlação cruzada, o residual entre o traço sísmico e o sintético, e as velocidades intervalares de entrada e saída (*input/output interval velocity*) foram monitoradas cuidadosamente. Esses parâmetros são fundamentais para identificar possíveis distorções introduzidas durante os ajustes, tais como variações abruptas ou fisicamente inconsistentes nas velocidades intervalares, que poderiam indicar problemas de calibração ou posicionamento de marcadores.

A combinação dos dados de entrada e saída, incluindo os perfis condicionados, a wavelet utilizada, as posições de ancoragem selecionadas e os indicadores de qualidade do ajuste, é apresentada

nas Figura 8.7 a Figura 8.10, correspondentes às amarrações realizadas nos poços, 2-CP-1-MA, 1-OGX-16-MA, 1-OGX-22-MA e 3-OGX-57-MA, respectivamente.

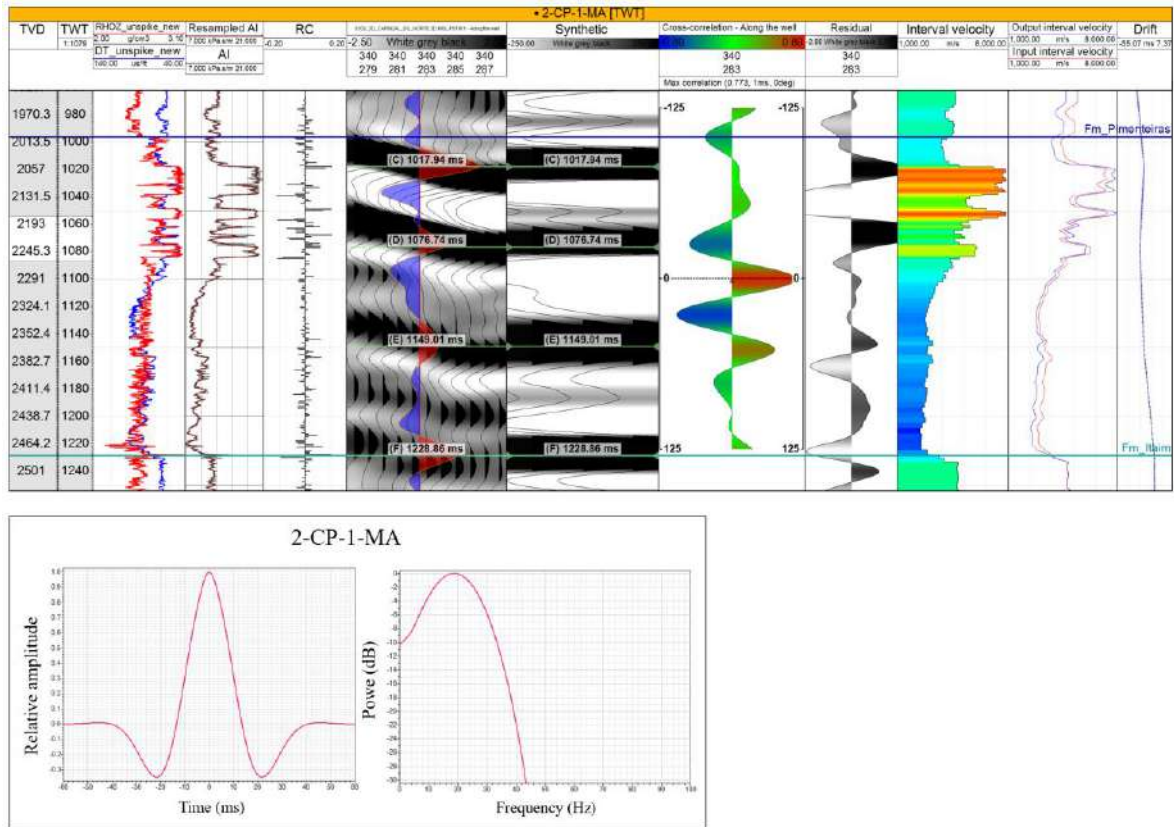


Figura 8.7. Amarração sísmica-poço para o poço 2-CP-1-MA utilizando o dado sísmico original.

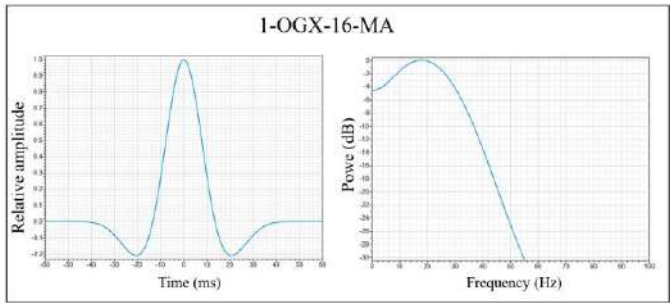
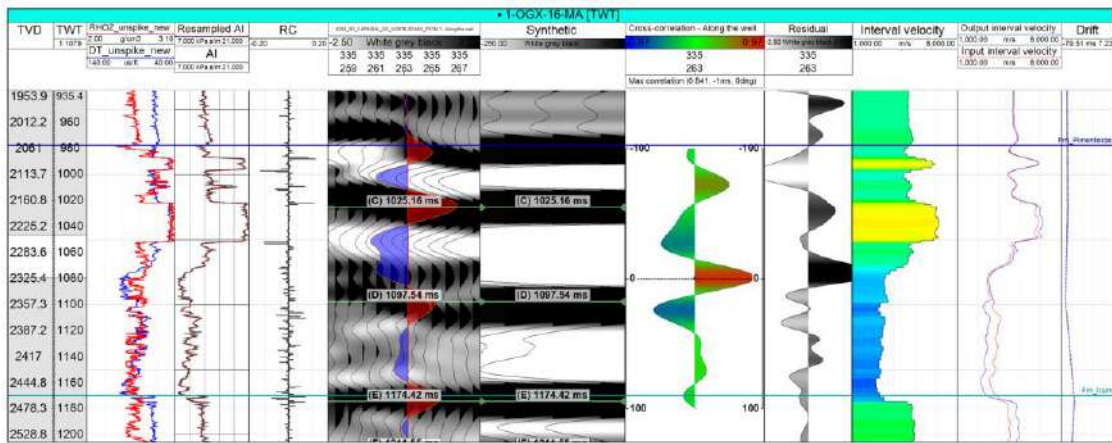


Figura 8.8. Amarração sísmica-poço para o poço 1-OGX-16-MA utilizando o dado sísmico original.

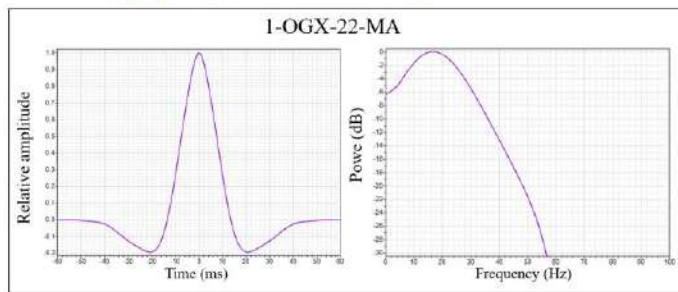
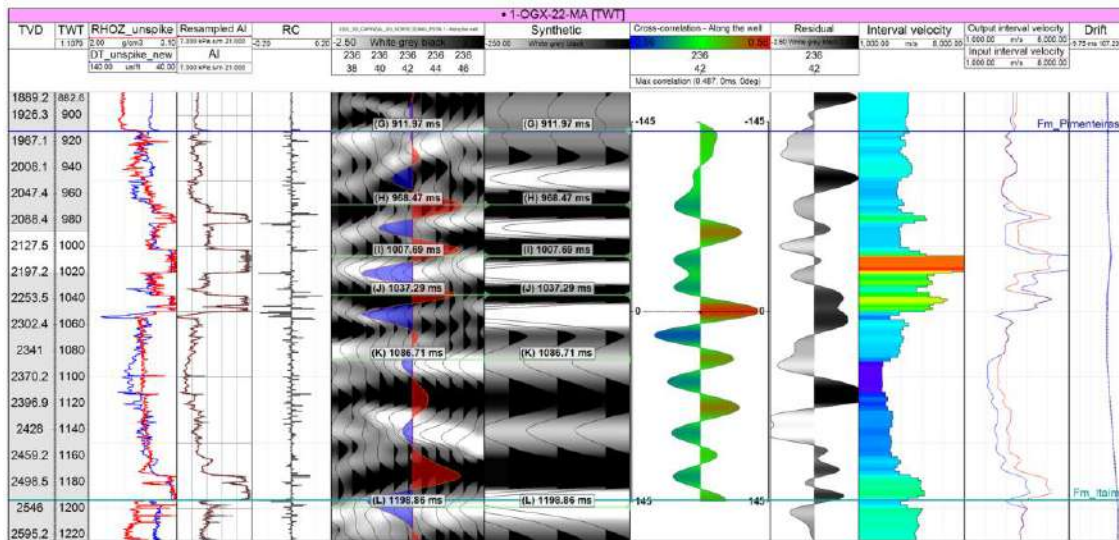


Figura 8.9. Amarração sísmica-poço para o poço 1-OGX-22-MA utilizando o dado sísmico original.

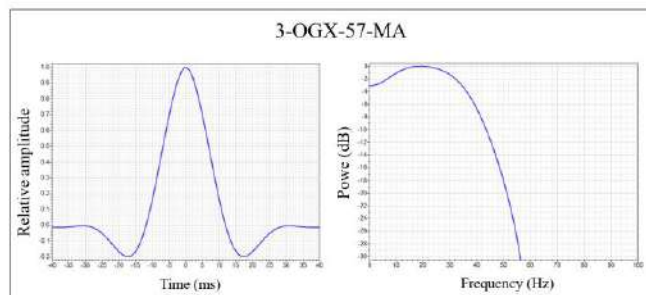
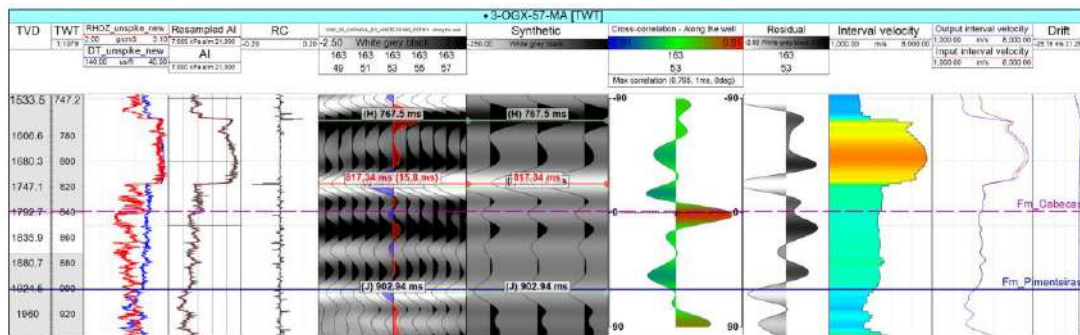


Figura 8.10. Amarração sísmica-poço para o poço 3-OGX-57-MA utilizando o dado sísmico original.

8.3.2. Dado Sísmico após o *Dip Steered Median Filter*

O mesmo procedimento metodológico descrito para o dado sísmico original foi aplicado à amarração dos poços ao dado sísmico após a primeira etapa de pré-condicionamento, que consistiu na aplicação do Dip-Steered Median Filter, que tem como principal objetivo a remoção de ruídos aleatórios.

Após a aplicação do DSMF, foram novamente extraídas as wavelets estatísticas na posição de cada poço, empregando o mesmo método descrito na seção anterior, isto é, o método estatístico com taper do tipo Papoulis. As extrações foram realizadas para o mesmo intervalo temporal utilizado no processo de amarração anterior, garantindo comparabilidade direta entre as duas etapas (dado original e dado filtrado).

Com as novas wavelets extraídas, os sismogramas sintéticos foram recalculados e ajustados aos dados sísmicos DSMF. Durante esse processo, observou-se que os pontos de ancoragem (âncoras) e os parâmetros de stretch e squeeze apresentaram variações em relação à amarração anterior, refletindo as diferenças de amplitude e continuidade introduzidas pelo pré-condicionamento. Os critérios adotados para o estabelecimento das âncoras, os ajustes aplicados e os indicadores de correlação cruzada e residual podem ser visualizados nas Figura 8.11 e Figura 8.14, que apresentam um panorama geral da amarração dos poços 2-CP-1-MA, 1-OGX-16-MA, 1-OGX-22-MA e 3-OGX-57-MA, respectivamente, após a aplicação do DSMF.

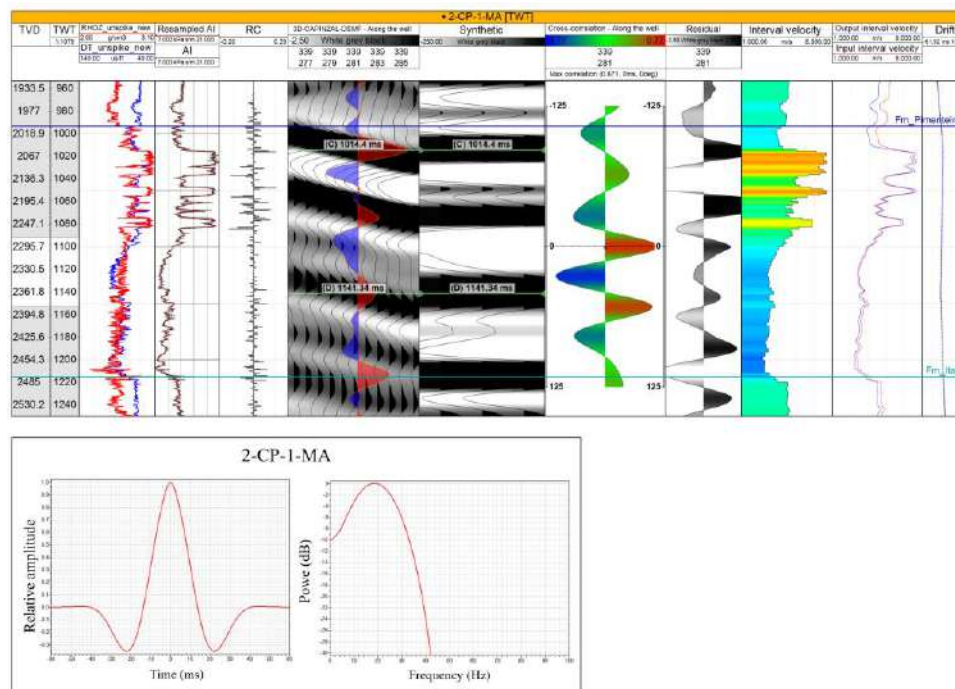


Figura 8.11. Amarração sísmica-poço para o poço 2-CP-1-MA após aplicação do Dip-Steered Median Filter.

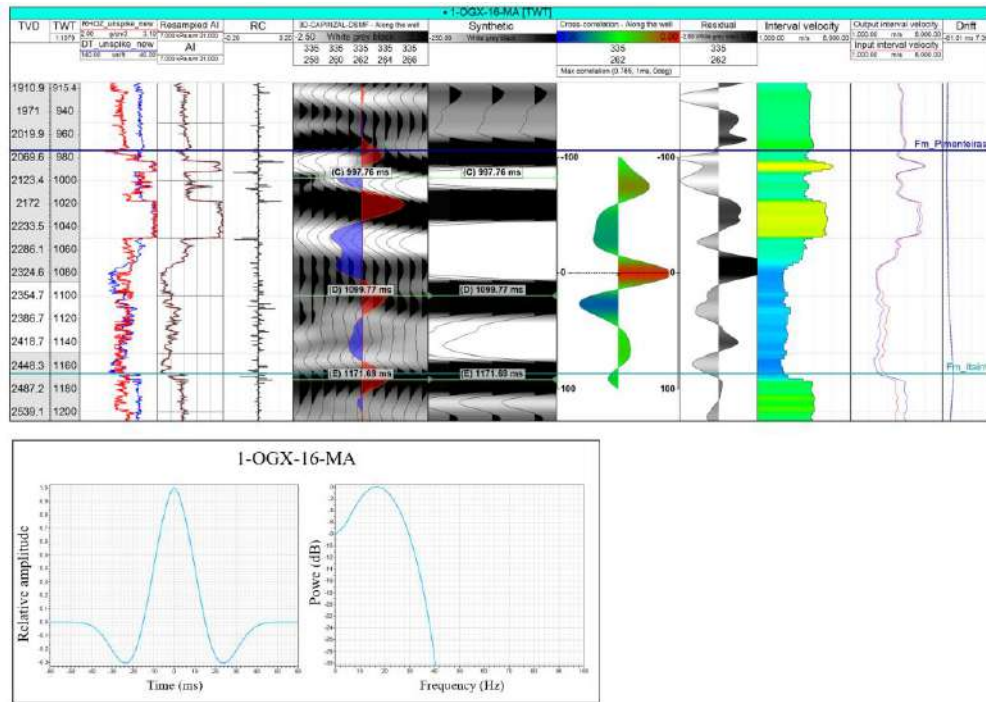


Figura 8.12. Amarração sísmica-poço para o poço 1-OGX-16-MA após aplicação do Dip-Steered Median Filter.

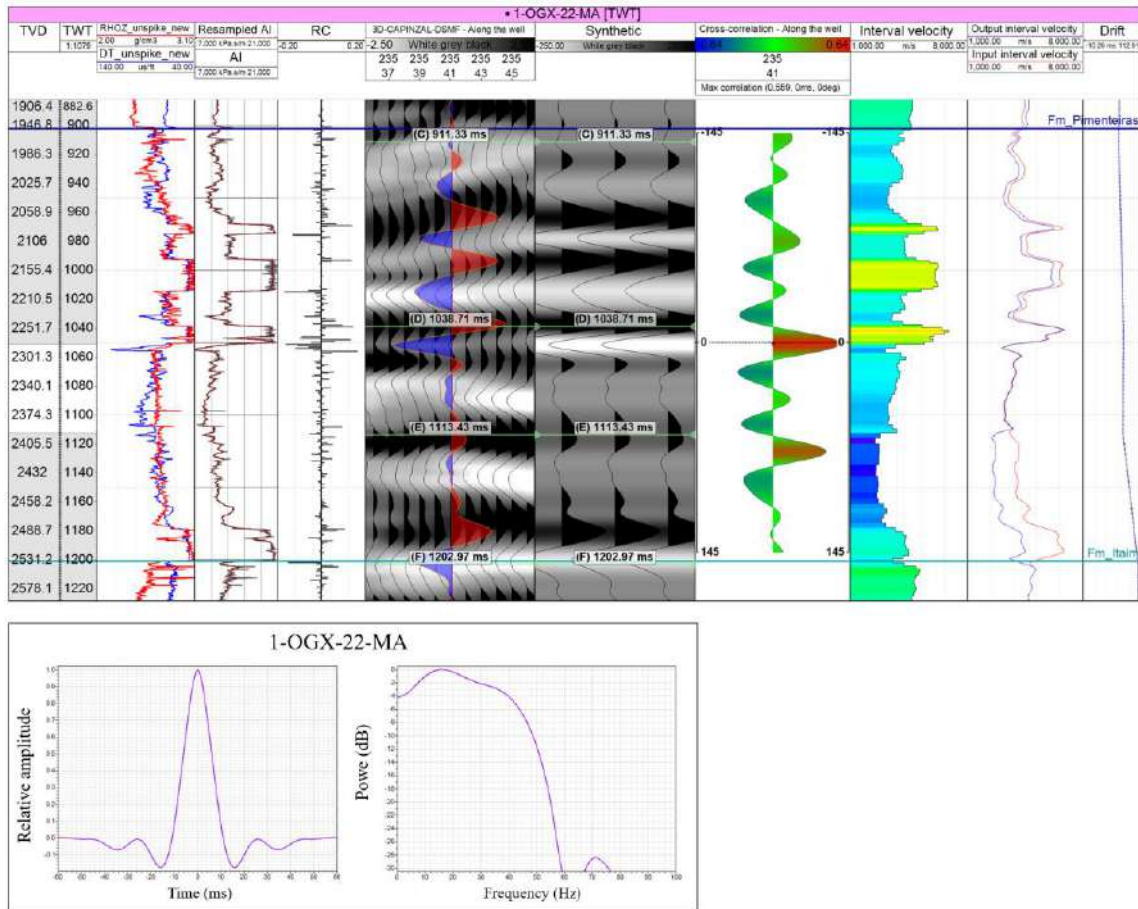


Figura 8.13. Amarração sísmica-poço para o poço 1-OGX-22-MA após aplicação do Dip-Steered Median Filter.

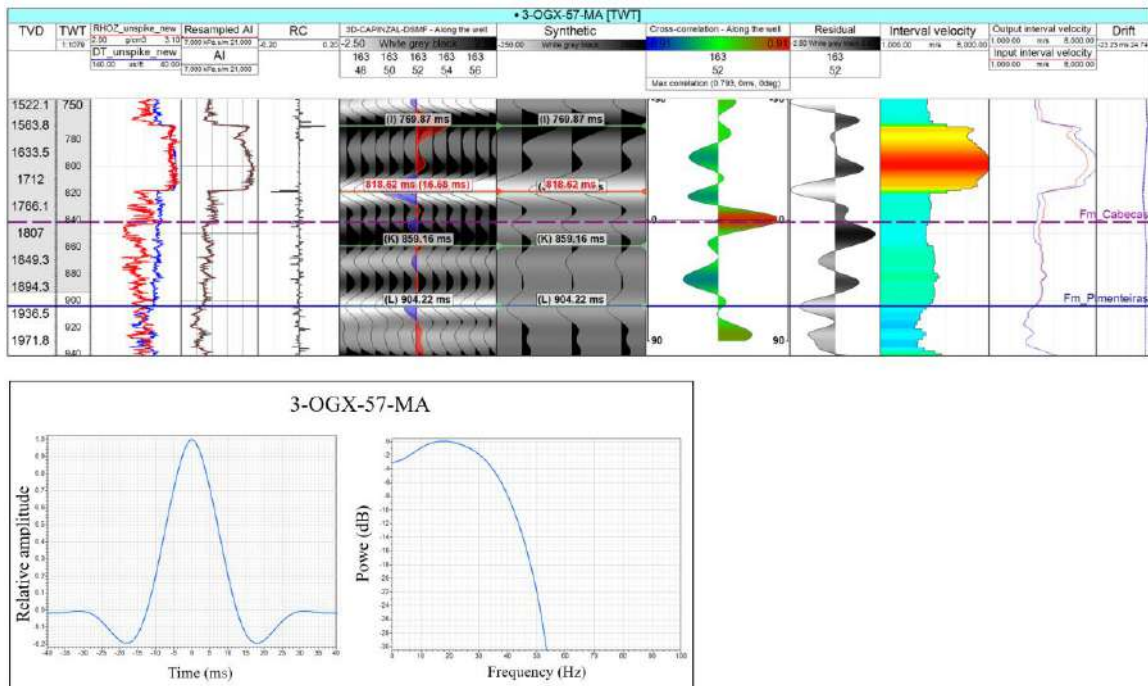


Figura 8.14. Amarração sísmica-poço para o poço 3-OGX-57-MA após aplicação do Dip-Steered Median Filter.

8.3.1. Dado Sísmico após o *Spectral Blueing*

Por fim, o mesmo procedimento metodológico foi aplicado à amarração dos poços utilizando o dado sísmico após o pré-condicionamento completo, que combinou o filtro de ruídos Dip Steered Median Filter com o ganho de frequência pelo método Spectral Blueing. Após o pré-condicionamento, as wavelets foram novamente extraídas pelo método estatístico, empregando o mesmo taper do tipo Papoulis, e calculadas na posição de cada poço para o intervalo de interesse previamente estabelecido. A comparação entre as wavelets extraídas antes e depois do Spectral Blueing evidencia o ganho de conteúdo espectral de alta frequência no dado pré-condicionado.

A geração dos sismogramas sintéticos seguiu os mesmos parâmetros adotados nas etapas anteriores, e o processo de amarração sísmica-poço foi conduzido com base na definição de âncoras de correlação, aplicando-se correções pontuais de *stretch* e *squeeze* apenas quando necessário para alinhar o sintético ao traço sísmico real. As Figura 8.15 a Figura 8.18 apresentam o panorama da amarração realizada para os poços 1-OGX-16-MA, 2-CP-1-MA, 1-OGX-22-MA e 3-OGX-57-MA, respectivamente, destacando as wavelets extraídas, os painéis de correlação cruzada, as curvas de velocidade intervalar e o controle de qualidade da amarração.

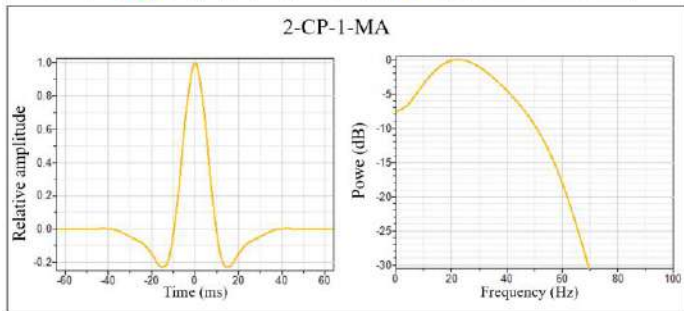
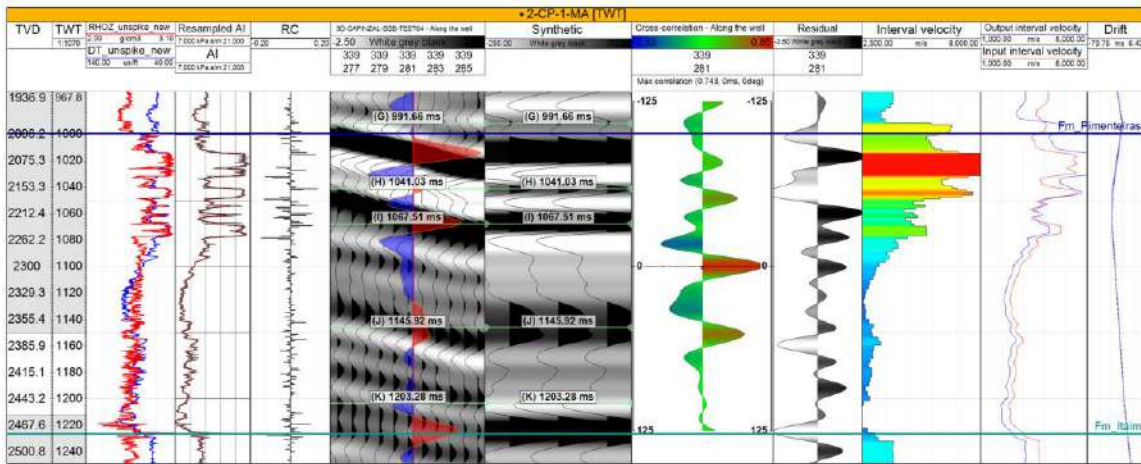


Figura 8.15. Amarração sísmica-poço para o poço 2-CP-1-MA após aplicação do Spectral Bluing.

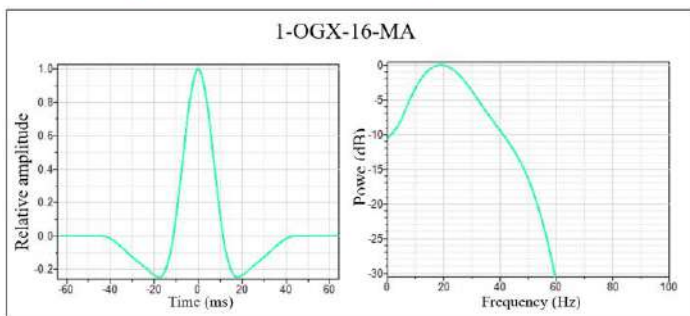
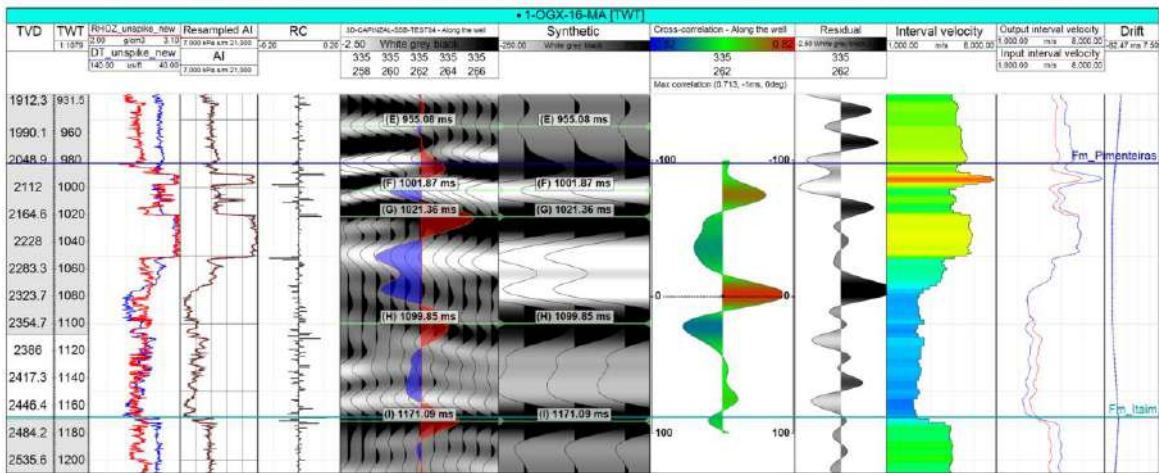


Figura 8.16. Amarração sísmica-poço para o poço 1-OGX-16-MA após aplicação do Spectral Bluing.

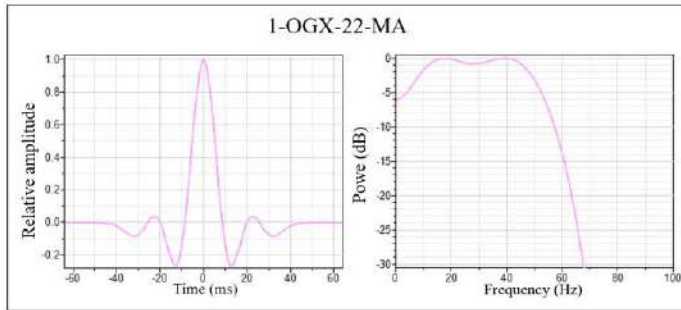
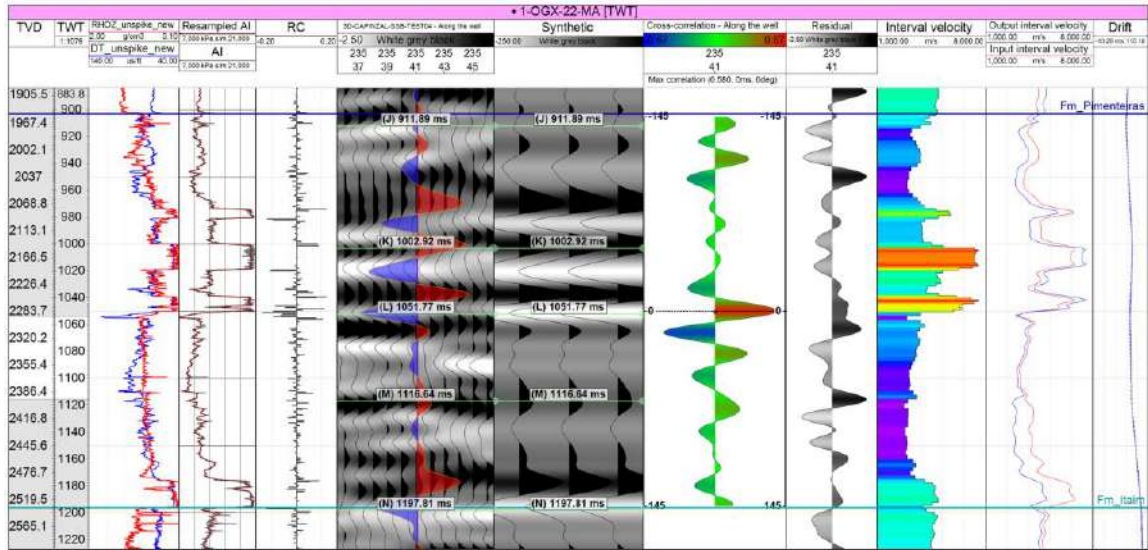


Figura 8.17. Amarração sísmica-poço para o poço 1-OGX-22-MA após aplicação do Spectral Blueing.

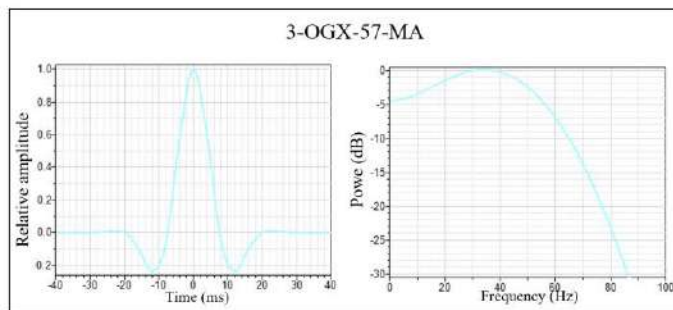
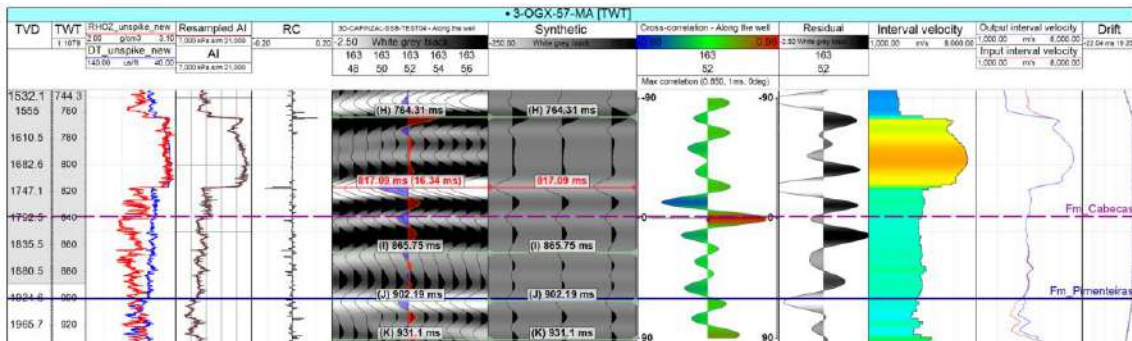


Figura 8.18. Amarração sísmica-poço para o poço 3-OGX-57-MA após aplicação do Spectral Blueing.

8.4. RESULTADOS E DISCUSSÕES

8.4.1. Amarração Sísmica-Poço

Os resultados da amarração sísmica-poço, obtidos conforme a metodologia descrita anteriormente, são apresentados na Figura 8.19 para os poços 1-OGX-16-MA e 2-CP-1-MA. Nessa figura, observa-se a seção sísmica condicionada (pós-DSMF e SSB) com os poços e os marcadores geológicos posicionados no domínio do tempo.

A Formação Pimenteiras apresenta uma resposta sísmica característica marcada por um pico negativo (pic branco) na entrada e um pico positivo (preto) na saída, que corresponde à interface com a Formação Itaim. Esse comportamento é consistente com o contraste de impedância entre os folhelhos ricos em matéria orgânica da Pimenteiras e os arenitos da Itaim.

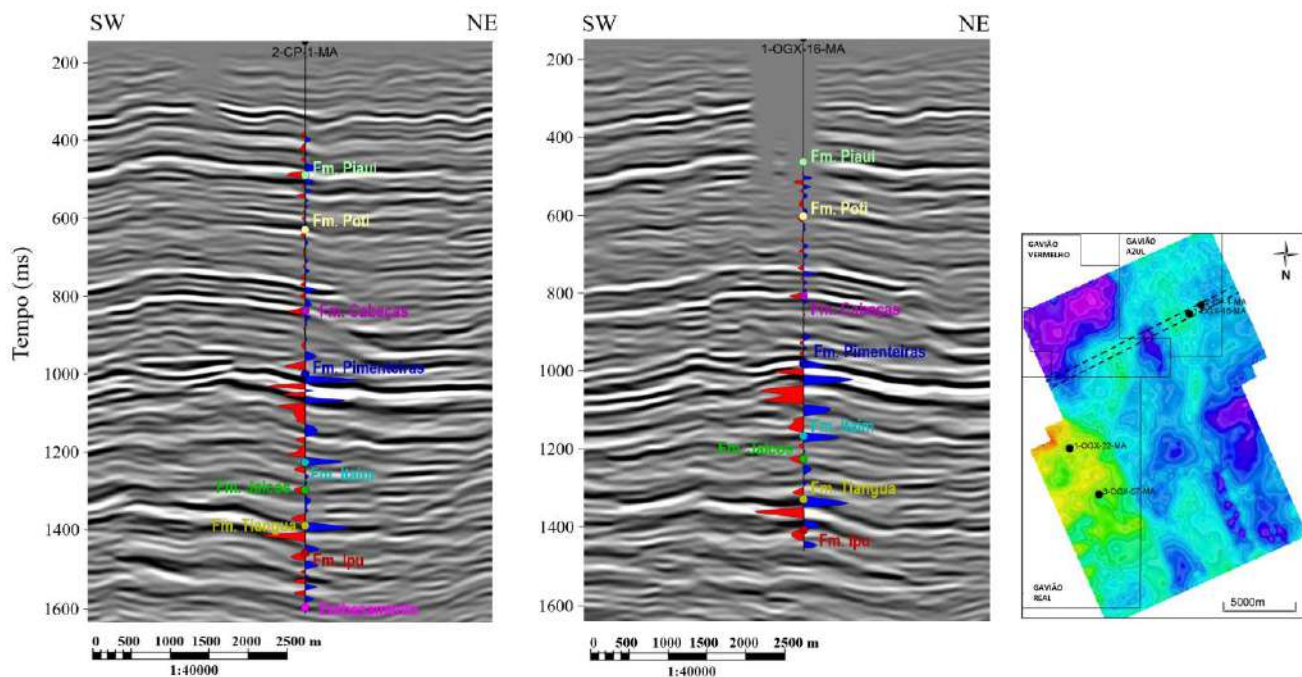


Figura 8.19. Amarração sísmica-poço para os poços 1-OGX-16-MA e 2-CP-1-MA mostrando o comportamento do refletor da Formação Pimenteiras.

Contudo, em intervalos onde há intrusões ígneas na base da Formação Pimenteiras, verifica-se uma inversão de polaridade, resultando em uma saída de pico negativo, como evidenciado na Figura 8.20, no poço 1-OGX-22-MA. Essa inversão está associada a uma diminuição abrupto de impedância na interface Pimenteiras/Itaim, causado pela presença de soleiras de diabásio de alta impedância. Ainda na Figura 8.20, é mostrado o resultado da amarração para o poço 3-OGX-57-MA, o qual, embora não

atravessa toda a Formação Pimenteiras, foi essencial para definir corretamente o topo da formação, etapa fundamental para a interpretação sísmica em uma área com escassez de poços.

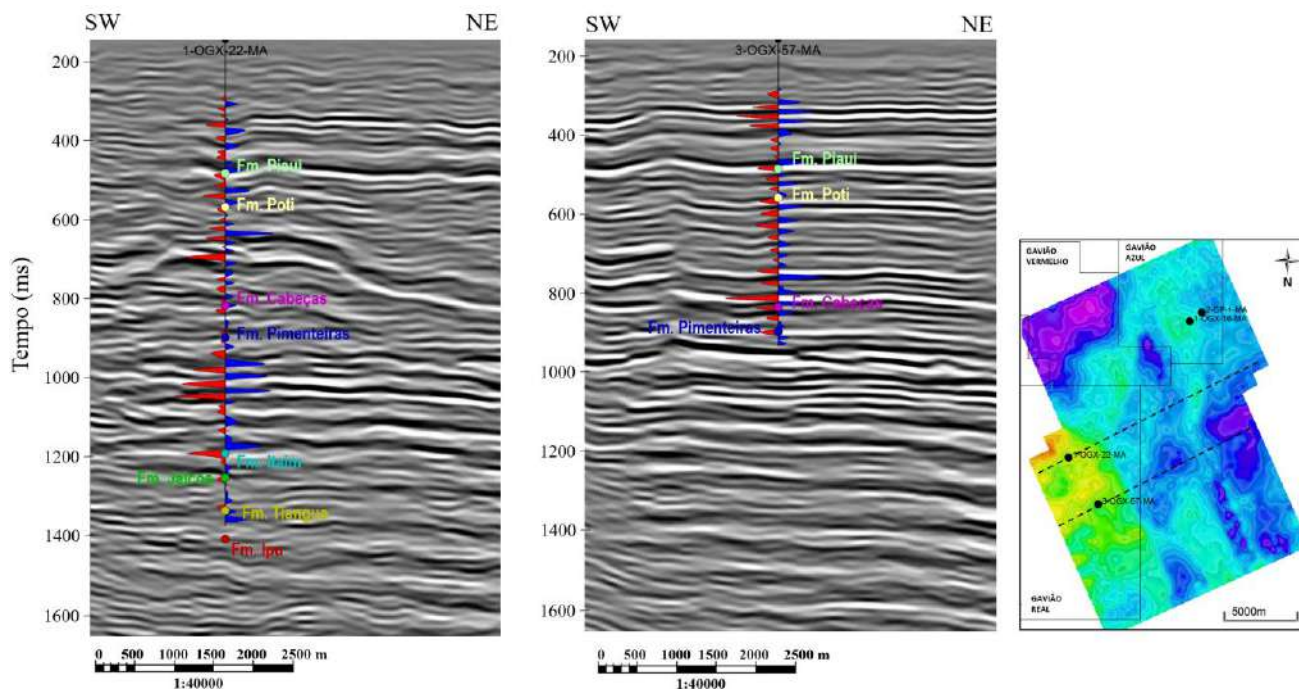


Figura 8.20. Amarração sísmica–poço para os poços 1-OGX-22-MA e 3-OGX-57-MA, evidenciando inversão de polaridade associada à presença de soleiras na base da Formação Pimenteiras.

A fim de avaliar a consistência da amarração sísmica–poço e a continuidade lateral dos principais refletores interpretados, em especial aquelas associadas ao topo da Formação Pimenteiras, foi construída uma seção sísmica arbitrária passando pelos poços CP-1, OGX-16, OGX-22 e OGX-57 (Figura 8.21). Nessa seção, os traços sísmicos sintéticos gerados a partir dos dados de poço foram sobrepostos aos dados sísmicos, permitindo a comparação direta entre os eventos refletivos observados e aqueles previstos a partir das propriedades elásticas medidas em poço. Observa-se uma boa correspondência entre os refletores do dado sísmico e os picos e vales do traço sintético, indicando uma amarração satisfatória entre os dados de poço e a sísmica.

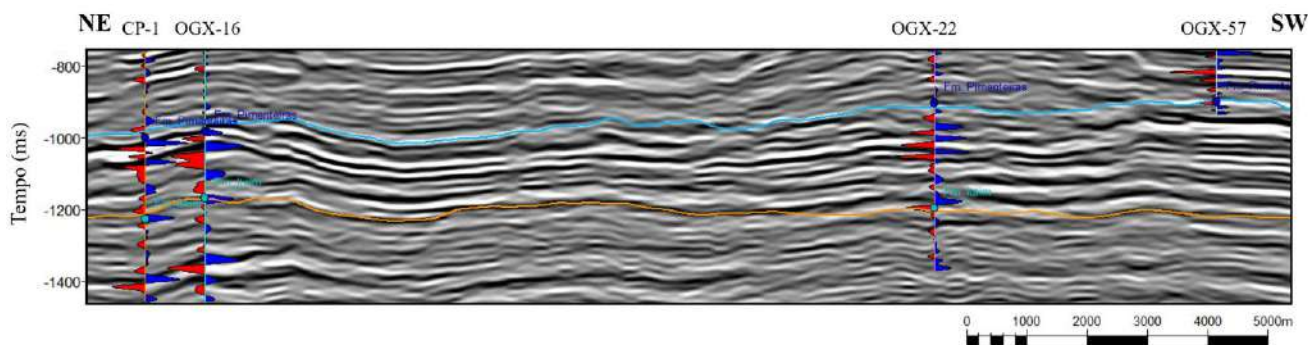


Figura 8.21. Seção sísmica arbitrária orientada NE–SW passando pelos poços CP-1, OGX-16, OGX-22 e OGX-57, após a amarração dos poços à sísmica.

Com o objetivo de avaliar o impacto do pré-condicionamento sísmico sobre a qualidade da amarração, foi realizada uma análise comparativa da correlação cruzada entre o sismograma sintético e o traço sísmico, considerando os três volumes: dado original, dado filtrado por DSMF, e dado pré-condicionado (DSMF + SSB). A Tabela 8.1 apresenta os valores de correlação obtidos. De modo geral, todos os poços apresentaram valores superiores a 0,6, o que indica uma boa coerência entre o sintético e o dado sísmico. A exceção foi o poço 1-OGX-22-MA, cuja correlação aumentou para aproximadamente 0.6 apenas após o pré-condicionamento completo, evidenciando a eficácia do processamento para esse caso.

Em contrapartida, para os poços 2-CP-1-MA e 1-OGX-16-MA, as maiores correlações foram obtidas com o dado original, o que sugere que a função de correlação cruzada é mais sensível a pequenas diferenças de resolução. Assim, mesmo leves desajustes temporais entre o sintético e o traço sísmico podem reduzir o coeficiente de correlação. Isso também indica que, em dados de alta resolução, métodos tradicionais de estimativa de wavelet podem não ser os mais adequados, e abordagens alternativas podem ser testadas em estudos futuros.

Tabela 8.1. Valores de correlação cruzada entre o sismograma sintético e o traço sísmico para os diferentes volumes testados (Original, DSMF e SSB)

	Correlação Cruzada			
	2-CP-1	1-OGX-16	1-OGX-22	1-OGX-57
SSB	0,743	0,713	0,580	0,850
DSMF	0,671	0,785	0,559	0,793
ORIGINAL	0,773	0,840	0,487	0,795

A Figura 8.22 mostra a distribuição dos resíduos (diferença entre o sismograma sintético e o traço sísmico) para o dado original e o dado pré-condicionado. Nota-se que, após o pré-

condicionamento, a distribuição se aproxima mais de uma curva normal, concentrando-se em torno de zero e apresentando menor dispersão, o que indica melhor ajuste entre o sintético e o dado sísmico real.

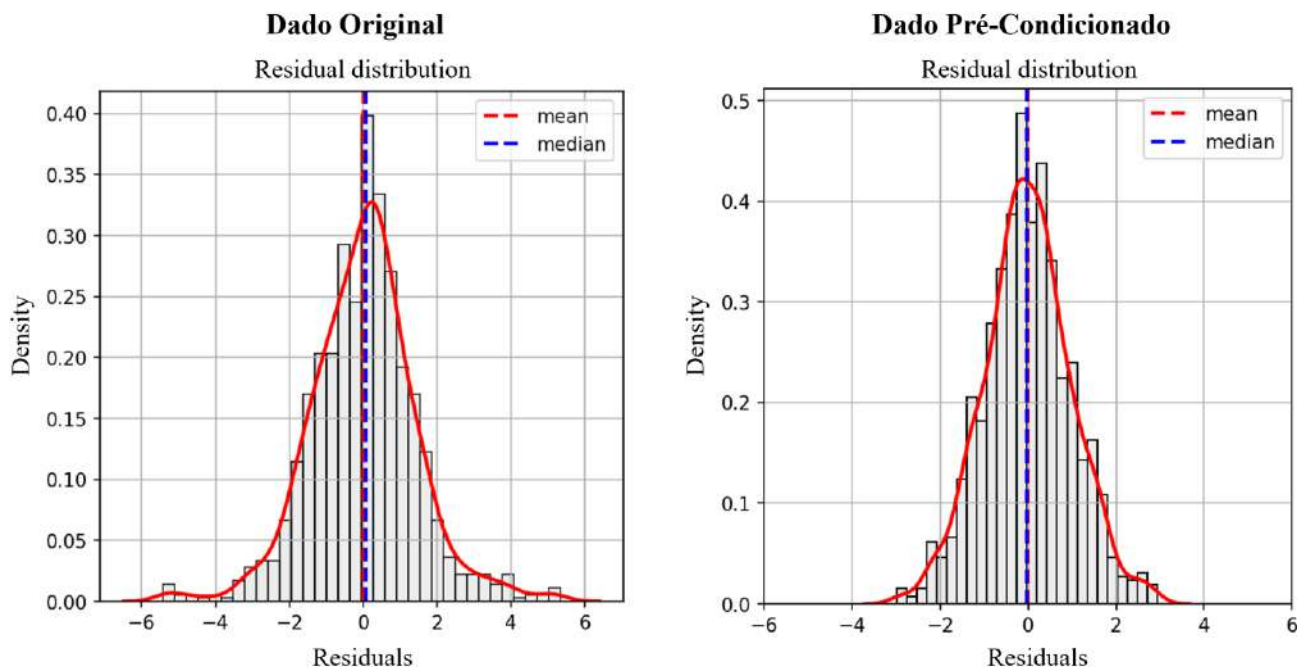


Figura 8.22. Distribuição dos resíduos entre o sismograma sintético e o traço sísmico para o dado original e o dado pré-condicionado.

As métricas estatísticas apresentadas na Tabela 8.2 corroboram essa observação, mostrando uma redução nos valores de erro médio, erro mediano e desvio-padrão. Embora as diferenças absolutas não sejam grandes, o ganho relativo de aproximadamente 30% na amarração com o dado condicionado representa uma melhora significativa na consistência global da calibração sísmica.

Tabela 8.2. Métricas estatísticas da análise residual entre poço e sísmica.

Métricas Estatísticas			
	Erro Absoluto Médio	Erro Absoluto Mediano	Desvio Padrão
Dado Original	1.07	0.85	1.44
Dado Condicionado	0.78	0.60	0.99

Ganho relativo de 30%

8.4.2. Interpretação Sísmica

Com o dado sísmico condicionado e os poços ajustados no domínio do tempo, iniciou-se o processo de interpretação sísmica a partir dos marcadores geológicos. A entrada da Formação Pimenteiras é representada por um pico negativo de alta amplitude, quando ocorre intrusões ígneas na

porção superior muito próximo a sua entrada, caso contrário, é caracterizada por um pico negativo de amplitude moderada. Na interface com a Formação Itaim, observa-se geralmente um pico positivo (preto) de amplitude moderada, resultado do contraste entre os folhelhos da Pimenteiras (baixa impedância) e os arenitos da Itaim (alta impedância). Quando uma soleira de diabásio intrude a base da Pimenteiras, esse padrão se inverte, gerando um pico negativo devido à alta impedância da intrusão ígnea.

Esses saltos de soleiras e inversões de polaridade em diferentes níveis estratigráficos aumentam a complexidade da interpretação sísmica. Assim, o mapeamento do topo e da base da Formação Pimenteiras foi realizado com espaçamento regular a cada quatro *inlines* e *xlines* como mostrado na Figura 8.23. Fica claro que devido a inversão de polaridade apresentada na base da formação, o nível de dificuldade para seu rastreamento é maior, e por isso foi dado prioridade para mapear principalmente na direção da *inline*, ainda que idealmente deva-se mapear nas duas direções.

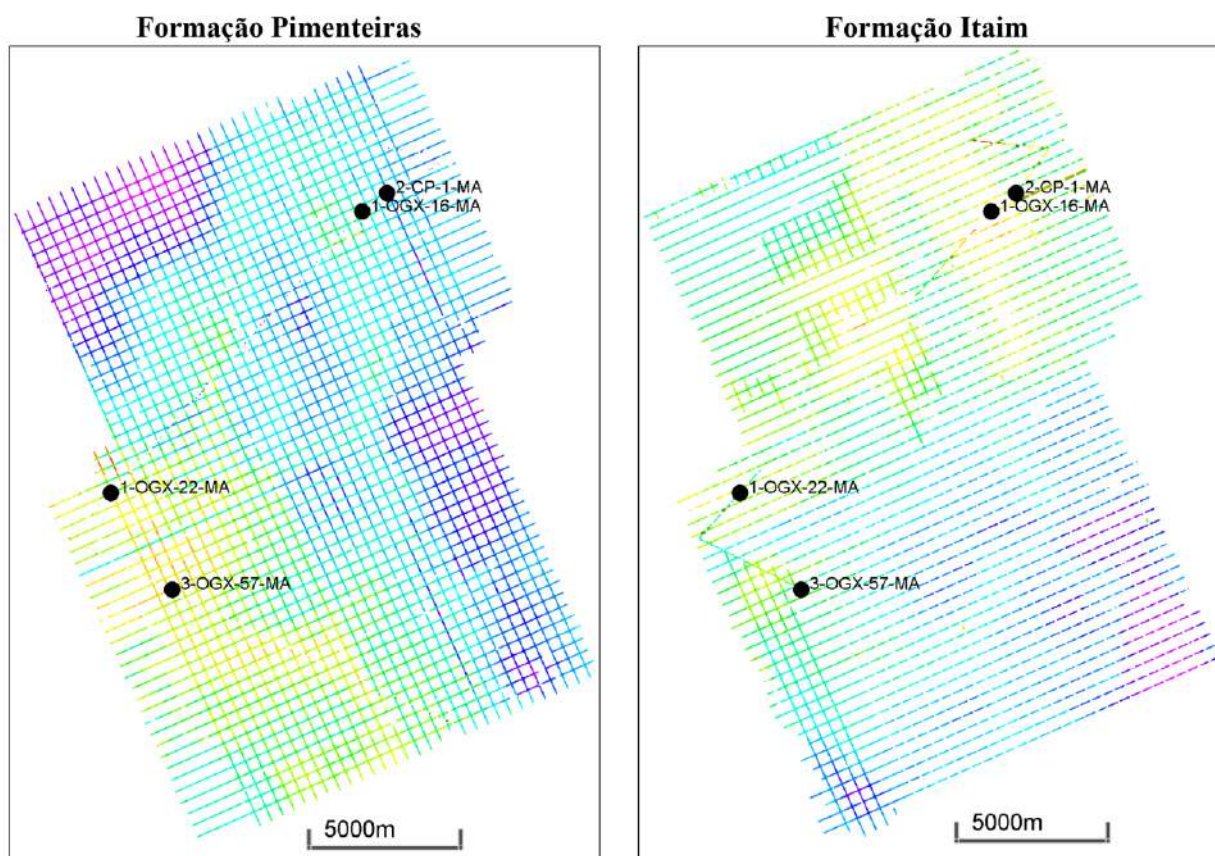


Figura 8.23. Mapeamento sísmico das superfícies de topo e base da Formação Pimenteiras.

As superfícies resultantes (topo e base) foram interpoladas pelo método *Convergent Interpolation*, com posterior suavização e remoção de *spikes*, como mostrado na Figura 8.24. Ambas

evidenciam o efeito estrutural das intrusões ígneas, que geram falhamentos locais e altos estruturais. Embora as falhas não tenham sido mapeadas neste estágio, as superfícies obtidas fornecem um arcabouço confiável para a construção do modelo 3D e para as etapas subsequentes de inversão sísmica e modelagem geoestatística do COT, descritas nos próximos capítulos.

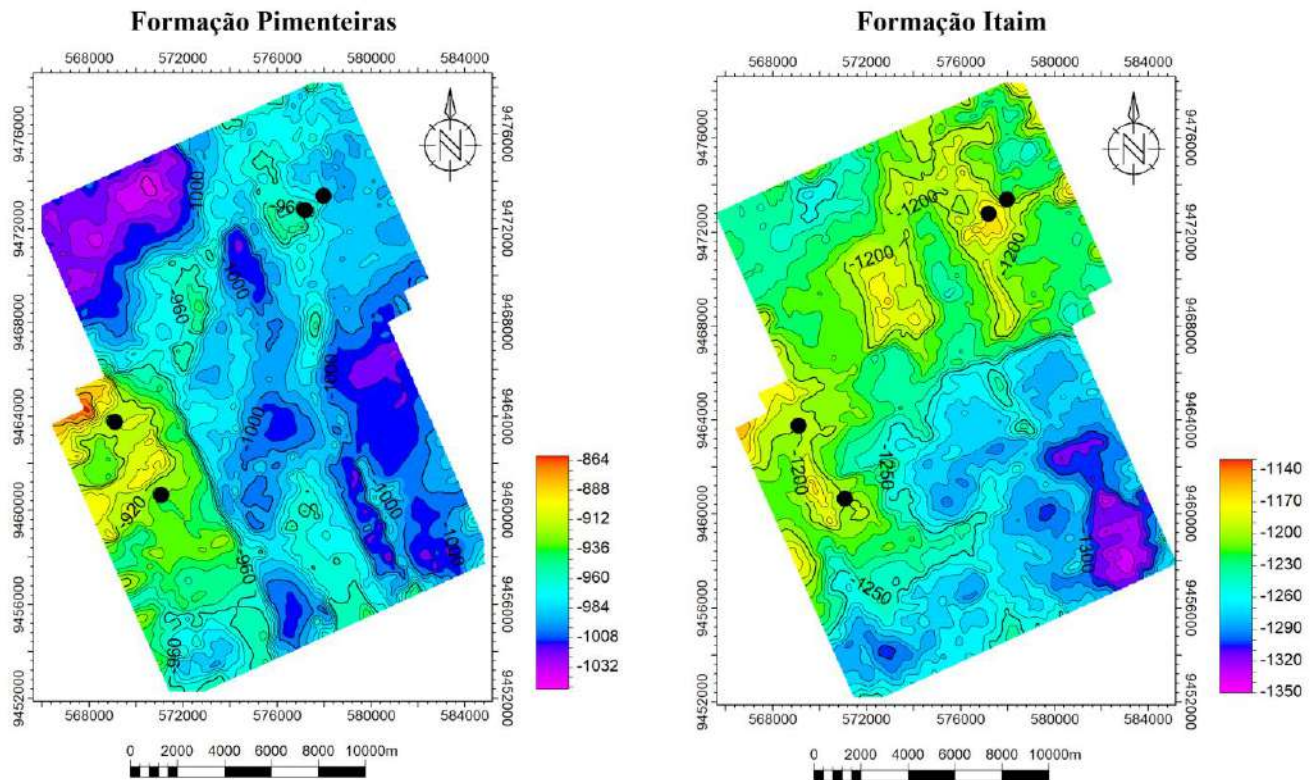


Figura 8.24. Superfícies de topo e base da Formação Pimenteiras, interpoladas e suavizadas, evidenciando o efeito estrutural das intrusões ígneas.

8.4.3. Construção do Grid 3D

As etapas seguintes do estudo (Capítulos 9 e 10) requerem um arcabouço estrutural tridimensional (Grid 3D) como base para a inversão sísmica e modelagem de propriedades. Para isso, foram utilizadas as superfícies interpretadas da Formação Pimenteiras como limites do modelo, seguindo os parâmetros apresentados na Tabela 8.3. O grid foi construído com resolução vertical de 4 ms, equivalente à amostragem temporal do dado sísmico, e com espaçamento médio de 50 m em X e Y. Para ampliar o intervalo modelado e evitar problemas de contorno nas simulações, foram aplicados deslocamentos verticais de 80 ms no topo e na base da Formação Pimenteiras. O grid resultante será utilizado para *upscaling* dos dados de poços (AI, COT e fácies) e para servir de suporte às simulações e inversões subsequentes.

Tabela 8.3. Parametrização adotada para a construção do grid 3D.

Descrição do Grid 3D	
Layering method	Follow horizon
Layering type	Layer thickness from base
Cell thickness	4 ms
Average Xinc	50 m
Average Yinc	50 m
Grid cells (nI x nJ x nK)	375 x 518 x 130

A Figura 8.25 ilustra o modelo final, mostrando a discretização vertical em camadas de 4 ms e o volume correspondente à zona da Formação Pimenteiras, que constitui a principal unidade de interesse para a modelagem geoquímica e sísmica deste trabalho.

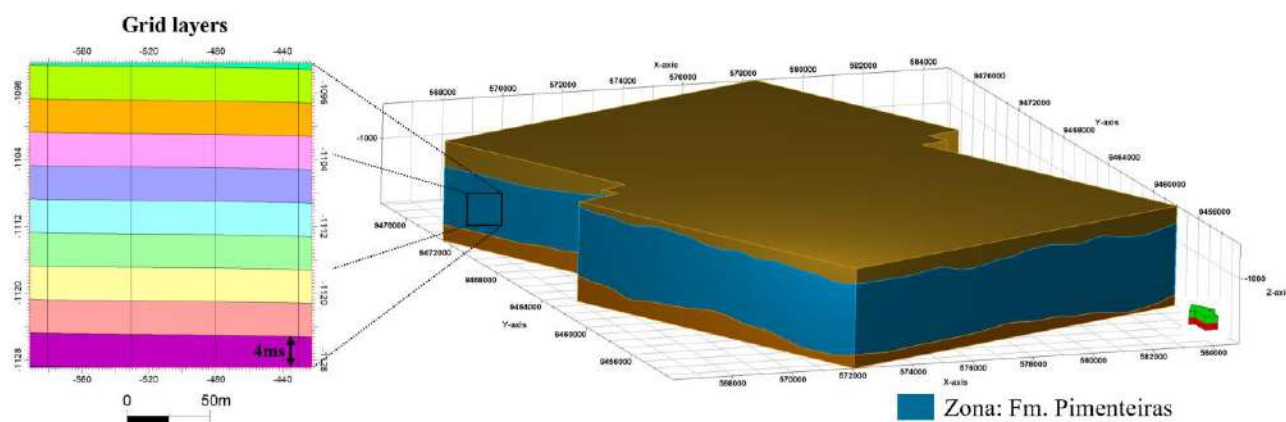


Figura 8.25. Ilustração da geometria do grid gerado para a zona do Pimenteiras bem como a disposição de camadas do grid.

8.5. CONCLUSÃO

O conjunto de procedimentos realizados neste capítulo consolidou uma etapa fundamental na integração entre dados sísmicos e de poço, estabelecendo a base estrutural e estratigráfica necessária para os processos de inversão e modelagem geoestatística subsequentes. A análise quantitativa da amarração sísmica-poço demonstrou que, de modo geral, os resultados apresentaram correlação satisfatória entre o traço sísmico e o sismograma sintético, com coeficientes superiores a 0,6 em todos os poços analisados. Observou-se ainda que o dado condicionado promoveu redução significativa nos resíduos e melhora nas métricas estatísticas, com ganho relativo de aproximadamente 30% em relação ao dado original. Esses resultados evidenciam que o pré-condicionamento contribuiu diretamente para

uma calibração sísmica mais consistente e representativa do comportamento das formações no domínio temporal.

Do ponto de vista geológico, a Formação Pimenteiras apresentou resposta sísmica heterogênea, refletindo a influência das intrusões ígneas que ocorrem em diferentes níveis estratigráficos. Essa característica resultou em inversões de polaridade e variações locais de amplitude, o que promoveu uma complexidade do arcabouço estrutural da área de estudo. O mapeamento do topo e da base da formação, realizado a partir dos poços calibrados e do dado sísmico condicionado, permitiu delinear com maior confiabilidade as superfícies estruturais. A partir dessas interpretações, foi possível construir um grid tridimensional coerente com as superfícies sísmicas, representando de forma realista a geometria da Formação Pimenteiras. O modelo gerado, com resolução vertical de 4 ms e espaçamento médio de 50 m, constitui o arcabouço estrutural que servirá como base para as etapas de inversão sísmica acústica e modelagem de COT, abordadas nos capítulos seguintes.

9. Inversões Sísmicas

9.1. INTRODUÇÃO

Este capítulo é dedicado à apresentação dos métodos de inversão sísmica empregados no presente trabalho, com foco em três abordagens distintas: Inversão Colorida, Inversão Determinística e Inversão Estocástica. A adoção desses métodos tem como objetivo central a obtenção de volumes de impedância acústica que proporcionem o maior nível possível de resolução e detalhamento, condição essencial para a caracterização da rocha geradora e a avaliação de seus níveis de geração.

A Inversão Colorida constitui a primeira etapa, resultando em um volume de impedância acústica relativa. Este resultado desempenha um papel fundamental não apenas no aprimoramento da interpretação de fácies sísmicas na formação estudada, mas também na construção do modelo de baixa frequência, elemento indispensável para a inversão da impedância acústica absoluta.

O segundo método considerado é a Inversão Determinística, implementada neste trabalho a partir do procedimento de Generalização de Tikhonov aplicado traço a traço. Este método fornece um volume de impedância absoluta, condicionado por um modelo de baixa frequência e por um pulso sísmico extraído do dado sísmico, possibilitando a calibração mais rigorosa entre os domínios temporal e a propriedade a ser modelada. Além disso, o volume obtido por este método serve como modelo inicial a ser perturbado e atualizado durante as iterações da inversão estocástica.

Por fim, será apresentada a Inversão Estocástica, a qual se apoia no resultado da inversão determinística. Essa abordagem, de caráter probabilístico, permite gerar múltiplas realizações equiprováveis da impedância acústica, preservando a variabilidade espacial e aumentando a resolução sísmica por meio da integração entre dados sísmicos, informações de poços e modelos estatísticos.

Assim, ao integrar métodos de natureza complementar, rápida e estável (Inversão Colorida), determinística e condicionada (Inversão Determinística), e probabilística e de alta resolução (Inversão Estocástica), a metodologia adotada busca alcançar um equilíbrio entre robustez, detalhamento e confiabilidade na caracterização da Formação Pimenteiras e na avaliação de seu potencial como rocha geradora.

9.2. FUNDAMENTAÇÃO

9.2.1. Inversão Colorida

A técnica de Inversão Colorida (*Coloured Inversion*, CI) foi introduzida por Lancaster e Whitcombe (2000) como uma alternativa rápida e robusta para estimar a Impedância Acústica (AI) a partir de dados sísmicos. Diferentemente dos métodos tradicionais de inversão, que geralmente exigem etapas detalhadas de estimativa da wavelet e calibrações, a abordagem é fundamentada em observações empíricas sobre a consistência da forma espectral dos perfis de AI em uma mesma área. Assim como observados por Walden & Hosken (1985), que demonstraram que as séries de coeficientes de reflexão da Terra apresentam espectros que seguem uma tendência característica, a qual pode ser descrita por uma lei de potência do tipo f^β , em que f corresponde à frequência e β é uma constante positiva. Segundo os autores, essa consistência permite modelar a inversão como uma operação de convolução utilizando um operador espectral fixo, com fase constante de -90° , simplificando substancialmente o processo (Lancaster & Whitcombe, 2000).

Do ponto de vista conceitual, a técnica assume que a relação entre dados sísmicos e AI pode ser representada pela aplicação de um operador convolucional. Esse operador é definido por duas propriedades principais: (i) o espectro de amplitude, que realiza o mapeamento entre o espectro médio sísmico e o espectro médio esperado da AI, “colorindo” os dados sísmicos de modo a reproduzir o conteúdo espectral da impedância; e (ii) a fase constante de -90° , coerente com a interpretação da inversão como um processo análogo à integração de uma série de coeficientes de reflexão (Lancaster & Whitcombe, 2000).

A construção do operador de inversão é baseada na análise espectral de múltiplos perfis de AI. Os autores observaram que, quando representados em escala logarítmica, os espectros de amplitude apresentam uma tendência linear que pode ser descrita por uma função do tipo:

$$A(f) \propto f^\alpha \quad (9.1)$$

onde f é a frequência e α é o expoente negativo que caracteriza o decaimento da amplitude da AI em altas frequências. Estudos de campo, como o realizado no Mar do Norte com 22 poços, indicaram que o valor de α é relativamente constante em uma mesma área, sendo possível derivar um único operador convolucional válido para toda a região (Lancaster & Whitcombe, 2000). Essa estabilidade de α é crucial, pois garante que a filtragem aplicada à sísmica produza resultados consistentes de AI sem a necessidade de estimar explicitamente a wavelet.

Lancaster & Whitcombe (2000) ao elaborarem estes métodos, realizaram comparações com métodos mais sofisticados, como a inversão por *sparse-spike* (irrestrito) e os resultados obtidos pela CI apresentam qualidade semelhante como mostrado na Figura 9.1, porém com significativa redução na complexidade e no tempo de processamento.

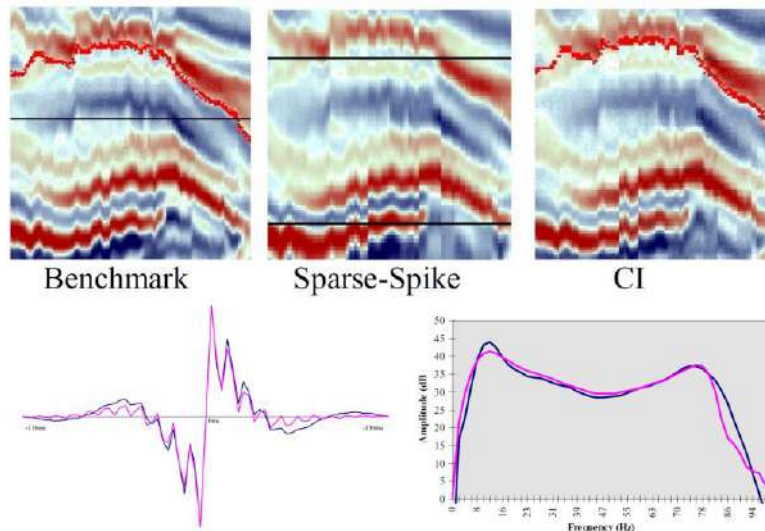


Figura 9.1. Comparação entre o resultado da inversão colorida e o método sparse-spike conforme Lancaster & Whitcombe (2000). Observa-se que a Coloured Inversion produz resultados de qualidade semelhante, com significativa redução da complexidade e do tempo computacional.

Os autores da técnica destacam que a CI deve ser entendida como uma solução de base, uma inversão rápida e estável que fornece uma primeira estimativa confiável da distribuição de AI. Essa inversão inicial pode servir como referência para aplicações posteriores que envolvam métodos mais sofisticados, restrições adicionais ou informações a priori. Ainda assim, a simplicidade, estabilidade e eficiência da técnica explicam sua ampla adoção em ambientes industriais e sua relevância como ferramenta de rotina para análise sísmica (Lancaster & Whitcombe, 2000).

9.2.2. Inversão Determinística

O avanço das pesquisas em geociências impulsionou significativamente o desenvolvimento da teoria dos problemas inversos, bem como de métodos e algoritmos voltados à sua resolução. Esses problemas são, em sua maioria, classificados como mal-postos, apresentando não unicidade, instabilidade e forte sensibilidade ao ruído nos dados. No contexto de dados sísmicos, tais dificuldades tornam-se ainda mais acentuadas, uma vez que se parte de respostas indiretas da Terra para inferir os parâmetros do modelo responsáveis por gerá-las, sob restrições impostas por limitações de largura de banda, resolução finita, ruído, erros de medição e aproximações numéricas inerentes ao problema direto (Tarantola, 2005; Bosch et al., 2010; Tompkins et al., 2011).

Mesmo admitindo um operador direto adequado, os modelos elásticos ótimos obtidos ao final do processo inverso representam apenas uma entre várias soluções possíveis que reproduzem de forma semelhante os dados observados. Desse modo, enquanto um mau ajuste entre dados sísmicos observados e sintéticos indica inequivocamente a inadequação do modelo, um bom ajuste não assegura proximidade com o modelo verdadeiro, já que a solução pode estacionar em mínimos locais afastados do mínimo global (Tarantola, 2005). Essa natureza não única e mal condicionada da inversão sísmica exige a avaliação e propagação contínua das incertezas associadas aos modelos obtidos, independentemente da metodologia empregada (Bosch et al., 2010; Grana e Della Rossa, 2010; Tompkins et al., 2011).

Para lidar com tais desafios, a regularização proposta por Andrei N. Tikhonov, na década de 1960, consolidou-se como uma das abordagens mais importantes na estabilização de soluções em problemas inversos. Desde então, a regularização de Tikhonov tornou-se uma metodologia amplamente aplicada em diferentes ramos da geofísica, bem como em outras áreas das ciências naturais. A aplicação desses conceitos permitiu o avanço dos métodos de inversão baseados em modelos (*model-based inversion*), e demonstraram que a incorporação de informações *a priori* é indispensável para a obtenção de soluções estáveis e geologicamente plausíveis. Ao longo das últimas décadas, a evolução dos métodos determinísticos de inversão, em particular os fundamentados na generalização da regularização de Tikhonov, abriu caminho para aplicações robustas e confiáveis na interpretação sísmica. Nesse contexto, a inversão *model-based* se apresenta como uma ferramenta essencial para a estimativa de propriedades físicas de subsuperfície, representando uma ponte entre teoria matemática, dados observacionais e modelos geológicos.

9.2.2.1. Inversão Linear Generalizada (GLI)

A inversão sísmica do tipo *model-based* pode ser entendida como um procedimento de ajuste iterativo: propõe-se um modelo inicial da Terra, simula-se a resposta sísmica correspondente e, a partir do desajuste entre o sintético e o observado, atualizam-se os parâmetros até alcançar compatibilidade aceitável com os dados (Russell, 1988). Esse arranjo contorna a inversão direta do traço e explicita um ponto conceitual central: o problema é não-único; diferentes combinações de velocidade e profundidade podem reproduzir tempos idênticos, de modo que um bom *fit* não garante veracidade geológica (Russell, 1988).

Na formulação sintetizada por Russell (1988), a GLI parte da representação vetorial de parâmetros e observações:

$M = (m_1, m_2, \dots, m_k)^T$: vetor de k parâmetros do modelo;

$T = (t_1, t_2, \dots, t_n)^T$: vetor de n observações.

A relação entre observações e modelo é escrita de forma funcional:

$$t_i = F(m_1, m_2, \dots, m_k), i = 1, \dots, n.$$

Dado um modelo inicial M_0 , lineariza-se F por expansão de primeira ordem:

$$F(M) = F(M_0) + \frac{\partial F(M_0)}{\partial M} \Delta M \quad (9.1),$$

onde M é o modelo verdadeiro, ΔM é a variação dos parâmetros e $\frac{\partial F(M_0)}{\partial M}$ é a matriz jacobiana (derivadas de F em relação a M , avaliada em M_0). Definindo o resíduo entre observado e calculado

$$\Delta F = F(M) - F(M_0) \quad (9.2)$$

obtem-se a forma matricial

$$\Delta F = A \Delta M \quad (9.3),$$

em que A é a matriz de derivadas com n linhas e k colunas. Embora a solução formal $\Delta M = A^{-1} \Delta F$ seja sugestiva, em geofísica $n > k$ é comum e A não é quadrada (caso superdeterminado). Assim, Russell (1988) demonstra a solução de mínimos quadrados (associada ao método de Marquardt–Levenberg; descrito por Lines & Treitel, 1984):

$$\Delta M = (A^T A)^{-1} A^T \Delta F \quad (9.4)$$

Para ligar o modelo às observações sísmicas, Russell (1988) toma como ponto de partida o modelo convolucional padrão:

$$s(t) = w(t) * r(t) \text{ onde } r(t) = \text{apenas primárias.}$$

Cooke e Schneider (1983), conforme discutido por Russell, ampliam essa formulação para incluir múltiplas e perdas por transmissão no cálculo da resposta sísmica. A vantagem é que esses efeitos são modelados no direto, sem entrar como parâmetros na inversão, mitigando a propagação de múltiplas para o modelo, um problema comum em métodos recursivos quando não há prévia atenuação.

A parametrização por blocos (impedância inicial, gradiente e espessura temporal por bloco) reduz a dimensionalidade, estabiliza a estimativa e mantém flexibilidade geológica (Russell, 1988). Exemplos sintéticos (Cooke; Schneider, 1983) mostram que, mesmo com ruído e múltiplas modeladas, a GLI recupera contrastes de impedância relevantes (e.g., arenitos portadores de gás em folhelhos).

Como a atualização de mínimos quadrados é sensível a ruído e mal-condicionamento, discutimos a seguir o problema mal-posto e regularização, preparando a incorporação explícita de penalização de Tikhonov.

9.2.2.2. Problema mal-posto

Segundo Zhdanov (2002), resolver diretamente um problema inverso mal-posto pode gerar modelos instáveis e pouco realistas. A regularização surge para contornar essa instabilidade, apoiada nos fundamentos de Tikhonov (1999). O problema inverso é formulado como a equação de operadores, a partir de:

$$d = A(m) \quad (9.5),$$

em que m são os parâmetros do modelo, d os dados observados e $A: M \rightarrow DA$ o operador direto agindo entre espaços de Banach (muitas vezes tratados como espaços de Hilbert para dotar o problema de uma geometria). O objetivo não é ajustar o ruído, mas obter um m_{pr} (modelo predito) tal que $d_{pr} = A(m_{pr})$ fique próximo de d dentro da precisão de observação, o que exige uma métrica $\mu_D(\cdot, \cdot)$ para medir o *misfit*. Um problema é bem-posto quando: (i) a solução existe, (ii) é única e (iii) depende continuamente dos dados; se alguma condição falha, é mal-posto. Nessa ótica, sensibilidade e resolução do método ligam-se ao comportamento de A^{-1} : quanto menor a norma de A^{-1} , maior a resolução e mais próximos são os modelos distinguíveis; se A^{-1} não é limitado, a resolução efetiva colapsa e pequenos erros em d dão grandes erros em m . Uma estratégia adicional é restringir o espaço de modelos a um subespaço mais simples/adequado (o conjunto de correção), tornando o problema condicionalmente bem-posto.

A regularização substitui o problema único $d = A(m)$ por uma família de problemas bem-postos

$$d = A_\alpha(m), \quad \alpha > 0 \quad (9.6),$$

cujos inversos A_α^{-1} são contínuos e aproximam A^{-1} quando $\alpha \rightarrow 0$. Define-se, assim, um operador de regularização $R(d, \alpha)$ com soluções $m_\alpha = R(d, \alpha)$ estáveis para dados ruidosos $d_\delta = d_t + \delta d$ e que tendem ao verdadeiro m_t (solução exata para o dado exato $d_t = A(m_t)$) sob hipóteses adequadas, à medida que $\alpha, \delta \rightarrow 0$. A construção de R apoia-se em funcionais estabilizadores $s(m)$ cujos subníveis $M_c = \{m \in M: s(m) < c\}$ são compactos (seleção de soluções plausíveis). Com o princípio da discrepância, define-se o conjunto admissível

$$Q_\delta = \{m \in M: \mu_D(A(m), d_\delta) \leq \delta\}.$$

A formulação paramétrica de Tikhonov será detalhada na seção seguinte, a qual combina o *misfit* e um funcional estabilizador $s(m)$ e fornece um mecanismo prático para escolher m_α dentro de Q_δ .

9.2.2.3. Regularização de Tikhonov

Em problemas inversos lineares, soluções de mínimos quadrados tendem a amplificar ruído e a produzir estimativas instáveis quando as matrizes normais são inexistentes ou mal-condicionadas, mesmo no caso ponderado $(A^T W_d^2 A)^{-1}$ (Zhdanov, 2002). A regularização de Tikhonov (Tikhonov; Arsenin, 1977) estabiliza a estimação ao incorporar informação *a priori* e controlar a complexidade do modelo via o parâmetro α .

A formulação introduz o funcional paramétrico

$$P^\alpha(m, d) = \|W_d A m - W_d d\|^2 + \alpha \|W_m m - W_m m_{apr}\|^2 \quad (9.7),$$

em que W_d e W_m são matrizes de ponderação nos espaços de dados e de modelo (não necessariamente diagonais), m_{apr} é um modelo a priori, $\|\cdot\|$ é a norma Euclidiana e $\alpha > 0$ é o parâmetro de Tikhonov (Zhdanov, 2002). Na prática, muitas aplicações usam $W_m = I$; alternativamente, tomam-se operadores de diferenças finitas (1ª ou 2ª ordem) para favorecer suavidade.

A quasi-solução é definida por

$$m_\alpha = \arg \min_m P^\alpha(m, d) \quad (9.8),$$

com α ajustado, por exemplo, pelo princípio da discrepância:

$$\|W_d(A m_\alpha - d)\| = \delta \quad (9.9),$$

onde δ estima o nível de ruído ponderado (Zhdanov, 2002). Anulando a primeira variação de P^α , obtêm-se as equações normais regularizadas:

$$(A^T W_d^T W_d A + \alpha W_m^T W_m) m_\alpha = A^T W_d^T W_d d + \alpha W_m^T W_m m_{apr} \quad (9.10),$$

e a solução correspondente

$$m_\alpha = (A^T W_d^T W_d A + \alpha W_m^T W_m)^{-1} (A^T W_d^T W_d d + \alpha W_m^T W_m m_{apr}) \quad (9.11)$$

Quando W_d e W_m são simétricas, $W_d^T W_d = W_d^2$ e $W_m^T W_m = W_m^2$, levando à forma compacta:

$$m_\alpha = (A^T W_d^2 A + \alpha W_m^2)^{-1} (A^T W_d^2 d + \alpha W_m^2 m_{apr}) \quad (9.12),$$

interpretada como a solução regularizada do problema de mínimos quadrados generalizados (Zhdanov, 2002). O primeiro termo controla o ajuste aos dados; o segundo estabiliza a resposta

ancorando o modelo em propriedades desejadas. O parâmetro α define o equilíbrio: valores altos privilegiam estabilidade; valores baixos privilegiam *fit* e podem reintroduzir sensibilidade ao ruído.

9.2.3. Inversão Estocástica

A inversão estocástica, também conhecida como inversão geoestatística, teve origem no início da década de 1990 a partir da aplicação da abordagem de simulação condicional à inversão de impedância acústica, com o objetivo de gerar múltiplas realizações tridimensionais condicionadas aos dados sísmicos e aos poços disponíveis (Dubrule, 2003). Essa formulação introduziu um novo paradigma na inversão sísmica: em vez de buscar uma única solução ótima, passou-se a descrever o conjunto de soluções equiprováveis, compatíveis tanto com as observações quanto com as propriedades estatísticas do modelo.

Diferentemente das abordagens determinísticas, que buscam uma única solução “ótima” ao minimizar uma função objetivo global. Em vez disso, as metodologias estocásticas procuram caracterizar a distribuição de possíveis soluções compatíveis com os dados sísmicos e com os modelos geológicos a priori (Simm & Bacon, 2014). Assim, ao invés de um único modelo, são geradas múltiplas realizações da impedância acústica que reproduzem o comportamento estatístico dos dados observados, respeitando tanto as medições nos poços quanto as restrições espaciais definidas pelo modelo geoestatístico.

A média de um grande número de realizações tende a aproximar-se da solução determinística, também chamada de “melhor estimativa”, enquanto o desvio padrão fornece uma medida direta da incerteza e revela o aumento da variabilidade longe do controle dos poços (Simm & Bacon, 2014). Essa capacidade de quantificar a incerteza e preservar a heterogeneidade natural do meio torna a inversão estocástica uma ferramenta essencial para o modelamento geológico e petrofísico (Figura 9.2).

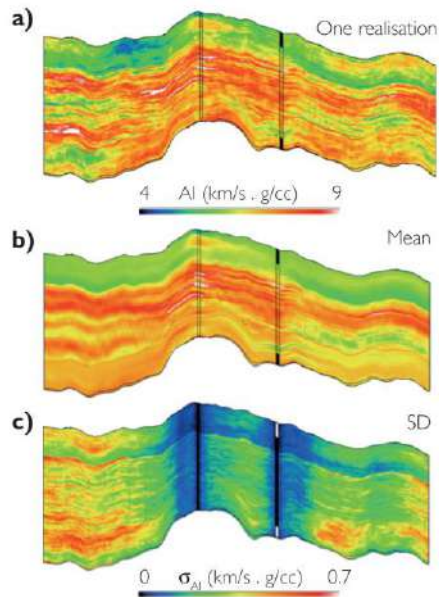


Figura 9.2. Exemplo ilustrativo dos produtos gerados pela inversão geostatística: (a) realização individual, (b) média das realizações e (c) desvio padrão. Observa-se que a média apresenta resposta suavizada em comparação à realização única, enquanto o desvio padrão aumenta nas regiões afastadas do controle dos poços, indicando a consistência da abordagem estocástica na reprodução da variabilidade e na honra aos dados de poço (adaptado de Lamy et al., 1999; Simm & Bacon, 2014).

A base metodológica da inversão estocástica é o algoritmo de Simulação Gaussiana Sequencial (SGS), um método de busca aleatória que integra as propriedades da krigagem e da amostragem de Monte Carlo. O processo consiste em selecionar aleatoriamente um traço sísmico do volume a ser invertido e gerar múltiplas realizações locais de impedância acústica, estimadas por krigagem condicional aos dados de poço. Para cada realização, calculam-se os coeficientes de refletividade correspondentes, que são então convoluídos com a wavelet sísmica. O traço cuja resposta sintética apresenta a melhor correspondência com o traço sísmico observado é incorporado ao modelo, e o procedimento é repetido em toda a área de estudo (Dubrule, 2003; Simm & Bacon, 2014).

A repetição do algoritmo (Figura 9.3) com diferentes sequências de amostragem aleatória gera um conjunto de realizações globais, representando o espaço de soluções possíveis. O número de realizações necessárias para capturar adequadamente a variabilidade estatística do sistema depende do comportamento cumulativo dos parâmetros derivados, sendo comum a geração de centenas de realizações para atingir significância estatística (Dubrule, 2003).

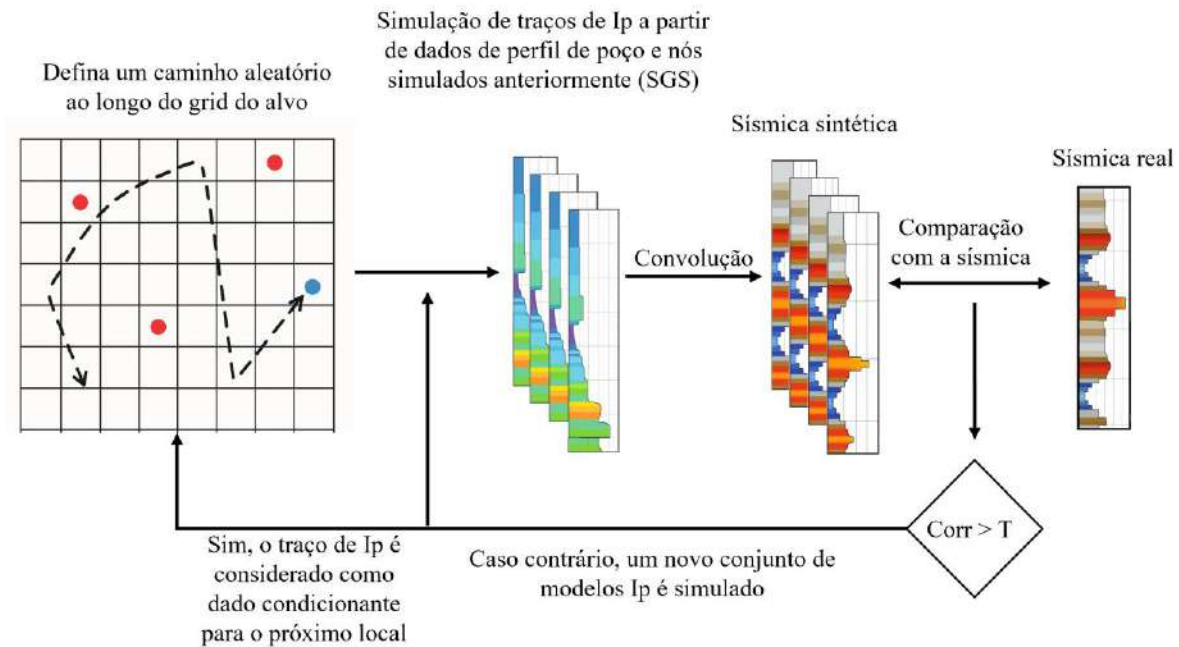


Figura 9.3. Fluxograma do algoritmo de Simulação Gaussiana Sequencial (SGS) conforme Azevedo & Soares (2017). O diagrama ilustra a sequência de amostragem aleatória e a incorporação condicional dos traços simulados.

Os parâmetros de controle, como os variogramas verticais e laterais, o histograma de impedâncias e a wavelet, constituem o modelo geoestatístico a priori, que define a continuidade espacial e a variabilidade vertical dos modelos simulados. Em especial, os variogramas controlam a variância e a continuidade lateral (Figura 9.4), sendo fundamentais para reproduzir padrões geológicos realistas, ainda que sua determinação lateral seja frequentemente limitada pelo número reduzido de poços (Simm & Bacon, 2014).

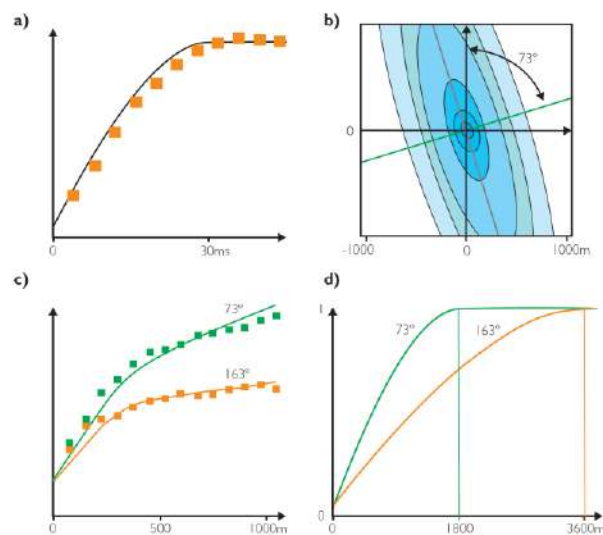


Figura 9.4. Exemplo de variograma anisotrópico utilizado no controle da continuidade lateral e vertical da impedância acústica. (Simm & Bacon, 2014).

As metodologias mais recentes de inversão geoestatística iterativa visam estimar um modelo de subsuperfície m que, após convolução com uma wavelet no modelo direto $F(m)$, produza dados sintéticos com alta correlação em relação à sísmica observada (Figura 9.5). A função objetivo, neste caso, mede a discrepância entre as amplitudes sísmicas reais e as simuladas, sendo iterativamente minimizada por técnicas de simulação estocástica condicional (Azevedo & Soares, 2017).

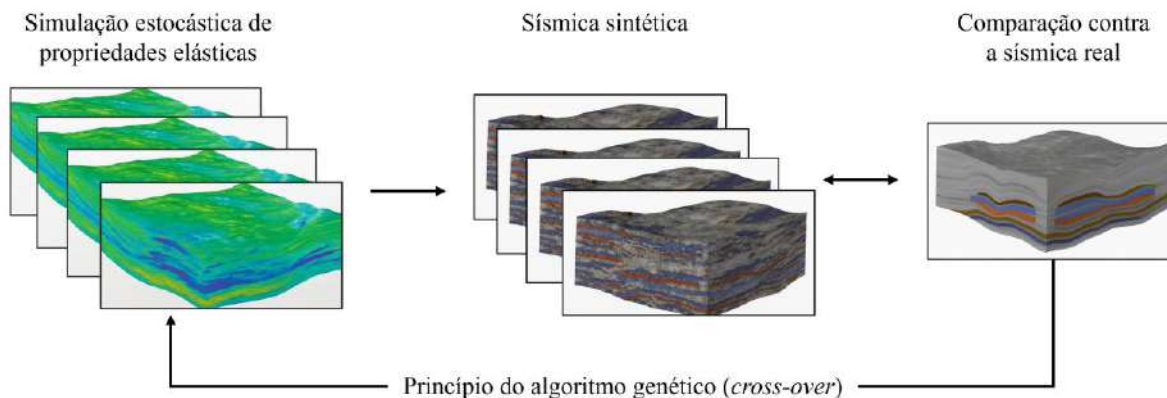


Figura 9.5. Esquema da inversão geoestatística iterativa conforme Azevedo & Soares (2017), destacando o processo de simulação condicional e atualização iterativa do modelo até o atingimento do critério de correlação global

A estrutura geral do procedimento envolve a geração de um conjunto de N_s modelos de impedância condicionados aos poços (via simulação sequencial direta e co-simulação) e a avaliação de cada realização por meio do coeficiente de correlação entre a sísmica sintética e a real. Os modelos que apresentam maior afinidade (melhor correlação) são utilizados como variáveis secundárias para guiar novas simulações nas iterações seguintes (Soares et al., 2007; Caetano, 2009). Esse processo iterativo se repete até que um critério de parada, geralmente, um valor mínimo de discrepância ou um coeficiente de correlação global, seja atingido.

O papel da co-krigagem colocada é fundamental neste processo. O estimador de co-krigagem empregado para gerar o modelo transformado $Z_t(x)$, com base em um conjunto de N_i modelos auxiliares $Z_1(x), Z_2(x), \dots, Z_{N_i}(x)$ é expresso por:

$$Z_1(x_0)^* - m(x_0) = \sum_{\alpha} \lambda_{\alpha}(x_0)[Z_t(x_{\alpha}) - m(x_{\alpha})] + \sum_{i=1}^{N_i} \lambda_i(x_0)[Z_i(x_0) - m(x_0)] \quad (9.13),$$

onde $Z_1(x_0)^*$ é o valor estimado da variável transformada no ponto x_0 , $m(x_0)$ representa a média local, e $\lambda_{\alpha}(x_0)$ e $\lambda_i(x_0)$ são os pesos atribuídos às variáveis primárias e auxiliares, respectivamente.

Considerando que os modelos compartilham os mesmos variogramas $\gamma_i(h)$ e $\gamma_t(h)$, aplica-se a aproximação de Markov (Almeida & Journel, 1994), em que o modelo de correlação entre variáveis é controlado pelos coeficientes $\rho_{t,i}(0)$. Esses coeficientes definem a afinidade entre o modelo transformado $Z_t(x)$ e os modelos auxiliares $Z_i(x)$, garantindo que as novas realizações mantenham coerência estatística e continuidade espacial com os dados sísmicos e geológicos (Azevedo & Soares, 2017). O processo iterativo continua até que o coeficiente de correlação global entre os volumes sísmicos simulados e os dados reais ultrapasse um limiar definido pelo usuário, indicando a convergência da solução.

Uma das principais vantagens da inversão geoestatística é sua capacidade de gerar modelos de alta resolução (Figura 9.6). Enquanto as inversões determinísticas são limitadas pela banda de frequência dos dados sísmicos, a abordagem estocástica incorpora informações de alta frequência provenientes dos poços, reproduzindo a variabilidade vertical e lateral que seria perdida em métodos lineares ou suavizados (Doyen, 2007; Filippova et al., 2011; Azevedo & Soares, 2017).

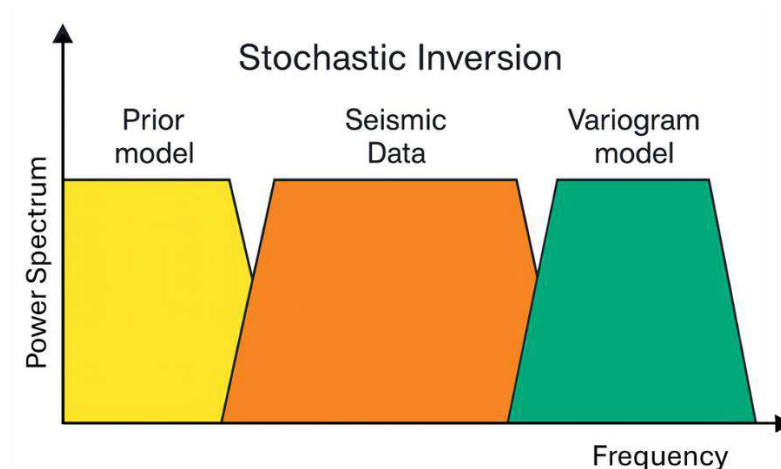


Figura 9.6. Comparação esquemática entre o conteúdo espectral dos modelos de impedância derivados das inversões determinística e estocástica. O aumento do conteúdo de alta frequência temporal na solução estocástica é decorrente da incorporação do modelo de continuidade espacial durante a simulação geoestatística (adaptado de Dubrule, 2003; Azevedo, L., & Soares, A. (2017))

Essa integração entre sísmica e dados de poço, conduzida dentro de uma grid geocelular tridimensional, permite que os modelos inversos resultantes apresentem detalhamento compatível com a escala deposicional, refletindo de forma mais realista a heterogeneidade e a complexidade geológica.

9.2.4. Modelagem Geoestatística

Como neste trabalho foi aplicados alguns métodos da geoestaticas, como por exemplo a Simulação Sequencial Gaussiana na inversão estocástica, a Co-krigagem para a construção do modelo

de baixa para a inversão determinística e por fim a modelagem de variogramas para ambas as abordagens, será descrito nesse tópico do trabalho brevemente a fundamentação desses três conceitos principais aplicadas.

9.2.4.1. Variogramas

O variograma é o elemento central da geoestatística, pois descreve quantitativamente a dependência espacial de uma variável regionalizada. Conforme Oliver e Webster (2015), sua correta construção é essencial para que métodos como o *kriging* produzam estimativas de variância mínima e representem adequadamente a continuidade espacial dos dados.

A estimativa do variograma baseia-se em amostras $z(x_1), z(x_2), \dots$, tratadas como realizações de um processo aleatório. O estimador clássico proposto por Matheron (1965) é dado por:

$$\hat{\gamma}(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} \{z(x_i) - z(x_i + h)\}^2 \quad (9.14)$$

em que h representa a distância (ou defasagem) entre pares de pontos e $m(h)$ é o número de pares comparados. A variação de h gera o variograma experimental, que expressa a semivariância média entre valores separados por diferentes distâncias.

Segundo Azevedo e Soares (2017), a modelagem do variograma consiste em ajustar uma função contínua e positiva definida ao variograma experimental, de modo a representar o comportamento espacial da propriedade em estudo. Essa modelagem sintetiza o grau de continuidade, anisotropias e dispersão dos dados em um conjunto reduzido de parâmetros.

Entre os modelos mais utilizados estão o esférico, exponencial, gaussiano e potência. O modelo esférico, utilizado neste trabalho, é definido por:

$$\gamma(h) = \begin{cases} C \left[1.5 \frac{h}{a} - 0.5 \left(\frac{h}{a} \right)^3 \right], & \text{para } h \leq a \\ C, & \text{para } h > a \end{cases} \quad (9.15)$$

em que C é o patamar (*sill*) e a o alcance (*range*), distância limite da correlação espacial.

Três parâmetros definem o formato do variograma (Figura 9.7): o efeito pepita (*nugget*), associado a variações de pequena escala; o patamar, que representa a variância total do conjunto; e o alcance, que indica o limite da dependência espacial. Semivariâncias pequenas em curtas distâncias refletem forte autocorrelação, enquanto valores crescentes com h indicam perda de similaridade.

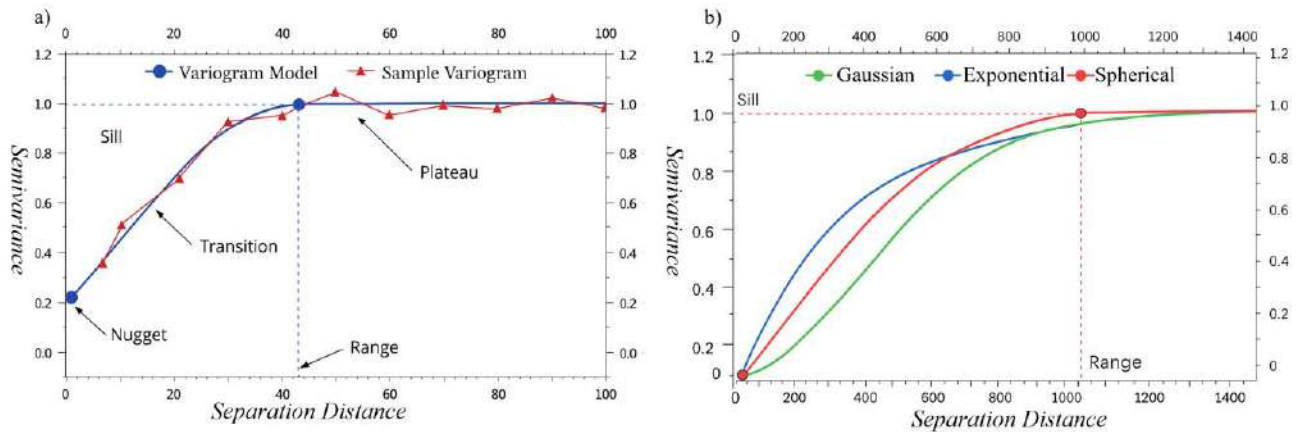


Figura 9.7. Modelos teóricos de variogramas e seus parâmetros principais (nugget, sill e range). Adaptado de Deutsch & Journel (1992).

9.2.4.2. Co-krigagem

A co-krigagem constitui uma extensão natural do método de krigagem tradicional, permitindo incorporar informações de uma variável secundária correlacionada à variável principal de interesse. Segundo Azevedo e Soares (2017), essa técnica é particularmente útil quando as variáveis estão amostradas em grades distintas e a variável principal apresenta amostragem esparsa, enquanto a secundária possui melhor cobertura espacial. Em estudos de reservatórios de hidrocarbonetos, por exemplo, é comum a estimativa conjunta da porosidade (variável primária) e da impedância acústica (variável secundária), aproveitando a correlação física existente entre ambas.

A estimativa da variável primária $Z_1(x_0)$ em uma posição desconhecida x_0 é obtida a partir de uma combinação linear ponderada dos valores amostrados das duas variáveis, conforme a equação:

$$[Z_1(x_0)]_{CK}^* = \sum_{i=1}^{N_1} a_i Z_1(x_i) + \sum_{j=1}^{N_2} b_j Z_2(x_j) \quad (9.16)$$

em que a_i e b_j são os pesos atribuídos às amostras da variável primária Z_1 e da secundária Z_2 , respectivamente. Esses pesos são determinados de modo a garantir que o estimador seja não tendencioso e apresente variância de erro mínima, obedecendo aos mesmos princípios da krigagem ordinária.

9.2.4.3. Simulação Sequencial Gaussiana (SGS)

A Simulação Sequencial Gaussiana (SGS) é um método geoestatístico amplamente utilizado para gerar múltiplas realizações equiprováveis de variáveis contínuas, com o objetivo de representar a incerteza espacial associada aos dados de amostragem (Azevedo & Soares, 2017). Baseia-se na

hipótese de que a variável regionalizada apresenta um comportamento multi-Gaussiano, o que permite aplicar modelos lineares de covariância e variograma.

O primeiro passo consiste em transformar os dados originais $Z(x)$ para o domínio Gaussiano por meio da transformação *normal-score*, resultando em $Y(x) = \varphi[Z(x)]$. Nesse domínio, a distribuição condicional em qualquer ponto x_0 é caracterizada pela média e variância obtidas via krigagem simples. A simulação segue uma trajetória aleatória sobre a malha, na qual, para cada nó, estima-se a média e variância condicionais e se sorteia um valor aleatório da distribuição $N(\mu, \sigma^2)$. Esse valor é incorporado ao conjunto de dados condicionantes, e o processo se repete até completar toda a grade. Os resultados são então reconvertidos ao domínio original pela transformação inversa $z^s(x) = \varphi^{-1}(y_x^s)$.

O método tende a reproduzir o modelo de variograma da variável transformada, garantindo consistência com a estrutura espacial inferida, embora a retrotransformação possa introduzir distorções quando a distribuição original se desvia da normalidade (Azevedo & Soares, 2017). A seleção do raio de busca e do subconjunto de dados condicionantes é crítica: raios muito curtos enfatizam variações locais, enquanto raios excessivos produzem suavização (Gómez-Hernández & Journel, 1993).

A SGS distingue-se da krigagem por gerar um conjunto de soluções possíveis, e não uma estimativa única, permitindo avaliar a incerteza e calcular estatísticas derivadas, como a média (valor esperado) e o desvio padrão (variabilidade local) das realizações.

9.3. MÉTODO

9.3.1. Inversão Colorida

A inversão colorida foi aplicada neste estudo no intervalo correspondente à Formação Pimenteiras, delimitada como Zona 3, conforme apresentado no Capítulo 6. A metodologia foi executada sobre uma janela temporal compreendida entre 750 ms e 1600 ms, abrangendo, portanto, uma extrapolação acima e abaixo do intervalo de maior interesse estratigráfico. A escolha desse intervalo de aplicação teve como principal motivação a necessidade de gerar um volume de impedância acústica relativa abrangente, o qual servirá como dado secundário na interpolação da impedância acústica absoluta entre os poços, para a construção do modelo de baixa frequência utilizado na inversão determinística. Essa abordagem tem como vantagem minimizar efeitos de borda e garantir continuidade vertical adequada na etapa subsequente de inversão determinística e estocástica.

Foram utilizados os cinco poços disponíveis na área de estudo, conforme apresentado na Figura 9.8. O poço 3-OGX-51-DP-MA, destacado em laranja, não atinge a Formação Pimenteiras; entretanto, foi incluído na análise para auxiliar na extrapolação espectral e calibração das porções superiores da janela temporal considerada.

A metodologia empregada seguiu o fluxo de trabalho esquematizado na Figura 9.9, que resume as etapas essenciais da inversão colorida conforme originalmente proposta por Lancaster e Whitcombe (2000). O procedimento pode ser descrito em três fases principais: estimativa do expoente α , modelagem do operador de inversão e aplicação convolucional sobre o dado sísmico.

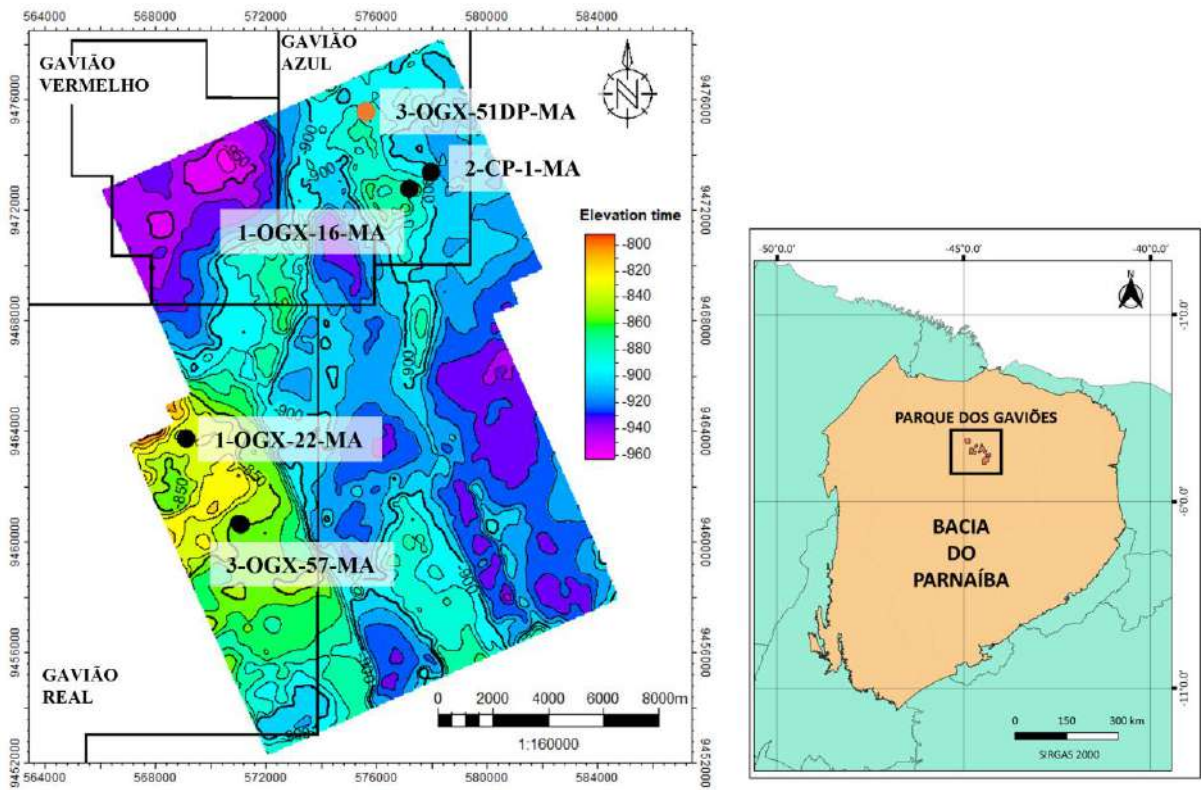


Figura 9.8. Localização e distribuição dos poços utilizados na área de estudo.

Inversão Colorida

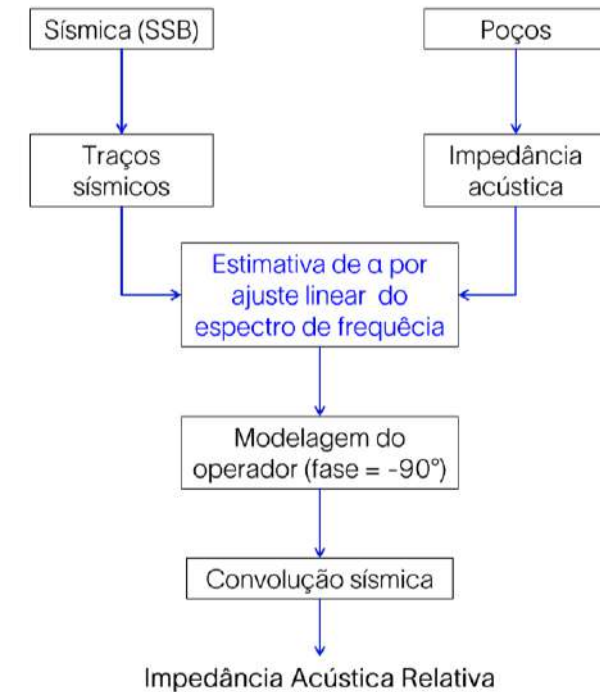


Figura 9.9. Fluxograma metodológico da inversão colorida (Lancaster & Whitcombe, 2000), mostrando as etapas de estimativa do expoente α , modelagem do operador e aplicação convolucional sobre o dado sísmico.

(i) Estimativa do expoente α

A estimativa do expoente α baseia-se na análise do comportamento espectral dos perfis de impedância acústica obtidos nos poços da área. Para cada poço, o espectro de amplitude do perfil de AI é analisado no domínio log-frequência, permitindo o ajuste de uma reta de regressão linear cujo coeficiente angular representa o valor de α . O valor médio resultante, calculado a partir de múltiplos poços, é considerado representativo do comportamento espectral médio da impedância acústica em toda a área de estudo (Figura 9.10). Esse parâmetro é fundamental, pois define a inclinação do espectro alvo a ser reproduzido pela inversão.

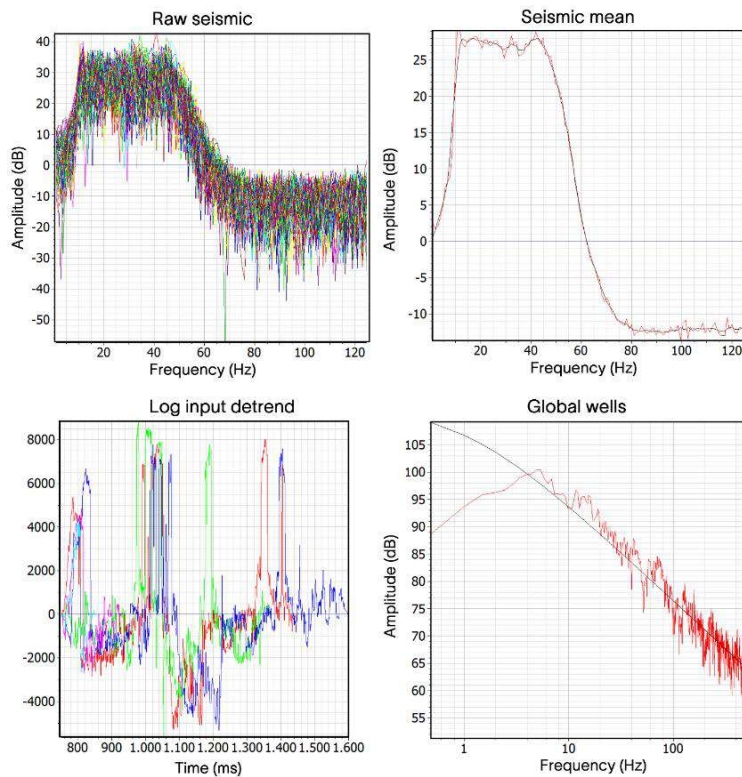


Figura 9.10. Determinação do expoente α a partir da regressão linear dos espectros log-frequência de impedância acústica dos poços, a partir da combinação do espectro médio dos traços e do espectro das curvas de impedância dos poços.

(ii) Construção do Operador de Inversão

Com o valor de α estimado, procede-se à construção do operador de inversão colorida. Esse operador é matematicamente definido para transformar o espectro médio do dado sísmico, que inclui a resposta da wavelet, em um espectro que siga a forma f^α , conforme a relação empírica observada por Walden e Hosken (1985).

A amplitude do operador é ajustada de modo a compensar as diferenças entre o espectro médio sísmico e o espectro alvo derivado do parâmetro α (Figura 9.11). Um aspecto relevante dessa metodologia é que não há necessidade de estimar explicitamente a wavelet, uma vez que o processo se fundamenta na relação estatística entre o espectro sísmico e o espectro de impedância. A fase do operador é assumida constante e igual a -90° , coerente com a interpretação da inversão como um processo de integração.

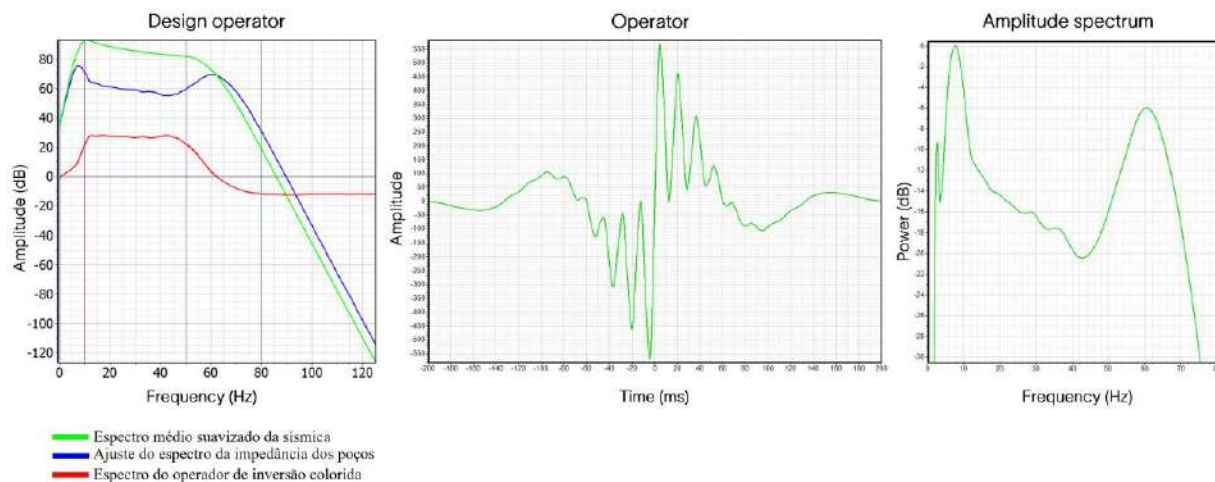


Figura 9.11. Ajustes do espectro de amplitude da impedância do poço e da sísmica a partir de cortes de frequência para modelar o operador que concolvido com o dado sísmico resultará no espectro de amplitude (curva vermelho) mostrado no gráfico *Design operator*.

Os parâmetros adotados para a modelagem do operador convolucional da inversão colorida estão sintetizados na Tabela 9.1, a qual apresenta os valores de corte de frequência inferior (*Low cut*) e superior (*High cut*) definidos a partir da análise conjunta dos espectros médio sísmico e de impedância acústica dos poços. A delimitação dessas faixas de frequência visa assegurar que o operador atue dentro da banda espectral efetiva, preservando a correspondência física entre a resposta sísmica e o comportamento espectral observado nos registros de poço, evitando assim, que aterfatos ou ruídos sejam adicionados.

Tabela 9.1. Parâmetros utilizados na modelagem do operador convolucional da inversão colorida, incluindo limites de frequência inferior (*Low cut*) e superior (*High cut*).

Low cut		High cut	
Full (F2):	10 Hz	Full (F3):	50 Hz
-60dB (F1):	0Hz	-60dB (F4):	80 Hz
Slope:	18 dB/oct.	Slope:	-88 dB/oct.
Attenuation:	-60 dB	Attenuation:	-60 dB

Os ajustes de frequência, responsáveis por adequar o espectro da sísmica ao espectro da impedância dos poços, foram realizados de forma interativa, permitindo o controle qualitativo, mediante a observação em tempo real da impedância acústica relativa no volume sísmico, e quantitativo, por meio da avaliação do coeficiente de correlação entre poço e sísmica. Os resultados desse processo estão apresentados na Tabela 9.2, que exhibe os coeficientes de correlação poço-sísmica individuais e médio, bem como os valores do parâmetro α estimados para cada poço. O valor médio

de α foi adotado como referência para a definição do operador da inversão colorida, representado na Figura 9.10, garantindo a consistência espectral do operador e sua representatividade em relação ao comportamento médio da impedância acústica na área de estudo.

Tabela 9.2. Valores dos coeficientes de correlação poço-sísmica e dos expoentes α estimados para cada poço. O valor médio de α foi adotado para representar o comportamento espectral da impedância acústica na área de estudo.

Well Name	Well Log	Well to Seismic Corr. Coef.	Alfa
3-OGX-57-MA	AI	0.74	-0.979
3-OGX-51DP-MA	AI	0.70	-0.909
1-OGX-16-M	AI	0.60	-0.871
2-CP-1-MA	AI	0.60	-0.890
1-OGX-22-MA	AI	0.45	-0.793
Average Correlation Coefficient	0.62	Average Alfa	-0.906

9.3.2. Construção do modelo de baixa frequência

Devido à distribuição espacial restrita e ao número limitado de poços disponíveis na área de estudo (Figura 9.8), a construção de um volume de impedância acústica exclusivamente por interpolação direta dos dados de poço não se mostrou adequada. Os métodos de interpolação mais comumente empregados, como a krigagem e o método dos vizinhos mais próximos (closest), tendem, nessas condições, a produzir distribuições concêntricas e altamente localizadas em torno dos poços, além de interpolações geologicamente inconsistentes, como ilustrado na Figura 9.12.

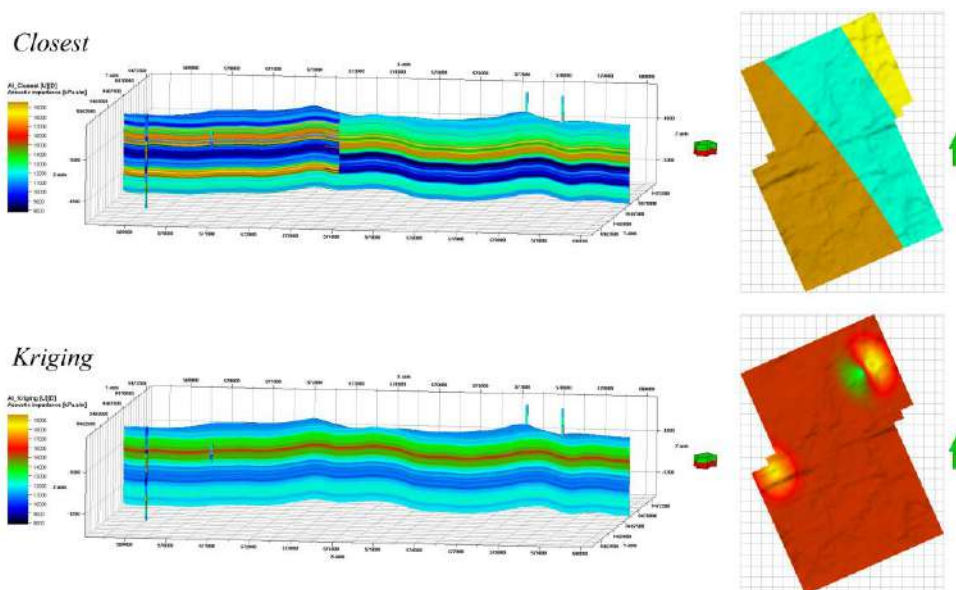


Figura 9.12. Exemplo ilustrativo das limitações da interpolação direta por krigagem e vizinhos mais próximos (closest), mostrando padrões concêntricos e inconsistentes em torno dos poços.

Além disso, o intervalo estratigráfico de interesse apresenta múltiplas intrusões ígneas que cortam a sequência sedimentar em diferentes níveis, resultando em fortes contrastes de impedância acústica entre as rochas encaixantes e as soleiras basálticas. Tal heterogeneidade litológica dificulta a correta representação da variação vertical e lateral de impedância, tornando necessária a incorporação de uma variável secundária capaz de guiar a distribuição dos valores de impedância no grid 3D.

Dessa forma, a construção do modelo de baixa frequência, etapa essencial para o condicionamento inicial da inversão determinística model-based, seguiu o fluxo de trabalho apresentado na Figura 9.13. O processo incluiu as seguintes etapas principais:

1. Upscaling dos perfis de impedância acústica dos poços para o grid 3D, assegurando compatibilidade espacial e resolução adequada à amostragem sísmica;
2. Modelagem do variograma a partir dos dados de poço, definindo os parâmetros de continuidade vertical e horizontal da propriedade;
3. Co-krigagem utilizando como variável secundária o volume de impedância acústica relativa, derivado do resultado da inversão colorida;
4. Conversão do modelo interpolado em volume sísmico; e
5. Aplicação de um filtro passa-baixa com frequência de corte em 6 Hz, de modo a reter apenas as componentes de baixa frequência, coerentes com o objetivo de representar as tendências estruturais e estratigráficas de longo comprimento de onda do modelo.

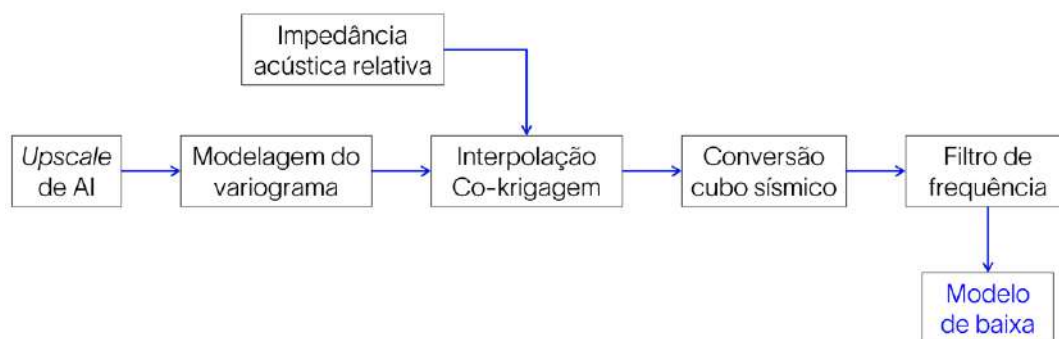


Figura 9.13. Fluxograma da construção do modelo de baixa frequência, destacando as etapas de upscaling, modelagem de variograma, co-krigagem e filtragem passa-baixa.

Os parâmetros utilizados na modelagem do variograma empregado na etapa de co-krigagem estão ilustrados na Figura 9.14. O peso atribuído à variável secundária (impedância acústica relativa) foi de 0,5, valor definido com o intuito de equilibrar a influência entre os dados de poço dado pelo variograma e o volume sísmico (controle regional e melhor amostrado).

Variograma experimental

Type	Spherical
Sill	1,6
Major range	4000
Minor range	3000
Vertical range	45

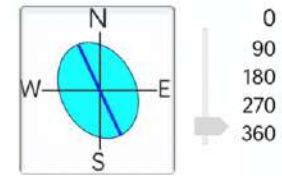
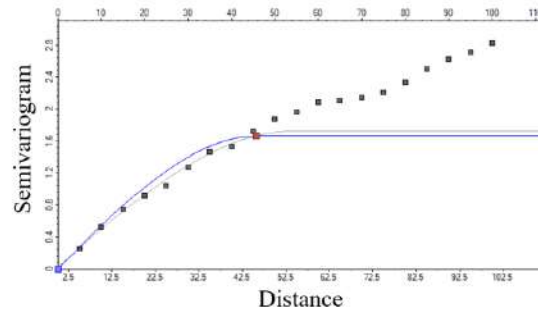


Figura 9.14. Variogramas verticais e laterais utilizados na co-krigagem para o modelo de baixa frequência.

9.3.3. Inversão Determinística

A aplicação da inversão sísmica determinística constitui uma das etapas centrais do presente estudo, sendo a responsável pela conversão quantitativa dos dados sísmicos em impedância acústica absoluta. Após o conjunto de fluxos de trabalho apresentados nos capítulos anteriores, em especial o Capítulo 6, dedicado ao pré-condicionamento dos dados sísmicos, e o Capítulo 8, que abordou a amarração sísmica-poço, a estimativa de wavelets e a interpretação sísmica, estabeleceu-se o conjunto de dados fundamentais necessários à implementação da inversão determinística.

O procedimento metodológico aqui empregado baseia-se na inversão *model-based*, no qual a inversão é guiada por um modelo inicial (ou modelo de baixa frequência) que representa as tendências geológicas de larga escala e fornece o componente de baixa frequência ausente nos dados sísmicos. A construção deste modelo foi detalhado neste capítulo, na seção extamente anterior. A Figura 9.15 ilustra esquematicamente o fluxo de trabalho adotado, destacando as etapas de construção do modelo inicial, aplicação da regularização e geração do cubo de impedância acústica.

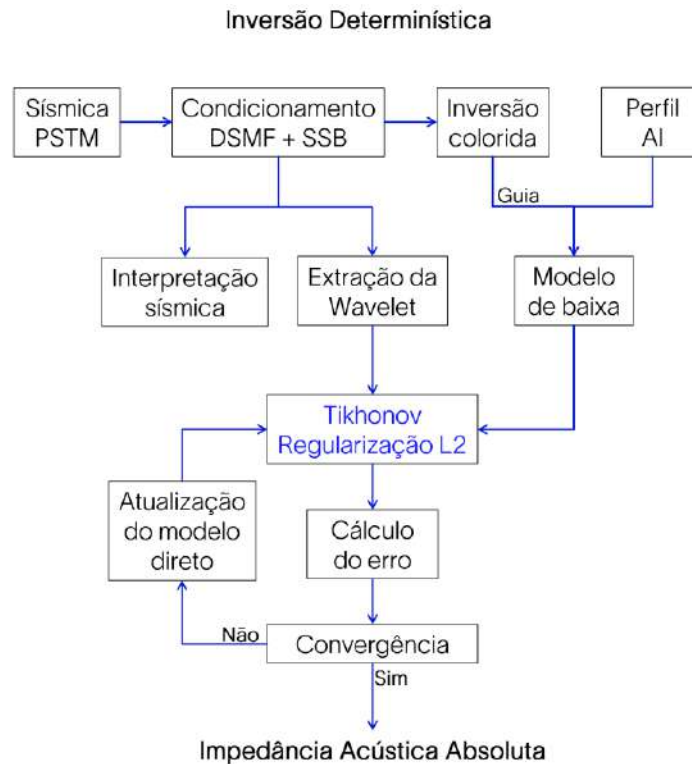


Figura 9.15. Esquema geral do fluxo de trabalho da inversão determinística model-based por generalização de Tikhonov, indicando o uso do modelo de baixa frequência e da wavelet média.

Para o presente estudo, foi gerada uma wavelet média representativa obtida a partir da média aritmética de quatro wavelets extraídas do volume sísmico condicionado, nas posições dos poços de interesse. O método de extração empregado foi o estatístico, considerando uma janela temporal definida e parâmetros de comprimento e amostragem especificados na Tabela 9.3.

Tabela 9.3. Parâmetros de extração das wavelets sísmicas e cálculo da wavelet média utilizada como operador de convolução na inversão determinística.

Poço	Intervalo (ms)	Comprimento (ms)	Amostragem (ms)
3-OGX-57-MA	-780/-937	80	
1-OGX-16-M	-780/-1415	128	4
2-CP-1-MA	-780/-1415	128	
1-OGX-22-MA	-780/-1377	128	

A Figura 9.16 apresenta as wavelets individuais obtidas para cada poço e a wavelet média resultante, que foi utilizada como operador de convolução no processo de inversão.

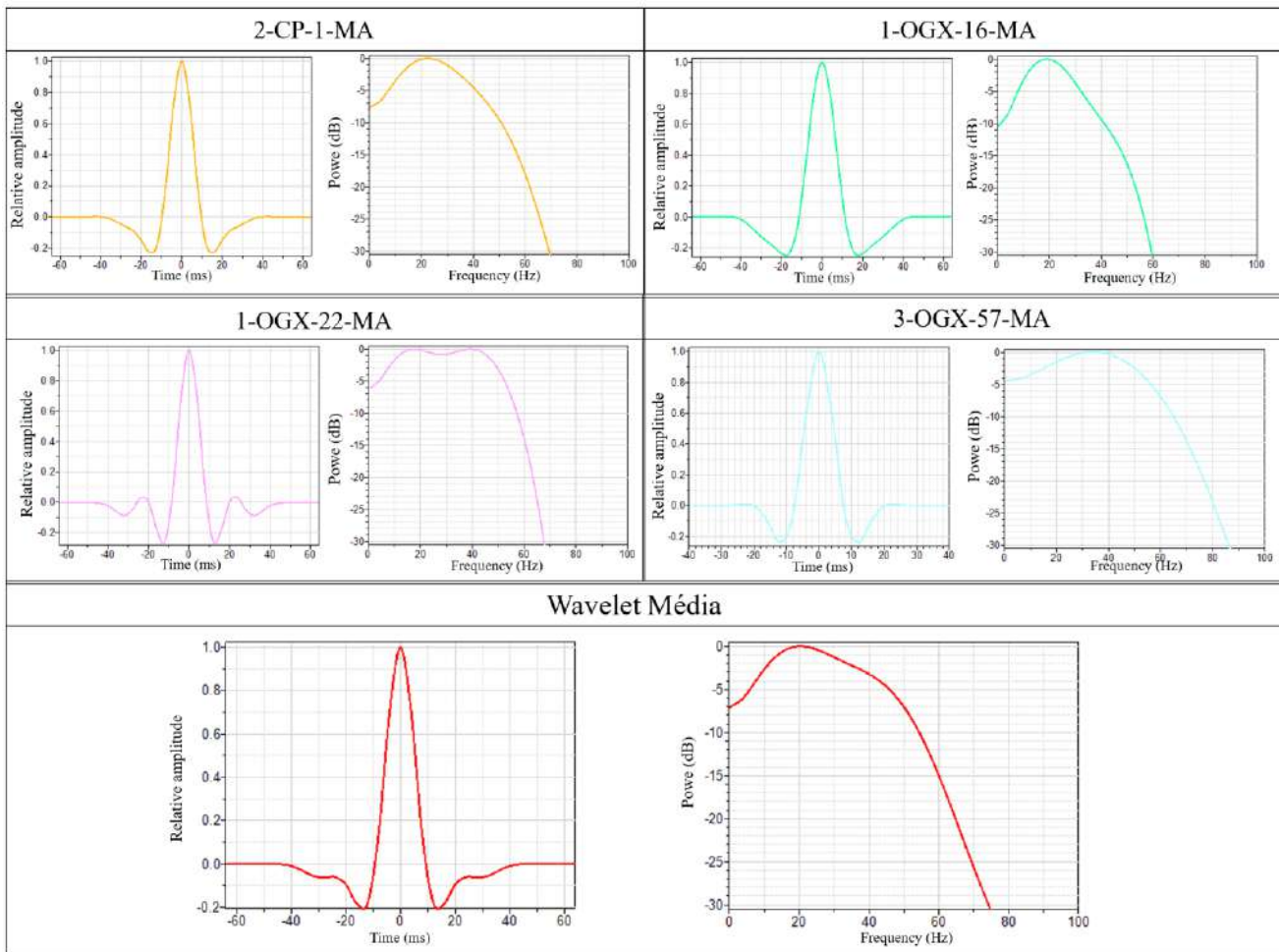


Figura 9.16. Wavelets extraídas individualmente em cada poço e wavelet média resultante empregada no processo de inversão determinística.

A implementação computacional, conduzida pelo grupo de pesquisa GIECAR/UFF, da inversão determinística model-based por Generalização de Tikhonov foi realizada empregando a biblioteca PyLops (*Python Linear Operators*), um framework *open-source* desenvolvido para a manipulação eficiente de operadores lineares em problemas inversos e de regularização. Essa biblioteca é amplamente utilizada em geofísica, pois permite representar os operadores como transformações lineares implícitas, evitando a construção explícita de grandes matrizes densas que tornam o problema computacionalmente custoso.

No contexto deste trabalho, o problema inverso determinístico é formulado segundo o funcional de Tikhonov generalizado:

$$J_{GT} = \min_m \|Gm - d\|_2^2 + \alpha^2 \|Lm\|_2^2 \quad (9.17),$$

em que G representa o operador direto de modelagem (convolução da wavelet com o modelo), d é o vetor de dados sísmicos observados, L é o operador de estabilização responsável por impor

suavidade ao modelo, e α é o parâmetro de regularização que controla o compromisso entre a fidelidade aos dados e a estabilidade da solução.

A função objetivo J_{GT} é resolvida de forma *trace-by-trace*, ou seja, cada traço sísmico é invertido independentemente, mantendo a mesma wavelet média e o modelo inicial de baixa frequência. Essa estratégia reduz o custo computacional e possibilita a calibração local do parâmetro α , ajustando o grau de suavização conforme a relação sinal-ruído de cada traço.

No ambiente Python, o método foi implementado utilizando a biblioteca *PyLops*, que oferece uma coleção de operadores lineares prontos, como convoluções e derivadas, aplicáveis diretamente sobre os vetores de dados.

O operador de convolução G é implementado pela classe *pylops.signalprocessing.Convolve1D*, enquanto o operador de regularização L é construído a partir de derivadas finitas, por meio das classes *pylops.FirstDerivative* ou *pylops.SecondDerivative*.

O uso da biblioteca *PyLops* apresenta diversas vantagens metodológicas e computacionais:

1. Eficiência de memória: evita a montagem de matrizes densas de grande porte, operando apenas sobre vetores.
2. Estabilidade numérica: os *solvers* implementados são otimizados para sistemas mal-postos e operadores não quadrados.
3. Flexibilidade: permite combinar diferentes tipos de regularização (0ª, 1ª ou 2ª ordem) e operadores personalizados.
4. Escalabilidade: facilmente adaptável para inversões 2D e 3D por empilhamento de operadores.
5. Integração direta com *NumPy/SciPy*: compatível com estruturas de dados e métodos padrão da linguagem, permitindo rápida prototipagem e experimentação.

O parâmetros de regularização α utilizado na inversão foi de 0.1, para priorizar o ajuste aos dados e evitar super-suavização, preservando contrastes no volume invertido.

9.3.4. Inversão Estocástica

Conforme discutido na seção de fundamentação teórica, este trabalho empregou a inversão estocástica, cuja base conceitual está fundamentada no método da Simulação Sequencial Gaussiana. Essa abordagem permite integrar de forma consistente diferentes tipos de informações, incluindo a wavelet extraída do dado sísmico, os perfis de impedância acústica medidos nos poços, os variogramas

vertical e horizontal e um modelo inicial de impedância. Diferentemente da inversão determinística, o modelo inicial na abordagem estocástica não precisa ser necessariamente de baixa frequência, uma vez que o objetivo principal é gerar múltiplas realizações equiprováveis que representem a variabilidade espacial do modelo de subsuperfície. O fluxograma ilustrado na Figura 9.17 sintetiza as principais etapas do processo, desde a geração das simulações até o cálculo do volume médio e a análise dos cenários equiprováveis.

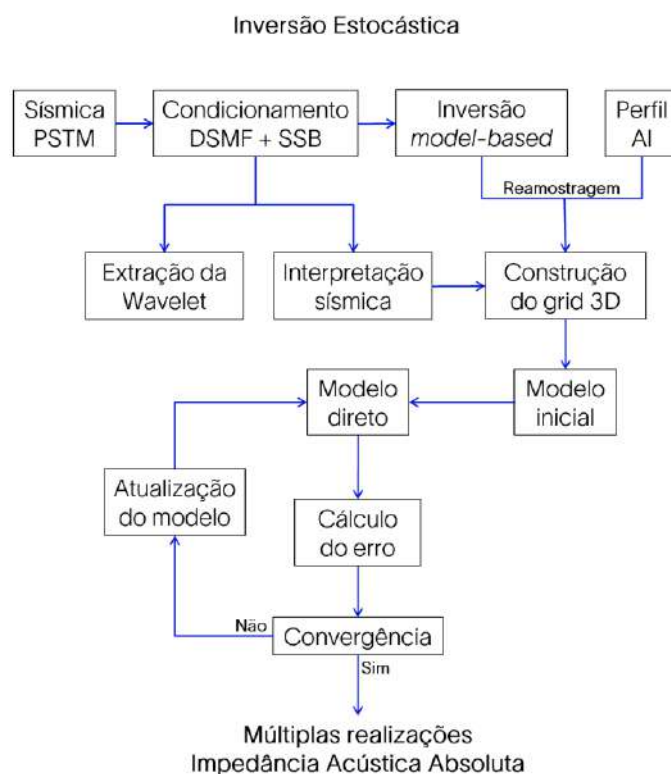


Figura 9.17. Fluxograma geral da inversão estocástica.

A inversão estocástica foi executada sobre um grid tridimensional (3D), conforme detalhado no Capítulo 8. O volume modelado abrangeu, como limites, o topo da Formação Pimenteiras e o topo da Formação Itaim, deslocados em 80 ms para cima e para baixo, respectivamente, em relação às superfícies interpretadas. A resolução vertical do grid foi configurada para 4 ms, equivalente à amostragem temporal dos dados sísmicos.

O passo seguinte consistiu na transferência e preparação dos dados necessários para o grid, etapa que envolveu a reamostragem dos perfis de impedância acústica dos poços e a realização de um controle de qualidade, para verificar se o método aritmético utilizado na reamostragem manteve a tendência geral do dado. Essa verificação foi conduzida principalmente por meio da avaliação do

histograma entre os valores medidos e reamostrados (Figura 9.18), assegurando a consistência estatística dos dados utilizados como entrada.

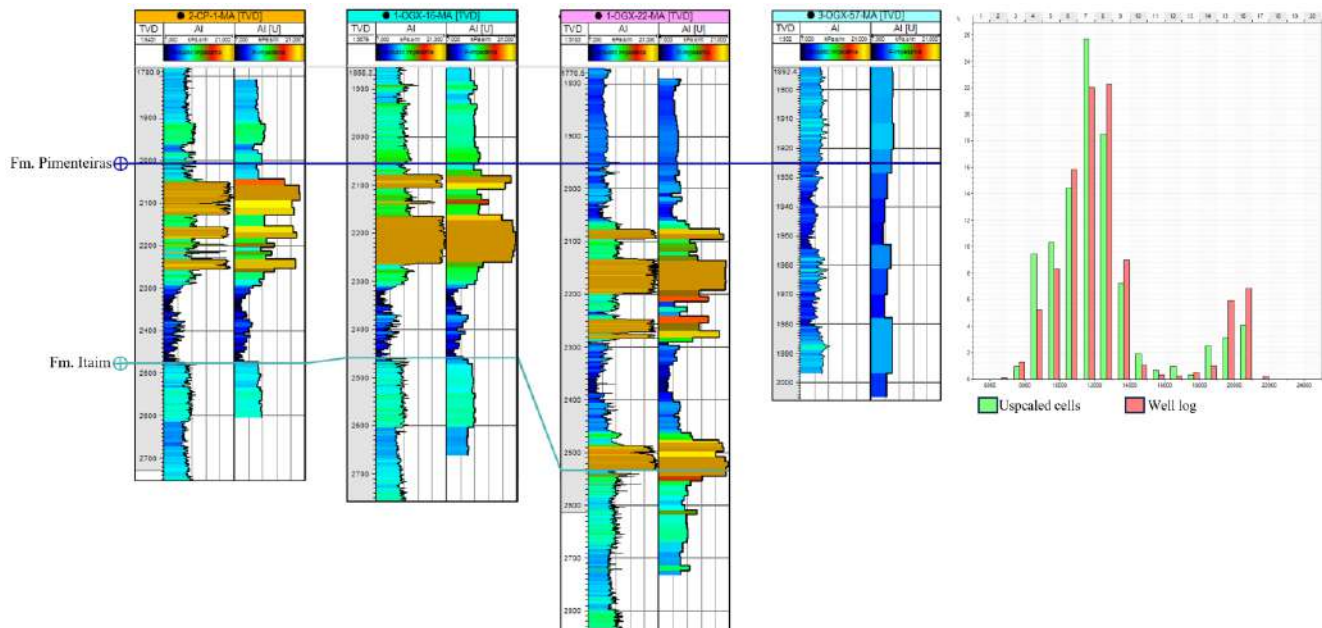


Figura 9.18. Controle de qualidade após reamostragem dos perfis de AI dos poços para o grid 3D, e comparação entre os histogramas dos valores medidos e reamostrados para validação do método escolhido.

O variograma vertical foi obtido a partir da análise dos perfis de impedância acústica dos poços. O modelo experimental do variograma foi ajustado aos dados conforme ilustrado na Figura 9.19, adotando-se um modelo esférico com alcance vertical (range) de aproximadamente 25 m.

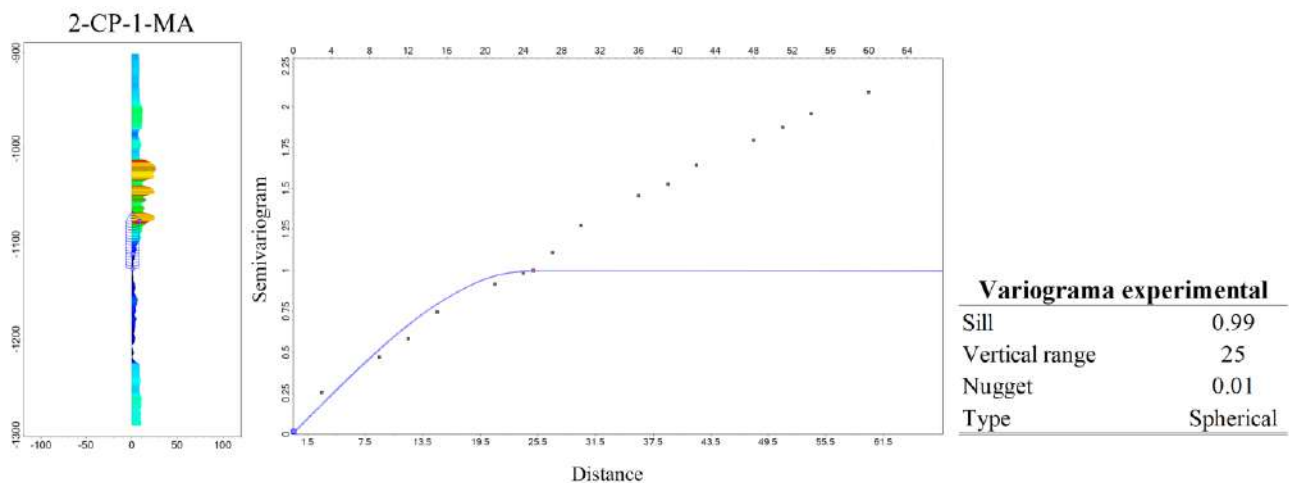


Figura 9.19. Modelo experimental do variograma vertical ajustado aos dados de impedância acústica dos poços, adotando função esférica com alcance de aproximadamente 25 m.

Já o variograma horizontal foi derivado de uma camada representativa do volume de impedância resultante da inversão determinística. Essa escolha justifica-se pela escassa cobertura de

poços na área de estudo, que inviabiliza a determinação robusta do variograma horizontal apenas a partir dos dados de poço. As medições nas direções de maior e menor eixo do variograma, teve como base os eixos com variância menor que 0.5, as quais são apresentadas na Figura 9.20, bem como a camada utilizada para sua extração e os parâmetros de busca para o cálculo do variograma.

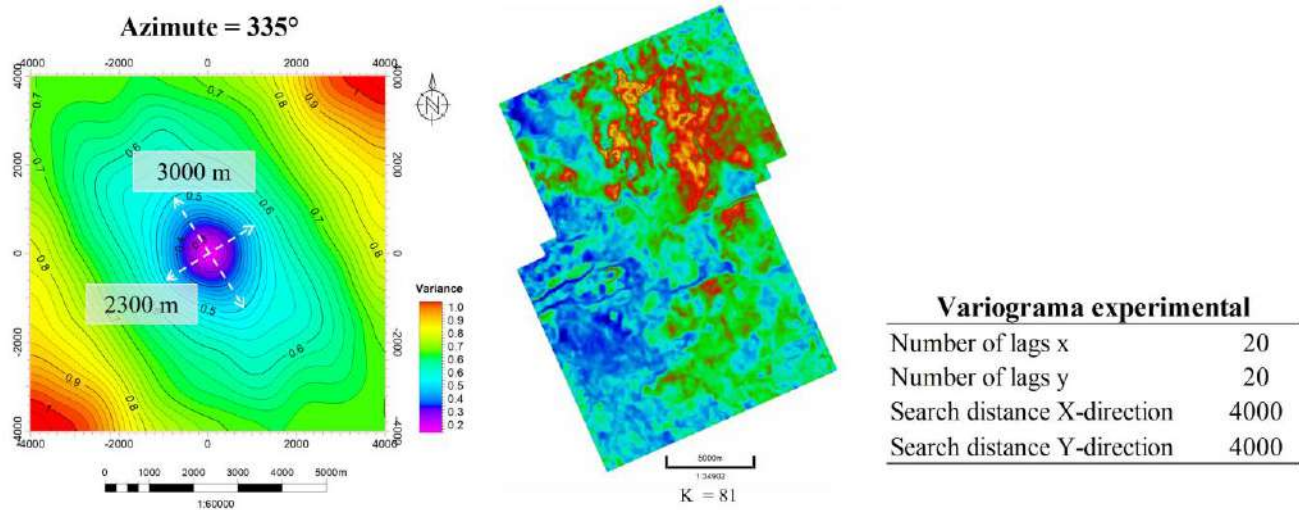


Figura 9.20. Variogramas horizontais derivados do volume de impedância determinística e parâmetros de busca adotados para o cálculo da continuidade lateral.

Neste estudo, o resultado da inversão determinística foi empregado como modelo inicial, mais especificamente, como dado secundário na co-krigagem. O volume foi reamostrado para o mesmo grid utilizado na inversão estocástica, o qual foi imposto uma alta correlação ($r = 0,85$) para que resultado da inversão determinística exercesse forte influência na distribuição de impedância e interpolação entre os poços, mantendo coerência geológica e continuidade lateral (Figura 9.21). O poço 1-OGX-16-MA, por sua vez, foi excluído da interpolação e utilizado como poço de validação para o controle de qualidade da inversão estocástica.

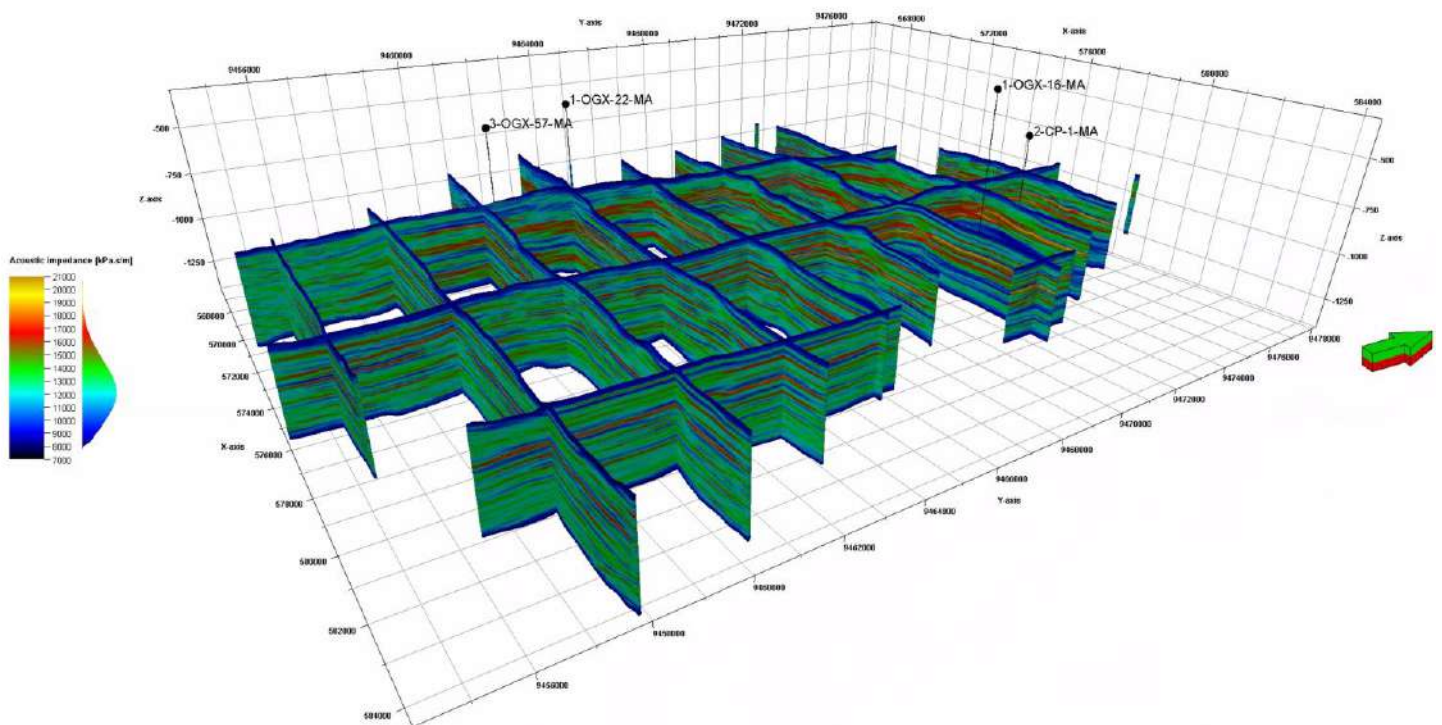


Figura 9.21. Modelo inicial derivado da inversão determinística e os poços, evidenciando coerência lateral e continuidade geológica.

De modo geral, a metodologia de inversão estocástica consiste em um processo iterativo de otimização, no qual um conjunto de realizações de impedância acústica é gerado por meio do SGS e ajustado progressivamente até que a função objetivo atinja o critério de convergência definido. Neste trabalho, adotou-se um valor de 0,8 para o parâmetro de convergência e 200 iterações como limite máximo do processo.

Cada modelo simulado é então convolvido com a wavelet sísmica estimada estatisticamente, a mesma utilizada na inversão determinística (Figura 9.16), resultando em traços sísmicos sintéticos que são comparados com os traços sísmicos observados. O processo é repetido iterativamente, atualizando as simulações até que o melhor ajuste entre os traços sintéticos e observados seja alcançado. O resultado final é um conjunto de realizações equiprováveis de impedância acústica, que reflete a incerteza inerente ao processo de inversão e fornece uma descrição probabilística mais completa da propriedade elásticas da Formação Pimenteiras.

Após a geração de 50 realizações estocásticas, foram selecionados para análise apenas os resultados que apresentaram coeficiente de correlação superior a 0,70 entre a impedância acústica invertida e a impedância medida no poço teste. A partir dessa seleção, realizou-se a análise do volume médio e do desvio-padrão das simulações, com o objetivo de quantificar a variabilidade e a consistência estatística dos modelos obtidos.

O objetivo final desta etapa foi comparar os resultados da inversão determinística com o volume médio da inversão estocástica e com o modelo individual que apresentou a maior correlação com o poço teste. Essa comparação visa identificar qual abordagem proporciona maior nível de detalhamento e fidelidade, sobretudo em relação à resolução vertical do intervalo de interesse, permitindo uma interpretação mais precisa e geologicamente consistente das variações de impedância na Formação Pimenteiras.

9.4. RESULTADOS

Após a fundamentação teórica e o detalhamento metodológico das técnicas de inversão aplicadas, Inversão Colorida, Inversão Determinística e Inversão Estocástica, esta seção apresenta os resultados obtidos a partir da aplicação desses métodos ao conjunto de dados sísmicos e de poço analisados.

O objetivo principal desta etapa é avaliar o desempenho e a consistência das diferentes abordagens de inversão na estimativa da impedância acústica, considerando tanto a resolução vertical quanto a correlação com aos dados observados. Enquanto a Inversão Colorida permite recuperar a impedância acústica relativa de maneira rápida, servindo também como guia na construção do modelo de baixa, as inversões Determinística e Estocástica possibilitam a obtenção de volumes de impedância absoluta, cada qual com níveis distintos de suavização e detalhamento.

Os resultados são apresentados em subseções específicas, dedicadas a cada método. Em cada subseção, são discutidos os parâmetros empregados, os principais produtos gerados (modelos e seções de impedância) e a análise comparativa entre as respostas obtidas e observadas, de modo a destacar as diferenças e complementaridades entre as abordagens.

9.4.1. Inversão Colorida

O dado sísmico utilizado como entrada para a inversão colorida corresponde ao produto do pré-condicionamento obtido pela aplicação conjunta do *Dip-Steered Median Filter* e do *Spectral Blueing*, resultando em um volume com maior resolução e melhor relação sinal/ruído. Além disso, os perfis de impedância acústica (AI) disponíveis na área foram utilizados em conjunto com os traços sísmicos para o ajuste do expoente α , que neste estudo alcançou o valor de $-0,906$. Esse valor definiu o operador espectral rotacionado em -90° , o qual, ao ser convolvido com o dado sísmico, gerou os resultados que serão apresentados a seguir.

Por se tratar de um volume de impedância acústica relativa, nesta seção é apresentada uma avaliação qualitativa dos resultados, com ênfase na interpretação das fácies sísmicas e no ganho de resolução obtido após a aplicação da inversão colorida. A análise é conduzida em correlação direta com as fácies identificadas nos poços, buscando compreender de que forma o produto de inversão pode enriquecer a caracterização da formação de interesse.

Um dos principais objetivos deste estudo consiste em avaliar qualitativamente o ganho de resolução vertical e gerar uma propriedade sísmica secundária de maior acurácia, a ser utilizada como variável guia na construção do modelo de baixa frequência empregado nas inversões subsequentes. Assim, análises mais detalhadas e avaliações quantitativas serão apresentadas nas seções seguintes, referentes às inversões determinística e estocástica.

Com o intuito de facilitar a identificação dos poços nas seções sísmicas que serão apresentadas a seguir, os mesmos foram renomeados conforme a Tabela 9.4.

Tabela 9.4. Identificação e renomeação dos poços utilizados nas seções sísmicas apresentadas neste capítulo

Nome dos poços	Abreviação
3-OGX-57-MA	OGX-57
1-OGX-16-M	OGX-16
2-CP-1-MA	CP-1
1-OGX-22-MA	OGX-22

A Figura 9.22 exibe uma seção sísmica arbitrária interceptando os poços de estudo, juntamente com uma *time slice*, evidenciando o resultado da inversão colorida em comparação à amplitude sísmica original. Observa-se, na área de estudo, a presença de múltiplas intrusões ígneas, as quais constituem um desafio recorrente à interpretação sísmica.

A partir da inversão colorida, torna-se evidente a ocorrência de um salto de soleira na base da Formação Pimenteiras, caracterizado por uma alta impedância acústica relativa positiva, responsável pela mudança de polaridade do refletor que delimita a formação. Na interface entre as Formações Pimenteiras e Itaim, na ausência de intrusões, observa-se um aumento gradual nos valores de impedância, o que sugere que a base da Formação Pimenteiras tende a se expressar como um pico positivo.

Entretanto, quando intrusões ígneas ocorrem dentro do intervalo da Formação Pimenteiras, especialmente em sua base, como observado no poço OGX-22, pode ocorrer inversão de polaridade devido ao forte contraste de impedância entre as rochas sedimentares e o corpo ígneo. Dessa forma,

embora a Formação Pimenteiras apresente, em geral, uma configuração estrutural simples e plano-paralela, a presença de intrusões ígneas introduz complexidades interpretativas significativas, as quais são evidenciadas de maneira mais clara nos volumes de impedância acústica relativa resultantes da inversão colorida.

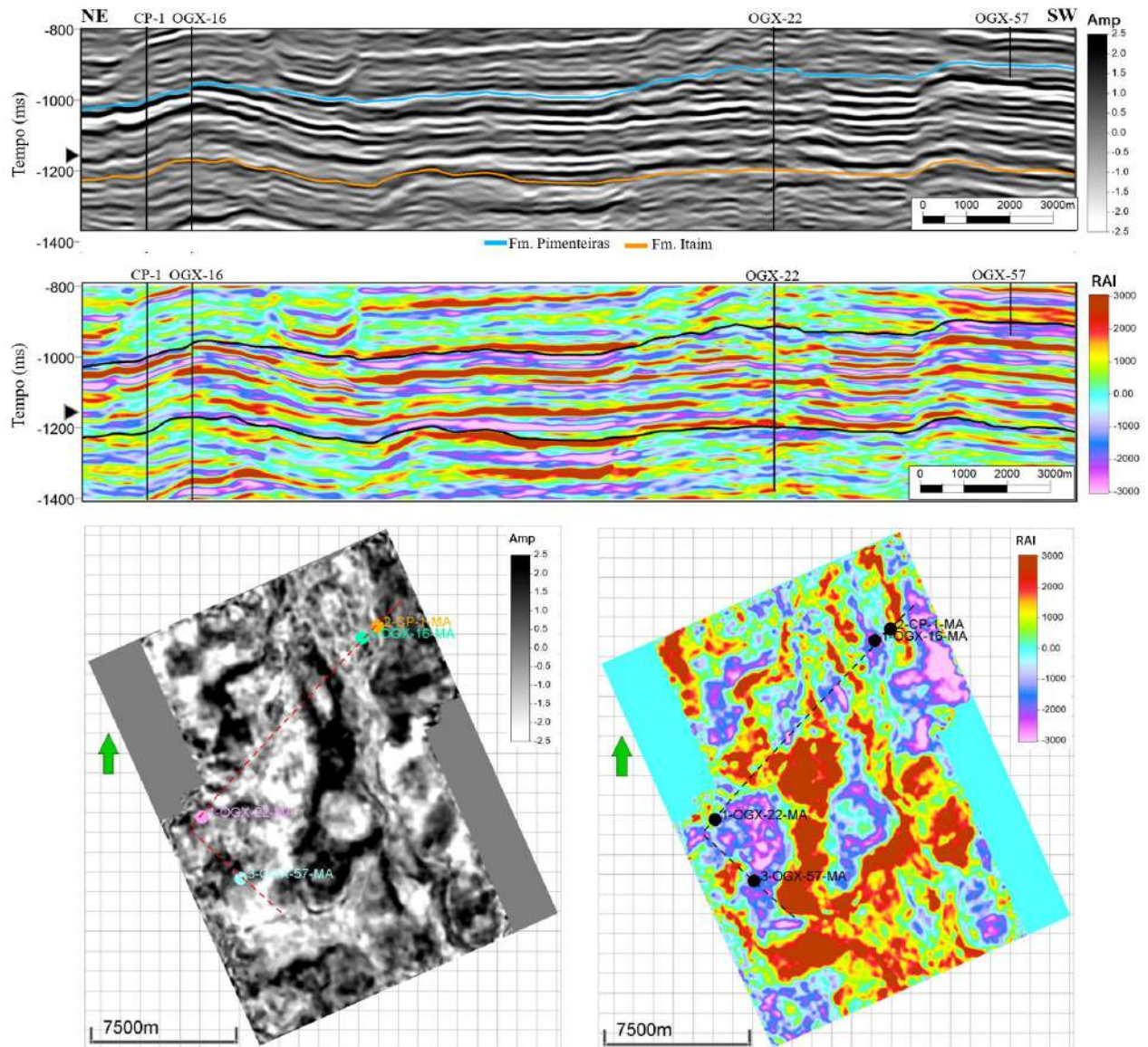


Figura 9.22. Seção arbitrária e *time slice* evidenciando o resultado da inversão colorida em comparação à amplitude sísmica original.

As seções sísmicas (*inlines*) que atravessam os poços evidenciam a presença de falhas normais, frequentemente associadas aos saltos de soleiras ígneas, que se tornam mais nítidos com o auxílio do volume de impedância acústica relativa. Embora essas estruturas ainda possam ser identificadas no dado de amplitude sísmica convencional, a aplicação da inversão colorida promoveu um ganho significativo de resolução, permitindo uma melhor individualização e continuidade dos refletores.

A Figura 9.23 apresenta as seções sísmicas que interceptam os poços OGX-16 e CP-1, nas quais é possível observar com maior definição as intrusões ígneas e as camadas de folhelho adjacentes. Nota-se que essas camadas apresentam características de impedância distintas, cuja variação está diretamente relacionada à proximidade das soleiras: quanto mais próximas das intrusões, mais acentuadamente negativos tendem a ser os valores de impedância acústica relativa, refletindo o impacto térmico dessas intrusões sobre as rochas sedimentares encaixantes.

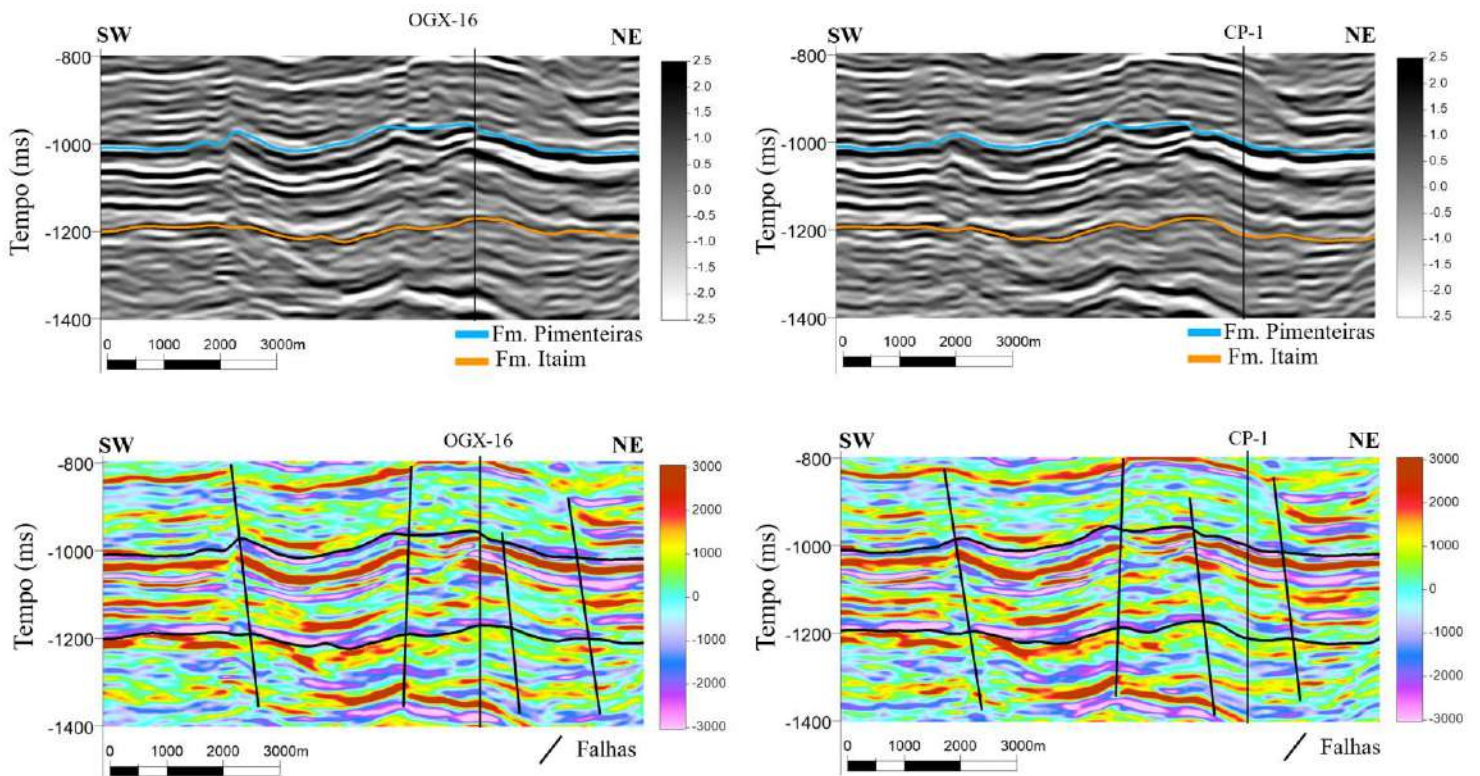


Figura 9.23. Seções sísmicas interceptando os poços OGX-16 e CP-1, mostrando a definição aprimorada das intrusões ígneas e dos folhelhos adjacentes após a inversão colorida.

A mesma análise pode ser realizada a partir da Figura 9.24, que apresenta as seções sísmicas interceptando os poços OGX-22 e OGX-57. Nessa região, os saltos de soleiras não se mostram tão expressivos quanto os observados na seção anterior; ainda assim, exercem influência estrutural significativa, promovendo falhamentos locais e afetando a continuidade dos refletores da formação de forma moderada.

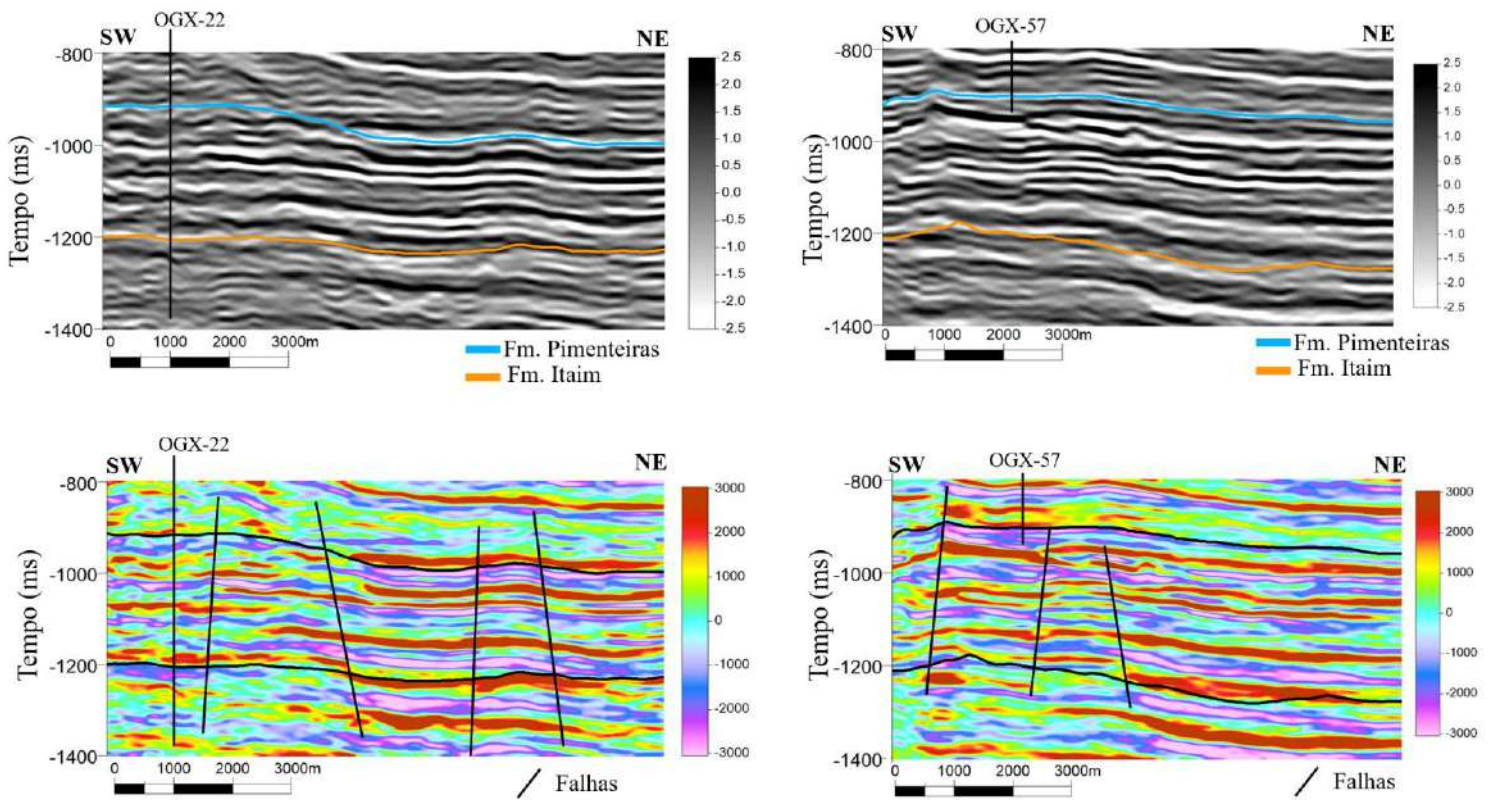


Figura 9.24. Seções sísmicas atravessando os poços OGX-22 e OGX-57, evidenciando a influência estrutural das soleiras ígneas e o impacto na continuidade dos refletores.

Conforme ilustrado nas Figura 9.25 e Figura 9.26 que destacam em maior detalhe os resultados da inversão colorida, é possível estabelecer uma correlação direta com os perfis de fácies dos poços, evidenciando o ganho de resolução e a melhor compreensão da heterogeneidade interna da Formação Pimenteiras. O produto de inversão revela um maior detalhamento estratigráfico das camadas de folhelho, as quais não são plenamente resolvidas no dado de amplitude sísmica convencional devido às limitações de resolução vertical.

Nas seções correspondentes à Figura 9.25, observa-se de forma mais nítida a intercalação das fácies 3 e 4 na porção inferior da Formação Pimenteiras, representadas no volume de impedância relativa por duas camadas de baixa impedância separadas por uma camada de maior impedância. Essa configuração pode estar associada às principais sequências estratigráficas interpretadas para esse intervalo através de estudos realizados em afloramentos (Caputo 1984). Além disso, observa-se na porção superior da Formação Pimenteiras um alto contraste de impedância acústica relativa, caracterizado por valores negativos associados aos folhelhos metassedimentares da fácies 2, em contraste com as soleiras ígneas correspondentes à fácies 1, que exibem altos valores positivos de impedância relativa.

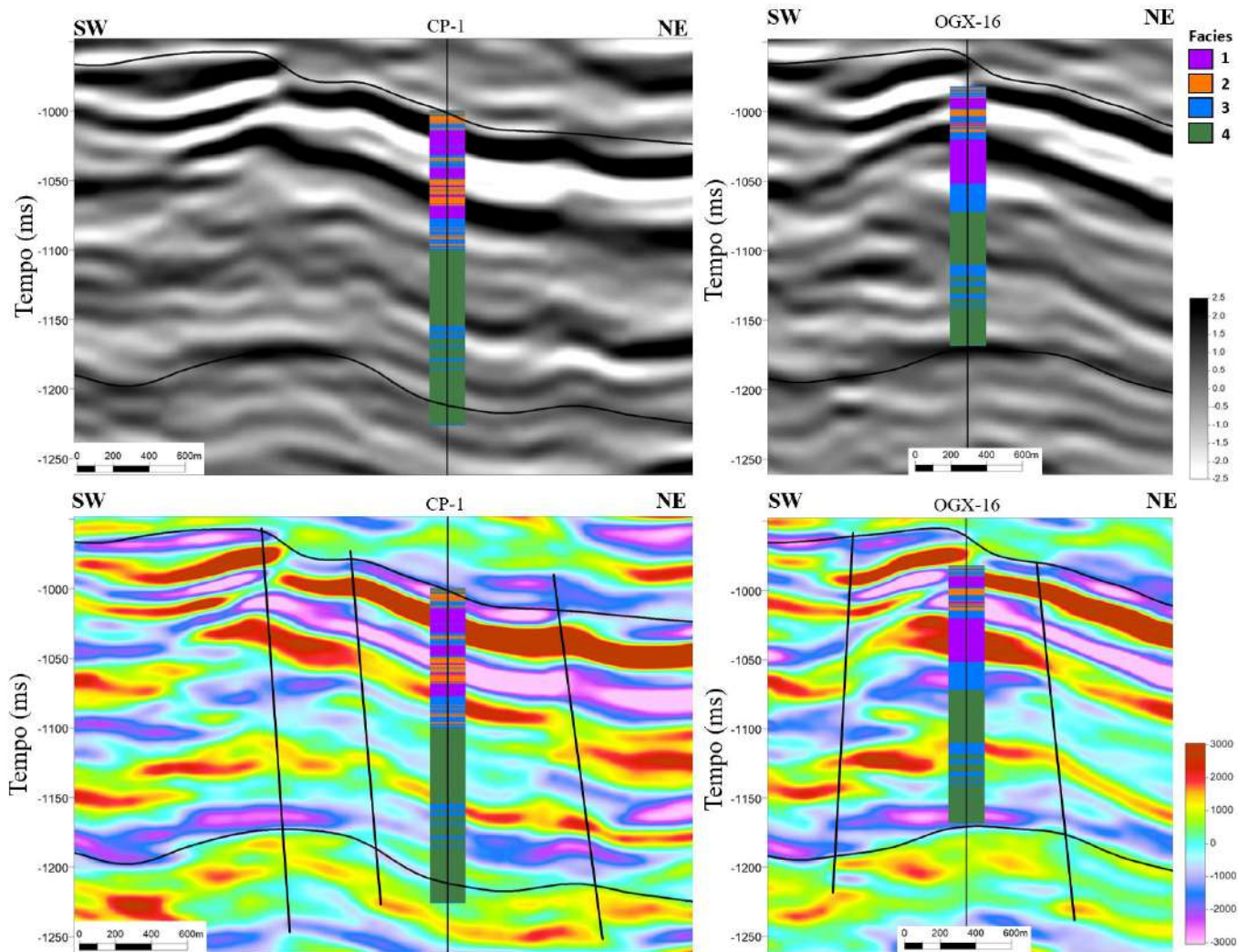


Figura 9.25. Detalhe das seções sísmicas destacando a correlação entre as fácies sedimentares e as variações de impedância acústica relativa na Formação Pimenteiras.

De modo análogo, a camada de maior impedância identificada na porção inferior da formação também é observada no poço OGX-22, onde corresponde à fácies 3, caracterizada por altos valores de impedância acústica relativa (Figura 9.26). Mesmo no poço OGX-57, que apresenta menor espessura sedimentar, é possível reconhecer contrastes significativos no volume de inversão colorida, sugerindo correspondência com as fácies classificadas nos poços, o que reforça a eficácia do método na diferenciação de unidades estratigráficas finas.

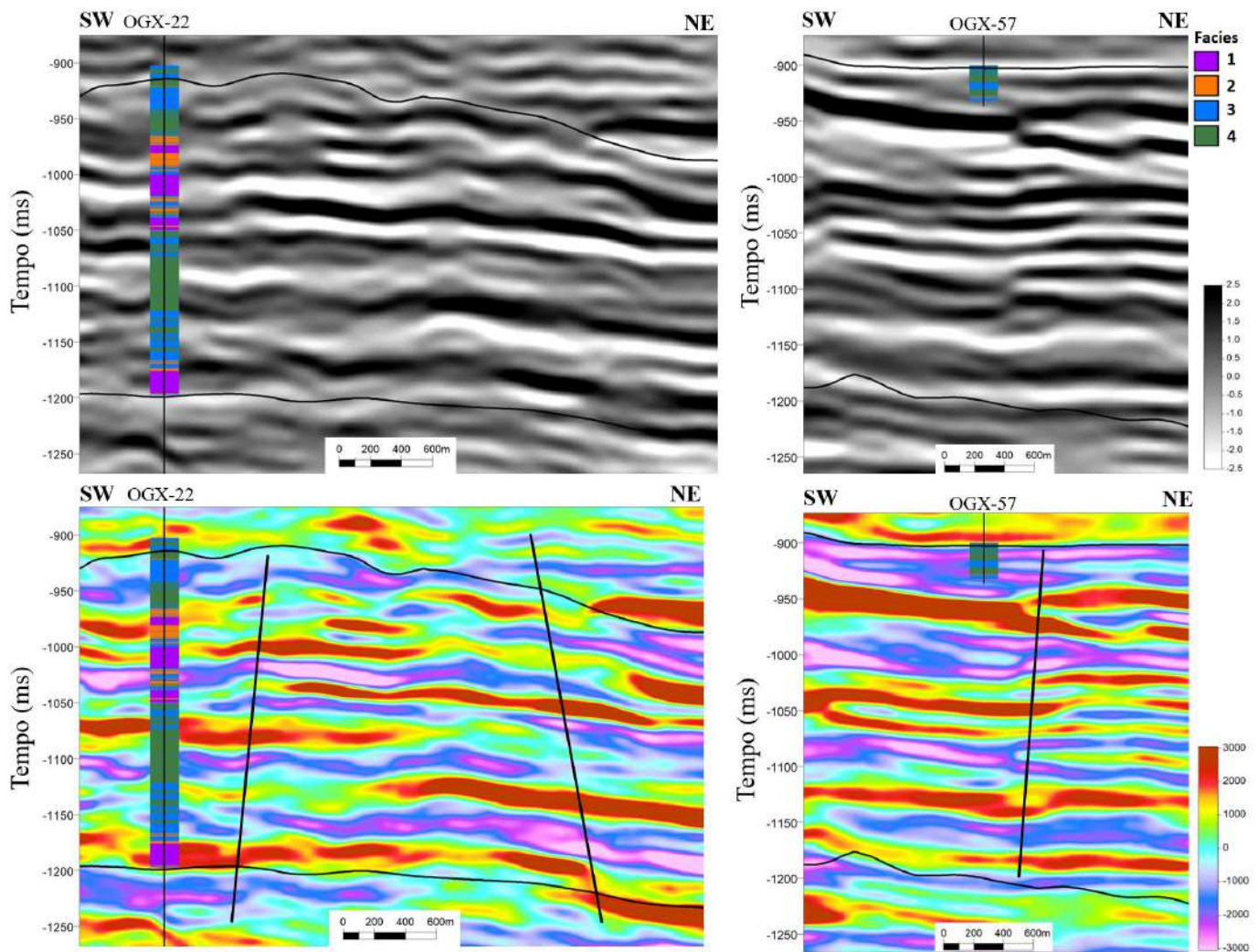


Figura 9.26. Seções sísmicas detalhadas mostrando o contraste de impedância associado às fácies sedimentares e às intrusões ígneas, com destaque para a fácies 3 no poço OGX-22.

9.4.2. Inversão Determinística

Conforme descrito na seção metodológica, o fluxo de trabalho desenvolvido para esta etapa envolveu a construção de um modelo de baixa frequência, elemento essencial para a aplicação da abordagem determinística. Inicialmente, é apresentado o resultado do modelo de baixa, elaborado com base no volume de impedância relativa obtido pela Inversão Colorida, o qual foi utilizado como guia para a interpolação dos valores de impedância acústica entre os poços da área de estudo. Em seguida, o modelo foi submetido a um filtro passa-baixa, com o objetivo de preservar apenas os componentes de baixa frequência, necessária à inversão (Figura 9.27).

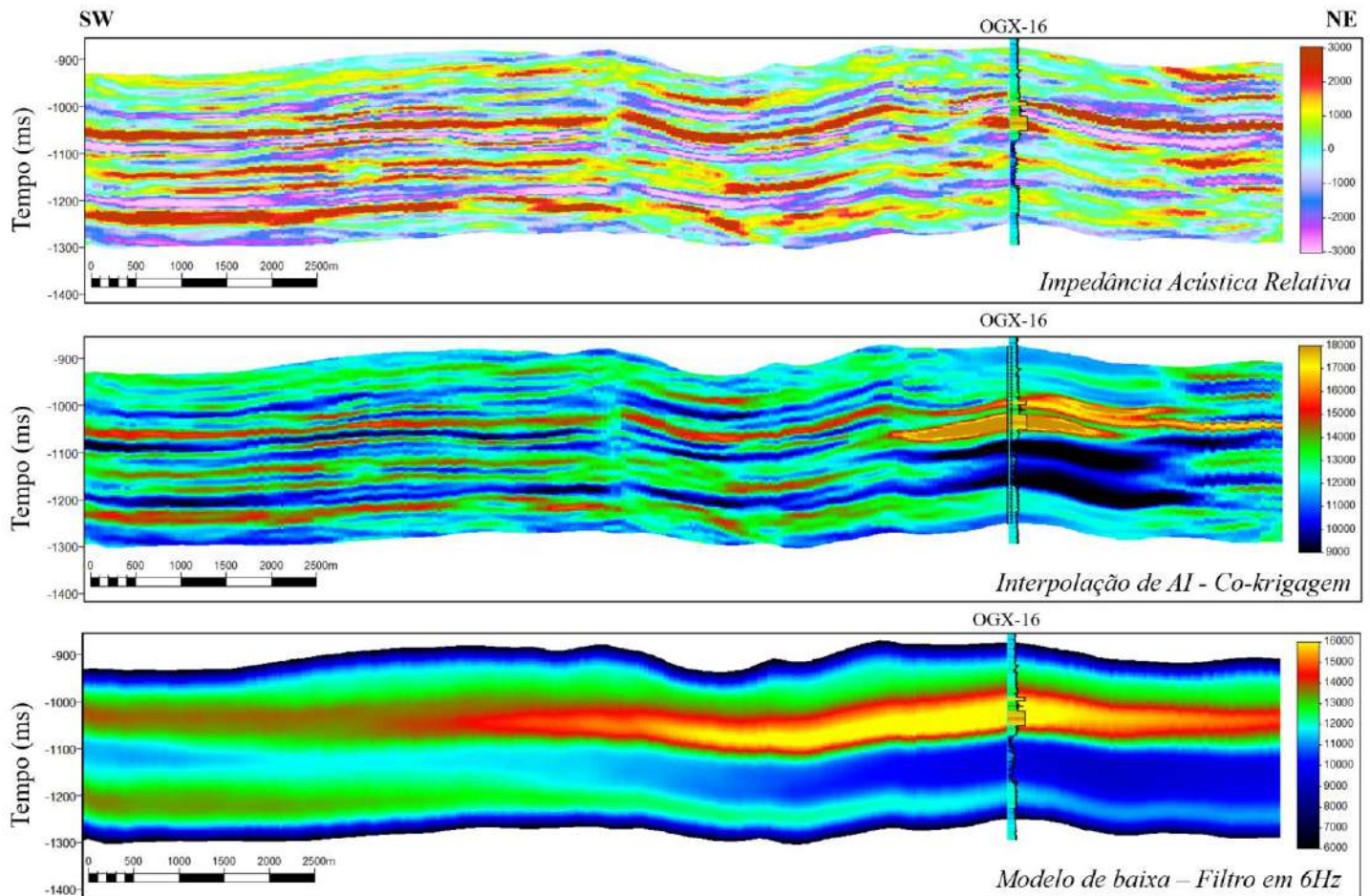


Figura 9.27. Modelo de baixa frequência obtido por co-krigagem e filtragem passa-baixa, utilizado como dado inicial para a inversão determinística.

Com o modelo de baixa frequência definido e a wavelet média estimada a partir dos poços, foi realizada a inversão determinística para impedância acústica absoluta. A avaliação quantitativa da performance do método é apresentada na Figura 9.28, que mostra o crossplot entre os valores de impedância acústica dos poços e os valores invertidos, além de uma tabela com os coeficientes de correlação obtidos individualmente por poço.

De modo geral, os coeficientes de correlação linear no domínio do tempo, calculados entre os perfis de impedância acústica reamostrados (*upscale*) e os resultados da inversão, indicam altos níveis de concordância, com valores superiores a 0,7, atingindo um máximo de 0,94 para o poço OGX-16. Esses resultados evidenciam a consistência da abordagem determinística empregada.

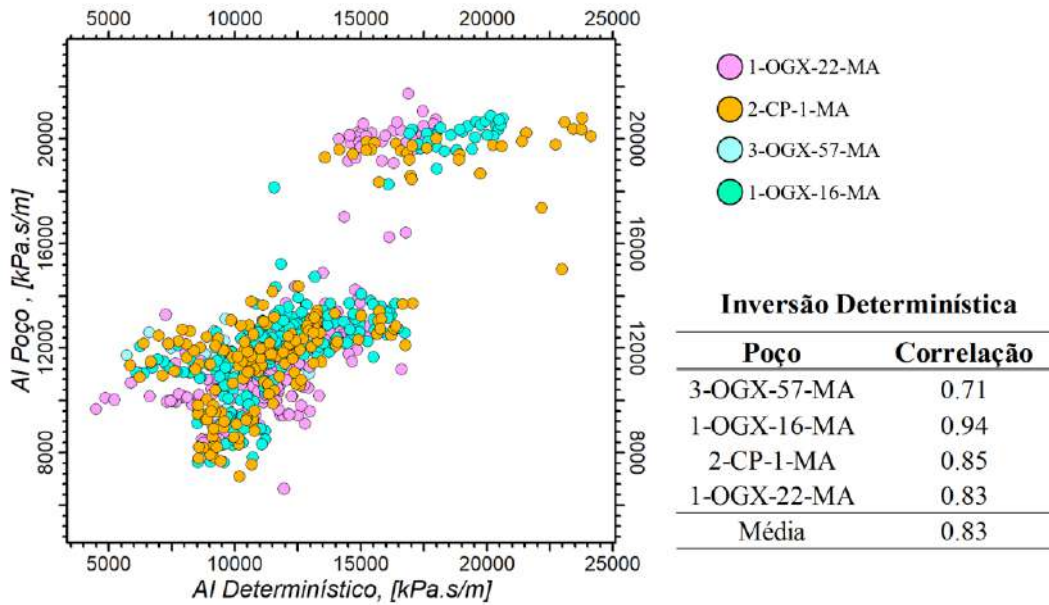


Figura 9.28. *Crossplot* entre a impedância acústica medida nos poços e a derivada da inversão determinística, mostrando os coeficientes de correlação individuais e médio.

A avaliação qualitativa dos resultados é ilustrada na Figura 9.29, que apresenta uma seção sísmica do volume de impedância acústica invertida atravessando os poços OGX-16 e CP-1, juntamente com as curvas de impedância acústica extraídas de diferentes fontes: o modelo de baixa (curva verde), a impedância reamostrada do poço (curva vermelha) e a impedância resultante da inversão determinística (curva azul). A análise comparativa dessas curvas, ao longo da profundidade temporal, permite identificar, ainda que qualitativamente, os intervalos de maior e menor correlação entre os dados observados e os modelados. Observa-se que os maiores desvios ocorrem nos valores extremos de impedância, tanto nos altos valores associados às soleiras de diabásio, quanto nos baixos valores característicos dos folhelhos.

Cabe ressaltar que a inversão sísmica foi executada em um intervalo temporal ampliado, abrangendo uma faixa que extrapola verticalmente os limites da Formação Pimenteiras, conforme indicado pelos marcadores estratigráficos na Figura 9.29. Dessa forma, as regiões correspondentes aos topos e bases dos poços, onde se observam baixas correlações entre a impedância medida e a invertida, não devem ser consideradas na avaliação, por se situarem fora do intervalo de interesse geológico.

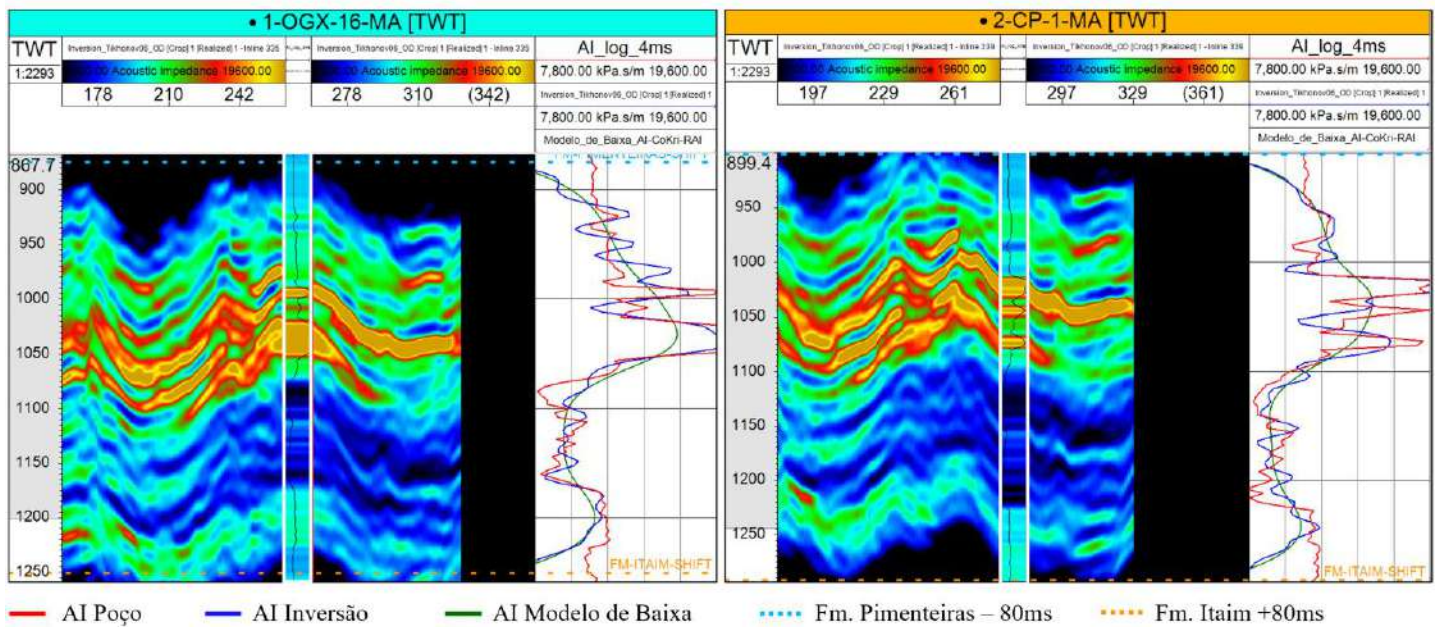


Figura 9.29. Seção sísmica ilustrando o ajuste entre o modelo de baixa, a impedância reamostrada dos poços e o volume resultante da inversão determinística.

O mesmo comportamento anteriormente descrito pode ser observado na Figura 9.30, especialmente em relação aos valores extremos de impedância, nos quais se concentram os maiores desvios. Observa-se que a inversão determinística não alcançou plenamente os valores máximos de impedância, sobretudo no poço OGX-22, no intervalo correspondente às soleiras de diabásio de menor espessura. Por outro lado, é importante ressaltar que, no principal intervalo de interesse, correspondente às camadas de folhelhos ricos em carbono orgânico total (COT), a inversão determinística apresentou excelentes resultados, evidenciados por altíssima correlação entre os dados de poço e os valores invertidos.

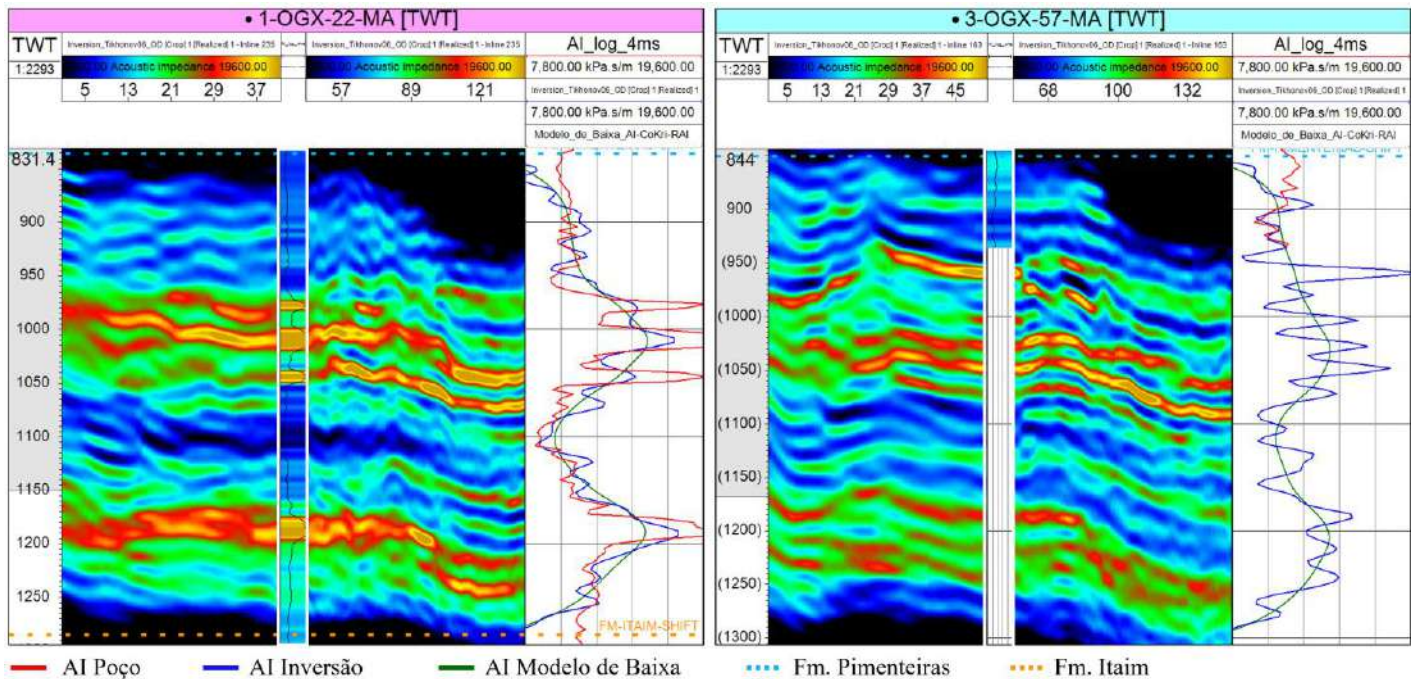


Figura 9.30. Comparação entre as curvas de impedância acústica dos poços e os resultados da inversão determinística, destacando desvios nas zonas de altos contrastes.

Na seção sísmica arbitrária apresentada na Figura 9.31, que intercepta os poços da área de estudo, observa-se que, na porção nordeste, onde se localizam os poços CP-1 e OGX-16, a Formação Pimenteiras apresenta maiores espessura caracterizadas por baixos valores de impedância acústica. Em direção à porção sudoeste, abrangendo os poços OGX-22 e OGX-57, os níveis de baixa impedância tornam-se gradualmente mais delgados, acompanhados por um aumento progressivo de impedância acústica. Esse aumento é associado à presença de uma intrusão ígnea que cortam a Formação Pimenteiras em sua base, conforme observado no poço OGX-22. À medida que se aproxima do poço OGX-57, nota-se que essa intrusão ocorre em profundidades maiores, possivelmente abaixo da Formação Pimenteiras, o que permite novamente a ocorrência de um intervalo mais espesso com baixas impedâncias, representando uma zona de maior preservação dos folhelhos ricos em matéria orgânica.

Do ponto de vista dos objetivos deste trabalho, esses níveis estratigráficos de baixa impedância são de extrema relevância, pois podem indicar intervalos com maior potencial gerador, associados às camadas mais ricas em COT. Assim, compreender a distribuição espacial das rochas de baixa impedância e o modo como as soleiras de diabásio influenciam suas propriedades elásticas é fundamental, uma vez que tais intrusões podem impactar significativamente o grau de maturidade térmica dos folhelhos geradores.

De forma semelhante ao observado na seção sísmica, o *time slice* apresentado na Figura 9.31 também evidencia um ganho considerável de resolução, o que torna viável a caracterização detalhada da formação. Observa-se, ainda, como as intrusões ígneas impactaram estruturalmente a área de estudo, manifestando-se não apenas por meio de falhas normais, mas também por geometrias elipsoidais associadas ao fechamento estrutural das soleiras. Nota-se, adicionalmente, uma tendência geral de aumento da impedância acústica à medida que as rochas sedimentares se aproximam das rochas ígneas, refletindo o efeito térmico e composicional dessas intrusões sobre as propriedades elásticas do meio.

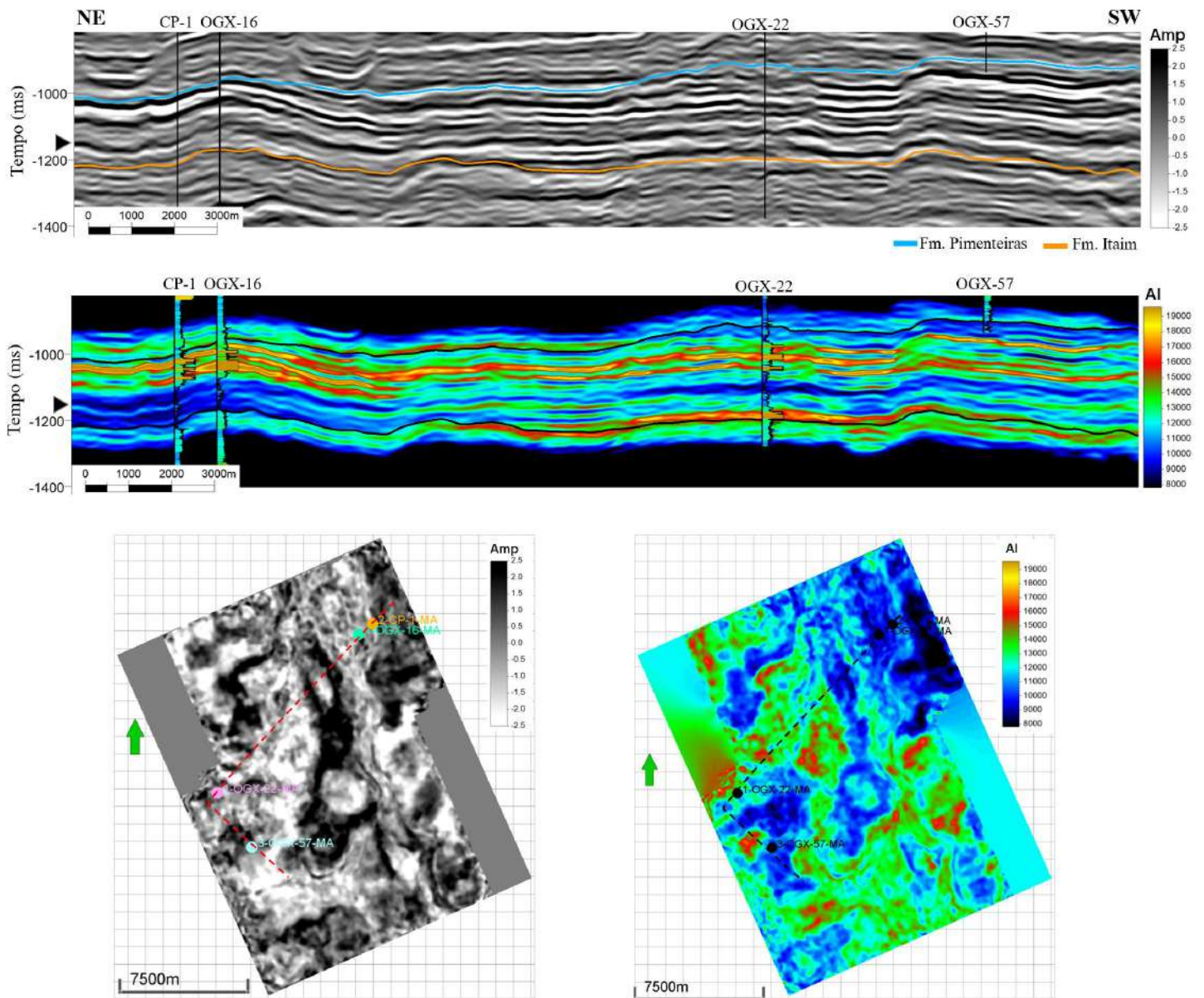


Figura 9.31. Seção sísmica arbitrária e *time slice* evidenciando o ganho de resolução e heterogeneidade na distribuição de impedância na Formação Pimenteiras.

As seções sísmicas (*inlines*) que cruzam os poços são apresentadas nas Figura 9.32 e Figura 9.33, as quais mostram em detalhe a estratificação da Formação Pimenteiras e sua heterogeneidade acústica, interpretadas como respostas às parassequências inseridas no contexto do Trato de Sistemas Transgressivo (TST) que originou essa formação (Della Fávera, 1990; Rodrigues, 1995). Na Figura 9.32, observa-se uma continuidade lateral de zonas de alta impedância acústica, para os folhelhos, em torno de 11.000 m/s·g/cm³, localizadas nas profundidades temporais de 1160 ms e 1120 ms, nos poços CP-1 e OGX-16, respectivamente. Essa alta impedância pode estar associada a um limite entre conjuntos de parassequências, representando mudanças sutis na taxa de sedimentação, concentração de matéria orgânica e na energia do ambiente deposicional.

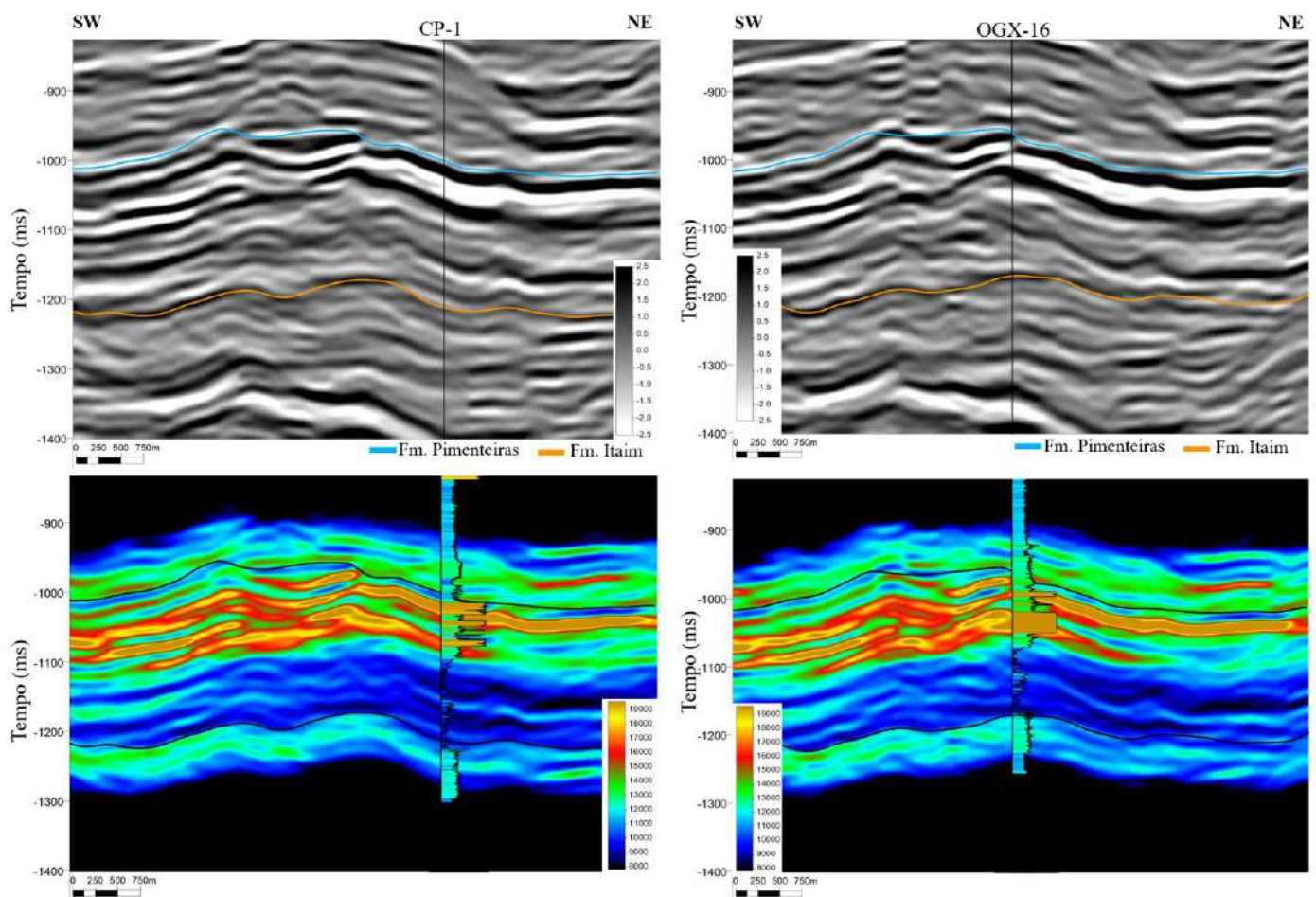


Figura 9.32. Seção sísmica atravessando os poços CP-1 e OGX-16, evidenciando os contrastes internos no intervalo de folhelhos que podem está associados aos limites entre parassequências do Trato de Sistemas Transgressivo.

Na Figura 9.33, o poço OGX-22 apresenta, em aproximadamente 1100 ms, uma baixa impedância (cerca de 8.000 m/s·g/cm³), que coincide com um pico de radioatividade natural (GR) da ordem de 200 API. Essa correlação sugere um momento de afogamento máximo, marcado pelo

enriquecimento em elementos radioativos e pela deposição de folhelhos finos e ricos em matéria orgânica. O mesmo comportamento de baixa impedância também é observado no topo da formação nos poços OGX-57 e OGX-22, com valores de GR próximos de 180 API, o que corrobora a interpretação de Della Fávera (1990) e René Rodrigues (1995), que descreveram essa superfície como a Superfície de Máxima Inundação, a qual marca o topo da Formação Pimenteiras.

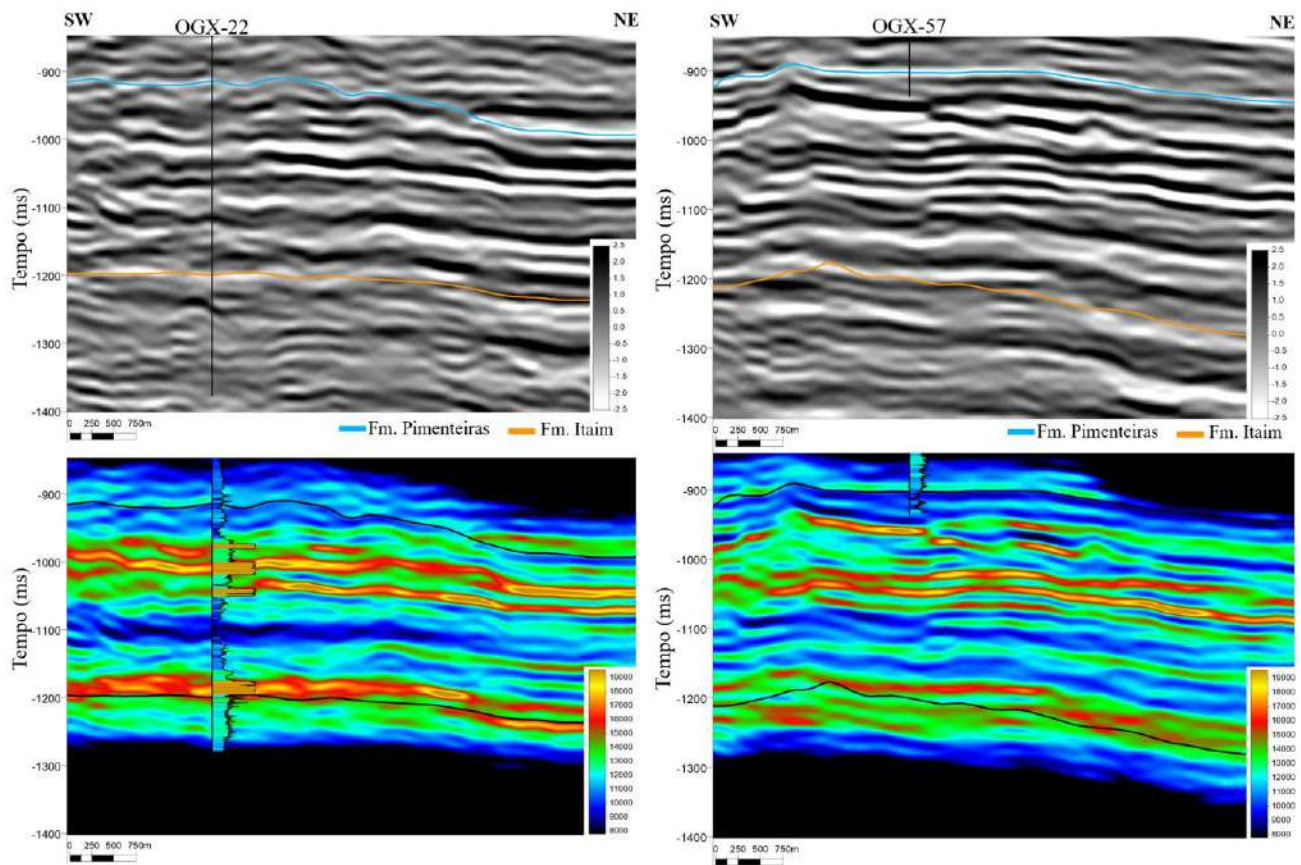


Figura 9.33. Seção sísmica interceptando os poços OGX-22 e OGX-57, mostrando intervalos de baixa impedância correlacionados a altos valores de GR e à Superfície de Máxima Inundação no topo da Fom. Pimenteiras.

Por fim, foi realizada uma comparação entre os resultados da inversão colorida e da inversão determinística ao longo de uma seção sísmica arbitrária que atravessa os poços, bem como em um *time slice* correspondente ao tempo de 1160 ms (Figura 9.34). Essa comparação evidencia a diferença de resolução e detalhamento entre os métodos: a inversão colorida, de natureza mais ágil e voltada à interpretação qualitativa, permite realçar a continuidade e a variação espectral dos refletores; já a inversão determinística, de caráter quantitativo, fornece valores absolutos de impedância acústica, conferindo maior robustez e confiabilidade aos modelos geológicos. Assim, ambas as abordagens se mostram complementares, permitindo uma interpretação sísmica integrada e consistente do intervalo de interesse.

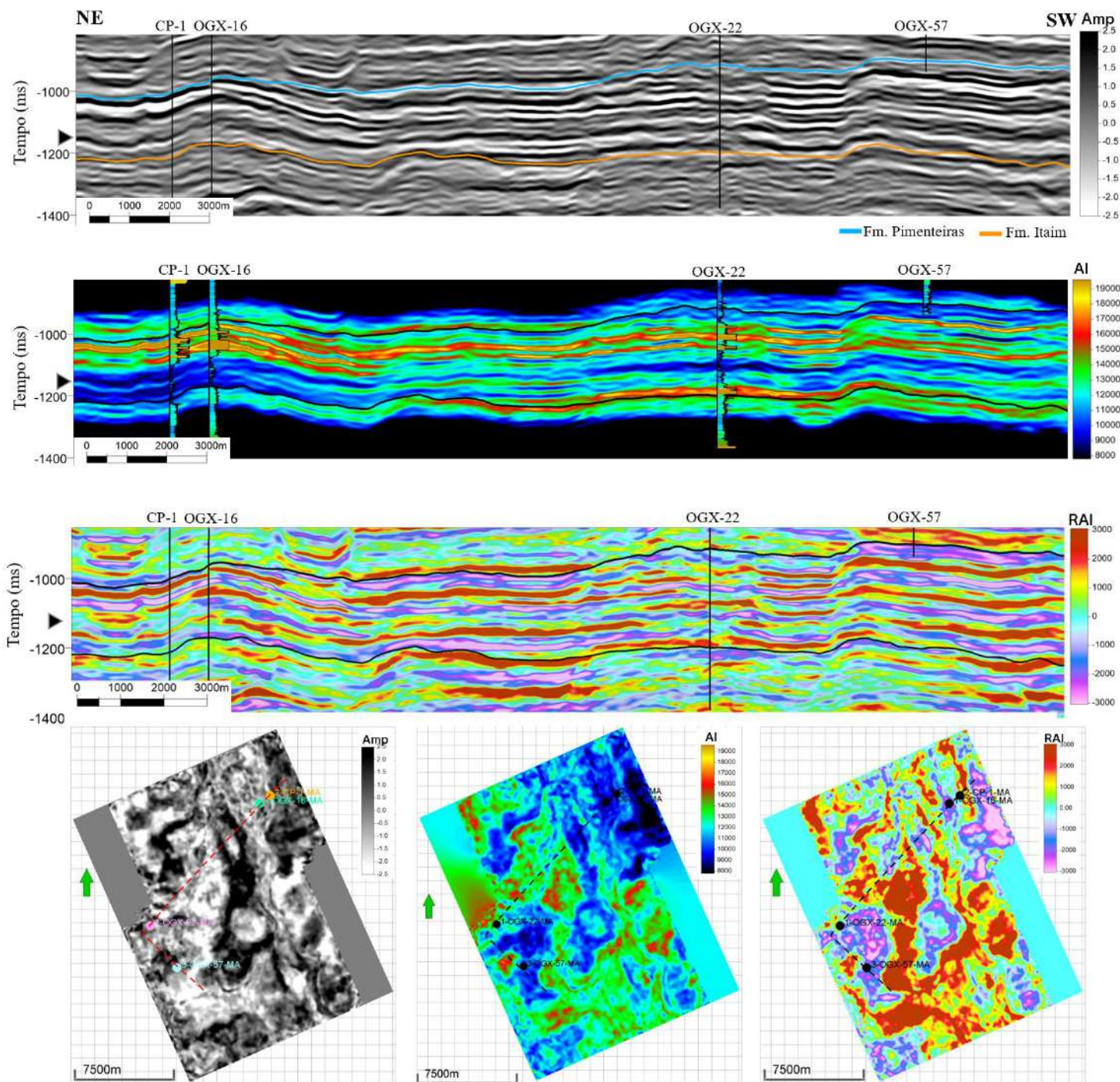


Figura 9.34. Seção sísmica arbitrária e *time slice* comparando os resultados obtidos pela inversão colorida e pela inversão determinística.

9.4.3. Inversão Estocástica

A partir deste ponto, são apresentados os resultados obtidos por meio da Inversão Estocástica. Após a execução da inversão determinística, procedeu-se à inversão estocástica, utilizando como

insumos principais: (i) o modelo inicial derivado do resultado da inversão determinística, (ii) a mesma wavelet empregada anteriormente, e (iii) os variogramas previamente calculados e descritos na seção metodológica. Embora a inversão estocástica permita a geração de múltiplas realizações equiprováveis, o objetivo principal desta etapa foi a obtenção de um volume de impedância acústica com resolução aprimorada, capaz de representar com maior detalhe as camadas finas de interesse no contexto do objetivo geral deste trabalho. A motivação para o uso dessa abordagem está relacionada ao fato de que o resultado final desta etapa servirá como dado de entrada para a simulação estocástica do volume de carbono orgânico total, a qual também poderá gerar múltiplas simulações. Assim, ainda que seja selecionada apenas uma das realizações como modelo de referência, serão avaliadas, como controle de qualidade, a média e o desvio padrão das simulações geradas, de modo a quantificar a variabilidade e a incerteza associadas ao processo inverso.

Com o intuito de estabelecer um poço de controle para avaliar o desempenho das simulações, foi construído um modelo inicial de impedância por meio de co-krigagem entre os poços, utilizando a inversão determinística como variável secundária (guia) com alto peso de correlação. Nesse procedimento, o poço OGX-16 foi excluído do processo de inversão estocástica, a fim de ser utilizado como poço de validação independente.

Como discutido na seção anterior, o coeficiente de correlação entre a inversão determinística e a impedância do poço OGX-16 foi de 0,94, indicando uma forte concordância. No entanto, ao ser excluído do modelo e utilizado como teste, o modelo inicial obtido apresentou uma correlação de 0,84, conforme ilustrado na Figura 9.35. Apesar de ainda representar um valor elevado, essa diferença é esperada e desejável, uma vez que o poço de controle não participa do processo de calibração da inversão estocástica, permitindo assim avaliar de forma independente o desempenho das simulações e estabelecer um critério objetivo para a seleção do melhor resultado.

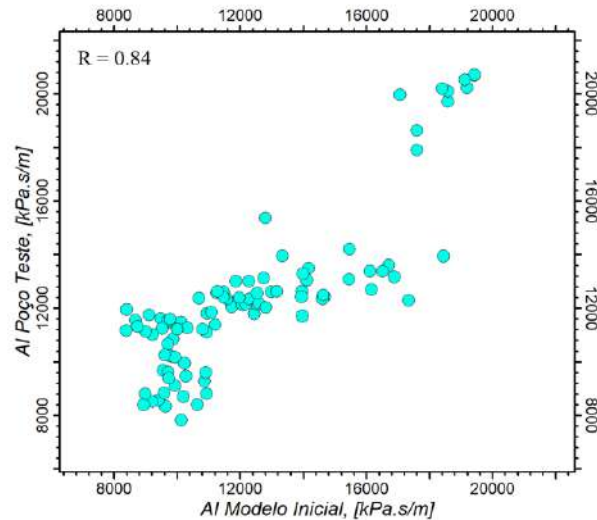


Figura 9.35. *Crossplot* mostrando a correlação linear entre o modelo inicial utilizado na inversão estocástica e a impedância do poço.

Inicialmente, foram geradas 50 realizações estocásticas de impedância acústica, tendo como parâmetro de controle o coeficiente de correlação entre os resultados da inversão estocástica e a impedância acústica do poço teste (OGX-16). Conforme apresentado na Tabela 9.5, 24 simulações apresentaram coeficientes de correlação superiores a 0,7 (ressaltado pelas células em verde), valor considerado satisfatório para os objetivos deste estudo. A partir desse conjunto, foram calculados o volume médio de impedância acústica e o respectivo desvio padrão, com o intuito de avaliar a consistência estatística e a variabilidade espacial das realizações.

Entre as simulações geradas, a simulação nº 4 foi selecionada como resultado final, por apresentar o maior coeficiente de correlação com o poço de controle, atingindo 0,81. Observa-se, contudo, que, apesar de o modelo inicial ter exibido alta correlação com o poço teste, nenhuma das realizações superou o valor de 0,84, e aproximadamente metade das simulações apresentaram coeficientes de correlação inferiores a 0,7 (ressaltadas pelas células em vermelho). Esse comportamento reflete a natureza probabilística e não determinística do método, no qual a variabilidade entre as realizações representa a incerteza inerente ao processo de inversão estocástica.

Tabela 9.5. Resultados da inversão estocástica: coeficientes de correlação entre a impedância acústica dos poços e as 50 realizações simuladas. As células em verde destacam as realizações com correlação superior a 0,7, e as em vermelho indicam valores inferiores a esse limite.

Coeficiente de Correlação		Coeficiente de Correlação	
Inversão Estocástica	Poço teste: OGX-16	Inversão Estocástica	Poço teste: OGX-16
SSI_Acoustic Impedance_1	0,75	SSI_Acoustic Impedance_26	0,66
SSI_Acoustic Impedance_2	0,70	SSI_Acoustic Impedance_27	0,63
SSI_Acoustic Impedance_3	0,60	SSI_Acoustic Impedance_28	0,55
SSI_Acoustic Impedance_4	0,81	SSI_Acoustic Impedance_29	0,73
SSI_Acoustic Impedance_5	0,79	SSI_Acoustic Impedance_30	0,72
SSI_Acoustic Impedance_6	0,67	SSI_Acoustic Impedance_31	0,70
SSI_Acoustic Impedance_7	0,77	SSI_Acoustic Impedance_32	0,76
SSI_Acoustic Impedance_8	0,74	SSI_Acoustic Impedance_33	0,74
SSI_Acoustic Impedance_9	0,70	SSI_Acoustic Impedance_34	0,66
SSI_Acoustic Impedance_10	0,70	SSI_Acoustic Impedance_35	0,70
SSI_Acoustic Impedance_11	0,73	SSI_Acoustic Impedance_36	0,71
SSI_Acoustic Impedance_12	0,65	SSI_Acoustic Impedance_37	0,67
SSI_Acoustic Impedance_13	0,74	SSI_Acoustic Impedance_38	0,55
SSI_Acoustic Impedance_14	0,55	SSI_Acoustic Impedance_39	0,58
SSI_Acoustic Impedance_15	0,79	SSI_Acoustic Impedance_40	0,65
SSI_Acoustic Impedance_16	0,70	SSI_Acoustic Impedance_41	0,68
SSI_Acoustic Impedance_17	0,61	SSI_Acoustic Impedance_42	0,66
SSI_Acoustic Impedance_18	0,54	SSI_Acoustic Impedance_43	0,60
SSI_Acoustic Impedance_19	0,62	SSI_Acoustic Impedance_44	0,65
SSI_Acoustic Impedance_20	0,64	SSI_Acoustic Impedance_45	0,73
SSI_Acoustic Impedance_21	0,78	SSI_Acoustic Impedance_46	0,68
SSI_Acoustic Impedance_22	0,68	SSI_Acoustic Impedance_47	0,77
SSI_Acoustic Impedance_23	0,68	SSI_Acoustic Impedance_48	0,53
SSI_Acoustic Impedance_24	0,71	SSI_Acoustic Impedance_49	0,76
SSI_Acoustic Impedance_25	0,59	SSI_Acoustic Impedance_50	0,68

A Figura 9.36 apresenta, em perfil, a comparação entre a impedância acústica do poço teste (OGX-16), o resultado da inversão estocástica com maior coeficiente de correlação e a média das simulações que apresentaram correlação superior a 0,7. Além disso, é exibido o perfil de desvio padrão, que permite identificar as profundidades associadas aos maiores desvios entre as simulações selecionadas, destacando os intervalos de maior variabilidade entre as realizações estocásticas. Na mesma figura, observa-se também o crossplot entre a impedância acústica resultante da inversão estocástica e a impedância medida no poço teste, colorido de acordo com os valores de desvio padrão. A distribuição dos pontos no gráfico indica que não há uma faixa específica de valores de impedância associada a maiores desvios, sugerindo que as diferenças entre as simulações são espacialmente e estatisticamente distribuídas ao longo do intervalo analisado, sem concentração em uma faixa particular de impedância acústica.

Do ponto de vista interpretativo, o desvio padrão reflete o grau de incerteza local do processo de inversão estocástica. Valores mais baixos de desvio indicam maior consistência entre as realizações e, portanto, maior confiabilidade do resultado médio. Por outro lado, valores elevados de desvio padrão

correspondem a zonas de menor controle sísmico ou de maior heterogeneidade litológica, nas quais pequenas variações nas condições de contorno (modelo inicial, variograma ou wavelet) produzem diferenças mais significativas entre as simulações. Assim, a análise conjunta do perfil de desvio padrão e do crossplot permite identificar os intervalos estratigraficamente mais estáveis e aqueles onde a incerteza associada à estimativa de impedância é mais significativa, fornecendo um importante diagnóstico para a propagação de incertezas nas etapas subsequentes de modelagem de COT.

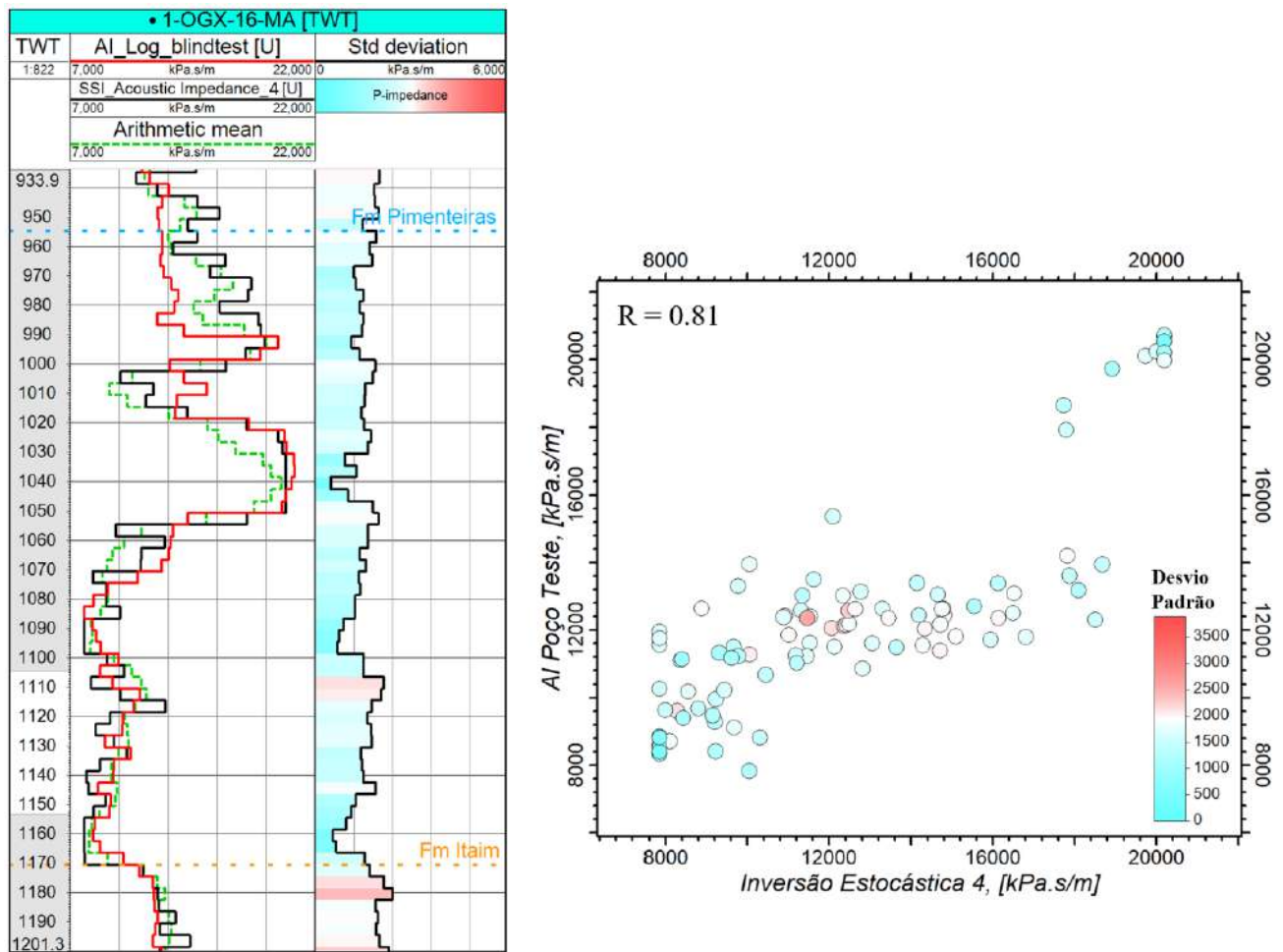


Figura 9.36. Comparação entre o perfil de impedância acústica do poço teste, a inversão estocástica de maior correlação e a média das realizações com correlação > 0,7. O perfil do segundo track exibe o desvio padrão entre as simulações. Também é mostrado o crossplot entre impedâncias invertida e medidas, colorido pelo desvio padrão.

A avaliação quantitativa dos resultados da inversão estocástica é apresentada na Figura 9.37, que exibe o crossplot entre a impedância acústica dos poços e a impedância derivada da inversão estocástica, bem como uma tabela com os coeficientes de correlação individuais por poço e o valor médio obtido. Ao comparar esses resultados com os da inversão determinística, observa-se que apenas para o poço OGX-16 o coeficiente de correlação da inversão determinística (0,94) superou o valor

obtido pela inversão estocástica (0,81). A diferença percentual entre as médias dos coeficientes de correlação dos dois métodos é pequena, em torno de 0,03, indicando um desempenho quantitativo semelhante entre ambas as abordagens.

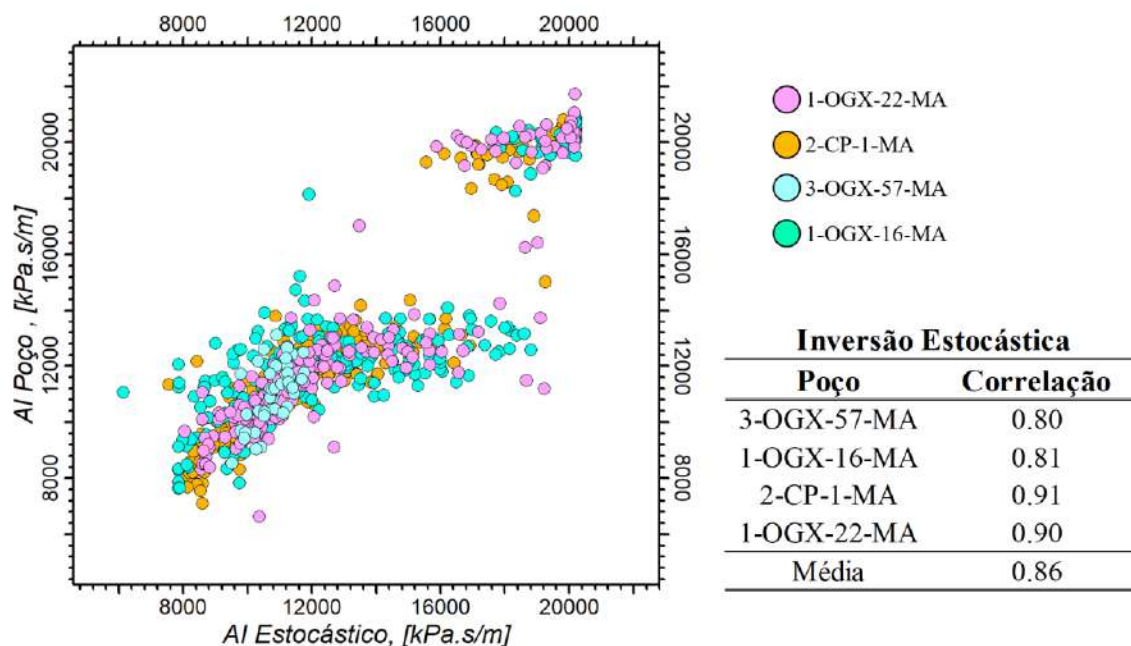


Figura 9.37. Avaliação quantitativa da inversão estocástica: *crossplot* entre impedâncias acústicas de poços e volume invertido, com tabela de coeficientes de correlação individuais e valor médio.

Ainda que as médias de correlação obtidas pelas inversões determinística e estocástica sejam próximas, o comportamento individual por poço revela diferenças importantes. Conforme ilustrado nas Figura 9.38 e Figura 9.39, os valores de impedância acústica estimados pela inversão estocástica apresentam maior correspondência com os valores medidos nos poços, especialmente nos extremos da distribuição, que são honrados de forma mais consistente. Diferentemente, a inversão determinística mostra limitações em reproduzir os valores máximos de impedância associados às soleiras de diabásio, bem como os valores mínimos correspondentes aos folhelhos, em determinados intervalos estratigráficos.

Outra diferença notável, evidenciada nas mesmas figuras, refere-se ao ganho de resolução vertical obtido com a inversão estocástica quando comparada à determinística. Esse aumento de detalhamento, contudo, vem acompanhado da incorporação de um efeito pepita moderado, resultante da parametrização do variograma e das condições de discretização do grid 3D utilizado nas simulações de 50 x 50 m (x, y) e 4 ms (z). Esse efeito é esperado, uma vez que a simulação condicional estocástica

introduz variabilidade aleatória controlada para preservar a estatística espacial definida pelos variogramas e, ao mesmo tempo, honrar a resposta sísmica e os dados de poço.

Do ponto de vista geológico e interpretativo, o ganho de resolução vertical proporcionado pela inversão estocástica tem implicações diretas para a caracterização interna da Formação Pimenteiras, especialmente no reconhecimento das camadas de folhelhos ricos em COT e na delimitação das intrusões ígneas que cortam a sequência. A melhor preservação dos valores extremos de impedância indica que o método foi capaz de representar de forma mais realista os contrastes acústicos entre os folhelhos de baixa impedância e as soleiras de diabásio de alta impedância, relações que a inversão determinística tende a suavizar.

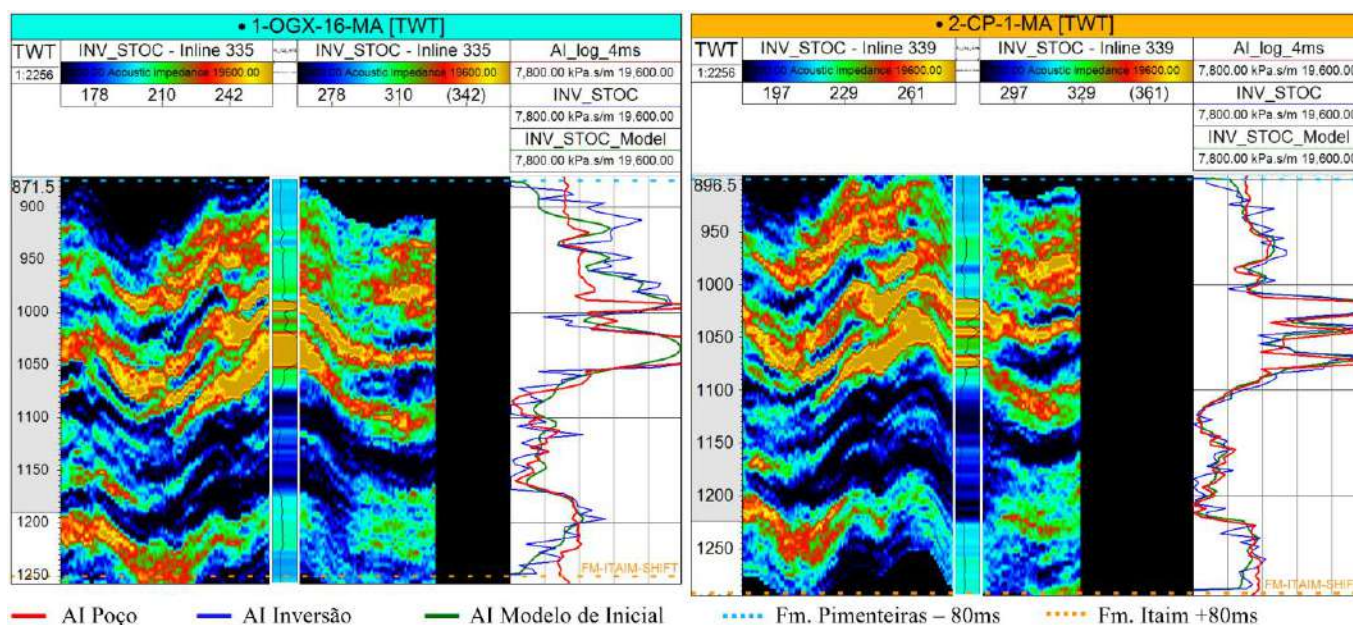


Figura 9.38. Seções sísmicas mostrando o resultado da inversão estocástica, e a comparação entre as curvas de impedância reamostradas dos poços com a inversão estocástica e o modelo inicial utilizado. Nota-se maior detalhamento vertical e preservação dos contrastes de impedância nas camadas finas para a solução estocástica, refletindo o ganho de resolução obtido.

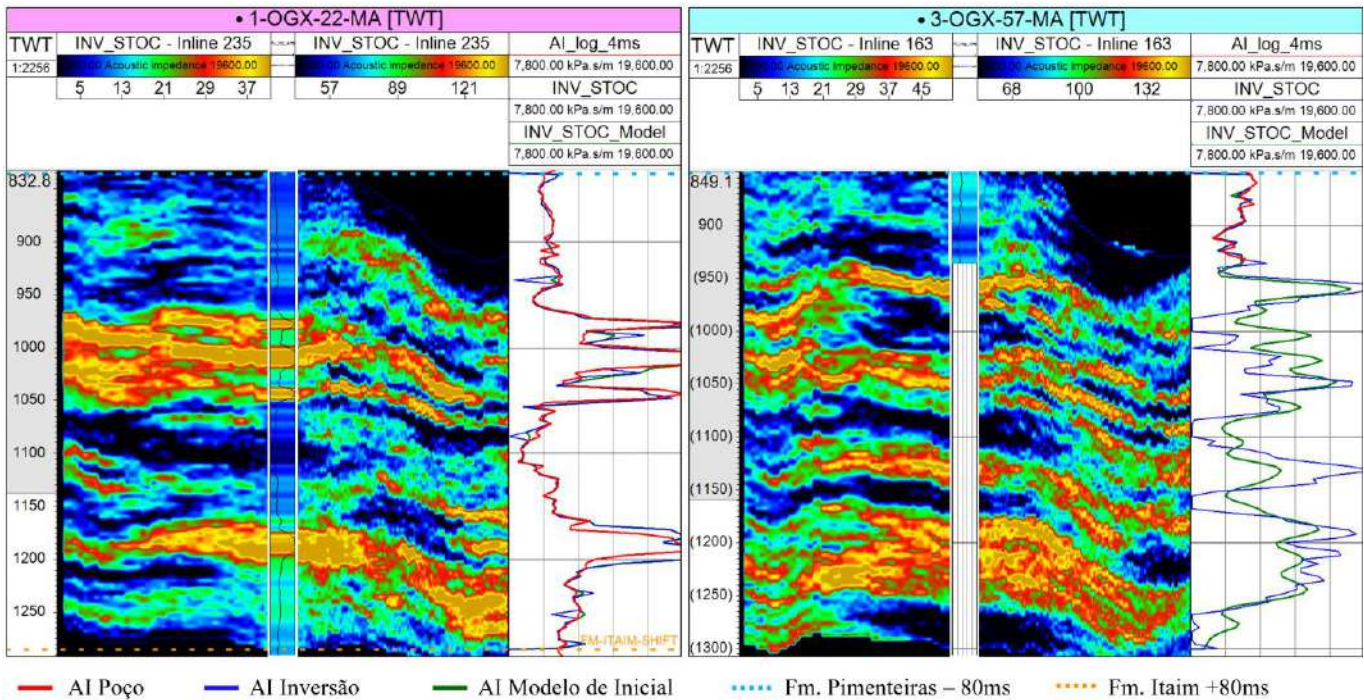


Figura 9.39. Seções sísmicas mostrando o resultado da inversão estocástica, e a comparação entre as curvas de impedância reamostradas dos poços com a inversão estocástica e o modelo inicial utilizado. Nota-se maior detalhamento vertical e preservação dos contrastes de impedância nas camadas finas para a solução estocástica, refletindo o ganho de resolução obtido.

A mesma avaliação pode ser estendida ao volume médio obtido a partir das realizações consideradas ótimas da inversão estocástica. Nesse caso, o coeficiente de correlação entre a média das realizações e a impedância acústica do poço teste apresentou um valor ligeiramente superior ao da simulação individual de maior correlação entre as cinquenta geradas, atingindo 0,83 (Figura 9.40).

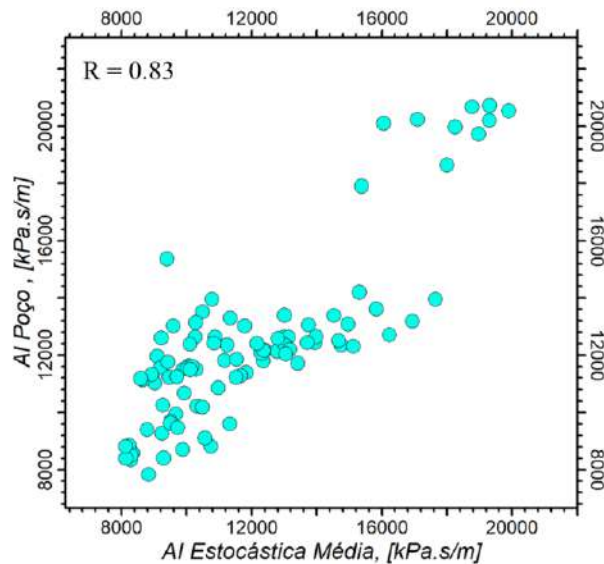


Figura 9.40. Crossplot e perfis de impedância mostrando a correlação entre a média das realizações estocásticas e a impedância acústica medida no poço teste (OGX-16)

Nas Figura 9.41 e Figura 9.42, é apresentado o controle de qualidade correspondente, onde se observa uma alta correlação geral na posição dos poços, evidenciando a consistência espacial do resultado médio. Além disso, nota-se que o volume médio apresenta um aspecto mais suavizado quando comparado às realizações individuais, refletindo o efeito de atenuação das variações locais promovido pela média estatística das simulações estocásticas.

Esse comportamento é esperado, pois o processo de média entre as realizações estocásticas tende a reduzir a variabilidade aleatória local introduzida pelo termo estocástico, resultando em um modelo mais estável e de maior correlação global com os dados observados. Em termos estatísticos, a média das realizações atua como um filtro de ruído espacial, suprimindo as flutuações de alta frequência que não são sistematicamente reproduzidas entre as simulações independentes (A. G., & Huijbregts, C. H. J., 1978).

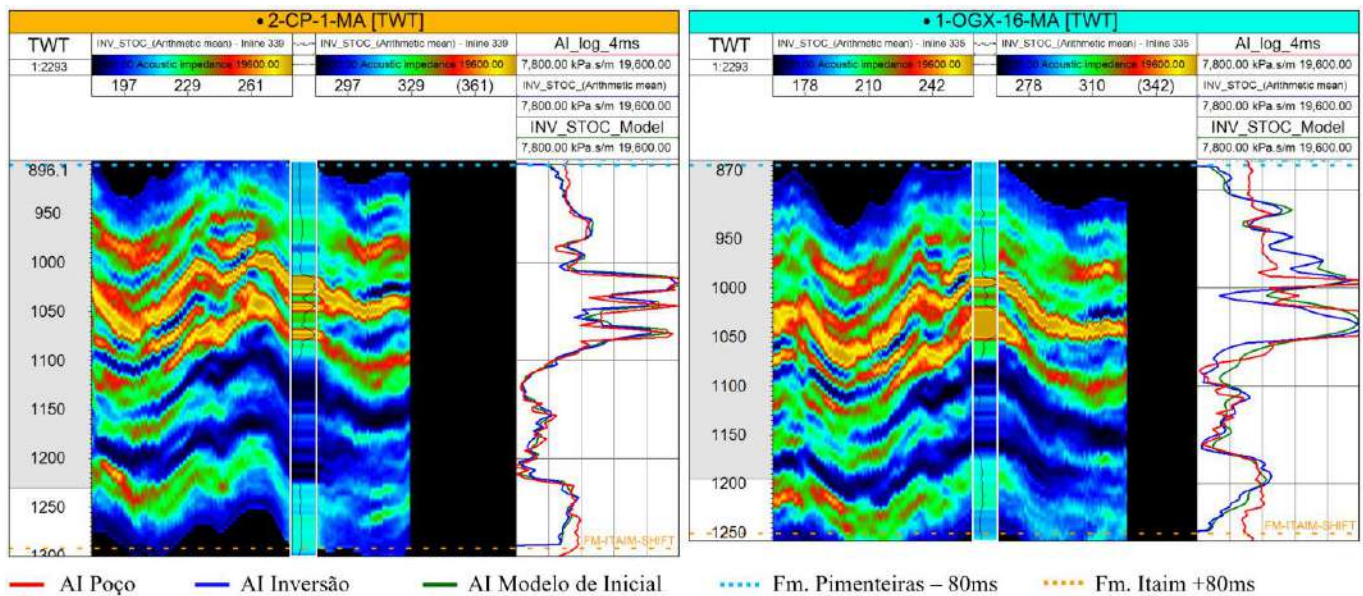


Figura 9.41. Controle de qualidade do volume médio da inversão estocástica, destacando a elevada correlação nas posições dos poços e a suavização do modelo médio em comparação às realizações individuais.

Consequentemente, o volume médio de impedância mantém a tendência estrutural e estratigráfica dominante do conjunto de simulações, apresentando valores suavizados, porém fisicamente coerentes, com os intervalos de maior continuidade geológica. Embora essa suavização implique uma redução da resolução local, ela também confere maior robustez estatística ao resultado, uma vez que minimiza os efeitos de anomalias isoladas e incertezas condicionais associadas à amostragem e ao modelo de variograma.

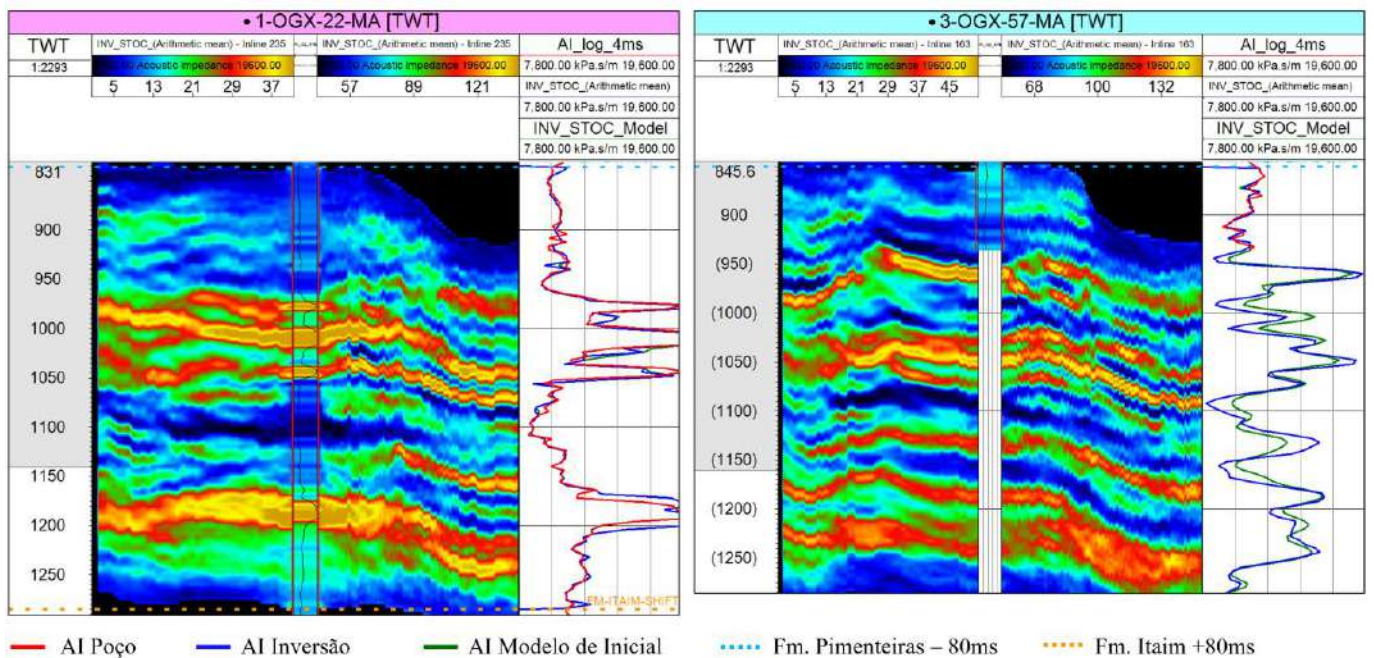


Figura 9.42. Comparação entre as realizações individuais e o volume médio da inversão estocástica, mostrando a preservação das tendências estruturais e o aumento da suavidade estatística no modelo médio.

Os resultados da inversão estocástica de maior correlação, da média das realizações de melhor desempenho e do desvio padrão são apresentados em maior detalhe nas Figura 9.43 e Figura 9.44, respectivamente, em cortes tridimensionais que permitem uma visão espacial integrada da distribuição de impedância acústica. Esses resultados evidenciam as amplas continuidades das intrusões ígneas, caracterizadas por altos valores de impedância acústica, e reforçam a heterogeneidade interna dos folhelhos pertencentes ao intervalo da Formação Pimenteiras, conforme já demonstrado no estudo de classificação de fácies apresentado no Capítulo 4.

A grande variação de impedância observada nessa formação reflete as diferenças composicionais entre as fácies, as quais estão associadas a distintos níveis de concentração de carbono orgânico total (COT), conforme discutido no estudo de estimativa de COT desenvolvido no Capítulo 5. Esses resultados destacam a importância da integração entre a inversão sísmica e as análises geoquímicas, fornecendo uma base interpretativa robusta para a discussão aprofundada que será conduzida no Capítulo 10, onde serão exploradas as implicações geológicas e petrofísicas dessas variações de impedância e sua relação com o potencial gerador da Formação Pimenteiras.

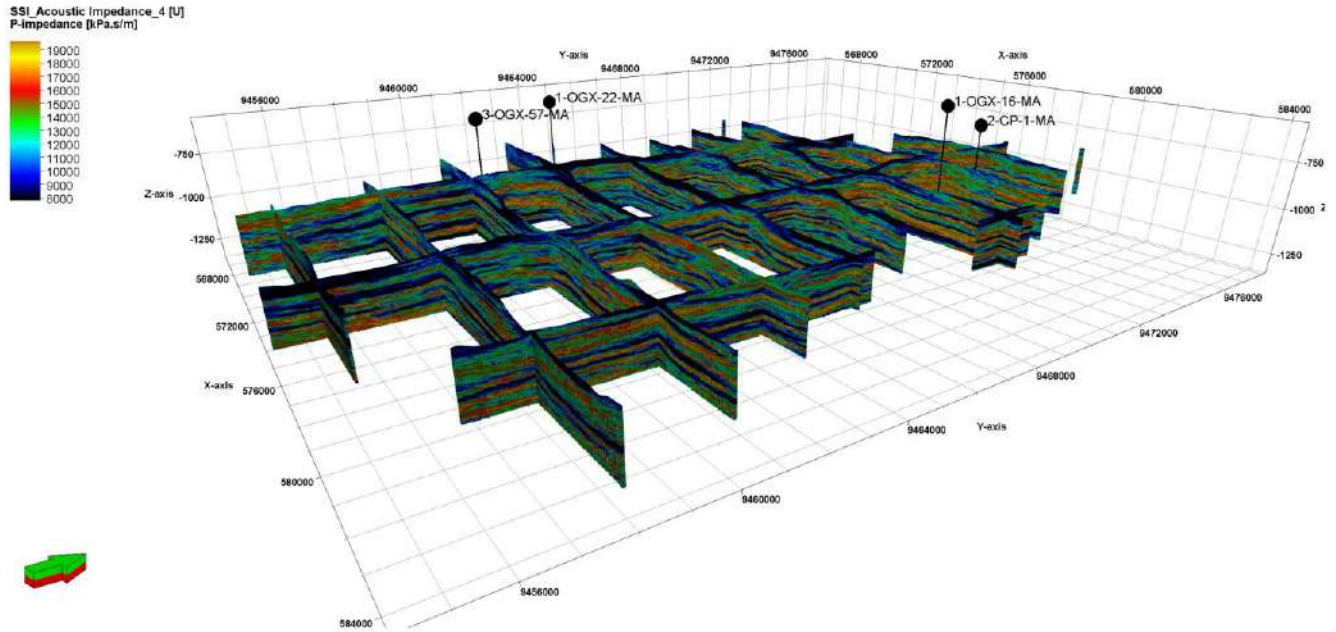


Figura 9.43. Resultado tridimensional da inversão estocástica de maior correlação, mostrando a continuidade lateral das intrusões ígneas e os contrastes de impedância acústica na Formação Pimenteiras.

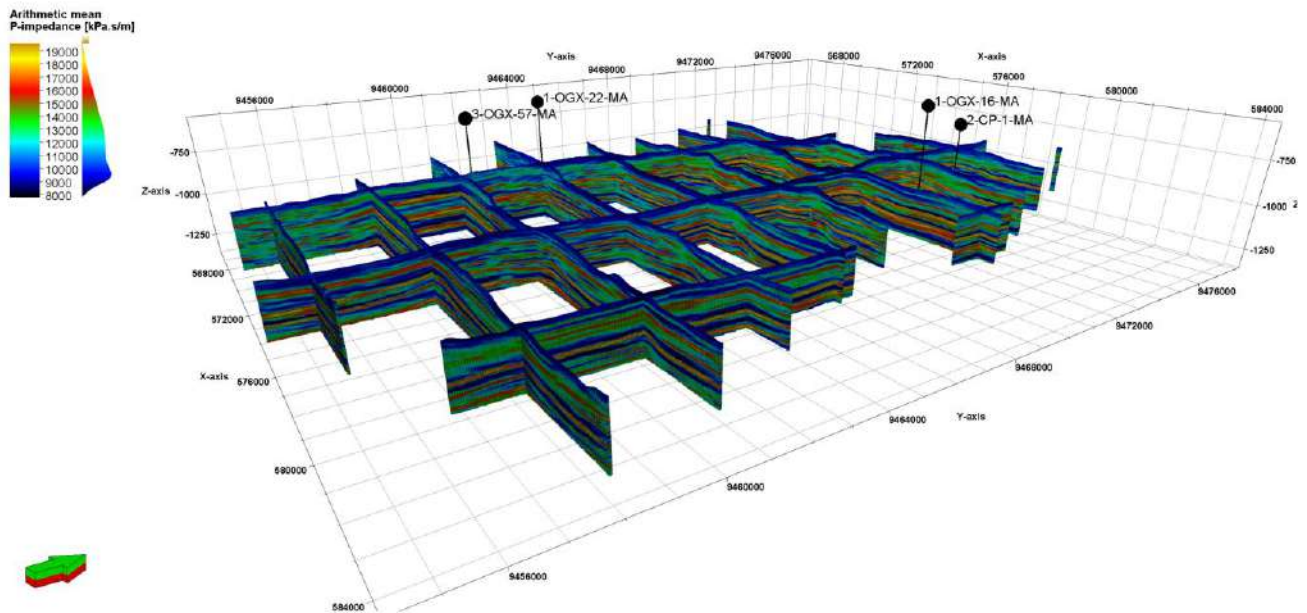


Figura 9.44. Volume médio das realizações da inversão estocástica, evidenciando a coerência espacial e a atenuação das flutuações locais de impedância.

A Figura 9.45 evidencia que, nas proximidades dos poços, os valores de desvio padrão são menores, como esperado em função do condicionamento direto aos dados de poço, variando até aproximadamente $800 \text{ m/s} \cdot \text{g/cm}^3$. À medida que se afasta dos poços, o desvio padrão aumenta gradualmente, alcançando valores entre 800 e $2000 \text{ m/s} \cdot \text{g/cm}^3$, e chegando a valores superiores

principalmente na porção superior da Formação Pimenteiras. Nessa região, observam-se os maiores valores de impedância acústica, associados predominantemente às soleiras de diabásio, cuja presença acentua a heterogeneidade acústica e estrutural do intervalo. Essa distribuição espacial do desvio padrão reflete o grau de incerteza local da inversão estocástica, as regiões com menor desvio indicam maior consistência entre as realizações e maior confiabilidade do modelo de impedância, correspondendo às zonas bem condicionadas por dados de poço e maior continuidade estratigráfica. Por outro lado, os altos desvios padrão observados nas áreas mais afastadas dos poços e nas porções ígneas superiores representam maior incerteza associada à ausência de controle direto e às variações abruptas de propriedades elásticas impostas pelas intrusões de diabásio.

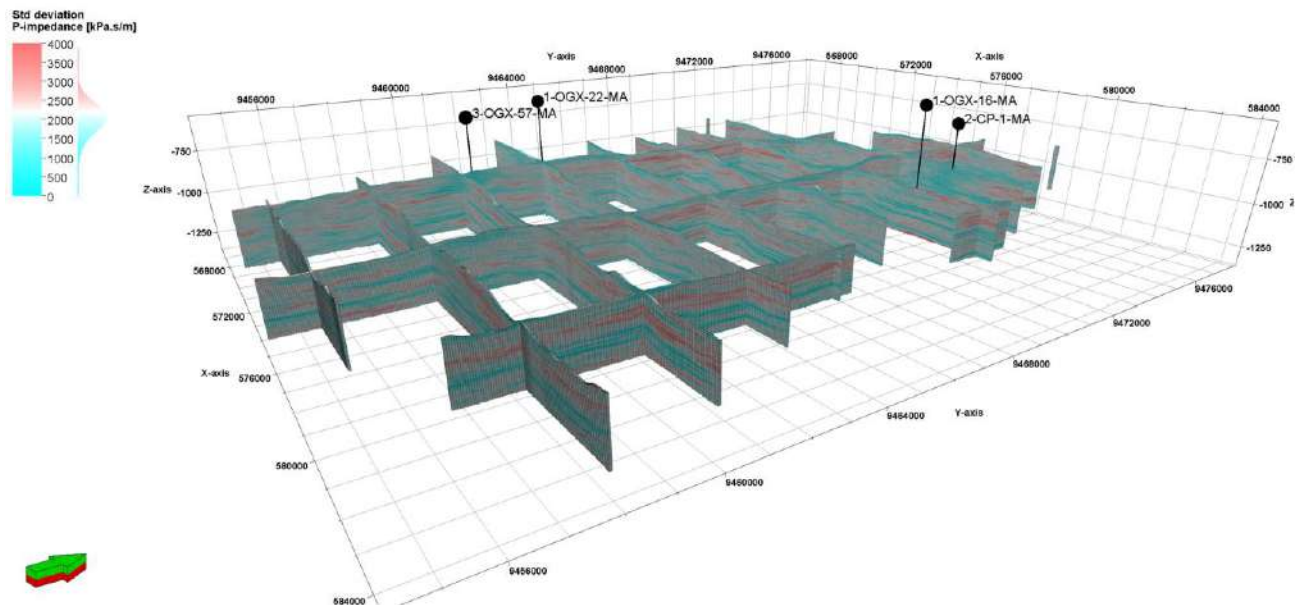


Figura 9.45. Volume do desvio padrão das realizações estocásticas, mostrando menores valores próximos aos poços e aumento progressivo em áreas distantes, especialmente na porção superior da formação.

Os resultados da inversão estocástica são apresentados em maior detalhe nas Figura 9.46, Figura 9.47, Figura 9.48, Figura 9.49, por meio de seções sísmicas (*inlines*) que cruzam os poços da área de estudo. Essas seções evidenciam o ganho de resolução vertical proporcionado pela abordagem estocástica, que se traduz em contrastes mais acentuados de impedância acústica nos folhelhos da Formação Pimenteiras, quando comparados aos resultados obtidos pela inversão determinística.

Esses contrastes não apenas refletem as fácies identificadas nos poços, mas também podem estar relacionados à combinação de parassequências sedimentares que se formaram durante a evolução deposicional da unidade. Dessa forma, os resultados reforçam a capacidade da inversão estocástica em

capturar variações de alta frequência e heterogeneidades estratigráficas sutis, fundamentais para a interpretação geológica detalhada da Formação Pimenteiras.

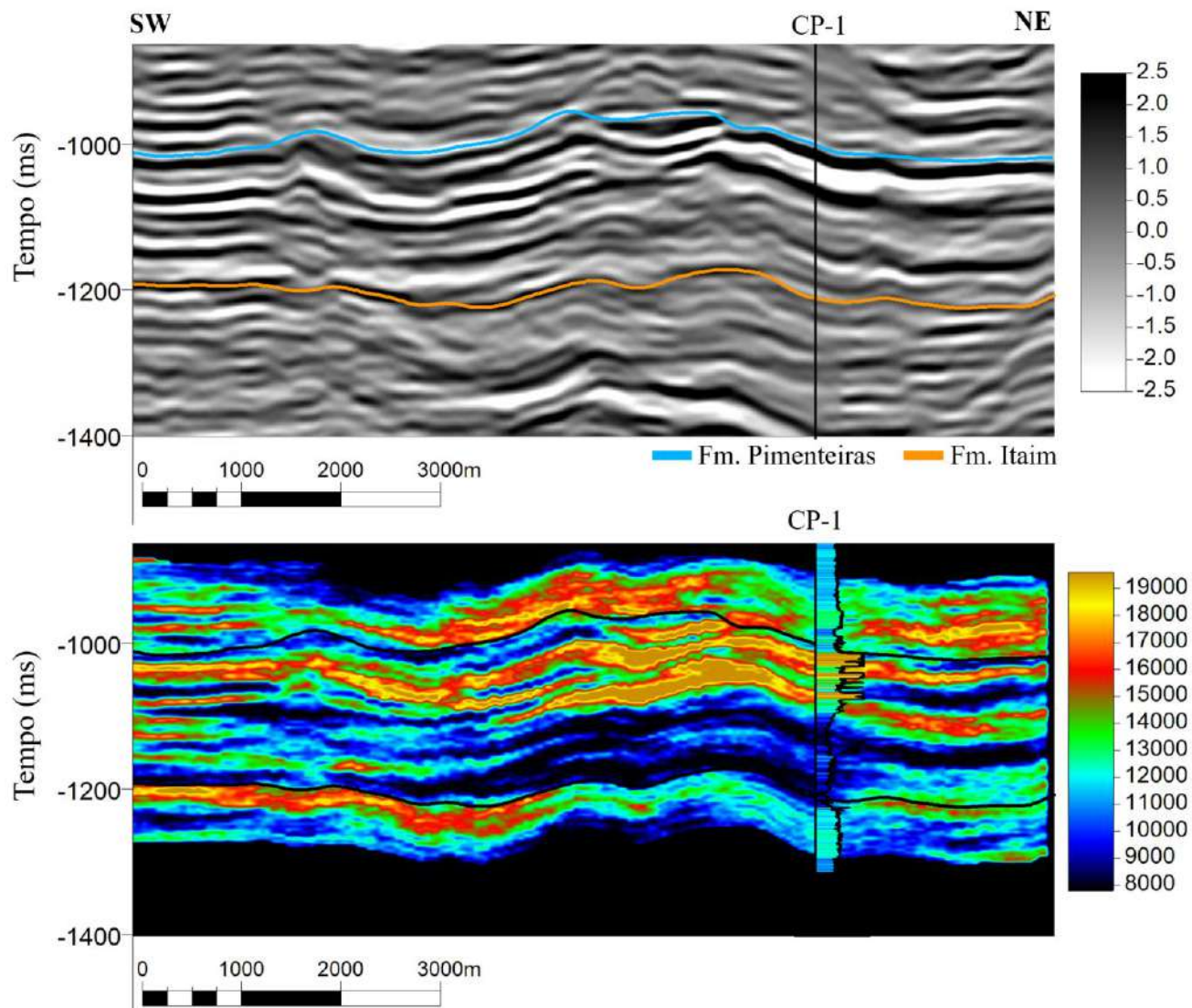


Figura 9.46. Seção *inline* atravessando o poço CP-1, mostrando os resultados da inversão estocástica com resolução aprimorada, ressaltando contrastes de impedância mais definidos em comparação à inversão determinística.

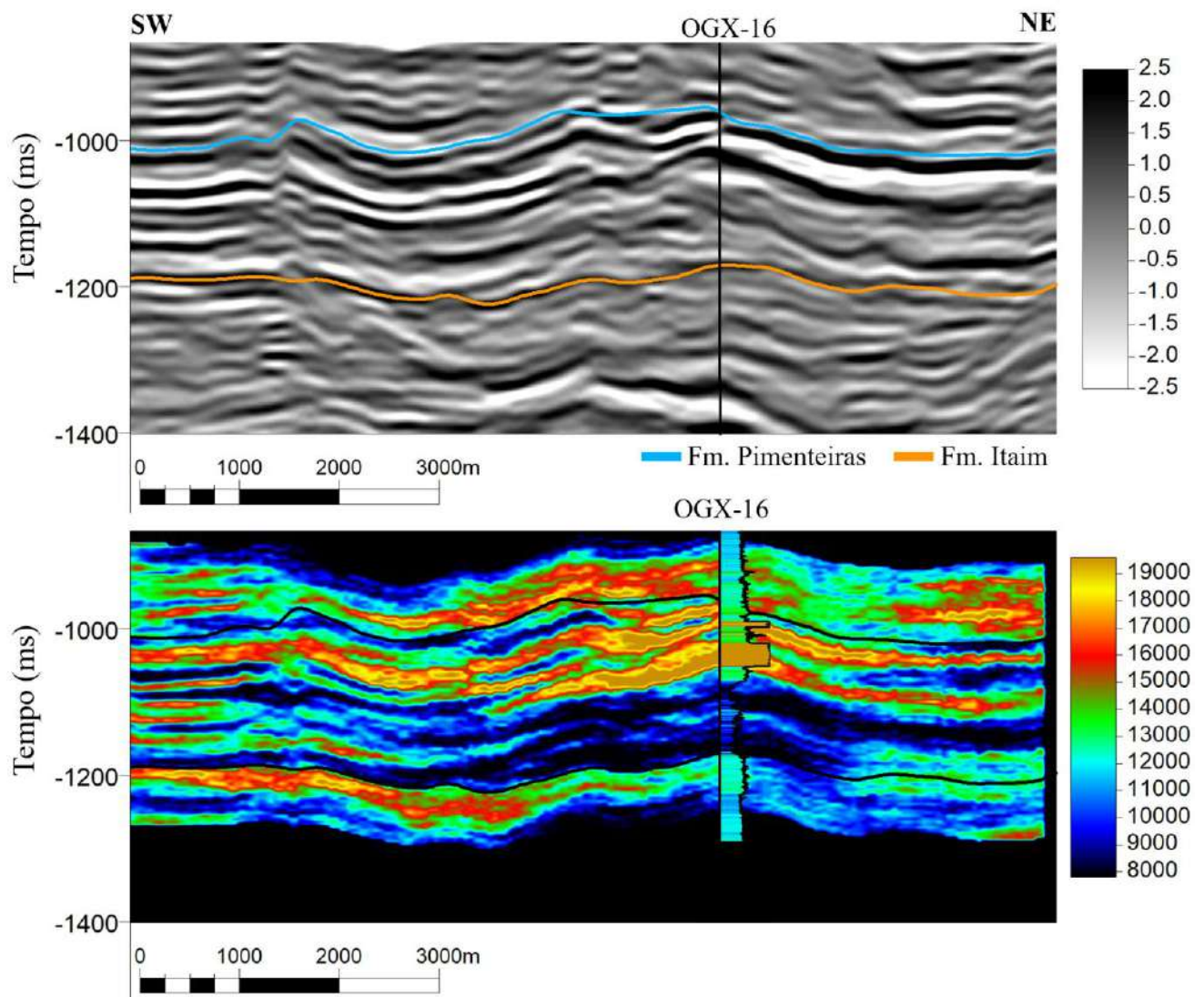


Figura 9.47. *Inline* passando pelo poço OGX-16 destacando a variação vertical e lateral da impedância acústica estimada pela inversão estocástica, ilustrando o ganho de detalhamento e a coerência com as fácies identificadas nos poços.

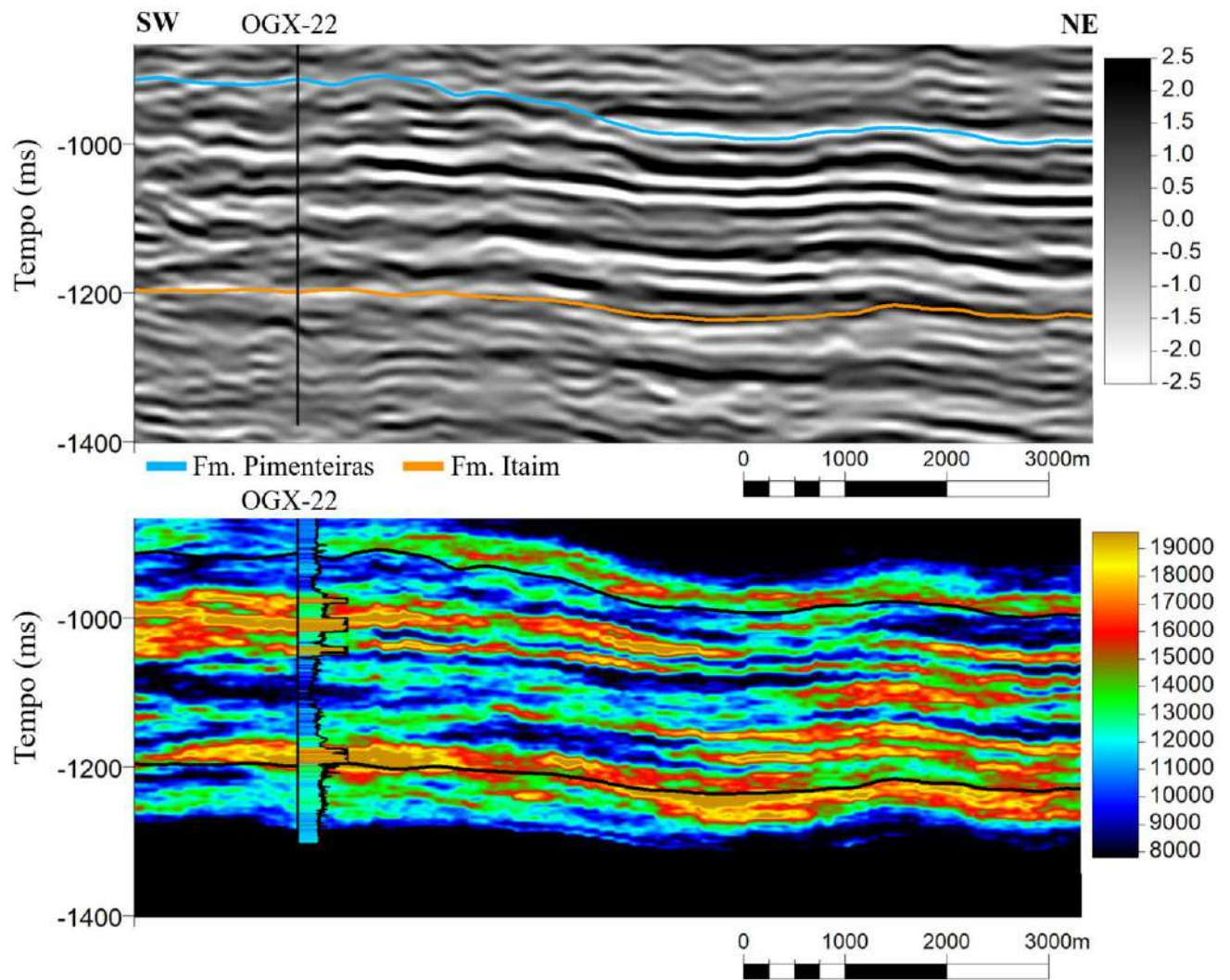


Figura 9.48. *Inline* passando pelo poço OGX-22 evidenciando o maior poder de resolução e a preservação de camadas delgadas no resultado estocástico.

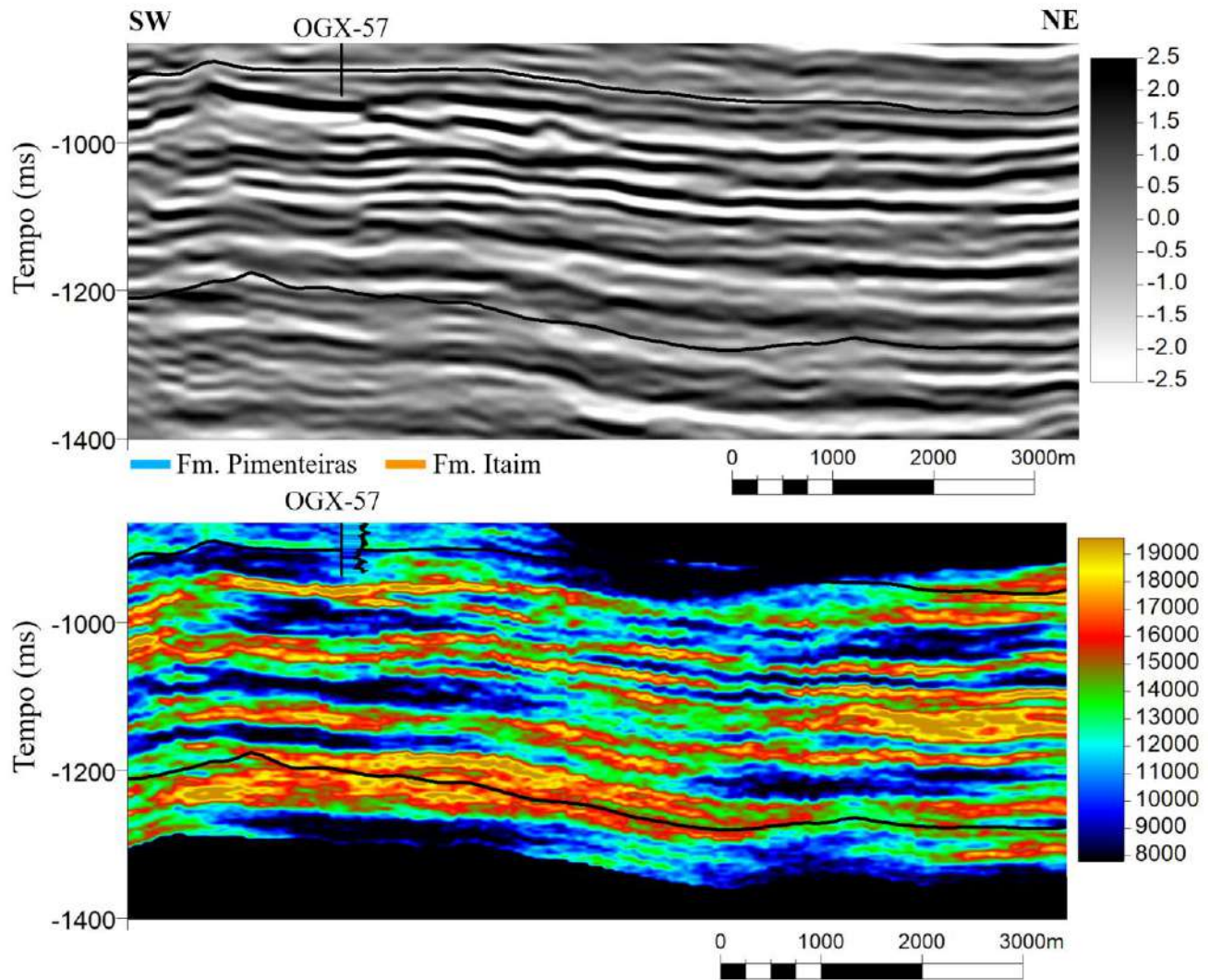


Figura 9.49. *Inline* passando pelo poço OGX-57 destacando a variação vertical e lateral da impedância acústica estimada pela inversão estocástica, ilustrando o ganho de detalhamento e a coerência com as fácies identificadas nos poços.

Por fim, com o objetivo de realizar uma comparação geral entre os resultados obtidos pelas inversões determinística e estocástica, tanto em termos qualitativos quanto quantitativos, são apresentadas na Figura 9.50 uma seção sísmica arbitrária atravessando os poços e um *time slice* correspondente, nos quais são exibidos os volumes de amplitude sísmica, inversão estocástica, média das realizações estocásticas e inversão determinística.

Do ponto de vista qualitativo, observa-se que a inversão estocástica apresenta maior capacidade de resolução vertical, permitindo distinguir com mais clareza os contrastes de camadas delgadas, enquanto a inversão determinística fornece um resultado mais suavizado. Entretanto, ao comparar a média das realizações estocásticas com a inversão determinística, nota-se o comportamento inverso: a inversão determinística evidencia maior detalhamento das camadas internas da formação, ao passo que

a média das realizações apresenta um aspecto mais homogêneo, decorrente da atenuação estatística das variações locais.

Sob uma perspectiva quantitativa, contudo, a média das realizações estocásticas demonstra maior compatibilidade com os valores de impedância medidos nos poços, refletindo sua robustez estatística e a capacidade de reproduzir de forma mais fidedigna as propriedades acústicas observadas nos poços. Assim, as duas abordagens se revelam complementares: enquanto a inversão estocástica privilegia a resolução e a variabilidade, a inversão determinística oferece estabilidade e coerência estrutural dos refletores sísmicos, e a combinação de ambas fornece uma interpretação sísmica mais completa e confiável da Formação Pimenteiras.

De modo geral, os resultados obtidos demonstram que a integração entre as diferentes abordagens de inversão sísmica, Colorida, Determinística e Estocástica, proporciona uma visão complementar e abrangente da distribuição de impedância acústica na Formação Pimenteiras. Essas diferenças metodológicas revelam não apenas as limitações e potencialidades de cada técnica, mas também a importância de uma abordagem integrada para a interpretação sísmica e geológica do intervalo. A comparação entre os métodos indica que a combinação entre previsibilidade determinística e variabilidade estocástica constitui uma estratégia eficaz para capturar o comportamento acústico complexo da Formação Pimenteiras, fortemente influenciado pela presença de soleiras ígneas e pela alternância de fácies sedimentares ricas em matéria orgânica.

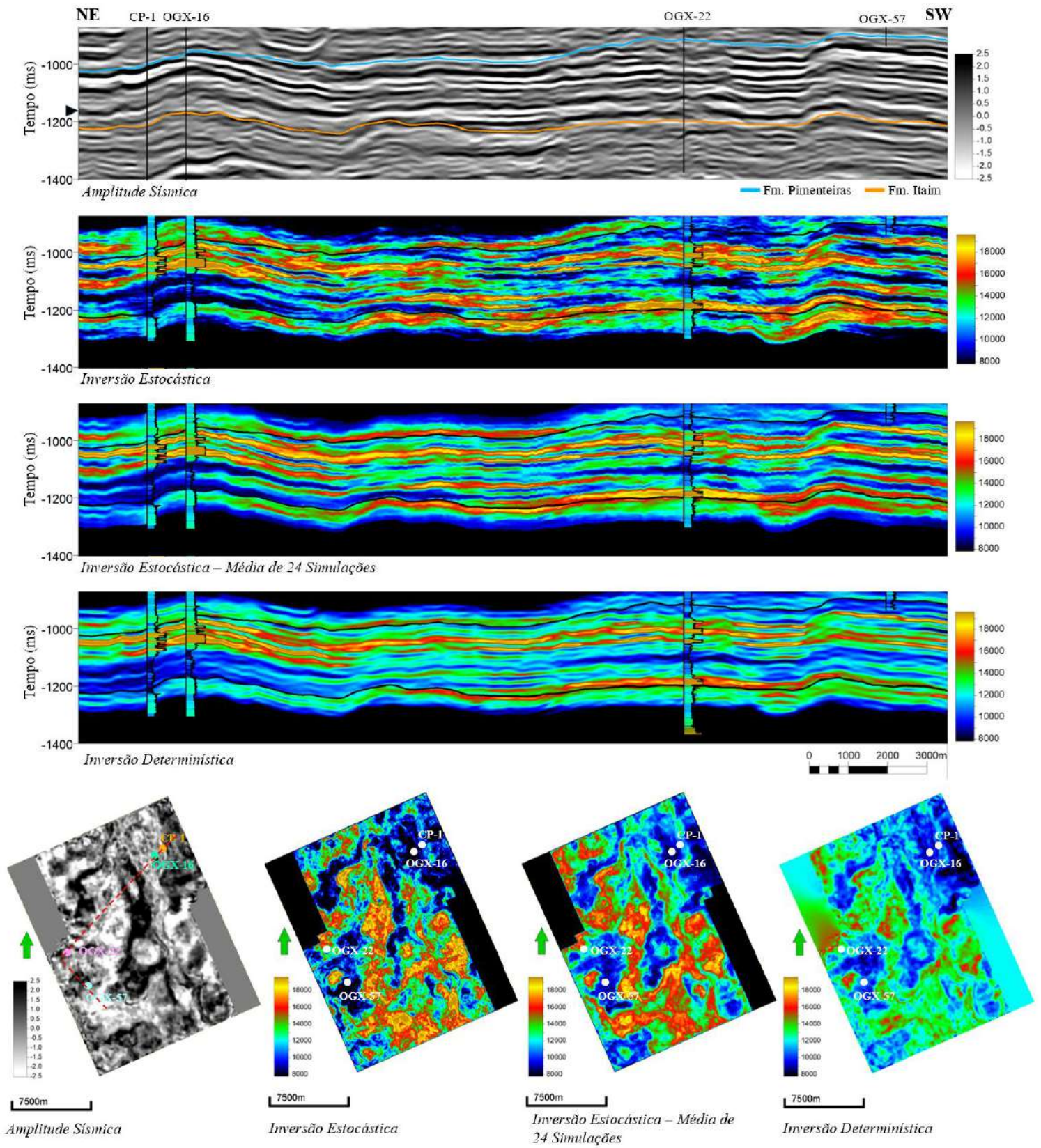


Figura 9.50. Comparação integrada entre amplitude sísmica, inversão estocástica, média das realizações e inversão determinística. A inversão estocástica apresenta maior capacidade de resolver contrastes de camadas finas, enquanto a determinística produz um volume

mais suavizado e geologicamente contínuo. Em termos quantitativos, a média das realizações mostra valores de impedância mais compatíveis com os medidos nos poços.

9.5. DISCUSSÕES

Desde sua proposição, a CI tem sido aplicada e expandida em diferentes contextos. Yang et al. (2013) demonstraram sua eficácia na resolução de camadas de carvão delgadas, tradicionalmente abaixo do limite de resolução sísmica. Nesse estudo, a inversão colorida foi combinada ao *spectral blueing* (assim como realizado no presente trabalho, e mostrado no Capítulo 6), técnica que desloca a energia espectral para frequências mais altas. Essa combinação permitiu recuperar detalhes estratigráficos de camadas de carvão muito finas, evidenciando que o *spectral blueing* aliada à CI amplia o poder de detecção de interfaces litológicas sutis em reservatórios complexos (Yang et al. 2013).

De forma semelhante, Blache-Fraser e Neep (2004) aplicaram o método no campo Cannonball, em Trinidad, também associado ao *spectral blueing*. Nesse caso, a aplicação elevou a frequência dominante dos dados sísmicos em aproximadamente 20 Hz, resultando em melhor separação de unidades arenosas, identificação de canais incisos e definição mais clara do contato gás-água. A CI, ao converter os dados sísmicos de banda limitada em impedância relativa, mostrou-se essencial para melhorar a correlação sísmica com poços e reduzir incertezas na modelagem do reservatório.

Além de sua aplicação em arenitos e carvões, a técnica também mostrou relevância em reservatórios carbonáticos. Ansari (2014) utilizou a inversão sísmica colorida para derivar impedância relativa em um campo heterogêneo, avaliando sua correlação com a porosidade. O estudo evidenciou que a impedância derivada pela CI apresenta correlação significativamente superior com a porosidade em relação à obtida por métodos tradicionais, como a *Generalized Linear Inversion*. Essa melhora deve-se ao fato de a CI ressaltar contrastes locais de impedância, eliminando parcialmente efeitos globais de heterogeneidade.

A aplicação sequencial das inversões determinística e estocástica, apoiada pelo pré-condicionamento do dado sísmico obtido na etapa anterior, resultou em volumes de impedância acústica com elevado grau de consistência geológica e correlação quantitativa com os dados de poço. Essa abordagem metodológica segue uma lógica semelhante à proposta por Maulana et al. (2022) que em seu trabalho também foram aplicados os métodos, *Spectral Blueing*, Inversão Determinística *model-based* e Inversão Estocástica, que demonstrou ganhos expressivos de resolução e fidelidade geológica em ambientes flúvio-deltaicos complexos.

De forma análoga ao estudo de Maulana e colaboradores, a inversão determinística model-based aplicada à Formação Pimenteiras mostrou-se fundamental para estabelecer uma base quantitativa coerente a partir da qual as variações de impedância puderam ser calibradas e interpretadas em termos de heterogeneidade litológica e potencial gerador, além de funcionar como modelo inicial para a inversão estocástica. Tanto o modelo de baixa frequência quanto o modelo inicial construídos com suporte da inversão colorida e da inversão determinística e controle do poço garantiu a continuidade geológica e minimizou distorções de tendência, resultando em valores de correlação cruzada acima de 0,7, comparáveis aos obtidos por del Monte et al. (2018) em sua aplicação à Formação Tanezzuft, na Tunísia.

del Monte et al. (2018) e Ouadfeul e Aliouane (2016), descreveram em seus trabalhos as relações encontradas entre a impedância acústica reduzida como uma assinatura elástica típica de rochas geradoras com alto teor de querogênio. Essa relação, de natureza inversa, decorre das propriedades intrínsecas do material orgânico, caracterizado por baixa densidade e menor velocidade de propagação da onda P, fatores que reduzem a impedância total do meio rochoso. Nos volumes obtidos no presente trabalho, as zonas de baixa impedância apresentaram correlação direta com os intervalos de folhelhos mais ricos em matéria orgânica, evidenciando esse mesmo comportamento e que será o tema central e descrito mais profundamente no próximo capítulo.

Entretanto, observou-se que a inversão determinística isolada apresentou limitações na representação da variabilidade acústica e da heterogeneidade espacial, principalmente nas zonas de maior contraste sísmico (valores extremos de impedância associados às soleiras de diabásio e aos folhelhos altamente orgânicos). Comportamento semelhante é descrito por Maulana et al. (2022) e Cyz & Azevedo (2020), que destacam que os métodos determinísticos tradicionais, embora forneçam modelos consistentes, tendem a suavizar as respostas e não capturam adequadamente as variações finas e a incerteza geológica.

A inversão estocástica, por sua vez, mostrou-se eficaz em superar essas limitações, incorporando a variabilidade geológica e o controle estatístico dos variogramas. Conforme demonstrado por Cyz e Azevedo (2020), esse tipo de inversão geostatística permite gerar múltiplas realizações que honram simultaneamente a sísmica e os dados de poço, produzindo um conjunto de modelos equiprováveis que representam a incerteza natural do sistema deposicional. Na Formação Pimenteiras, a aplicação dessa metodologia resultou em um melhor ajuste aos valores de impedância medidos, especialmente nas zonas de folhelhos finos e nas interfaces ígneas, além de evidenciar uma melhoria perceptível na resolução vertical e na capacidade de distinção de camadas delgadas,

comportamento também observado no estudo de Maulana et al. (2022) após a incorporação do *Spectral Blueing*.

O conjunto das realizações estocásticas apresentou baixa dispersão nas regiões próximas aos poços e maior variabilidade nas porções menos amostradas, refletindo de maneira realista a distribuição da incerteza espacial. Essa característica é consistente com a abordagem de Cyz e Azevedo (2020) e reforça o papel da inversão estocástica como ferramenta para quantificar a confiança dos resultados de impedância e preparar o terreno para a modelagem de propriedades derivadas, como o COT.

A discussão dos resultados também deve considerar o contexto apresentado por Thiéblemont et al. (2022), que destaca a complexidade das relações entre propriedades elásticas e conteúdo orgânico em rochas geradoras. Esses autores demonstraram que a impedância acústica absoluta pode não ser um parâmetro discriminante universal para o COT, especialmente em teores abaixo de 4–5 wt%, e que a razão de Poisson e a mineralogia local exercem papel fundamental na resposta elástica. Os resultados da Formação Pimenteiras corroboram essa visão: embora a tendência geral de baixa impedância associada a altos teores de COT seja evidente, variações locais relacionadas à presença de rochas ígneas, devido ao contato térmico, podem gerar alterações nas propriedades elásticas que devem ser levadas em consideração.

Além disso, a integração entre inversões determinística e estocástica permite abordar precisamente essa variabilidade. A primeira fornece um modelo físico estável e calibrado, enquanto a segunda incorpora a dispersão natural das propriedades elásticas e a heterogeneidade deposicional, aspectos críticos para a modelagem geoquímica subsequente. Assim como proposto por Ouadfeul e Aliouane (2016), a impedância acústica resultante dessas inversões pode ser transformada em mapas contínuos de COT, calibrados com dados de poço e respaldados pela consistência sísmica. Essa abordagem, que será detalhada no próximo capítulo, constitui a etapa final do fluxo de trabalho integrado, no qual a inversão sísmica atua como elo quantitativo entre a geofísica e a geoquímica.

9.6. CONCLUSÕES

A integração sequencial entre as técnicas de inversão sísmica colorida, determinística e estocástica demonstrou ser fundamental para o aprimoramento da resolução e da consistência geológica dos modelos de impedância acústica aplicados à Formação Pimenteiras. A inversão colorida, associada ao *spectral blueing*, proporcionou o condicionamento espectral necessário para expandir a faixa de frequências da sísmica e recuperar detalhes de alta resolução espacial, estabelecendo a base espectral adequada para a inversão *model-based*.

A inversão determinística *model-based*, apoiada por modelos de baixa frequência geologicamente coerentes, produziu volumes quantitativos de impedância com boa aderência aos dados de poço e elevada estabilidade numérica. No entanto, observou-se que, por sua natureza suavizadora, o método apresenta limitações na reprodução de valores extremos de impedância, especialmente nas soleiras de diabásio e nos folhelhos de mais baixa impedância que podem está associados aos níveis mais ricos em matéria orgânica.

A subsequente inversão estocástica, empregando os resultados determinísticos como tendência, incorporou de forma mais realista a variabilidade espacial e as incertezas associadas à heterogeneidade deposicional. O conjunto de realizações estocásticas produziu volumes com maior fidelidade aos dados de poço, permitindo a preservação de contrastes acústicos locais e a quantificação explícita da incerteza por meio da análise do desvio padrão.

Do ponto de vista geológico, os resultados destacam o potencial da inversão estocástica para caracterizar de maneira mais refinada as descontinuidades associadas às intrusões ígneas e para discriminar as variações internas dos folhelhos da Formação Pimenteiras, que potencialmente reflete às diferenças composicionais e variações no teor de carbono orgânico total. Essa resposta acústica heterogênea, quando analisada em conjunto com dados geoquímicos, fornece uma base robusta para o estudo das relações entre impedância e COT, que será aprofundado no capítulo subsequente.

Por fim, o fluxo de trabalho desenvolvido, integrando inversão colorida, inversão determinística e inversão estocástica, consolidou-se como uma abordagem eficiente para representar simultaneamente a continuidade estratigráfica e a variabilidade local das propriedades elásticas, oferecendo um modelo de subsuperfície quantitativo e geologicamente consistente, aplicável a estudos de caracterização de rochas geradoras e de sistemas petrolíferos complexos.

10. Modelagem 3D do Carbono Orgânico Total

10.1. INTRODUÇÃO

Ao longo das últimas décadas, inúmeros autores têm se dedicado a compreender a resposta sísmica de rochas geradoras e a maneira como suas propriedades petrofísicas, mineralógicas e orgânicas influenciam a impedância acústica e o comportamento das amplitudes sísmicas. Esses trabalhos consolidaram uma base conceitual e metodológica que fundamenta o uso de dados sísmicos quantitativos na caracterização indireta do carbono orgânico total (COT) em rochas geradora.

Estudos pioneiros, como o de Vernik e Nur (1992), revelaram que folhelhos negros ricos em querogênio apresentam densidade e velocidades ultrassônicas extremamente baixas quando medidas perpendicularmente às laminações, refletindo uma forte anisotropia elástica. Esse comportamento foi posteriormente detalhado por Vernik e Liu (1997), que demonstraram que a anisotropia aumenta com a compactação e o conteúdo de querogênio. Essas observações indicam que o querogênio exerce influência não apenas composicional, mas também estrutural e mecânica sobre as propriedades elásticas e, conseqüentemente, sobre a resposta sísmica.

Em paralelo, Sayers (2013) investigou o efeito direto do querogênio na resposta de amplitude com offset (AVO) em folhelhos ricos em matéria orgânica, demonstrando que a presença de querogênio reduz a impedância acústica e origina uma resposta AVO típica de Classe IV, com amplitudes negativas que decrescem com o aumento do ângulo de incidência. Esse efeito é reforçado pela anisotropia intrínseca das rochas, decorrente da orientação preferencial do material orgânico e dos minerais laminares.

De forma complementar, Løseth et al. (2011) propuseram uma abordagem inovadora para a identificação e mapeamento de rochas geradoras espessas utilizando exclusivamente dados sísmicos. Os autores demonstraram que a impedância acústica decresce de forma não linear com o aumento do COT e que as interfaces superior e inferior dessas camadas exibem amplitudes características (negativa no topo e positiva na base), que podem ser calibradas com dados de poço para estimar

quantitativamente o teor orgânico. Essa relação entre impedância e COT possibilitou mapear lateralmente zonas ricas em matéria orgânica, ampliando o potencial da sísmica na avaliação do sistema petrolífero.

Carcione e Avseth (2015), por sua vez, integraram modelagem físico-rocha e evolução geoquímica para gerar *templates* elásticos que correlacionam o COT, a saturação de fluidos e os parâmetros elásticos em rochas argilosas. Aplicando médias de Backus e a equação de Gassmann generalizada para meios anisotrópicos, os autores analisaram o comportamento sísmico de rochas geradoras em diferentes estágios de maturação, destacando assinaturas sísmicas diagnósticas relacionadas à geração de hidrocarbonetos.

Nos anos seguintes, o avanço das técnicas de inversão sísmica e de aprendizado de máquina consolidou novas abordagens para a estimativa direta do COT em escala sísmica. Ouadfeul e Aliouane (2016) aplicaram um processo de inversão genética pós-empilhamento baseado em uma rede neural *Multi-Layer Perceptron* (MLP) na formação Barnett Shale, utilizando dados sísmicos 3D e impedância acústica derivada de poços horizontais para treinar o modelo. A partir da impedância invertida, os autores estabeleceram uma relação linear com o teor de carbono orgânico total, gerando um volume tridimensional de COT. Essa metodologia, validada pela comparação com o modelo empírico de Schmoker, demonstrou elevada precisão na identificação de *sweet spots* e destacou o potencial das redes neurais artificiais como ferramentas preditivas na caracterização de rochas não convencionais.

Seguindo a mesma linha de integração, del Monte et al. (2018) combinaram inversão de impedância acústica, análise AVO e atributos sísmicos para estimar o COT na Formação Tanezzuft (Tunísia). O estudo demonstrou que o modelo Backus–Yan fornece as estimativas mais realistas dentro das zonas ricas em matéria orgânica, enquanto abordagens puramente empíricas podem gerar anomalias artificiais fora dessas zonas. A análise AVO reforçou o caráter diagnóstico das respostas Classe IV e Base Classe IV, típicas de refletores associados a rochas geradoras orgânicas.

Mais recentemente, Cyz e Azevedo (2020) propuseram um método de inversão sísmica geoestatística direta, que permite estimar simultaneamente propriedades do *shale*, como índice de fragilidade (BI), COT e porosidade, a partir de dados de reflexão pré-empilhados. Diferentemente do procedimento tradicional em duas etapas, o método realiza a inversão diretamente nas propriedades do *shale*, promovendo melhor propagação das incertezas e maior consistência entre dados sísmicos e de poço. A aplicação em um reservatório Paleozoico na Polônia demonstrou a robustez e a aplicabilidade da abordagem em sistemas não convencionais.

Em conjunto, esses estudos evidenciam a evolução conceitual e metodológica no entendimento da relação entre matéria orgânica e resposta sísmica: de observações laboratoriais e modelos físico-rochosos a inversões quantitativas e algoritmos de aprendizado de máquina. Tal progressão sustenta a hipótese de que a impedância acústica constitui uma variável-chave para a estimativa indireta do teor de carbono orgânico total, especialmente quando associada a metodologias de calibração geoestatística e modelagem sísmica integrada.

Neste contexto, o presente capítulo tem como objetivo desenvolver e aplicar um modelo de predição de COT a partir da impedância acústica derivada da inversão sísmica estocástica, buscando compreender a distribuição espacial e a variabilidade lateral da matéria orgânica nas unidades geradoras da área de estudo, bem como avaliar a eficácia do método na caracterização sísmica quantitativa de rochas geradora.

A modelagem do COT neste trabalho será conduzida por meio do *Embedded Model Estimator* (EMBER), um algoritmo desenvolvido por Daly (2020) voltado à modelagem espacial não estacionária em contextos geológicos complexos. O EMBER integra princípios da geoestatística e do aprendizado de máquina, combinando modelos geoestatísticos incorporados (*embedded*) com o algoritmo *Quantile Random Forests*, de modo a capturar relações não lineares entre a variável alvo e múltiplas variáveis secundárias. Essa combinação permite que o método produza simultaneamente resultados de estimação e simulação estocástica, representando com maior realismo a continuidade espacial da variável modelada. Essa abordagem híbrida é particularmente vantajosa para ambientes onde a variabilidade lateral e vertical das propriedades é significativa, como ocorre nas rochas geradoras ricas em matéria orgânica afetadas por intrusões ígneas. Aplicações recentes do EMBER em modelagem petrofísica, caracterização sísmica e integração multivariada podem ser encontradas em Daly (2020), Daly et al. (2021), Daly (2021) e Daly (2022), evidenciando seu potencial para representar de forma robusta a heterogeneidade e a incerteza associadas à distribuição espacial das propriedades petrofísicas.

10.2. FUNDAMENTAÇÃO

10.2.1. Contexto e evolução conceitual

O *Embedded Model Estimator* (EMBER) é um algoritmo de modelagem espacial não estacionária proposto por Daly (2020, 2021, 2022), desenvolvido no contexto da integração entre geoestatística clássica e aprendizado de máquina. Seu principal objetivo é produzir, em cada ponto do domínio, uma distribuição condicional completa da variável regionalizada de interesse $Z(x)$,

denominada envelope, permitindo derivar estimativas pontuais, quantis, medidas de incerteza e realizar simulações estocásticas condicionais de forma consistente e não estacionária (Daly, 2021; Daly, 2022).

O método distingue-se das abordagens tradicionais de co-krigagem ou simulação gaussiana por inverter a ordem do processo de modelagem: em vez de assumir previamente um campo aleatório estacionário e, a partir dele, estimar as distribuições condicionais, o EMBER estima diretamente as distribuições condicionais locais, utilizando uma floresta de decisão não paramétrica, e somente depois impõe uma estrutura de correlação espacial para a geração de realizações (Daly, Hardy & McNamara, 2021).

Conforme salientado por Daly (2022), o algoritmo combina a robustez da teoria dos campos aleatórios condicionais (*Conditional Random Fields*, CRF; Lafferty et al., 2001) com a flexibilidade das florestas de decisão (*Random Forests*; Breiman, 2001; Ho, 1998), incorporando ainda modelos geoestatísticos embutidos (tipicamente krigagem) que fornecem informação sobre a continuidade espacial descrita pelo variograma (Chiles & Delfiner, 2012; Wackernagel, 2003).

10.2.2. Formulação em campos aleatórios condicionais (CRF)

A formulação fundamental do EMBER é expressa pela hipótese de Markov aplicada aos Conditional Random Fields (Lafferty et al., 2001). Seja $Z(x)$ a variável de interesse em uma posição x , e $Y(x)$ o vetor de variáveis secundárias conhecidas nesse mesmo ponto. Dado o conjunto de observações $\{Z_i, Y_i\}$, e um vetor $M(x)$ contendo estimadores prévios (por exemplo, modelos de krigagem de curto e longo alcance), Daly (2021) define a relação:

$$\hat{F}(z | Y(x), \{Z_i, Y_i\}) = \mathbb{E}[\mathbb{I}_{Z(x) < z} | Y(x), \{Z_i, Y_i\}] \approx \mathbb{E}[\mathbb{I}_{Z(x) < z} | Y(x), M(x)] \quad (10.1)$$

Essa aproximação expressa que a distribuição condicional de $Z(x)$, dado todo o conjunto de dados, pode ser reduzida a uma distribuição condicional dependente apenas das variáveis locais $Y(x)$ e dos modelos embutidos $M(x)$. Em outras palavras, o EMBER substitui a necessidade de uma modelagem explícita da lei multivariada do campo aleatório, como ocorre em co-simulações Gaussianas (Chiles & Delfiner, 2012; Wackernagel, 2003), por uma estimativa local, não paramétrica e baseada em aprendizado estatístico.

Essa estratégia segue o espírito das discussões de Matheron (1989) sobre a objetividade e parcimônia na modelagem espacial: evitar a inferência completa da distribuição multivariada quando o interesse prático recai sobre o comportamento condicional local (Matheron, 1989; Daly, 2022).

10.2.3. Estimação da distribuição condicional via *Quantile Random Forests*

A estimação da função distribuição condicional $\hat{F}(z|y)$ é realizada por meio de um *Quantile Random Forest*, que generaliza o método das florestas de regressão para estimar distribuições completas em vez de valores médios. Assim, Daly (2021) define o estimador:

$$\hat{F}(z | Y(x), \{Z_i, Y_i\}) = \sum_{i=1}^n \omega_i(y) \mathbb{I}_{\{Z_i < z\}} \quad (10.2)$$

onde os pesos $\omega_i(y)$ representam a frequência relativa com que a amostra Z_i é selecionada como vizinha do ponto x dentro das árvores da floresta. Sob condições de regularidade (Györfi et al., 2002), este estimador é consistente, ou seja, $\hat{F}(z|y) \rightarrow F(z|y)$ quando $n \rightarrow \infty$ (Daly, 2022).

A partir de $\hat{F}(z|y)$ obtém-se a média (valor esperado condicional), os quantis P10, P50 e P90, as probabilidades de excedência e o *spread* (P90–P10), permitindo caracterizar, para cada ponto, o envelope completo de incerteza associado à variável modelada.

10.2.4. Modelos embutidos (*embedding*) e validação cruzada

Os modelos embutidos $M(x)$ correspondem a preditores geoestatísticos (por exemplo, krigagens com variogramas de curto e longo alcance) que carregam a informação de continuidade espacial. Para evitar viés nos pontos amostrados, Daly (2020) recomenda que cada árvore da floresta seja treinada com estimativas validadas cruzadamente, isto é, M_{-i} calculadas com exclusão da amostra i :

Treino por árvore: $\{Z_i; Y_i, M_{-i}\}$.

Esse procedimento garante que o modelo aprenda a usar as krigagens embutidas apenas como variáveis auxiliares (ou “oráculos” de continuidade lateral), sem colapsar a estimativa nos valores observados. Essa técnica deriva da discussão de Fouedjio (2017) sobre modelagem não estacionária e mitigação de dependência entre amostras.

Além disso, o algoritmo pode utilizar múltiplos modelos embutidos (por exemplo, uma krigagem de longo alcance e outra de curto alcance), permitindo que a floresta combine essas informações de forma não linear e adaptativa, conforme a relevância local de cada preditor (Daly, Hardy & McNamara, 2021).

10.2.5. Envelope e simulação condicional

Uma vez estimado o envelope $\hat{F}(z|Y(x))$ em todo o domínio, Daly (2020, 2021, 2022) propõe método para gerar realizações condicionais que honrem os dados e reproduzam a heterogeneidade local. Seja

$$U(x) = F(Z(x) | Y(x) = y) \quad (10.3),$$

a variável de amostragem uniforme associada a cada ponto. Assume-se que $U(x) = G(X(x))$, em que $X(x)$ é um campo aleatório normal padronizado com correlação $\rho(h)$. Assim, a transformação inversa é dada por:

$$Z(x) = \varphi_x(X(x)) \equiv \psi_x(G(X(x))) \quad (10.4),$$

onde φ_x é a função quantil local associada ao envelope. Essa construção permite amostrar o envelope de modo correlato, produzindo realizações que preservam as relações espaciais e cruzam fronteiras geológicas sem artefatos.

Para compatibilizar o campo $X(x)$ com a variabilidade residual local, Daly (2022) demonstra que a correlação $\rho(h)$ deve satisfazer:

$$\rho(x_1 - x_2) = \mathbb{E} \left[\frac{Z(x_1) - \hat{\mu}(x_1|y)}{\sigma_{x_1}} \frac{Z(x_2) - \hat{\mu}(x_2|y)}{\sigma_{x_2}} \right] \quad (10.5)$$

em que $\hat{\mu}(x|y)$ é a média *Ember* e σ_x o desvio-padrão local do envelope. Essa relação fornece a ligação entre o envelope e o variograma, possibilitando ajustar a estrutura de correlação utilizada na simulação estocástica.

Daly (2022) demonstra que o estimador $\hat{F}(z|y)$ é assintoticamente consistente e que, em casos onde os modelos embutidos possuem forte poder preditivo, a média *Ember* tende ao valor obtido por krigagem. Entretanto, diferentemente da krigagem clássica, o EMBER não é exato nos dados, pois seu objetivo é estimar distribuições locais completas, e não apenas valores pontuais.

A principal vantagem do EMBER reside na sua capacidade de capturar heteroscedasticidade e não-estacionaridade espacial, ajustando automaticamente a importância das variáveis secundárias em diferentes regiões do modelo. Essa característica o torna particularmente adequado para formações geológicas complexas, em que as relações entre atributos sísmicos e propriedades petrofísicas variam localmente (Daly, 2022).

Assim, o método posiciona-se como uma extensão natural das técnicas de simulação condicional gaussiana, incorporando os avanços do *machine learning* não paramétrico e mantendo

coerência com os princípios de modelagem espacial descritos por Matheron (1981, 1989, *apud* Daly, 2022).

O fluxo operacional do EMBER pode ser resumido nas seguintes etapas (Daly, 2020; Daly, 2021; Daly, 2022):

1. Definição das variáveis secundárias $Y(x)$ e dos modelos embutidos $M(x)$ (krigagens de curto e longo alcance).
2. Treinamento da QRF para estimar $\hat{F}(z|Y, M)$ com cross-validation nos embutidos.
3. Geração do envelope de distribuições condicionais (média, quantis, probabilidades, *spread*).
4. Ajuste da correlação $\rho(h)$ a partir dos resíduos normalizados.
5. Simulação condicional via campo $X(x)$ correlato e transformação inversa $\varphi_x(X(x))$.

O resultado é um modelo espacial capaz de honrar os dados duros (*hard data*), reproduzir continuidade lateral e capturar variações locais nas relações entre variáveis, com rapidez, consistência e robustez estatística.

10.3. MÉTODO

A trajetória metodológica desenvolvida até este ponto integrou diferentes disciplinas e etapas de avaliação dos dados disponíveis em múltiplas escalas e resoluções. De forma sintética, o trabalho iniciou-se com o estudo de fácies no intervalo da Formação Pimenteiras, a partir dos quatro poços disponíveis na área de estudo. Essa etapa foi conduzida por meio do algoritmo de *machine learning K-Means*, a partir da clusterização não supervisionada dos perfis básicos dos poços, o que permitiu identificar distintas eletrofácies no intervalo de interesse. Em seguida, empregou-se a Classificação Bayesiana, com o objetivo de realizar um estudo de viabilidade para a inversão sísmica.

Considerando que esse intervalo é predominantemente composto por folhelhos ricos em matéria orgânica, investigou-se a relação entre as eletrofácies identificadas e o carbono orgânico total. Para obter perfis contínuos dessa propriedade nos poços, desenvolveu-se um estudo específico voltado à estimativa de COT a partir de diferentes técnicas de *machine learning*, integrando dados externos a área de estudo.

Posteriormente, foram realizados estudos de pré-condicionamento sísmico abrangendo uma área de aproximadamente 300 km² (20 km × 15 km), localizada na região de Capinzal do Norte,

também conhecida como Parque dos Gaviões. O objetivo dessa etapa foi aprimorar a resolução vertical e a relação sinal-ruído dos dados sísmicos, produzindo um volume adequado para a inversão sísmica, etapa responsável por gerar o volume de impedância acústica, propriedade elástica tridimensional mais importante para o estudo, o qual será mostrada a sua relação com o COT, embora de forma não linear. Assim, a impedância acústica foi empregada como uma variável secundária fundamental na modelagem 3D do COT, que constitui o principal objetivo deste estudo.

Todas essas etapas, bem como outras complementares, foram detalhadas nos capítulos anteriores e estão sintetizadas no fluxo de trabalho apresentado na Figura 10.1.

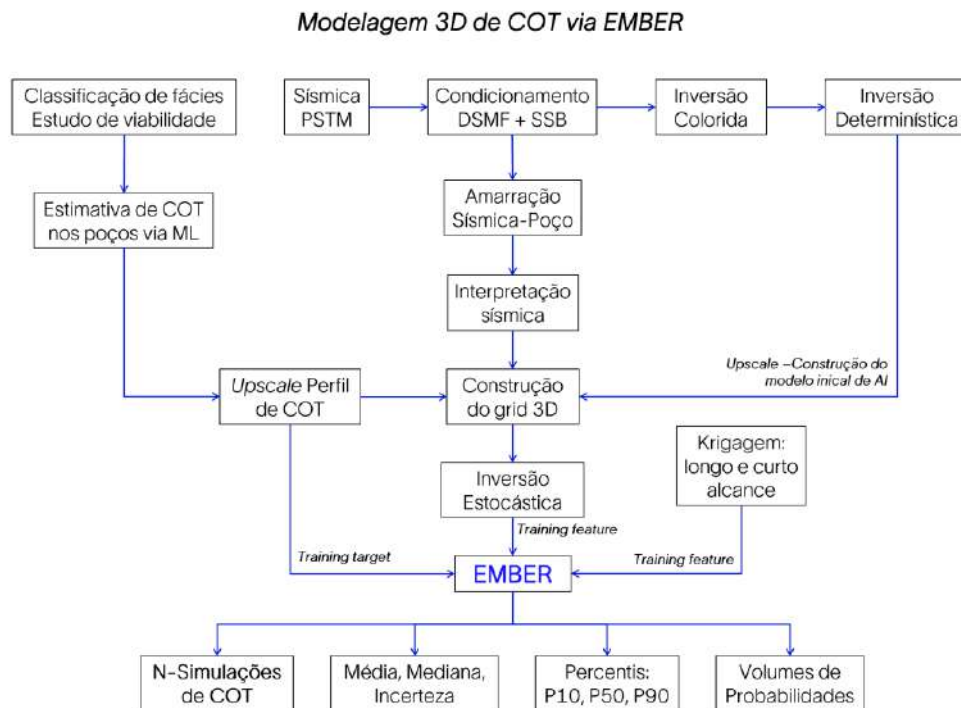


Figura 10.1. Fluxo de trabalho para predição de COT via EMBER: síntese das principais etapas do trabalho geral da tese.

10.3.1. Definição das variáveis de entrada

Após as etapas descritas, o primeiro passo na aplicação do EMBER consiste em compreender e quantificar a relação entre o COT estimado nos poços e a impedância acústica invertida, relação essa que orienta, de maneira criteriosa, a seleção da variável secundária a ser utilizada na modelagem do volume 3D de COT. A Figura 10.2 ilustra a correlação entre COT e AI antes e após o *upscaling*, com os pontos coloridos conforme as fácies identificadas nos poços.

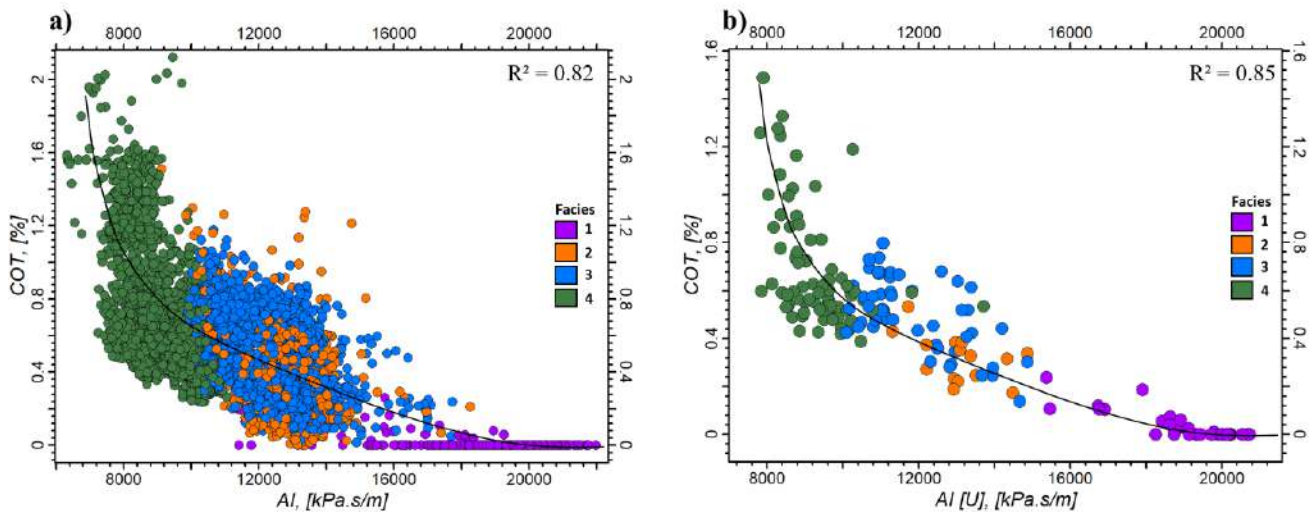


Figura 10.2. Correlação entre COT e AI antes e após o upscaling (pontos coloridos por fácies), destacando a tendência não linear negativa ajustada por polinômio de 2º grau.

A relação entre o COT e a AI não segue uma tendência linear simples, mas sim uma tendência não linear negativa, que pode ser descrita adequadamente por uma função polinomial de segundo grau. Nesse tipo de relação, o aumento de AI está associado a uma redução do COT, mas essa redução não ocorre de forma constante ao longo do intervalo, sendo mais acentuada em valores baixos de impedância e mais suave em valores altos.

Em situações como essa, o coeficiente de correlação linear de Pearson (r) tende a subestimar a força da associação entre as variáveis, pois ele mede apenas dependências lineares. Portanto, para caracterizar adequadamente a relação observada, é mais apropriado utilizar métodos de ajuste não linear e analisar o coeficiente de determinação (R^2) do modelo ajustado, que expressa a proporção da variabilidade dos dados explicada pela função polinomial.

No caso apresentado na Figura 10.2, a função polinomial de grau 2 apresenta um ajuste superior às alternativas lineares, evidenciando que a relação entre as propriedades segue um comportamento curvilíneo contínuo coerente com o comportamento físico esperado: rochas com maior impedância tendem a apresentar menor teor de matéria orgânica (fácies pobre em COT), e rochas com menor impedância podem refletir rochas com maior teor de matéria orgânica (fácies rica em COT)

Dada a relação não linear observada entre COT e AI, torna-se evidente, conforme discutido anteriormente na fundamentação teórica do EMBER, a motivação principal para a escolha desse método como ferramenta para a modelagem do COT. A Figura 10.3 apresenta o resultado da inversão estocástica (descrita detalhadamente no Capítulo 9), que representa a variável secundária (*training feature*) de maior relevância neste estudo. A variável selecionada a partir da inversão estocástica

representa um volume de impedância acústica de alta resolução, capaz de realçar contrastes sutis associados a camadas delgadas dentro do intervalo estudado. Observa-se que os níveis de menor impedância correspondem, de modo geral, às fácies mais ricas em COT, enquanto os valores intermediários de impedância estão relacionados a fácies com teores moderados a baixos de COT. Por sua vez, as altas impedâncias respondem de forma coerente às intrusões ígneas identificadas na área, caracterizadas pela fácies 1 (em roxo).

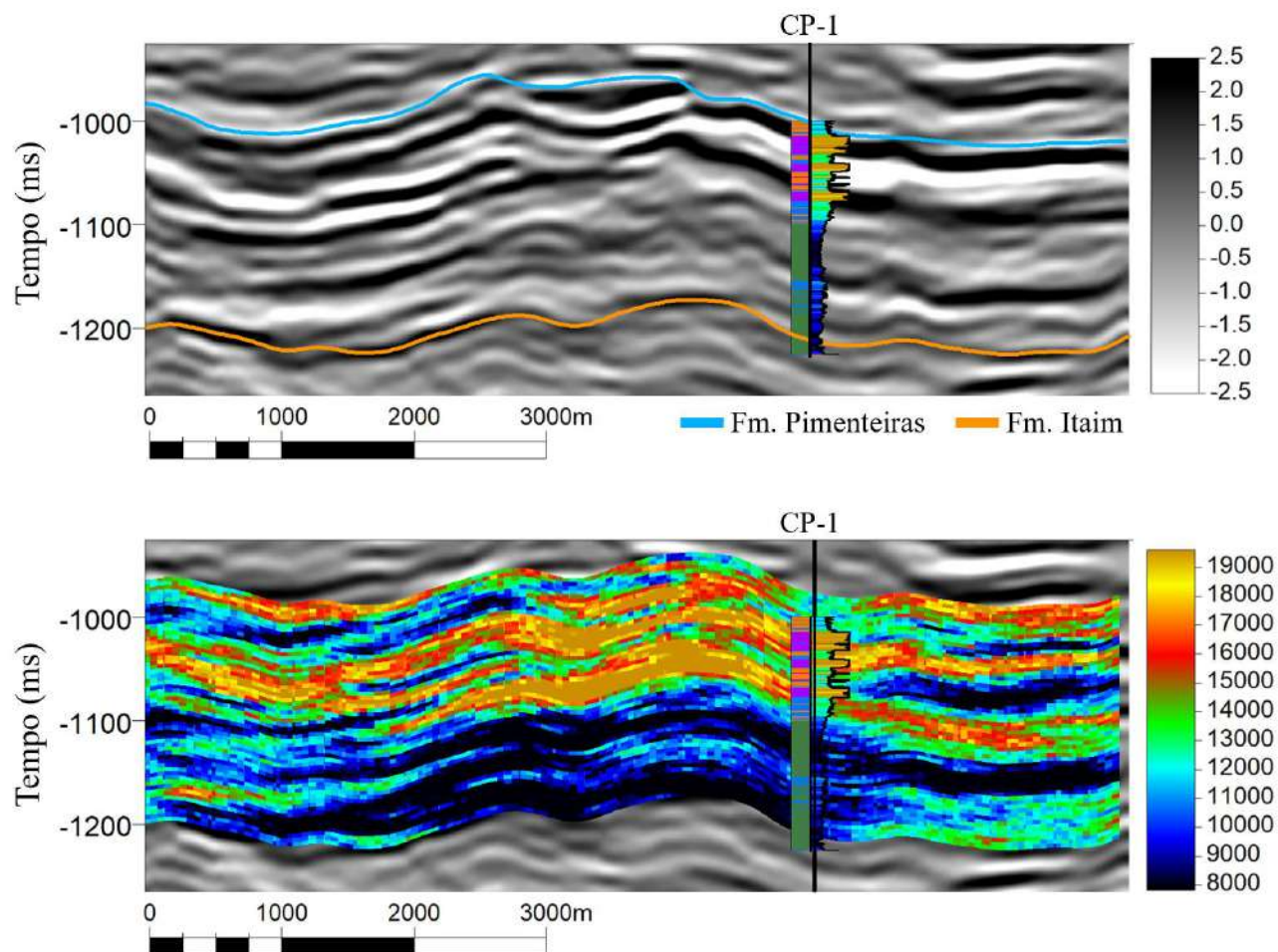


Figura 10.3. Volume de impedância acústica resultante da inversão estocástica (*training feature*): realce de camadas delgadas; associação entre baixas impedâncias e fácies mais ricas em COT; altas impedâncias compatíveis com intrusões ígneas (fácies 1).

Adicionalmente, a Figura 10.4 ilustra o arcabouço tridimensional (grid 3D) utilizado para as simulações do COT, com dimensões de 50 x 50 m (x, y) e 4 ms (z), configurado de modo a preservar a resolução espacial compatível entre os volumes sísmicos e os poços.

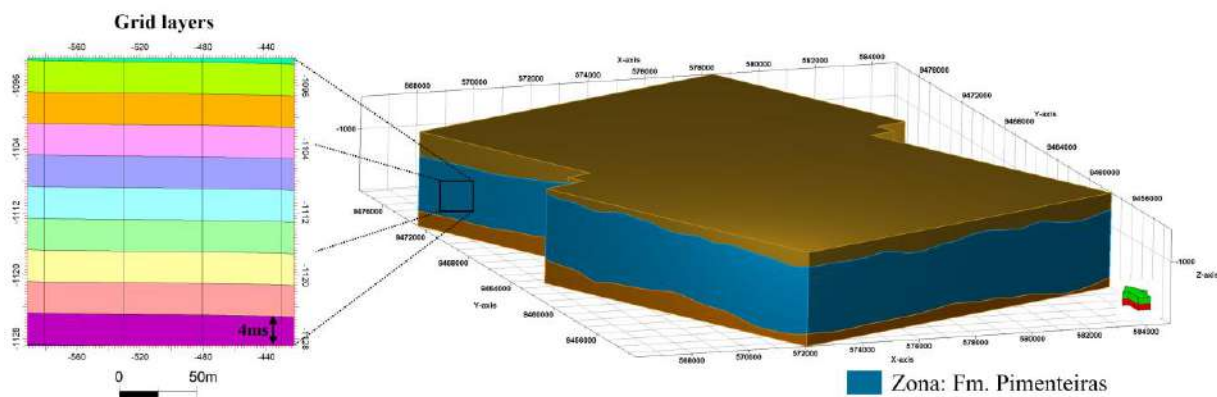


Figura 10.4. Arcabouço tridimensional (grid 3D) utilizado nas simulações (50×50 m em x-y; 4 ms em z), configurado para compatibilizar a resolução entre volumes sísmicos e poços.

10.3.2. Etapas de aplicação do EMBER

A aplicação do algoritmo EMBER para a modelagem tridimensional do COT seguiu as etapas principais descritas a seguir:

1. Seleção do *Training Target*: perfis de COT reamostrados e compatibilizados com o *grid* tridimensional, por meio do processo de *upscaling*.
2. Seleção do *Training Feature*: volume de impedância acústica resultante da inversão estocástica, utilizado como variável secundária principal. Além dessa variável explicitamente definida pelo usuário, o EMBER calcula internamente, de forma automática, dois modelos geoestatísticos embutidos, de krigagem de longo e de curto alcance, que são incorporados como modelos embutidos (*embedded models*) durante o treinamento. Esses modelos fornecem informações adicionais de continuidade espacial, atuando como *priors* de tendência e suavização, fundamentais para capturar variações regionais e locais do COT.
3. Definição do *Blind Test*: o poço CP-1, considerado poço teste, foi utilizado como referência independente para validação do modelo, sendo o único na área de estudo com medições laboratoriais de COT.
4. Configuração das simulações: definição do número de realizações e dos parâmetros de controle de continuidade horizontal e vertical, razão de aspecto, orientação estrutural e níveis de suavização aplicados à distribuição da propriedade. Essa etapa envolveu uma série de testes iterativos e análises estatísticas (média, variância, percentis, e desvio padrão) até a obtenção dos parâmetros ideais para a área de estudo, conforme resumido na Tabela 10.1.

- Controle de qualidade: avaliação quantitativa do desempenho do modelo por meio das métricas correlação (R), erro médio absoluto (MAE), raiz do erro quadrático médio (RMSE), além da análise da variância, desvio-padrão, média e probabilidades de ocorrência em diferentes faixas de COT.

Tabela 10.1. Parâmetros numéricos empregados nas simulações do EMBER.

Parametrização - EMBER	
Interwell texture	6
Vertical texture	6
Aspect ratio	1
Orientation	0
Continuity	Rough
N° Simulation	20

* *Rough* : rápia mudança local, suavização diminui com a distância

Esse conjunto de etapas constitui a espinha dorsal do fluxo de trabalho de predição do COT via EMBER, integrando informações provenientes da inversão sísmica, dos estudos petrofísicos e das análises geoestatísticas e *machine learning*.

As seções seguintes apresentam os resultados obtidos com a aplicação desse fluxo e as respectivas discussões e interpretações geológicas, abordando o desempenho do EMBER na representação tridimensional da distribuição do COT na Formação Pimenteiras.

10.4. RESULTADOS

Após a geração de 20 simulações no EMBER, utilizando exclusivamente a impedância acústica como variável de treinamento (*training feature*) e considerando os modelos de krigagem de curto e longo alcance processados automaticamente pelo algoritmo, foram conduzidas análises detalhadas dos resultados, seguidas por uma série de procedimentos de controle de qualidade, os quais serão apresentados nesta seção.

Inicialmente, avaliou-se a correlação entre os valores estimados pelo EMBER e os dados medidos no poço cego (CP-1), com o objetivo de identificar a realização com melhor desempenho estatístico (Tabela 10.2). Essa etapa foi essencial para selecionar a simulação mais representativa em relação à curva real de COT, garantindo a escolha de um modelo geologicamente consistente para as análises subsequentes.

As simulações estocásticas resultaram em um conjunto de modelos com média global de 0.48, desvio padrão de 0.22 e variância de 0.05, evidenciando que a abordagem adotada foi capaz de incorporar de maneira adequada a heterogeneidade intrínseca da Formação Pimenteiras. Essa dispersão

controlada demonstra a eficiência do método em representar as variações espaciais do teor de matéria orgânica, evitando dois comportamentos indesejáveis frequentemente observados em metodologias tradicionais: (i) a suavização excessiva dos modelos determinísticos, que tende a mascarar variações relevantes das propriedades modeladas, e (ii) a geração de padrões ruidosos ou não geológicos, decorrentes de simulações mal condicionadas. Assim, o modelo estocástico obtido preserva simultaneamente as tendências regionais e os contrastes locais, resultando em uma representação mais realista e contínua do COT no volume 3D.

No teste de poço cego, a realização SIM5 apresentou o melhor desempenho, com RMSE = 0.188, MAE = 0.144 e correlação R = 0.798, indicando elevada capacidade de reproduzir a distribuição do COT ao longo do intervalo de interesse. Esses resultados confirmam a robustez da modelagem estocástica implementada no EMBER para capturar a variabilidade espacial e fornecer modelos tridimensionais consistentes para interpretação geológica e avaliação da qualidade das rochas geradoras.

Tabela 10.2 . Avaliação do desempenho das 20 simulações no poço cego.

Poço cego: 2-CP-1-MA			
Propriedade	RMSE	MAE	R
COT_Blind mean	0.196	0.148	0.787
COT_Blind median	0.206	0.144	0.756
COT_Blind SIM1	0.234	0.17	0.661
COT_Blind SIM2	0.224	0.159	0.707
COT_Blind SIM3	0.244	0.177	0.694
COT_Blind SIM4	0.233	0.148	0.702
COT_Blind SIM5	0.188	0.144	0.798
COT_Blind SIM6	0.236	0.187	0.728
COT_Blind SIM7	0.276	0.187	0.564
COT_Blind SIM8	0.295	0.184	0.510
COT_Blind SIM9	0.225	0.157	0.710
COT_Blind SIM10	0.209	0.158	0.757
COT_Blind SIM11	0.255	0.192	0.663
COT_Blind SIM12	0.240	0.192	0.652
COT_Blind SIM13	0.269	0.178	0.542
COT_Blind SIM14	0.261	0.191	0.637
COT_Blind SIM15	0.221	0.151	0.697
COT_Blind SIM16	0.202	0.139	0.779
COT_Blind SIM17	0.219	0.164	0.710
COT_Blind SIM18	0.219	0.153	0.739
COT_Blind SIM19	0.226	0.163	0.729
COT_Blind SIM20	0.272	0.197	0.629
Média		0.48	
Desvio padrão		0.22	
Variância		0.05	

A **Error! Reference source not found.** apresenta a comparação entre a curva de COT reamostrada no grid do poço cego e o COT estimado pela realização SIM5, permitindo observar tanto os intervalos de maior correlação quanto aqueles de maior dispersão (**Error! Reference source not found.a**). Nota-se que a fácies 1 (em roxo), correspondente às soleiras de diabásio, exibe em determinados níveis valores ligeiramente diferentes de zero. Esse comportamento decorre da combinação entre o método de *upsampling* e a resolução do grid, que, ao realizar o processo de média volumétrica, gerou valores residuais próximos de zero, mas não exatamente nulos como na curva original.

Nesses intervalos associados às intrusões ígneas, observam-se maiores discrepâncias entre o COT do poço e o COT estimado pela SIM5, incluindo um pico isolado de alto COT em torno de 1205 ms, possivelmente relacionado a artefatos do processo de interpolação local. Ainda assim, a tendência geral do perfil permanece coerente, apresentando boa aderência entre as curvas, conforme evidenciado pelo coeficiente de correlação ($R = 0.80$) mostrado no *crossplot* da **Error! Reference source not found.b**.

A **Error! Reference source not found.c** apresenta o *crossplot* entre o COT do poço cego e o COT médio representado pelo P50 das simulações, o qual também demonstra elevada correlação ($R = 0.76$), confirmando a consistência estatística global dos resultados obtidos pelo EMBER.

A **Error! Reference source not found.d** apresenta o histograma comparativo das distribuições de COT obtidas a partir dos dados de poço (barra magenta), do modelo correspondente ao percentil 50 (P50, barra azul-escuro) e da simulação estocástica 5 (SIM 5, barra azul-claro). Observa-se que as distribuições simuladas seguem, de modo geral, o comportamento estatístico dos dados dos poços, preservando a forma e o posicionamento modal da distribuição original. Por outro lado, é possível analisar que a realização SIM5 reproduz de forma mais completa a amplitude dos valores de COT dos poços, cobrindo todo o intervalo observado, enquanto que o modelo P50 apresenta distribuição mais concentrada em torno dos valores modais, subestimando os teores mais elevados de COT que, embora menos frequentes, são geologicamente relevantes. Essa correspondência indica que o modelo reproduz de forma satisfatória as características estatísticas e a tendência central do COT, com pequenas variações atribuídas à aleatoriedade inerente ao processo de simulação geoestatística, que busca representar diferentes realizações possíveis dentro do envelope de incerteza.

A **Error! Reference source not found.e** mostra o histograma do *spread* das simulações, calculado como a diferença entre os percentis 90 e 10 ($P90 - P10$) para cada célula do modelo. Essa métrica expressa a amplitude de variação ou da incerteza associada às estimativas de COT, sendo que

valores mais altos indicam maior variabilidade entre as simulações e, portanto, menor confiabilidade local. Nota-se que a maior parte dos valores se concentra em torno de 0,4 a 0,5 %, sugerindo que a incerteza global do modelo é moderada e que as previsões de COT apresentam boa estabilidade estatística. Os intervalos com *spread* mais elevados ($> 0,8$ %) ocorrem de forma pontual e podem estar relacionados a regiões com menor controle de dados, variações alta de impedância associadas a valores semelhantes de COT, ou podem está associadas influência de intrusões ígneas que afetam a continuidade lateral e a geoestatística do modelo. De forma geral, a análise conjunta dos histogramas indica que o modelo gerado reproduz adequadamente o comportamento global dos dados de poço e mantém níveis aceitáveis de incerteza, demonstrando consistência estatística e robustez no processo de modelagem do COT.

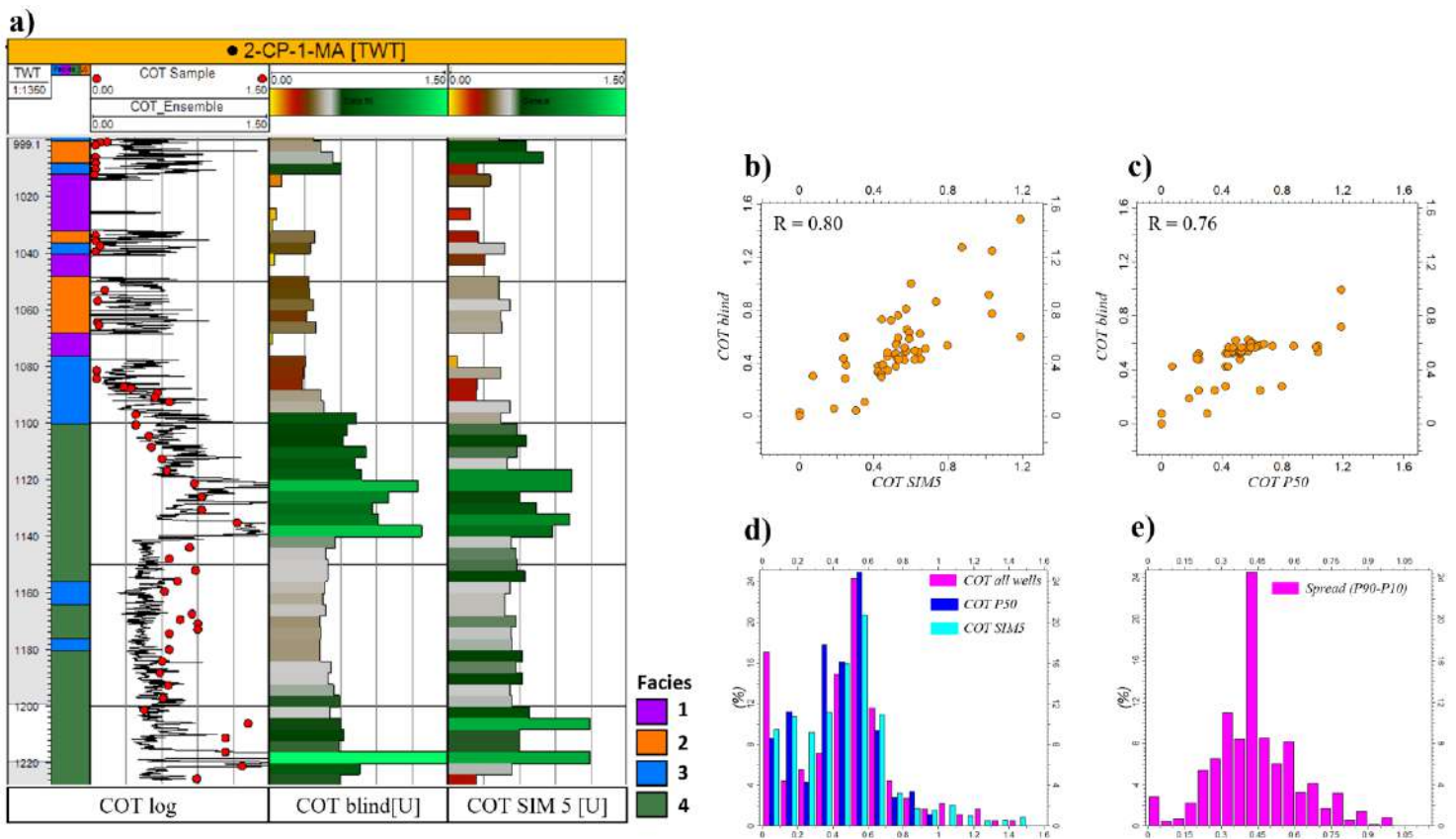


Figura 10.5. Avaliação no poço cego a partir da simulação do EMBER: (a) perfis de COT reamostrado vs. COT estimado (SIM5); (b) crossplot entre o COT do poço e COT estimado pela SIM5; (c) crossplot entre o COT do poço e COT estimado a partir do P50; (d) histogramas de COT (distribuição de todos os poços, P50 e SIM5); (e) histograma do *spread* (P90–P10).

A Figura 10.6 apresenta os resultados da simulação 5 (SIM 5) do COT em seção e em mapa, juntamente com o volume de impedância acústica utilizado como dado de entrada e o volume de incerteza (P90–P10) obtido a partir do EMBER. Na seção sísmica que intercepta os poços, é possível

identificar claramente as intrusões ígneas de diferentes espessuras, caracterizadas pelos tons mais quentes (valores elevados) tanto no volume de impedância acústica quanto na distribuição de COT, nas quais se observam valores nulos ou próximos de zero de carbono orgânico total. Essas intrusões, ao atravessarem as camadas sedimentares, interrompem a continuidade estratigráfica e podem ter atuado como fontes locais de calor, ocasionando maturação térmica da matéria orgânica nas rochas encaixantes, o que explica a redução dos teores de COT nessas zonas.

Os níveis mais ricos em matéria orgânica destacam-se pelos tons de verde-claro no volume de COT, distribuídos em diferentes horizontes estratigráficos e associados a zonas de menor impedância acústica. Essa correlação inversa entre impedância acústica e COT é fisicamente consistente e reforça a confiabilidade da calibração entre o modelo de inversão sísmica e o modelo geoestatístico gerado pelo EMBER. Observa-se ainda que as zonas mais enriquecidas em COT ocorrem preferencialmente em níveis contínuos e paralelos, compatíveis com depósitos pelíticos formados em condições de mar calmo e baixo aporte detrítico, característicos de tratos de sistemas transgressivos ou de mar alto, o que confirma a coerência sedimentológica e paleoambiental da simulação.

A seção de incerteza evidencia que os valores de *spread* são predominantemente moderados, o que demonstra boa estabilidade estatística do modelo. Contudo, zonas localizadas de maior incerteza são observadas na porção noroeste da área, próximas aos poços CP-1 e OGX-16, coincidindo com regiões de COT mais elevado, podendo ser um comportamento associadas a menor ocorrência desses níveis mais ricos em COT quando comparados com o maior volume de rochas com valores de COT moderados.

Nos mapas exibidos na parte inferior da figura, uma camada representativa do *grid* ilustra a distribuição espacial das propriedades de impedância acústica, COT e incerteza, revelando uma coerência espacial entre os padrões verticais e horizontais. Nota-se a relação inversa entre impedância e COT, bem como o aumento da incerteza em áreas com menor condicionamento por dados de poço, o que confirma que o modelo final do EMBER preserva a continuidade espacial esperada e reflete de forma realista a interação entre as propriedades sísmicas, petrofísicas e geoestatísticas da Formação Pimenteiras.

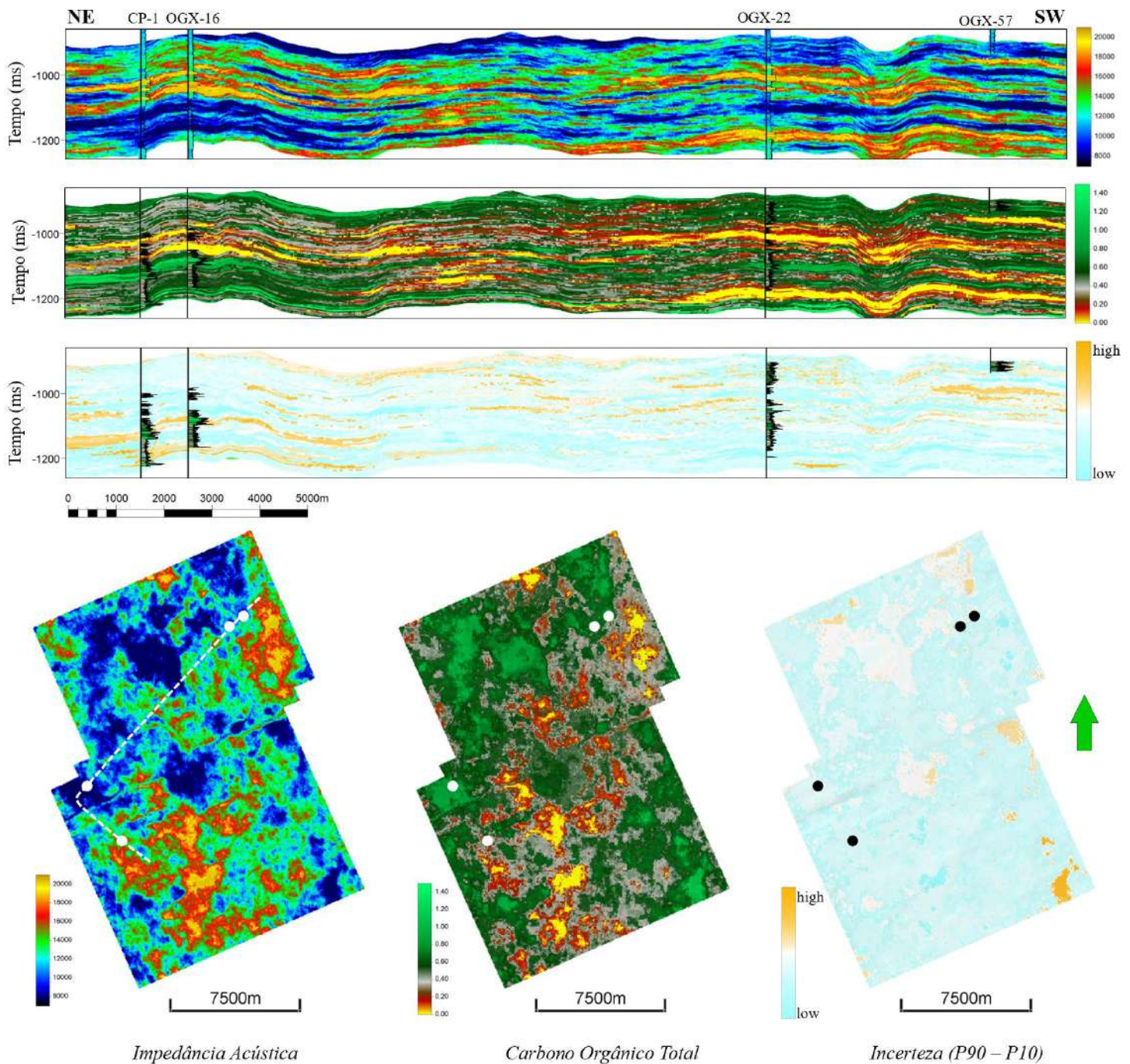


Figura 10.6. Resultados da SIM5 em seção e mapas: AI (entrada), COT e incerteza (P90–P10). Evidenciam-se altos valores de COT nas mais baixas impedâncias (relação inversa AI–COT) e a dispersão espacial da incerteza.

O modelo tridimensional de COT resultante da simulação 5 é apresentado na Figura 10.7, visualizado em volume e em três camadas representativas do *grid*. Na parte superior da figura, o volume 3D é exibido com cortes ortogonais nas direções *i* e *j* do *grid*, permitindo observar a distribuição espacial e a continuidade vertical e lateral dos níveis ricos em matéria orgânica ao longo da Formação Pimenteiras.

Os mapas apresentados na parte inferior da figura correspondem a três fatias horizontais extraídas do volume de COT (níveis K = 73, 89 e 110), que ilustram a heterogeneidade e a variação lateral da concentração de carbono orgânico. Os tons de verde representam valores mais elevados de COT, indicando intervalos ricos em matéria orgânica e potencialmente geradores, enquanto os tons amarelados e avermelhados correspondem a valores mais baixos de COT, associados intrusões ígneas. Além disso os tons mais próximos do cinza e marrom claro podem refletir a folhelhos mais maduros ou empobrecidas em COT.

A variação lateral observada nos mapas reflete o caráter heterogêneo da distribuição de COT, podendo ser resultado de flutuações paleoambientais, das variações no aporte sedimentar e na taxa de preservação da matéria orgânica e mais importante neste contexto, a relação de proximidade dos folhelhos potenciais geradores com as intrusões ígneas. Essa representação tridimensional permite visualizar a conectividade e a extensão dos horizontes mais ricos em COT, que ajuda compreender a arquitetura interna dos intervalos geradores e a extensão da qualidade da rocha geradora ao longo da área de estudo.

Outros produtos gerados a partir da modelagem pelo EMBER incluem os volumes de probabilidade de ocorrência de determinadas faixas de valores da propriedade simulada — neste caso, COT. Essa abordagem probabilística permite estimar, para cada célula do modelo, a probabilidade de a rocha apresentar teores de COT dentro de intervalos específicos, conforme a classificação proposta por Peters & Cassa (1994), que define:

1. valores inferiores a 0,5 % como pobres em COT,
2. valores entre 0,5 % e 1,0 % como moderados ou razoáveis, e
3. valores superiores a 1,0 % como ricos em matéria orgânica.

A Figura 10.8 apresenta os três volumes de probabilidade resultantes dessa classificação. Observa-se que a probabilidade de $COT > 1,0\%$ (painel superior) é restrita a intervalos específicos, com boa continuidade lateral, correspondendo aos níveis mais ricos em matéria orgânica da Formação Pimenteiras. Esses horizontes coincidem com as fácies 4 (em verde na **Error! Reference source not found.**), interpretadas nos poços como folhelhos ricos em COT, o que reforça a coerência entre a modelagem probabilística e as interpretações de eletrofácies.

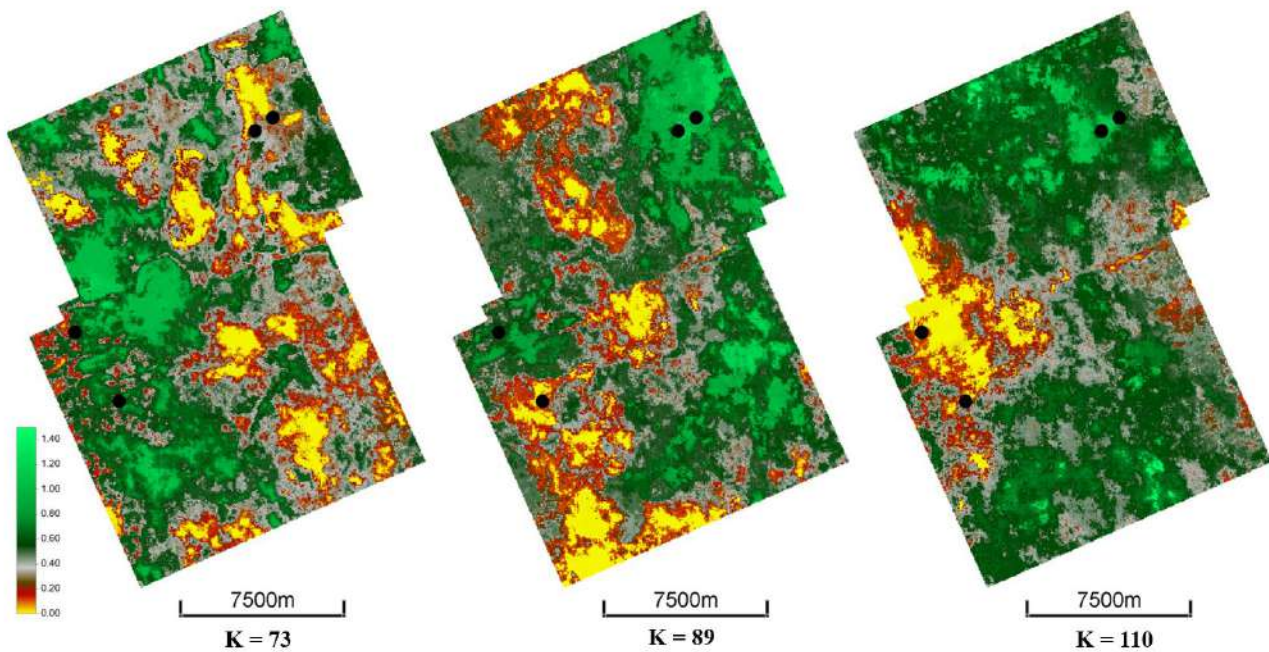
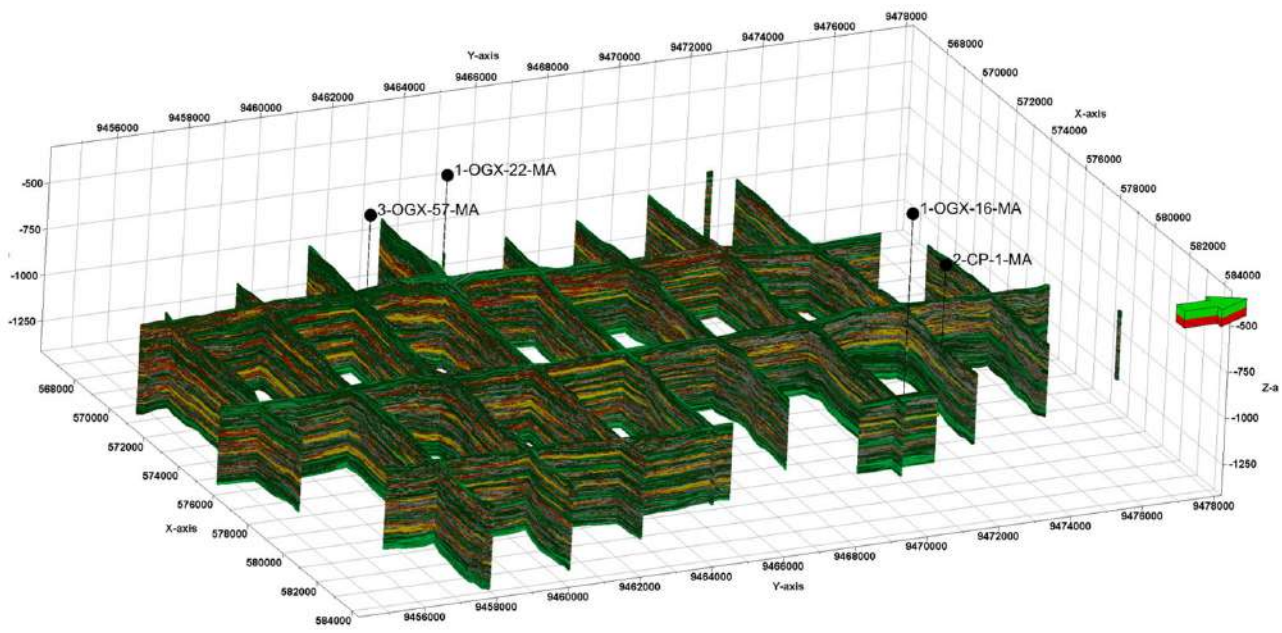


Figura 10.7. Modelo 3D de COT (SIM5) em volume com cortes ortogonais (i, j) e três fatias do grid (K = 73, 89 e 110); destaque para a continuidade lateral dos níveis enriquecidos em COT e a heterogeneidade espacial associada às intrusões.

O intervalo intermediário ($0,5 \% < \text{COT} < 1,0 \%$) (painel central) apresenta distribuição mais ampla e contínua, dominando grande parte do intervalo modelado. Essa tendência indica que a maior parte da Formação Pimenteiras é composta por rochas com teores moderados de COT, o que sugere um potencial gerador razoável a bom, que podem está associadas com as fácies 3 e 4 classificadas nos poços como discutido com detalhe no Capítulo 5.

Em contrapartida, o volume de probabilidade de COT < 0,5 % (painel inferior) destaca-se nas regiões associadas às intrusões ígneas, onde a probabilidade de ocorrência se aproxima de 1. Esses intervalos refletem tanto as soleiras de diabásio (nas quais o COT é nulo) quanto os folhelhos termicamente alterados nas zonas de contato com as intrusões, associados a fácies 2 (em laranja na **Error! Reference source not found.**), onde o aquecimento magmático pode ter promovido a maturação térmica e a degradação parcial da matéria orgânica, reduzindo significativamente o teor de COT.

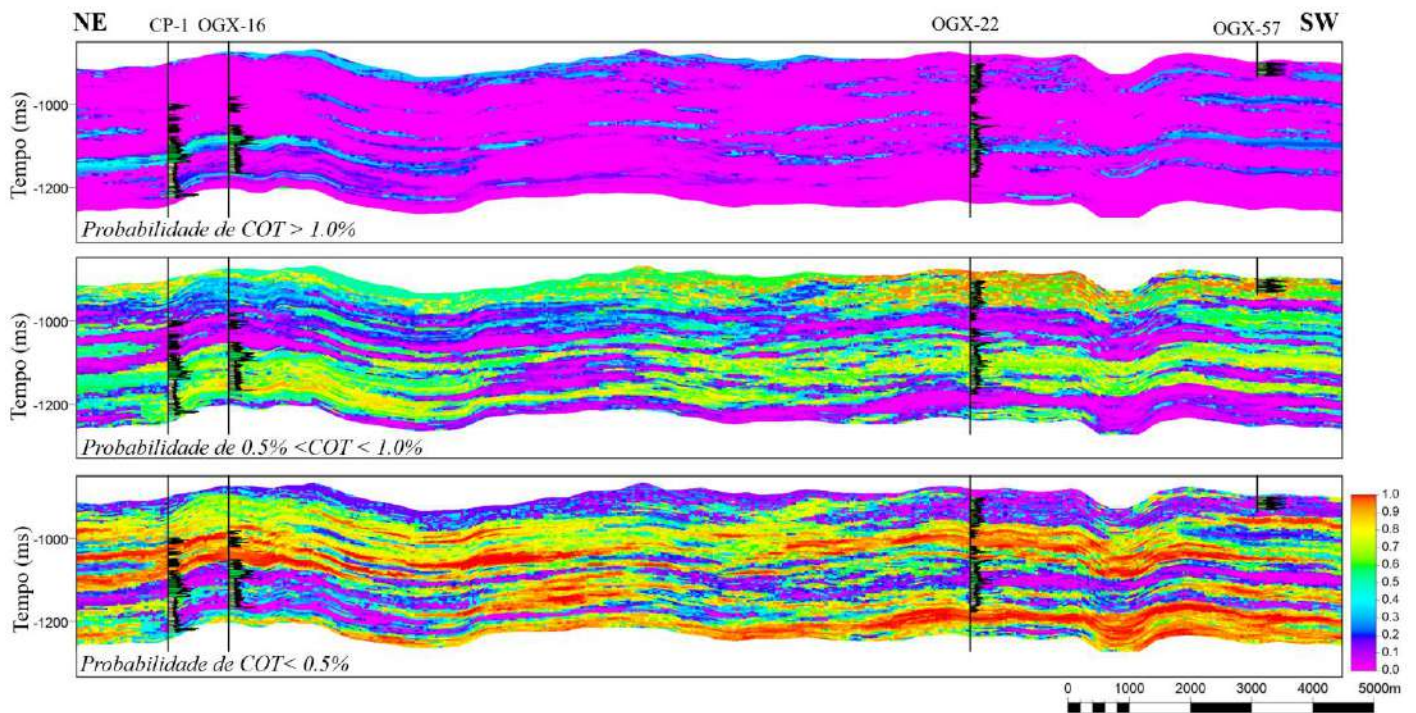


Figura 10.8. Volumes de probabilidade de ocorrência por classes de COT segundo Peters & Cassa (1994): $P(COT > 1,0 \%)$, $P(0,5 \% < COT < 1,0 \%)$ e $P(COT < 0,5 \%)$; padrões coerentes com as eletrofácies identificadas nos poços e com a distribuição esperada de rochas ricas, moderadas e pobres em matéria orgânica.

A mesma análise probabilística apresentada na Figura anterior foi aplicada a três camadas representativas do modelo de COT, correspondentes aos níveis $K = 73$, $K = 89$ e $K = 110$, cujos resultados são mostrados na Figura 10.9. Essa representação em mapa permite avaliar a variação lateral da probabilidade de ocorrência das diferentes faixas de COT e compreender a heterogeneidade espacial que ocorre ao longo do intervalo de interesse.

Nos mapas da probabilidade de $COT > 1,0 \%$ (coluna à esquerda), maior probabilidade de ocorrência de rochas ricas em matéria orgânica, entre 50 a 60% de chance, se dá de maneira isolada e concentrada nas três camadas selecionadas.

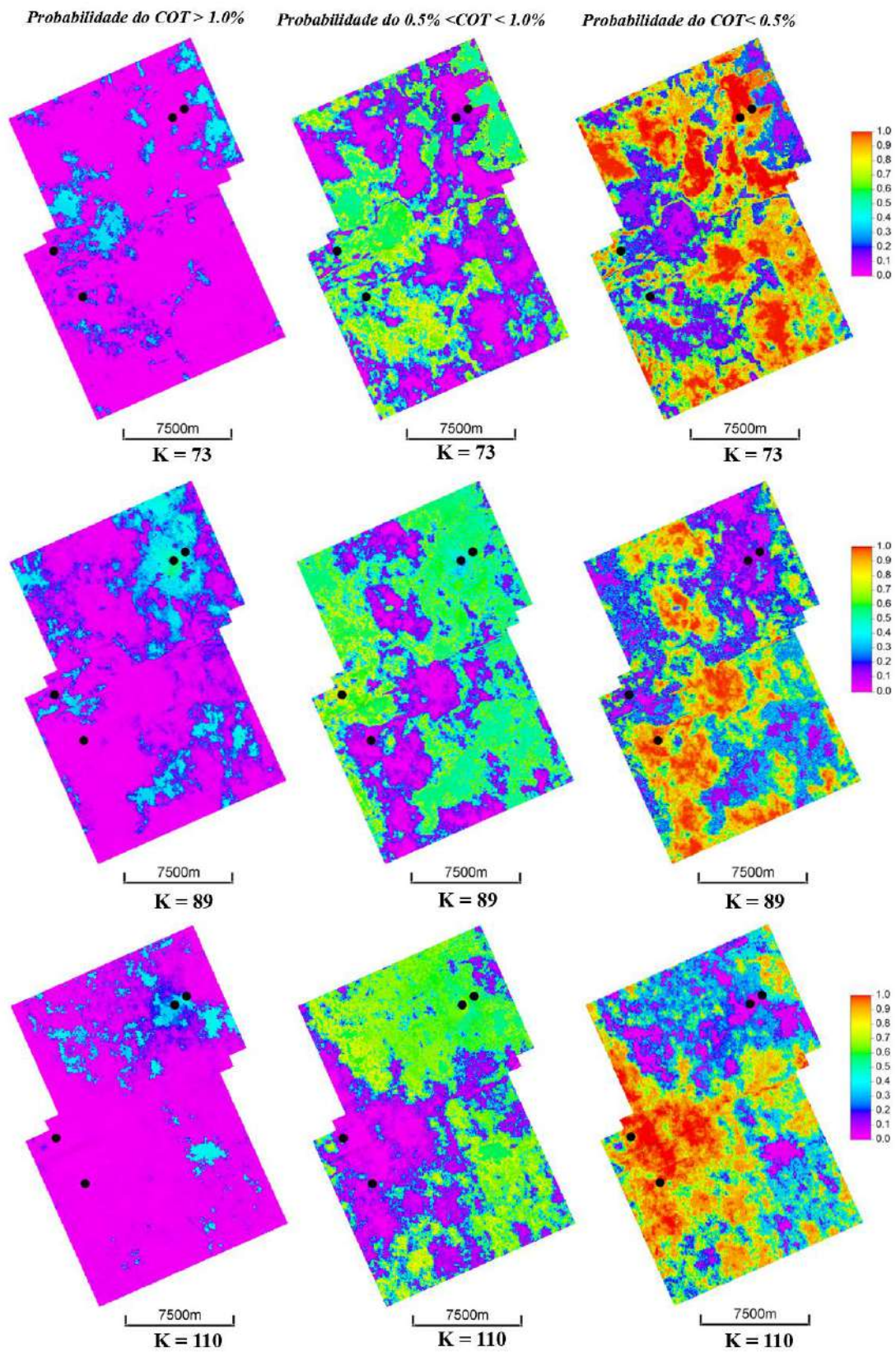


Figura 10.9. Mapas de probabilidade por classe de COT (mesmas faixas da Figura 8) em três camadas representativas (K = 73, 89 e 110), evidenciando variação lateral, predominância do intervalo moderado (0,5–1,0 %) e concentração de baixos COT próximos às intrusões.

Para o intervalo $0,5 \% < \text{COT} < 1,0 \%$ (coluna central), as zonas de probabilidade alta (entre 60% a 80%) se distribuem amplamente ao longo de todas as camadas analisadas. Essa predominância reflete que a maior parte da Formação Pimenteiras é composta por rochas com teores moderados de COT, indicando um potencial gerador razoável a bom. A variação lateral revela uma heterogeneidade significativa, controlada por variações de fácies e pela influência local de eventos magmáticos.

Nos mapas de probabilidade de $\text{COT} < 0,5 \%$ (coluna à direita), as zonas em vermelho e amarelo marcam regiões onde há maior probabilidade de ocorrência de baixos teores de COT. Essas áreas se concentram principalmente nas proximidades das intrusões ígneas que como fica evidente na Figura 10.9, ocorrem em diferentes áreas de acordo com a profundidade, devido aos saltos das soleiras, que irão coincidir com zonas de aquecimento térmico localizado, onde o calor das soleiras de diabásio pode ter promovido a maturação e degradação térmica da matéria orgânica, reduzindo o COT original das rochas encaixantes.

Por fim, as Figura 10.10 e Figura 10.13 apresentam as seções verticais que atravessam os poços CP-1, OGX-16, OGX-22 e OGX-57, exibindo a distribuição de COT obtida a partir da simulação 5 (SIM 5) e dos volumes de percentis P10, P50 e P90 calculados pela modelagem EMBER. Essa análise comparativa fornece uma visão detalhada sobre a distribuição vertical e lateral da matéria orgânica, permitindo avaliar a consistência do modelo geoestatístico, a incerteza associada e a conectividade dos horizontes geradores na Formação Pimenteiras.

De maneira geral, observa-se que os maiores valores de COT ($> 1,0 \%$), representados pelos tons de verde-claro, estão concentrados em intervalos estratigraficamente bem definidos, com boa continuidade lateral entre os poços. Esses horizontes podem estar associados às fácies 4, interpretadas nas eletrofácies como folhelhos ricos em matéria orgânica, e correspondem aos principais níveis geradores da Formação Pimenteiras. A simulação 5 (SIM 5) e o P50 apresentam comportamento semelhante, reproduzindo adequadamente a tendência central do modelo e indicando que o algoritmo EMBER conseguiu integrar coerentemente os dados de poço, impedância acústica e variabilidade espacial.

O percentil P10, que representa o cenário pessimista (valores mínimos esperados de COT), evidencia um aumento expressivo de áreas com baixa concentração de carbono orgânico ($< 0,5 \%$), em tons amarelados e alaranjados, especialmente nas zonas próximas às intrusões ígneas (fácies 1). Assim,

o P10 destaca as áreas de menor potencial gerador e maior incerteza de preservação da matéria orgânica.

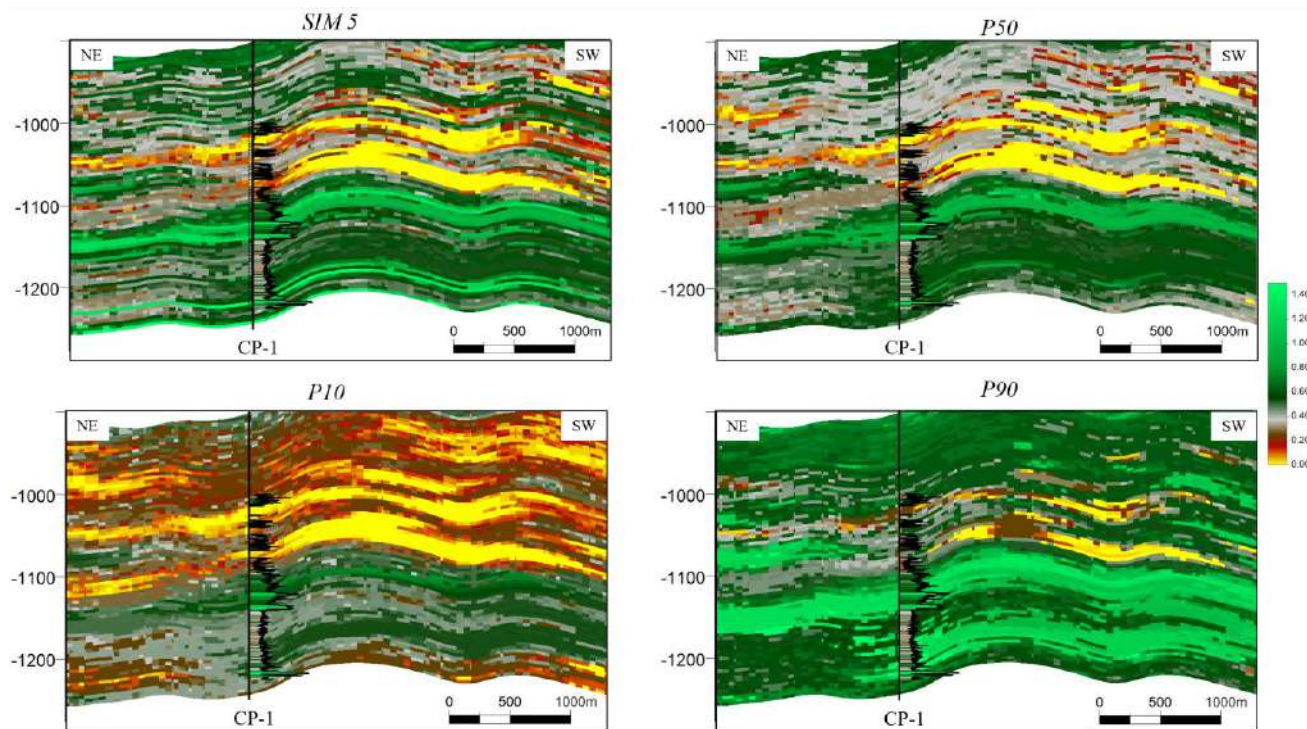


Figura 10.10. Seções dos modelos de COT passando pelo poço CP-1 mostrando o resultado da simulação de maior correlação (SIM5) e os cenários: provável (P50), pessimista (P10) e otimista (P90).

O percentil P50, correspondente ao cenário mais provável, apresenta uma distribuição equilibrada entre zonas ricas e moderadas em COT, refletindo a resposta mediana do modelo probabilístico. Essa seção evidencia a continuidade dos intervalos de COT intermediário (0,5–1,0 %), associados às fácies 3 e 4, que representam rochas potencialmente geradoras. O P50 também mostra que o modelo reproduz de forma realista a variação vertical do conteúdo orgânico, compatível com a sequência da Formação Pimenteiras, em que os teores de COT tendem a aumentar nas porções mais transgressivas e decrescer em direção às zonas mais oxidadas em momentos de tempestade (Lima e Leite, 1978) ou influenciadas por eventos magmáticos (Rodrigues, 1995; Mussa, 2020).

Já o percentil P90 corresponde ao cenário otimista, refletindo o limite superior da variabilidade do modelo. Nele, as zonas verdes tornam-se mais espessas e contínuas, expandindo as áreas com COT superior a 1,0 %. Esse comportamento pode ser interpretado como indicativo de alta probabilidade de preservação da matéria orgânica e de ambientes deposicionais mais redutores, sugerindo condições paleoambientais favoráveis à acumulação de material orgânico, possivelmente relacionadas a eventos de máxima inundação marinha descritos por Rodrigues (1995).

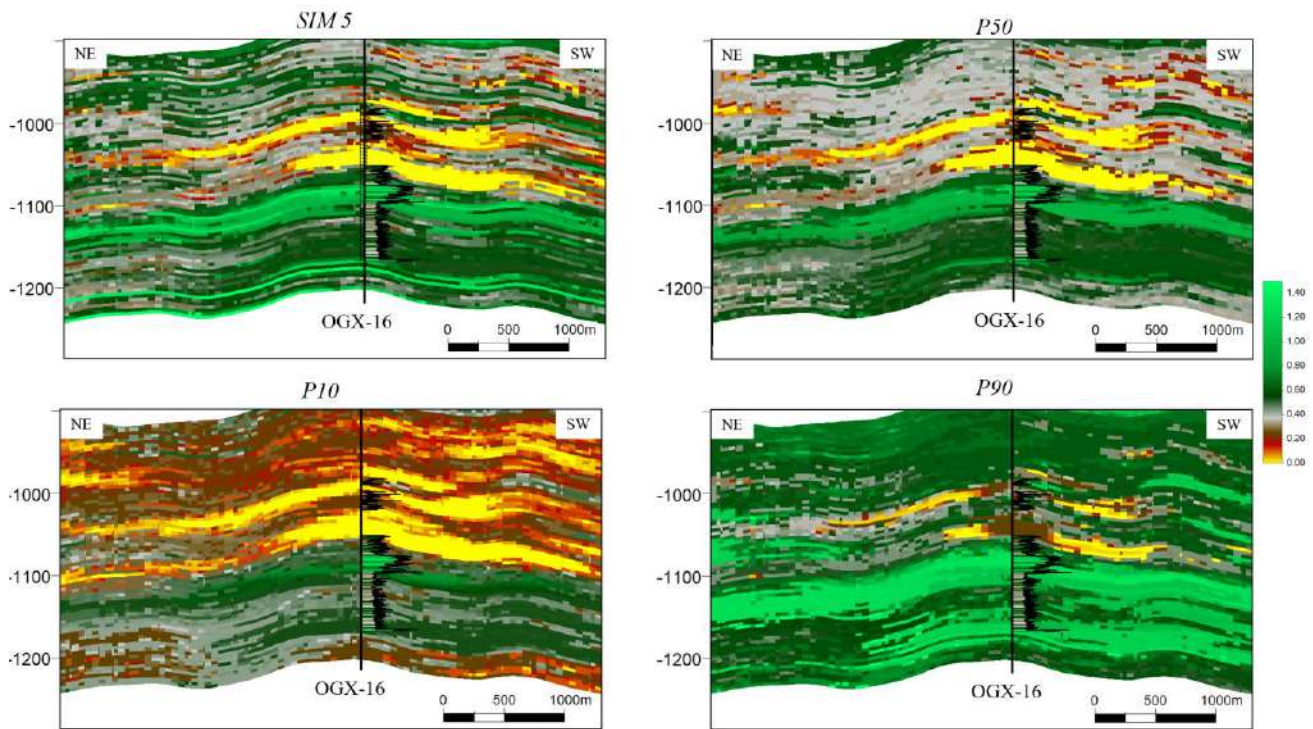


Figura 10.11. Seções dos modelos de COT passando pelo poço OGX-16 mostrando o resultado da simulação de maior correlação (SIM5) e os cenários: provável (P50), pessimista (P10) e otimista (P90).

A comparação entre os três percentis (P10, P50 e P90) revela o intervalo de incerteza geoestatística do modelo, permitindo identificar regiões mais confiáveis (onde as diferenças entre os percentis são pequenas) e zonas de maior variabilidade (onde o COT flutua significativamente entre P10 e P90). Essa análise evidencia que o EMBER apresenta maior estabilidade nos níveis mais rico em COT e nos intervalos que representam evidentemente as soleiras de diabásio (COT = 0.0%) da Formação Pimenteiras, enquanto as zonas de contato com as intrusões ígneas exibem maior incerteza e dispersão, possivelmente devido ao efeito térmico, ao gradiente de maturação e à heterogeneidade litológica local. De modo geral, a análise integrada das seções demonstra que o modelo reproduz de forma consistente a distribuição espacial da riqueza orgânica, que podem auxiliar na distinção das rochas ricas (fácies 4), moderadas (fácies 3 e 4 transicionais) e pobres (fácies 1 e 2) em COT. A consistência observada entre as simulações e as eletrofácies demonstra que o modelo de COT obtido pelo EMBER é geologicamente consistente e estatisticamente robusto, fornecendo uma base sólida para a avaliação integrada do potencial gerador e do grau de maturação térmica da Formação Pimenteiras na Bacia do Parnaíba, ao mesmo tempo em que incorpora de forma explícita as incertezas inerentes ao processo de modelagem geoestatística.

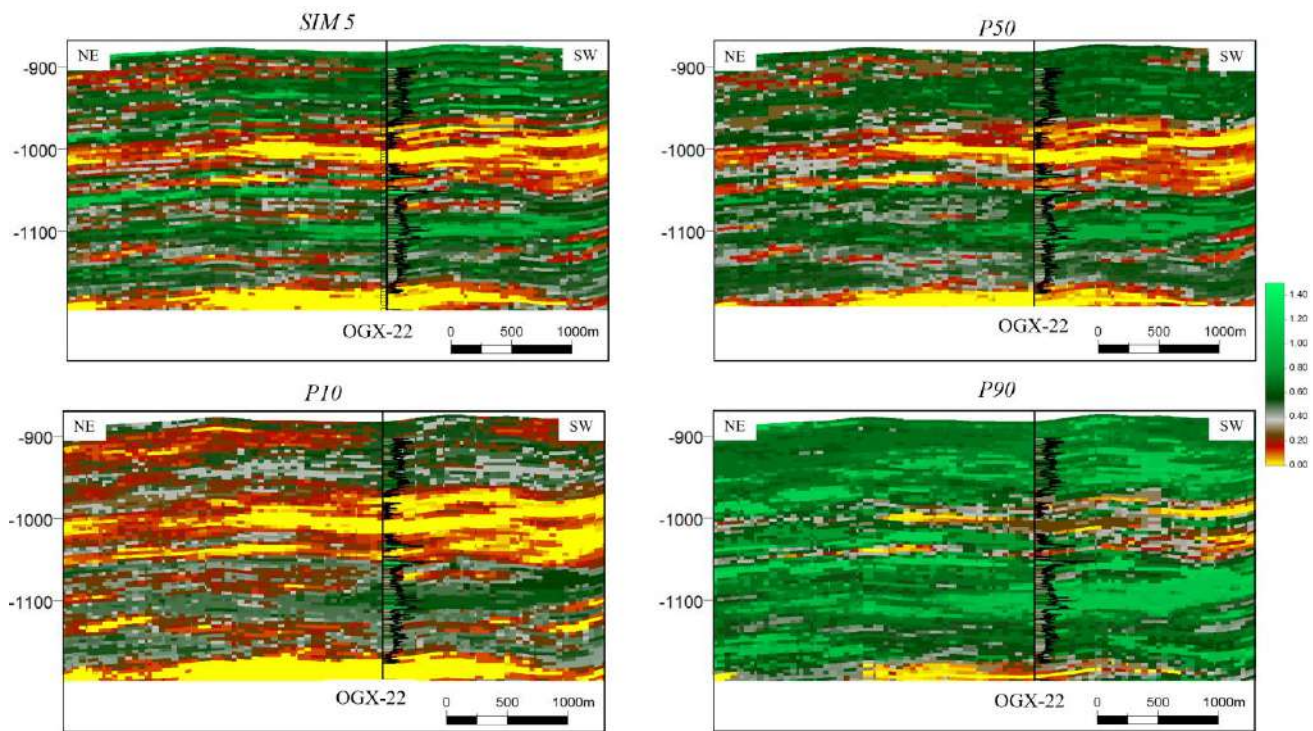


Figura 10.12. Seções dos modelos de COT passando pelo poço OGX-22 mostrando o resultado da simulação de maior correlação (SIM5) e os cenários: provável (P50), pessimista (P10) e otimista (P90).

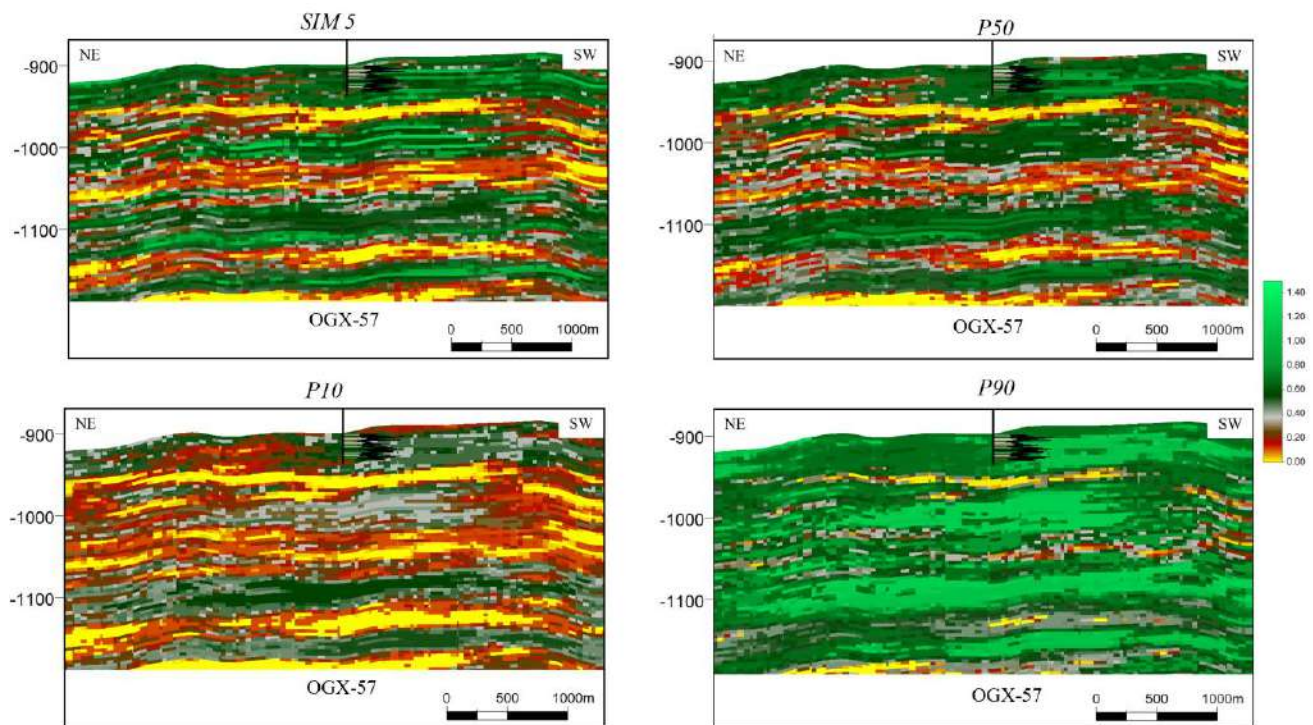


Figura 10.13. Seções dos modelos de COT passando pelo poço OGX-57 mostrando o resultado da simulação de maior correlação (SIM5) e os cenários: provável (P50), pessimista (P10) e otimista (P90).

10.5. DISCUSSÕES

A Figura 10.14 apresenta a seção sísmica atravessando os poços CP-1, OGX-16, OGX-22 e OGX-57, exibindo o modelo de COT gerado pelo algoritmo Embedded Model Estimator (EMBER), com e sem a interpretação dos principais intervalos enriquecidos em carbono orgânico. Três níveis de maior concentração de COT foram identificados e denominados N1, N2 e N3, os quais apresentam boa continuidade lateral e correspondem a zonas de baixa impedância acústica, indicando associação com folhelhos ricos em matéria orgânica da Formação Pimenteiras.

Esses três intervalos podem ser correlacionados com os folhelhos radioativos A, B e C descritos por René Rodrigues (1995) no perfil geoquímico da Figura 14a. Segundo o autor, o folhelho A corresponde a depósitos localizados na base do trato de sistemas transgressivos; o folhelho B integra a sequência transgressiva principal; e o folhelho C representa o momento de inundação marinha máxima do Devoniano, quando as condições anóxicas foram mais intensas. Essa correlação entre os níveis N1–N3 do modelo e os folhelhos A–C de Rodrigues evidencia consistência estratigráfica e geoquímica, sugerindo que o modelo de COT gerado pelo EMBER reproduz de forma coerente a distribuição vertical e lateral dos principais níveis geradores da Formação Pimenteiras.

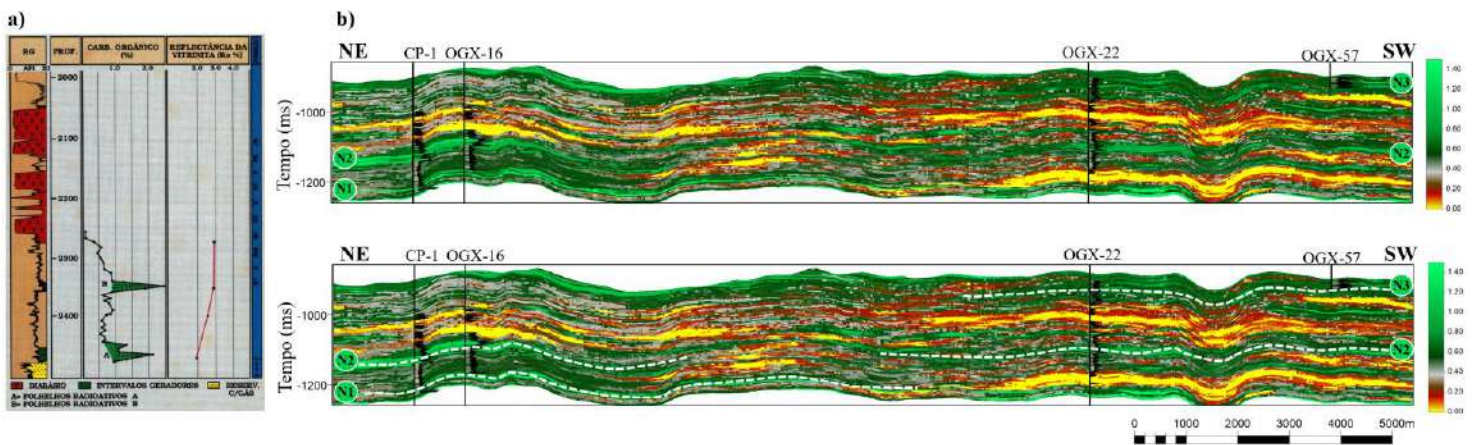


Figura 10.14. Seção arbitrária atravessando os CP-1, OGX-16, OGX-22 e OGX-57 com o modelo de COT (com e sem interpretação), destacando três níveis enriquecidos em COT (N1, N2, N3) de boa continuidade lateral; correlação estratigráfica proposta com os folhelhos radioativos A, B e C de Rodrigues (1995).

Do ponto de vista geoquímico, os intervalos com COT superior a 1.0 % representam as principais zonas de rocha geradora, associadas a ambientes de plataforma em momentos de baixa energia e com forte preservação da matéria orgânica. Já os horizontes com teores intermediários de COT (0.5 % - 1.0 %) indicam rochas de potencial gerador moderado, enquanto os valores inferiores a 0.5 % refletem empobrecimento orgânico por oxidação, diluição clástica nos períodos de tempestades ou efeito térmico das soleiras de diabásio. A distribuição probabilística obtida pelo EMBER reforça essa

interpretação, delineando zonas de alta probabilidade de $COT > 1 \%$, interpretadas como os principais *sweet spots* geradores, e faixas intermediárias com potencial secundário de geração.

A coerência entre os resultados de COT e os padrões sísmicos-geoquímicos descritos em Rodrigues (1995) atesta o potencial do EMBER em capturar o comportamento não estacionário e as heterogeneidades verticais e laterais do sistema gerador. O modelo obtido reproduz de forma consistente as variações associadas às fácies 4 dos poços (ricas em matéria orgânica) e demonstra elevada sensibilidade na detecção de zonas de baixo COT nas proximidades das intrusões ígneas, confirmando o efeito térmico local sobre a degradação do querogênio, associado a um aumento da impedância acústica nesses intervalos.

Os resultados deste trabalho corroboram as observações de Ferreira et al. (2021), que aplicaram o algoritmo EMBER na modelagem 3D da porosidade e permeabilidade da Formação Barra Velha (Campo de Búzios, Bacia de Santos). Nesse estudo, o EMBER demonstrou alta capacidade preditiva, preservando as distribuições estatísticas originais das propriedades e mapeando padrões verticais e espaciais coerentes com a estratigrafia, com alta correlação linear (73% a 82%) entre os dados originais e simulados. Além disso, os autores avaliaram as incertezas por meio das distribuições condicionais do *spread* (P90–P10), destacando que os intervalos de melhor qualidade do reservatório apresentavam também maior variabilidade e incerteza local. De forma análoga, no presente trabalho, o EMBER permitiu quantificar a incerteza na distribuição do COT e identificar zonas mais confiáveis, reforçando sua eficiência como ferramenta integrada de modelagem geoestatística assistida por *machine learning* tanto no contexto do reservatório carbonático da Barra Velha quanto na rocha geradora siliclástica da Formação Pimenteiras.

Os resultados obtidos neste estudo reforçam o papel da impedância acústica como variável-chave na caracterização indireta do conteúdo orgânico total (COT), corroborando os achados de Løseth et al. (2011) e del Monte et al. (2018), que identificaram uma correlação inversa robusta entre o COT e a impedância em rochas geradoras espessas e ricas em querogênio. Assim como observado na Formação Pimenteiras, a redução da impedância acústica reflete a maior presença de material orgânico.

Entretanto, a análise realizada neste trabalho demonstra que a relação entre COT e impedância não é linear, conforme também enfatizado por Thiéblemont et al. (2022). De fato, a resposta acústica das rochas geradoras é fortemente modulada por fatores como soterramento, maturação térmica e variações mineralógicas locais. Na Formação Pimenteiras, essa variabilidade é intensificada pela influência térmica das intrusões ígneas, que promovem maturação diferenciada e degradação do querogênio, elevando localmente a impedância e reduzindo o COT. Esse comportamento foi detectado

de maneira clara no modelo EMBER, que apresentou distribuições heterogeneidade associadas a processos de alteração térmica, como evidenciados pelos volumes de probabilidade e pelos percentis P10, P50 e P90, resultado consistente com a interpretação de Davison et al. (2018) para margens afetadas por intrusões magmáticas no Atlântico Sul, onde a presença de SDRs basálticos e soleiras limitou a deposição e preservação das rochas geradoras sin-rifte.

A abordagem utilizada neste estudo aproxima-se conceitualmente das estratégias propostas por del Monte et al. (2018) e Cyz & Azevedo (2020), que combinaram inversões sísmicas quantitativas e modelagens geoestatísticas diretas para estimar propriedades orgânicas e elásticas de folhelhos. No presente trabalho, a integração do volume de impedância acústica proveniente da inversão estocástica como variável secundária dentro do algoritmo EMBER permitiu obter uma representação espacial contínua do COT, evitando a suavização excessiva típica de métodos determinísticos. O resultado foi um modelo tridimensional capaz de honrar simultaneamente o condicionamento local dos poços e as tendências regionais do sistema deposicional.

Outro ponto de convergência entre o presente estudo e o de Ouadfeul & Aliouane (2016) é o uso de algoritmos de *machine learning* para predição indireta de propriedades orgânicas. Assim como a rede neural MLP empregada na inversão genética pós-stack do Barnett Shale, o EMBER utiliza aprendizado não paramétrico, neste caso, via *Quantile Random Forests*, para capturar relações não lineares complexas entre a variável alvo (COT) e a variável sísmica secundária (AI). No entanto, o EMBER supera as abordagens puramente neurais por integrar explicitamente a continuidade espacial através dos modelos de krigagem embutidos, conciliando a flexibilidade do *machine learning* com o rigor da geoestatística clássica. Essa característica explica a boa coerência entre as simulações estocásticas e os dados de poço observada no poço cego CP-1, com coeficiente de correlação de 0,80 e RMSE inferior a 0,19, o que indica excelente desempenho preditivo dentro do domínio estudado.

Assim, o presente estudo demonstra que a integração entre a impedância acústica e o algoritmo EMBER representa uma evolução metodológica em relação a caracterização de rochas geradoras. O método permite não apenas estimar o COT de forma quantitativa e espacialmente contínua, mas também quantificar a incerteza associada, fornecendo um envelope estatístico que traduz de forma explícita o grau de confiança em cada porção do modelo.

Apesar dos resultados satisfatórios obtidos, é importante reconhecer algumas limitações inerentes à modelagem apresentada. A principal delas refere-se à utilização de apenas uma variável secundária, a impedância acústica proveniente da inversão estocástica, sem a inclusão de outros atributos sísmicos potencialmente correlacionados ao COT, como amplitude instantânea, fase,

frequência dominante ou *sweetness*. A decisão metodológica de restringir o modelo a uma única variável teve como propósito central destacar, com maior resolução vertical e coerência geológica, os níveis mais ricos em matéria orgânica dentro da Formação Pimenteiras. No entanto, a ausência de uma análise sistemática da relação entre o COT e outros atributos sísmicos pode limitar o potencial do algoritmo utilizado, que pode gerar resultados ainda melhores a medida que um número maior de *training features* é inserido.

Adicionalmente, a quantidade reduzida de poços disponíveis na área de estudo constitui um fator de incerteza relevante, uma vez que o número restrito de *hard data* compromete o controle amostral necessário para a calibração estatística e espacial do modelo. Essa limitação é particularmente sensível em regiões com forte heterogeneidade litológica, como aquelas afetadas por intrusões ígneas. Ainda assim, ressalta-se que, embora as soleiras de diabásio aumentem a complexidade geológica local, a Formação Pimenteiras caracteriza-se por boa continuidade lateral das camadas, o que favoreceu a consistência da modelagem geoestatística.

Outro aspecto a ser considerado é a ausência de um modelo tridimensional de fácies sedimentares, que poderia atuar como uma variável secundária adicional para condicionar a distribuição do COT e aprimorar a representação espacial das zonas de enriquecimento orgânico. A incorporação futura desse tipo de informação, aliada ao uso de múltiplos atributos sísmicos tem potencial para reduzir as incertezas e aperfeiçoar a acurácia da modelagem, permitindo uma caracterização ainda mais detalhada e realista das rochas geradoras na Formação Pimenteiras.

10.6. CONCLUSÃO

Este trabalho demonstrou a viabilidade e a robustez da predição tridimensional do Conteúdo Orgânico Total (COT) na Formação Pimenteiras a partir da integração entre impedância acústica invertida e o algoritmo Embedded Model Estimator (EMBER). A relação inversa e não linear entre COT e impedância, capturada por um ajuste polinomial, foi adequadamente explorada pelo *framework* do EMBER, que gera simulações estocásticas consistentes com a continuidade espacial imposta pelos modelos de krigagem embutidos, além de distribuições condicionais (P10, P50, P90). O resultado é um modelo que honra os dados de poço, evita a suavização excessiva típica de abordagens determinísticas e quantifica explicitamente a incerteza via spread (P90–P10).

Do ponto de vista quantitativo, a realização selecionada (SIM5) apresentou desempenho elevado no poço cego CP-1 ($R \approx 0,80$; $RMSE \approx 0,19$; $MAE \approx 0,14$), ao mesmo tempo em que o conjunto de realizações preservou estatísticas realistas do campo (média $\approx 0,48$; desvio-padrão $\approx 0,22$). A

distribuição do spread concentrou-se em valores moderados ($\sim 0,4-0,5$), com aumentos localizados em áreas de menor controle de dados e nas proximidades de intrusões ígneas, zonas onde a heterogeneidade geológica e os efeitos térmicos tendem a ampliar a variabilidade preditiva. Em termos probabilísticos, os volumes de ocorrência por classes de COT (pobre $<0,5\%$; moderado $0,5-1,0\%$; rico $>1,0\%$) podem fornecer uma leitura operacional, permitindo destacar “*sweet spots*” geradores com continuidade lateral e identificar setores menos favoráveis, com utilidade direta para avaliação de sistemas petrolíferos e priorização exploratória.

A consistência geológica do modelo decorre de três linhas de evidência: (i) o alinhamento entre níveis de baixa impedância e enriquecimento orgânico, (ii) a correspondência estratigráfica entre os três intervalos mais ricos (N1–N3) e os folhelhos radioativos A–C descritos na literatura regional, e (iii) a resposta espacial do COT diante do condicionamento estrutural e térmico imposto pelas soleiras de diabásio, que reduzem o teor orgânico e elevam a impedância nas zonas de influência. Em conjunto, esses achados consolidam o papel da impedância acústica como variável-chave para caracterização sísmica indireta de rochas geradoras e evidenciam a vantagem do EMBER ao conciliar aprendizado não paramétrico com continuidade geoestatística.

Como toda modelagem baseada em dados reais, há limitações a reconhecer. A principal é o uso deliberado de uma única variável secundária (impedância), adotada para maximizar a resolução dos níveis ricos em COT. A avaliação sistemática de outros atributos sísmicos (p. ex., amplitude/fase instantânea, frequência dominante, *sweetness*, atributos pré-empilhamento/AVO) não foi objeto deste estudo e pode ampliar o poder preditivo. A escassez de poços na área reduz o controle amostral e tende a aumentar a incerteza local; além disso, a ausência de um modelo 3D de fácies sedimentares impediu o condicionamento estratigráfico explícito do COT, e o processo de *upscaling* pode introduzir valores residuais em fácies onde idealmente o COT deveria ser nulo.

Recomenda-se, portanto, como direções de avanço: (a) incorporar múltiplos atributos sísmicos, including pré-empilhamento, como *training features* adicionais, (b) desenvolver um modelo 3D de fácies para atuar como covariável/*soft constraint*, (c) integrar modelos de maturação térmica e proximidade às intrusões para capturar gradientes termo-estruturais, e (e) comparar o EMBER com abordagens complementares, com protocolos de validação cruzada espacial.

Em síntese, a abordagem aqui apresentada comprova que a combinação entre inversão estocástica de impedância e EMBER permite (i) estimar COT em escala sísmica com aderência estatística e geológica, (ii) mapear a conectividade lateral de horizontes geradores e (iii) explicitar a incerteza espacial de forma operacional. Esses resultados fornecem uma base técnica sólida para a

caracterização de rochas geradoras, ao mesmo tempo em que delineiam um caminho claro para o refinamento do modelo por meio, principalmente do condicionamento por fácies que foi realizado em 1D, nos poços, mas que seria interessante obter essa informação no volume 3D.

11. Conclusões

O desenvolvimento desta tese resultou na construção de um fluxo metodológico integrado voltado à caracterização sísmica e modelagem tridimensional do teor de Carbono Orgânico Total (COT) na Formação Pimenteiras, Bacia do Parnaíba. A pesquisa combinou diferentes domínios do conhecimento, geoquímica, geofísica e ciência de dados, com o propósito de aprimorar a compreensão sobre a distribuição espacial das rochas geradoras em um sistema petrolífero atípico, fortemente influenciado por intrusões ígneas.

11.1 SÍNTESE DOS PRINCIPAIS RESULTADOS

Os resultados demonstram que a metodologia proposta foi eficaz em todas as suas etapas, desde o pré-condicionamento sísmico até a modelagem 3D do COT, proporcionando uma representação espacialmente coerente das zonas ricas em matéria orgânica e evidenciando ganhos significativos de resolução e confiabilidade interpretativa.

O estudo da matéria orgânica e dos processos de formação e preservação das rochas geradoras permitiu consolidar o entendimento sobre o papel da Formação Pimenteiras como principal rocha geradora da Bacia do Parnaíba. Observou-se que o querogênio é predominantemente do tipo II–III, indicando origem mista marinha e continental, com potencial gerador para óleo e gás. A influência térmica das intrusões ígneas sobre a maturação orgânica foi identificada como fator determinante na heterogeneidade do potencial gerador.

As análises de *machine learning* não supervisionado (*K-Means*) e de classificação Bayesiana possibilitaram discriminar eletrofácies nos folhelhos da Formação Pimenteiras com base nas respostas dos perfis de poço. Essa diferenciação evidenciou a existência de padrões de impedância acústica correlacionados ao teor de COT, confirmando a viabilidade da utilização da impedância como variável preditiva para a modelagem.

Os modelos supervisionados, Random Forest, XGBoost, CatBoost, apresentaram desempenho satisfatório na predição do COT, com coeficientes de correlação superior a 0,7 no poço teste. Essa abordagem permitiu estimar curvas contínuas de COT em poços sem dados laboratoriais, assegurando maior representatividade espacial da variável. A integração de diferentes algoritmos mostrou-se eficaz para minimizar vieses individuais e aumentar a generalização dos resultados.

A combinação dos métodos Dip-Steered Median Filter e Spectral Blueing produziu melhoria significativa na qualidade do dado sísmico, com aumento na relação sinal/ruído e ganho de até 38% na resolução vertical. A filtragem orientada por mergulho manteve a continuidade estrutural dos eventos e reduziu artefatos associados ao ruído aleatório, enquanto o *Spectral Blueing* compensou a atenuação de altas frequências, tornando o dado mais adequado para aplicações quantitativas de inversão e modelagem.

Os resultados da análise espectral confirmaram a eficiência do pré-condicionamento, demonstrando ampliação do espectro de frequência útil e melhoria na definição dos refletores da Formação Pimenteiras. O estudo de resolução sísmica indicou que camadas com espessura inferior a 15 ms puderam ser parcialmente resolvidas, tornando possível a identificação de variações sutis associadas a níveis mais ricos em matéria orgânica.

A amarração sísmica-poço recalculada para o dado pré-condicionado apresentou correlações mais elevadas entre traço sintético e sísmico, validando o ganho de qualidade obtido no processamento. A reinterpretação do topo e base da Formação Pimenteiras e a construção do *grid* tridimensional garantiram a consistência geométrica necessária para a inversão sísmica e posterior modelagem do COT.

As inversões determinística (*model-based*), colorida e estocástica resultaram em volumes de impedância coerentes e complementares. O modelo determinístico produziu uma distribuição suavizada, adequada para controle de tendência, enquanto a inversão estocástica gerou um volume de alta resolução que preservou as variações locais associadas a camadas delgadas e heterogêneas. A impedância acústica absoluta apresentou alta correlação inversa de grau 2° com o teor de COT, confirmando sua aplicabilidade como dado secundário na modelagem 3D.

A aplicação do método EMBER permitiu integrar dados de poço e volume sísmico em um modelo espacial contínuo, reproduzindo de forma realista a distribuição do COT e suas incertezas. Os resultados mostraram coerência geológica e estatística, com boa correspondência entre valores simulados e observados. A análise probabilística (P10, P50 e P90) demonstrou maior incerteza em áreas afastadas dos poços e em regiões com interferência de soleiras, refletindo o impacto do controle amostral e das intrusões na confiabilidade da modelagem.

11.2 CONTRIBUIÇÕES E AVANÇOS METODOLÓGICOS

Esta pesquisa contribui de forma inédita para a caracterização de rochas geradoras em bacias intracratônicas com sistemas petrolíferos atípicos, ao propor um fluxo metodológico integrado que

combina técnicas de processamento sísmico avançado, *machine learning* e modelagem geoestatística. Entre as principais contribuições destacam-se:

1. Desenvolvimento de um pré-condicionamento sísmico híbrido (DSMF + SSB) em dados terrestres, com comprovação quantitativa do ganho de resolução;
2. Aplicação inédita do algoritmo EMBER para a modelagem tridimensional do COT na Bacia do Parnaíba, representando um avanço na integração entre *machine learning* e geoestatística;
3. Implementação de uma etapa robusta de validação e quantificação de incertezas (P10, P50, P90), proporcionando melhor avaliação da confiabilidade espacial dos resultados;
4. Adaptação de metodologias comumente aplicadas em reservatórios para o contexto de rochas geradoras, contribuindo para uma nova perspectiva de estudos integrados em sistemas fonte-reservatório.

11.3 LIMITAÇÕES E PERSPECTIVAS FUTURAS

Apesar dos resultados promissores, algumas limitações devem ser reconhecidas. O número reduzido de poços disponíveis impôs restrições ao controle direto do modelo, ampliando a incerteza em regiões menos amostradas. A ausência de atributos sísmicos adicionais, que poderiam atuar como variáveis secundárias na modelagem, também limitou a captura de algumas variações laterais sutis do COT.

Sugere-se que trabalhos futuros explorem:

1. A incorporação de atributos sísmicos derivados da inversão elástica e análise multiatributo, para refinar a sensibilidade da modelagem ao COT;
2. A utilização de modelos 3D de fácies para condicionar a simulação da modelagem do COT e reduzir incertezas em áreas com controle limitado, intervindo no modelo de acordo com o maior conhecimento da área de estudo;
3. A integração de dados térmicos e modelagem de maturação para melhor representar os efeitos térmicos das intrusões ígneas sobre o potencial gerador.

11.4 CONSIDERAÇÕES FINAIS

O fluxo metodológico desenvolvido demonstrou ser tecnicamente consistente e aplicável a diferentes contextos geológicos, constituindo uma ferramenta robusta para a predição e modelagem de

propriedades orgânicas em rochas geradoras. A aplicação na Formação Pimenteiras comprovou a viabilidade da abordagem e reforçou a importância da integração entre sísmica, dados de poço e inteligência artificial na avaliação quantitativa de sistemas petrolíferos complexos.

12. Referências

12.1. CAPÍTULO 1

- Almeida, F. F. M., de Brito Neves, B. B., & Carneiro, C. D. R. (2000). The origin and evolution of the South American Platform. *Earth-Science Reviews*, 50(1-2), 77-111.
- Carcione, J. M., & Avseth, P. (2015). Rock-physics templates for clay-rich source rocks. *Geophysics*, 80(5), D481-D500.
- Cioccari, G. M., & Mizusaki, A. M. P. (2019). Sistemas petrolíferos atípicos nas bacias paleozoicas brasileiras—uma revisão. *Geosciences= Geociências*, 38(2), 367-390.
- de Andrade, C. L. N., Cardoso, T. R. M., Santos, R. R., Dino, R., & de Jesus Machado, A. (2020). Organic facies and palynology from the middle to late Devonian of the Pimenteiras Formation, Parnaíba Basin, Brazil. *Journal of South American Earth Sciences*, 99, 102481.
- del Monte, A. A., Antonielli, E., De Tomasi, V., Luchetti, G., Papanozzi, E., & Gambacorta, G. (2018). Methods for source rock identification on seismic data: An example from the Tanezzuft Formation (Tunisia). *Marine and Petroleum Geology*, 91, 108-124.
- Løseth, H., Wensaas, L., Gading, M., Duffaut, K., & Springer, M. (2011). Can hydrocarbon source rocks be identified on seismic data?. *Geology*, 39(12), 1167-1170.
- Magoon, L. B., & Dow, W. G. (Eds.). (1994). *The Petroleum System—From Source to Trap*. AAPG Memoir 60. Tulsa, OK: American Association of Petroleum Geologists. 655 p.
- Miranda, F. S. D. (2014). *Caracterização geológica da Formação Pimenteiras como potencial reservatório do tipo shale-gas (Devoniano da Bacia do Parnaíba)*. Rio de Janeiro.
- Mussa, A. (2020). *O Potencial de geração de hidrocarbonetos na Formação Pimenteiras, Bacia do Parnaíba*. Tese de Doutorado em Geociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.
- Reis, M. A. A. D. A. D., Damasceno, A. C., Roriz, C. E. D., Korenchandler, A. L., Silva, A. M. D., Praxedes, E. D. S., & Silva, V. G. (2023). Source rock evaluation from rock to seismic data: an integrated machine-learning-based work flow and application in the Brazilian Presalt (Santos Basin). *Minerals*, 13(9), 1179.
- Rodrigues, R. (1995). A geoquímica orgânica na Bacia do Parnaíba. *Revista Brasileira de Geociências*, 25, 145-146.

Rudolph, K. W., & Goulding, F. J. (2017). Benchmarking exploration predictions and performance using 20+ yr of drilling results: One company's experience. *AAPG bulletin*, 101(2), 161-176.

Souza, A. C. B. D., Esteves, M. C. B., & Nascimento Junior, D. R. D. (2017). Geoquímica inorgânica e orgânica dos folhelhos da Formação Pimenteiras: implicações para um sistema petrolífero não convencional.

Tissot, B. P., & Welte, D. H. (1984). *Petroleum formation and occurrence—second revised and enlarged edition*. SpringerVerlag Berlin Heidelberg, New York, 160-198.

12.2. CAPÍTULO 2

Aguiar, G. D. (1971). Revisão geológica da bacia paleozóica do Maranhão. Congresso.

Albuquerque, O. D., & Dequech, V. I. C. T. O. R. (1946). Contribuição para a geologia do Meio-Norte, especialmente Piauí e Maranhão, Brasil. In Congresso Pan-Americano de Engenharia de Minas e Geologia (Vol. 1, No. 1946, pp. 69-108).

Almeida, F. D., & Carneiro, C. D. R. (2004). Inundações marinhas fanerozóicas no Brasil e recursos minerais associados. Mantesso Neto, V.; Bartorelli, A.; Carneiro, CDR, 43-60.

Almeida, F. F. M., de Brito Neves, B. B., & Carneiro, C. D. R. (2000). The origin and evolution of the South American Platform. *Earth-Science Reviews*, 50(1-2), 77-111.

Brito Neves, B. B., & Fuck, R. A. (2014). The basement of the South American platform: Half Laurentian (N-NW)+ half Gondwanan (E-SE) domains. *Precambrian Research*, 244, 75-86.

Campanha, V. A., and Mabesoone, J. M. (1976). Paleoambiente e Paleocologia do Membro Picos, Formação Pimenteiras (Devoniano do Piauí): *Soc. Bras. Geol., An. do XXVIII Congresso Bras. de Geol.*, v. 2, p. 221-235.

Caputo, M. V. (1984). Stratigraphy, tectonics, paleoclimatology and paleogeography of northern basins of Brazil. University of California, Santa Barbara.

Caputo, M. V., & Lima, E. C. (1984). Estratigrafia, idade e correlação do Grupo Serra Grande–Bacia do Parnaíba. In Congresso Brasileiro de Geologia (Vol. 33, No. 1984, pp. 740-753).

Castro, D. L., Fuck, R. A., Phillips, J. D., Vidotti, R. M., Bezerra, F. H., & Dantas, E. L. (2014). Crustal structure beneath the Paleozoic Parnaíba Basin revealed by airborne gravity and magnetic data, Brazil. *Tectonophysics*, 614, 128-145.

Chamani, M. A. C. (2015). Tectônica sinsedimentar no Siluro-Devoniano da Bacia do Parnaíba, Brasil: o papel de grandes estruturas do embasamento na origem e evolução de bacias intracratônicas.

Copper, P. (1977). Paleolatitudes in the Devonian of Brazil and the Frasnian-Famennian mass extinction: *Palaeogeography, Palaeoclimatology, Palaeoecology*, v. 21, p. 165-207.

- Cordani, U. G., & Sato, K. (1999). Crustal evolution of the South American Platform, based on Nd isotopic systematics on granitoid rocks. *Episodes Journal of International Geoscience*, 22(3), 167-173.
- Cordani, U. G., Pimentel, M. M., Araújo, C. E. G. D., & Fuck, R. (2013). The significance of the Transbrasiliiano-Kandi tectonic corridor for the amalgamation of West Gondwana. *Brazilian journal of Geology*, 43(3), 583-597.
- Cordani, U. G., Sato, K., Teixeira, W., Tassinari, C. C. G., & Basei, M. A. S. (2000). Crustal evolution of the South American platform. *Tectonic Evolution of South America*, 19-40.
- Daly, M. C., Fuck, R. A., Julià, J., Macdonald, D. I., & Watts, A. B. (2018). Cratonic basin formation: a case study of the Parnaíba Basin of Brazil.
- Della Fávera, J. C. (1990). *Tempestitos da bacia do Parnaíba*. Porto Alegre, 243.
- de Oliveira, D. C., & Mohriak, W. U. (2003). Jaibaras trough: an important element in the early tectonic evolution of the Parnaíba interior sag basin, Northern Brazil. *Marine and Petroleum geology*, 20(3-4), 351-383.
- Dow, W. G. (1994). *The petroleum system. I. B. magoon*; WG Dow (eds.).
- Ernesto, M., Bellieni, G., Piccirillo, E. M., Marques, L. S., De Min, A., Pacca, I. G., ... & Macedo, J. W. P. (2003). Paleomagnetic and geochemical constraints on the timing and duration of the CAMP activity in northeastern Brazil. *The Central Atlantic Magmatic Province: Insights from Fragments of Pangea*, 136, 129-149.
- Ferraz, N. C. (2015). *Análise estratigráfica da sequência mesodevoniana-eocarbonífera da Bacia do Parnaíba, nordeste do Brasil* (Master's thesis, Universidade Federal do Rio Grande do Norte).
- Góes, A. M. (1995). *Formação Poti (carbonífero inferior) da Bacia do Parnaíba* (Doctoral dissertation, Universidade de São Paulo).
- Góes, A. M. O., de Souza, J. M. P., & Teixeira, L. B. (1990). Estágio exploratório e perspectivas petrolíferas da Bacia do Parnaíba. *Boletim de Geociências da PETROBRAS*, 4(1), 55-64.
- Góes, A. M. O.; Feijó, F. J. (1994). *Bacia do Parnaíba*. *Boletim de Geociências da Petrobras*, Rio de Janeiro, v. 8, n. 1, p. 57-68.
- Goulart, J. P. D. M. (2019). *Caracterização de chaminés de gás nas bacias do Parnaíba, Paraná e Tacutu: uma nova ferramenta exploratória* (Master's thesis, Brasil).
- Jerram, D., Pertord, N., (2011). *Field description of Igneous Rocks*. Geological field guide series. Ed. John Wiley & Sons, 2º ed.
- Johnson, A. M.; Pollard, D. D. (1973) Mechanics of growth of some laccolithic intrusions in the Henry Mountains, Utah, I: field observations, Gilbert's model, physical properties and flow of the Magma. *Tectonophysics*, Amsterdam, v. 18, n. 3-4, p. 261-309.

- Junior, I. T., de Moraes Neto, J. M., Santos, S. F., & Portela Filho, C. V. (2014). Diques e soleiras na Bacia do Parnaíba: geometria e padrões de alojamento. *Boletim de Geociências da Petrobras*, 22(2), 261-287.
- Klein, G. de V. (1982). Probable sequential arrangement of depositional systems on cratons: *Geology*, v. 10, p. 17-22.
- Lima, E. D. A. M., & Leite, J. F. (1978). Projeto estudo global dos recursos Minerais da Bacia sedimentar do Parnaíba. Integração geológico-metalogenética. CPRM.
- Mabesoone, J. M. (1977). Paleozoic-Mesozoic deposits of the Piauí-Maranhão syncline (Brazil): Geological history of a sedimentary basin: *Sedim. Geol.* v, 19, p. 7-38.
- Marini, O. J., Fuck, R. A., Danni, J. C. M., Dardenne, M. A., Loguercio, S. O. C., & Ramalho, R. (1984). As faixas de dobramentos Brasília, Uruaçu e Paraguai-Araguaia e o maciço mediano de Goiás. *Geologia do Brasil*, 251-303.
- Miranda, F. S., Vettorazzi, A. L., Cunha, P. R. D. C., Aragão, F. B., Michelon, D., Caldeira, J. L., & Andreola, K. (2018). Atypical igneous-sedimentary petroleum systems of the Parnaíba Basin, Brazil: seismic, well logs and cores. *Geological Society, London, Special Publications*, 472(1), 341-360.
- Magoon, L. B., & Dow, W. G. (Eds.). (1994). *The Petroleum System—From Source to Trap*. AAPG Memoir 60. Tulsa, OK: American Association of Petroleum Geologists. 655 p.
- Planke, S., Rasmussen, T., Rey, S. S., & Myklebust, R. (2005). Seismic characteristics and distribution of volcanic intrusions and hydrothermal vent complexes in the Vøring and Møre basins.
- Pollard, D. D.; Johnson, A. (1973) Mechanics of growth of some laccolithic intrusions in the Henry Mountains, Utah, II: bending and failure of overburden layers and sill formation. *Tectonophysics*, Amsterdam, v. 18, n. 3-4, p. 311-354.
- Praxedes, I. F. (2015). Lineamento Transbrasiliano, contribuição aerogeofísica, tectônica e geocronológica no setor nordeste da Faixa Brasília.
- Ribeiro, C. C., and Dardenne (1978). O minério de Ferro da Formação Pimenteiras na Borda SW da Bacia do Maranhão (Goiás): *Soc. Bras. Geol., An. XXX Congr. Bras. Geol., Recife*, v.4, p. 1583-1595.
- Rodrigues, R. (1995). A geoquímica orgânica na Bacia do Parnaíba. *Revista Brasileira de Geociências*, 25, 145-146.
- Rossetti, D. F., & Góes, A. M. (2003). Caracterização paleoambiental de depósitos albianos na borda Sul da Bacia de São Luís-Grajaú: modelo de delta fluvial influenciado por tempestade. *Revista Brasileira de Geociências*, 33(3), 299-312.
- Rubin, A. M. (1995). Propagation of magma-filled cracks. *Annual Review of Earth and Planetary Science*, Palo Alto, v. 23, p. 287-336.

Senger, K., Millett, J., Planke, S., Ogata, K., Eide, C. H., Festøy, M., & Jerram, D. A. (2017). Effects of igneous intrusions on the petroleum system: a review. *First Break*, 35(6).

Small, H. L. (1914). *Geologia e Suprimento d'água subterrânea no Piauí e parte do Ceará* (No. 32). W. Ribeiro.

Tissot, B. P., & Welte, D. H. (2013). *Petroleum formation and occurrence*. Springer Science & Business Media.

Vaz, P. T., Wanderley Filho, J. R., & Travassos, W. A. S. (2007). Bacia do Parnaíba. *Boletim de Geociências da PETROBRAS*, 15(2), 253-263.

Vettorazzi, A. L. S. (2012). Caracterização sedimentológica dos arenitos da Formação Cabeças (Devoniano) na borda leste da Bacia do Parnaíba. Rio de Janeiro.

Zalán, P. V. (2004). Evolução fanerozóica das bacias sedimentares brasileiras. *Geologia do continente sul-americano: evolução da obra de Fernando Flávio Marques de Almeida*. São Paulo, Beca, 595-613.

12.3. CAPÍTULO 3

Allen, P. A., & Allen, J. R. (2013). *Basin analysis: Principles and application to petroleum play assessment*. John Wiley & Sons.

Bjorlykke, K. (2010). *Petroleum geoscience: From sedimentary environments to rock physics*. Springer Science & Business Media.

Gluyas, J. G., & Swarbrick, R. E. (2004). *Petroleum geoscience*. Blackwell Science Ltd.

Higley, D. K., Henry, M., Roberts, L. N., & Steinshouer, D. W. (2005). 1-D/3-D geologic model of the Western Canada sedimentary basin. *The Mountain Geologist*.

Huc, A. Y. (2003). Petroleum geochemistry at the dawn of the 21st century. *Oil & gas science and technology*, 58(2), 233-241.

Magoon, L. B., & Dow, W. G. (Eds.). (1994). *The Petroleum System—From Source to Trap*. AAPG Memoir 60. Tulsa, OK: American Association of Petroleum Geologists. 655 p.

Mussa, A. (2020). O Potencial de geração de hidrocarbonetos na Formação Pimenteiras, Bacia do Parnaíba. Tese de Doutorado em Geociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

Rodrigues, R. (1995). A geoquímica orgânica na Bacia do Parnaíba. *Revista Brasileira de Geociências*, 25, 145-146

Peters, K. E., Walters, C. C., & Moldowan, J. M. (2007). *The biomarker guide: Volume 1, Biomarkers and isotopes in the environment and human history*. Cambridge university press.

Peters, K. E., & Cassa, M. R. (1994). *Applied source rock geochemistry*.

Tissot, B. P., & Welte, D. H. (1984). *Petroleum formation and occurrence*. Springer Science & Business Media.

12.4. CAPÍTULO 4

Arthur, D., & Vassilvitskii, S. (2006). K-means++: The advantages of careful seeding. Stanford.

Bholowalia, P., & Kumar, A. (2014). EBK-means: A clustering technique based on elbow method and K-means in WSN. *International Journal of Computer Applications*, 105(9).

Bussab, W. D. O., & Morettin, P. A. (2010). *Estatística básica*. In *Estatística básica* (pp. xvi-540).

DeGroot, M. H., & Schervish, M. J. (2012). *Probability and statistics*. Addison-Wesley.

Duda, R. O., & Hart, P. E. (2006). *Pattern classification*. John Wiley & Sons.

Fernandes, F. J. D., Freire, A. F. M., & Lupinacci, W. M. (2022). Estudo de viabilidade para a classificação Bayesiana em reservatórios turbidíticos do Eoceno no Novo Campo de Jubarte, Bacia de Campos. IX SimBGf. Paraná, Brazil.

Glover, P. W., Mohammed-Sajed, O. K., Akyüz, C., Lorinczi, P., & Collier, R. (2022). Clustering of facies in tight carbonates using machine learning. *Marine and Petroleum Geology*, 144, 105828.

Iwayama, M., & Tokunaga, T. (1995, July). Cluster-based text categorization: a comparison of category search strategies. In *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 273-280).

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall, Inc..

Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognition Lett.* (2009), doi:10.1016/j.patrec.2009.09.011.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3), 264-323.

Kolmogorov, A. N. (2018). *Foundations of the theory of probability: Second English Edition*. Courier Dover Publications.

Madhulatha, T. S. (2012). An overview on clustering methods. arXiv preprint arXiv:1205.1117.

Mussa, A., Kalkreuth, W., Mizusaki, A. M. P., & da Silva, T. F. (2024). Geochemistry and petrography of shales from Brazilian Paleozoic basins: an example of the Parnaíba Basin, Devonian Pimenteiras Formation—A review. *Marine and Petroleum Geology*, 160, 106640.

Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.

Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.

Thorndike, R. L. (1953). Who belongs in the family?. *Psychometrika*, 18(4), 267-276.

Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2, pp. 131-160). Reading, MA: Addison-wesley.

12.5. CAPÍTULO 5

Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In *Noise reduction in speech processing* (pp. 1-4). Berlin, Heidelberg: Springer Berlin Heidelberg.

Binetti, M. S., Massarelli, C., & Uricchio, V. F. (2024). Machine learning in geosciences: A review of complex environmental monitoring applications. *Machine Learning and Knowledge Extraction*, 6(2), 1263-1280.

Bione, F. R., Venancio, I. M., Santos, T. P., Belem, A. L., Rangel, B. R., Souza, I. V., ... & Albuquerque, A. L. S. (2024). Estimating total organic carbon of potential source rocks in the Espírito Santo Basin, SE Brazil, using XGBoost. *Marine and Petroleum Geology*, 162, 106765.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford university press..

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: springer.

Bredesen, K., Jensen, E. H., Johansen, T. A., & Avseth, P. (2015). Seismic reservoir and source-rock analysis using inverse rock-physics modeling: A Norwegian Sea demonstration. *The Leading Edge*, 34(11), 1350-1355.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2), 121-167.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE). *Geoscientific model development discussions*, 7(1), 1525-1534.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.

da Costa Cabral, J., da Silva, C. A., de Souza, G., & Saporetto, C. M. (2025). Uso de Métodos de Aprendizado de Máquina e Algoritmo Genético para Predição de TOC e Classificação de Litologia. *VETOR-Revista de Ciências Exatas e Engenharias*, 35(1), e18357-e18357.

del Monte, A. A., Antonielli, E., De Tomasi, V., Luchetti, G., Paparozzi, E., & Gambacorta, G. (2018). Methods for source rock identification on seismic data: An example from the Tanezzuft Formation (Tunisia). *Marine and Petroleum Geology*, 91, 108-124.

- Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.
- Elkhatny, S. (2019). A self-adaptive artificial neural network technique to predict total organic carbon (TOC) based on well logs. *Arabian Journal for Science and Engineering*, 44(6), 6127-6137.
- Emelyanova, I., Pervukhina, M., Clennell, M. B., & Dewhurst, D. N. (2016). Applications of standard and advanced statistical methods to TOC estimation in the McArthur and Georgina basins, Australia. *The Leading Edge*, 35(1), 51-57.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc."
- Goliatt, L., Saporetti, C. M., & Pereira, E. (2023). Super learner approach to predict total organic carbon using stacking machine learning models based on well logs. *Fuel*, 353, 128682.
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1, No. 2). Cambridge: MIT press.
- Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1), 389-422.
- Hall, B. (2016). Facies classification using machine learning. *The Leading Edge*, 35(10), 906-909.
- Hassan, M., Hossin, A., & Combaz, A. (1976, June). Fundamentals of the differential gamma ray log-interpretation technique. In *SPWLA Annual Logging Symposium* (pp. SPWLA-1976). SPWLA.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.
- Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. Prentice Hall.
- Ho, T. K. (1995, August). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278-282). IEEE.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R* (Vol. 103). New York: springer.
- Khan, M. R., Kalam, S., Asad, A., & A. Abu-khamsin, S. (2023, March). Development of a Deterministic total organic carbon (Toc) predictor for shale reservoirs. In *SPE Middle East Oil and Gas Show and Conference* (p. D021S084R004). SPE.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

- Kleinberg, E. M. (1990). Stochastic discrimination. *Annals of Mathematics and Artificial intelligence*, 1(1), 207-239.
- Kleinberg, E. M. (1996). An overtraining-resistant stochastic modeling method for pattern recognition. *The annals of statistics*, 24(6), 2319-2349.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- Liu, W., Zhang, G., Cao, J., Zhang, J., & Yu, G. (2019). Combined petrophysics and 3D seismic attributes to predict shale reservoirs favourable areas. *Journal of Geophysics and Engineering*, 16(5), 974-991.
- Liu, X., Tian, Z., & Chen, C. (2021). Total organic carbon content prediction in lacustrine shale using extreme gradient boosting machine learning based on bayesian optimization. *Geofluids*, 2021(1), 6155663.
- Lopes, R. L., & Jorge, A. (2017). Mind the gap: a well log data analysis. arXiv preprint arXiv:1705.03669.
- Løseth, H., Wensaas, L., Gading, M., Duffaut, K., & Springer, M. (2011). Can hydrocarbon source rocks be identified on seismic data?. *Geology*, 39(12), 1167-1170.
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- Mandal, P. P., Rezaee, R., & Emelyanova, I. (2021). Ensemble learning for predicting TOC from well-logs of the unconventional goldwyer shale. *Energies*, 15(1), 216.
- Meissner, F. F. (1978). *Petroleum geology of the Bakken formation Williston basin, North Dakota and Montana*.
- Mitchell, T. M. (1997). Does machine learning really work?. *AI magazine*, 18(3), 11-11.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.
- Myers, K. J., & Jenkyns, K. F. (1992). Determining total organic carbon contents from well logs: an intercomparison of GST data and a new density log method. *Geological Society, London, Special Publications*, 65(1), 369-376.
- Nixon, R. P. (1973). Oil source beds in Cretaceous Mowry Shale of northwestern interior United States. *AAPG bulletin*, 57(1), 136-161.
- Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). Activation functions: Comparison of trends in practice and research for deep learning. arXiv preprint arXiv:1811.03378.

- Ogiesoba, O., & Hammes, U. (2014). Seismic-attribute identification of brittle and TOC-rich zones within the Eagle Ford Shale, Dimmit County, South Texas. *Journal of Petroleum Exploration and Production Technology*, 4, 133-151.
- Olden, J. D., & Jackson, D. A. (2002). Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecological modelling*, 154(1-2), 135-150.
- Passey, Q. R., Creaney, S., Kulla, J. B., Moretti, F. J., & Stroud, J. D. (1990). A practical model for organic richness from porosity and resistivity logs. *AAPG bulletin*, 74(12), 1777-1794.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- Polikar, R. (2012). Ensemble learning. In *Ensemble machine learning* (pp. 1-34). Springer, New York, NY.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems*, 31.
- Purificação, C. A. C. D. (2021). Predição de carbono orgânico total a partir de perfis geofísicos de Poços da Bacia de Santos.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088), 533-536.
- Schmoker, J. W. (1979). Determination of organic content of Appalachian Devonian shales from formation-density logs: Geologic notes. *AAPG bulletin*, 63(9), 1504-1509.
- Schmoker, J. W., & Hester, T. C. (1983). Organic carbon in Bakken formation, United States portion of Williston basin. *AAPG bulletin*, 67(12), 2165-2174.
- Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.
- Scornet, E. (2023, February). Trees, forests, and impurity-based variable importance in regression. In *Annales de l'Institut Henri Poincaré (B) Probabilités et statistiques* (Vol. 59, No. 1, pp. 21-52). Institut Henri Poincaré.
- Shalaby, M. R., Jumat, N., Lai, D., & Malik, O. (2019). Integrated TOC prediction and source rock characterization using machine learning, well logs and geochemical analysis: case study from the Jurassic source rocks in Shams Field, NW Desert, Egypt. *Journal of Petroleum Science and Engineering*, 176, 369-380.
- Shirmard, H., Farahbakhsh, E., Müller, R. D., & Chandra, R. (2022). A review of machine learning in processing remote sensing data for mineral exploration. *Remote Sensing of Environment*, 268, 112750.

- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929-1958.
- Sun, J., Dang, W., Wang, F., Nie, H., Wei, X., Li, P., ... & Li, F. (2023). Prediction of TOC content in organic-rich shale using machine learning algorithms: Comparative study of random forest, support vector machine, and XGBoost. *Energies*, 16(10), 4159.
- Sutton, R. S., & Barto, A. G. (2015). *Reinforcement learning: An introduction* Second edition. A Bradford Book, The MIT Press Cambridge.
- Thiéblemont, A., Castano-Urbe, C., Paternoster, B., Chaboureau, A. C., & Shipilova, E. (2022). Seismic characterization of source rocks: Why should it be assessed on a site-by-site basis?. *The Leading Edge*, 41(3), 177-183.
- Vapnik, V. (1998). *Statistical Learning Theory*. Wiley-Interscience
- Von Luxburg, U., & Schölkopf, B. (2011). Statistical learning theory: Models, concepts, and results. In *Handbook of the History of Logic* (Vol. 10, pp. 651-706). North-Holland.
- Wilcox, R. R. (2011). *Introduction to robust estimation and hypothesis testing*. Academic press.
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1), 79-82.
- Zhang, Y., Zhang, G., Zhao, W., Zhou, J., Li, K., & Cheng, Z. (2024). Total organic carbon content estimation for mixed shale using Xgboost method and implication for shale oil exploration. *Scientific Reports*, 14(1), 20860.
- Zhao*, T., Verma, S., Devegowda, D., & Jayaram, V. (2015). TOC estimation in the Barnett Shale from triple combo logs using support vector machine. In *SEG Technical Program Expanded Abstracts 2015* (pp. 791-795). Society of Exploration Geophysicists.

12.6. CAPÍTULO 6

- Blache-Fraser, G., & Neep, J. (2004). Increasing seismic resolution using spectral blueing and colored inversion: Cannonball field, Trinidad. In *SEG Technical Program Expanded Abstracts 2004* (pp. 1794-1797). Society of Exploration Geophysicists.
- Gonzalez, R. C. (1987). An overview of image processing and pattern recognition techniques. *Handbook of Geophysical Exploration, Pattern Recognition and Image Processing*, edited by F. Aminzadeh, Geophysical Press, 20.
- Haffner, S., & Cheadle, S. (1999). Colored reflectivity compensation for increased vertical resolution. In *SEG Technical Program Expanded Abstracts 1999* (pp. 1307-1310). Society of Exploration Geophysicists.

Kazemeini, S. H., Yang, C., Juhlin, C., Fomel, S., & Cosma, C. (2010). Enhancing seismic data resolution using the prestack blueing technique: An example from the Ketzin CO₂ injection site, Germany. *Geophysics*, 75(6), V101-V110.

Marfurt, K. J., Kirlin, R. L., Farmer, S. L., & Bahorich, M. S. (1998). 3-D seismic attributes using a semblance-based coherency algorithm. *Geophysics*, 63(4), 1150-1165.

Neep, J. P. (2007). Time variant coloured inversion and spectral blueing. In 69th EAGE Conference and Exhibition incorporating SPE EUROPEC 2007 (pp. cp-27). European Association of Geoscientists & Engineers.

Rodrigues, R. (1995). A geoquímica orgânica na Bacia do Parnaíba. Unpublished PhD thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre.

Saggaf, M. M., & Robinson, E. A. (2000). A unified framework for the deconvolution of traces of nonwhite reflectivity. *Geophysics*, 65(5), 1660-1676.

Tingdahl, K. M. (1999). Improving seismic detectability using intrinsic directionality. Paper B194. Earth Science Centre, Goteberg University, 00090-0.

Tingdahl, Kristofer & de Groot, Paul. (2003). Post-stack-dip- and azimuth processing. *Journal of Seismic Exploration*. 12. 113-126.

Walden, A. T., & Hosken, J. W. J. (1985). An investigation of the spectral properties of primary reflection coefficients. *Geophysical prospecting*, 33(3), 400-435.

Wu, S., & Wang, Y. (2022). Seismic image dip estimation by multiscale principal component analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1-10.

12.7. CAPÍTULO 7

Addison, P. S. (2002). *The illustrated wavelet transform handbook: Introductory theory and application in science, engineering, medicine and finance*. Napier University, 6–14. ISBN 0750306920.

Jenkins, F. A.; White, H. E.; Brukhard, D. G. (1958) *Fundamentals of optics*. *American Journal of Physics*, v. 26, n. 4, p. 272–272.

Kallweit, R. S.; Wood, L. C. The limits of resolution of zero-phase wavelets. *Geophysics* Vol. 47, P. 1035–1046. DOI: 10.1190/1.1441367, 1982.

Ricker, N. (1953) Wavelet contraction, wavelet expansion, and the control of seismic resolutions. *Geophysics* Vol. 18, issue 4, P. 769-792.

Simm, R.; Bacon, M. (2014). *Seismic Amplitude: An Interpreter's Handbook*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511984501

Widess, M. B. (1973) How thin is a thin bed? *Geophysics* Vol. 38, P. 1176-1180.

Yilmaz, Ö.; Yilmaz, O.; Doherty, S. (2001) *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data*. Society of Exploration Geophysicists.

12.8. CAPÍTULO 8

Simm, R., & Bacon, M. (2014). *Seismic amplitude: An interpreter's handbook*. Cambridge university press.

Stewart, R. R., Huddleston, P. D., & Kan, T. K. (1984). Seismic versus sonic velocities: A vertical seismic profiling study. *Geophysics*, 49(8), 1153-1168.

Walden, A. T., & White, R. E. (1998). Seismic wavelet estimation: A frequency domain solution to a geophysical noisy input-output problem. *IEEE transactions on Geoscience and Remote Sensing*, 36(1), 287-297.

White, R. E. (1980). Partial coherence matching of synthetic seismograms with seismic traces. *Geophysical Prospecting*, 28(3), 333-358.

White, R. E., & Simm, R. (2003). Tutorial: Good practice in well ties. *First break*, 21(10).

12.9. CAPÍTULO 9

Almeida, A. S., & Journel, A. G. (1994). Joint simulation of multiple variables with a Markov-type coregionalization model. *Mathematical geology*, 26(5), 565-588.

Ansari, H. R. (2014). Use seismic colored inversion and power law committee machine based on imperial competitive algorithm for improving porosity prediction in a heterogeneous reservoir. *Journal of Applied Geophysics*, 108, 61-68.

Azevedo, L., & Soares, A. (2017). *Geostatistical methods for reservoir geophysics* (p. 141). Berlin: Springer.

Blache-Fraser, G., & Neep, J. (2004, October). Increasing seismic resolution using spectral blueing and colored inversion: Cannonball field, Trinidad. In *SEG International Exposition and Annual Meeting* (pp. SEG-2004). SEG.

Bosch, M., Mukerji, T., & Gonzalez, E. F. (2010). Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review. *Geophysics*, 75(5), 75A165-75A176.

Caetano, H. (2012). *Integration of Seismic Information in Reservoir*. *New Technologies in the Oil and Gas Industry*, 119.

- Caputo, M. V. (1984). Stratigraphy, tectonics, paleoclimatology and paleogeography of northern basins of Brazil. University of California, Santa Barbara.
- Cooke, Dennis A.; Schneider, William A. Generalized linear inversion of reflection seismic data. *Geophysics*, v. 48, n. 6, p. 665–676, 1983. DOI: 10.1190/1.1441497.
- Cyz, M., & Azevedo, L. (2020). Direct geostatistical seismic amplitude versus angle inversion for shale rock properties. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6), 5335-5344.
- del Monte, A. A., Antonielli, E., De Tomasi, V., Luchetti, G., Paparozzi, E., & Gambacorta, G. (2018). Methods for source rock identification on seismic data: An example from the Tanezzuft Formation (Tunisia). *Marine and Petroleum Geology*, 91, 108-124.
- Della Fávera, J. C. (1990). *Tempestitos da bacia do Parnaíba*. Porto Alegre, 243.
- Deutsch, C. V., & Journel, A. G. (1992). *Geostatistical software library and user's guide*. New York, 119(147), 578.
- Doyen, P. M. (2007). *Seismic reservoir characterization: An earth modelling perspective*. EAGE.
- Dubrule, O. (2003). *Geostatistics for seismic data integration in earth models*. Society of Exploration Geophysicists and European Association of Geoscientists and Engineers.
- Filippova, K., Kozhenkov, A., & Alabushin, A. (2011). Seismic inversion techniques: choice and benefits. *First Break*, 29(5).
- Gómez-Hernández, J. J., & Journel, A. G. (1993). Joint sequential simulation of multigaussian fields. In *Geostatistics Tróia'92: volume 1* (pp. 85-94). Dordrecht: Springer Netherlands.
- Grana, D., & Della Rossa, E. (2010). Probabilistic petrophysical-properties estimation integrating statistical rock physics with seismic inversion. *Geophysics*, 75(3), O21-O37.
- Journel, A. G., & Huijbregts, C. H. J., 1978. *Mining geostatistics*. Breau De RecherchesGeologiques Et Miners, France Academic Pres Harcourt Brace & Company, Publishers London, San Diego, New York, Boston, Sidney, Toronto.
- Lancaster, Steve; Whitcombe, David. Fast-track 'coloured' inversion. In: *SEG International Exposition and Annual Meeting*. SEG, 2000. p. SEG-2000-1572.
- Lines, Larry R.; Treitel, Sven. A review of least-squares inversion and its application to geophysical problems. *Geophysical Prospecting*, v. 32, n. 2, p. 159–186, 1984. DOI: 10.1111/j.1365-2478.1984.tb00726.x.
- Matheron, G. (1965). Principles of geostatistics. *Economic Geology*, 58, 1246–1266.
- Maulana, A. D., Safira, N., Prayoga, O. A., Wijaksono, E., & Ascaria, A. (2022). Seismic Inversion Resolution Enhancement With (3S) Spectral Blueing, Spectral Balancing, and Stochastic Inversion on Fluvio Deltaic Environment.

Oliver, M. A., & Webster, R. (2015). *Basic steps in geostatistics: the variogram and kriging* (Vol. 106). Cham, Switzerland: Springer International Publishing.

Ouadfeul, S. A., & Aliouane, L. (2016). Total organic carbon estimation in shale-gas reservoirs using seismic genetic inversion with an example from the Barnett Shale. *The Leading Edge*, 35(9), 790-794.

Russell, Brian H. *Introduction to seismic inversion methods*. Tulsa: Society of Exploration Geophysicists, 1988. (Course Notes Series, n. 2). DOI: 10.1190/1.9781560802303.

Simm, R., & Bacon, M. (2014). *Seismic amplitude: An interpreter's handbook*. Cambridge university press.

Soares, A., Diet, J. D., & Guerreiro, L. (2007, September). Stochastic inversion with a global perturbation method. In *EAGE Conference on Petroleum Geostatistics* (pp. cp-32). European Association of Geoscientists & Engineers.

Tarantola, A. (2005). The least square criterion. *Inverse problem theory and methods for model parameter estimation*, 68-72.

Thiéblemont, A., Castano-Urbe, C., Paternoster, B., Chaboureau, A. C., & Shipilova, E. (2022). Seismic characterization of source rocks: Why should it be assessed on a site-by-site basis?. *The Leading Edge*, 41(3), 177-183.

Tompkins, M. J., Fernández Martínez, J. L., Alumbaugh, D. L., & Mukerji, T. (2011). Scalable uncertainty estimation for nonlinear inverse problems using parameter reduction, constraint mapping, and geometric sampling: Marine controlled-source electromagnetic examples. *Geophysics*, 76(4), F263-F281.

Walden, A. T., & Hosken, J. W. J. (1985). An investigation of the spectral properties of primary reflection coefficients. *Geophysical prospecting*, 33(3), 400-435.

Yang, Ruizhao & Zhao, Zhengguang & Ma, Yan-long & Huo, Chao & Li, Yang. (2013). Thin Coal Bed Resolution by Using Seismic Spectral Blueing and Colored Inversion. *natural gas geoscience*. 24. 156-161. 10.11764/j.issn.1672-1926.2013.01.156.

Zhdanov, M. S. (2002). *Geophysical inverse theory and regularization problems* (Vol. 36). Elsevier.

12.10. CAPÍTULO 10

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

Carcione, J. M., & Avseth, P. (2015). Rock-physics templates for clay-rich source rocks. *Geophysics*, 80(5), D481-D500.

Chiles, J. P., & Delfiner, P. (2012). *Geostatistics: modeling spatial uncertainty*. John Wiley & Sons.

Cyz, M., & Azevedo, L. (2020). Direct geostatistical seismic amplitude versus angle inversion for shale rock properties. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6), 5335-5344.

- Daly, C. (2020, November). Tight integration of decision forests into geostatistical modelling. In First EAGE Digitalization Conference and Exhibition (Vol. 2020, No. 1, pp. 1-5). European Association of Geoscientists & Engineers.
- Daly, C. (2021). An application of an embedded model estimator to a synthetic nonstationary reservoir model with multiple secondary variables. *Frontiers in Artificial Intelligence*, 4, 624697.
- Daly, C. (2022). An embedded model estimator for non-stationary random functions using multiple secondary variables. *Mathematical Geosciences*, 54(5), 979-1015.
- Daly, C., Hardy, M., & McNamara, K. (2021). Leveraging machine learning for enhanced geostatistical modelling of reservoir properties. In 82nd EAGE Annual Conference & Exhibition (Vol. 2021, No. 1, pp. 1-5). European Association of Geoscientists & Engineers.
- Davison, I., Rodriguez, K., & Eastwell, D. (2018). Seismic detection of source rocks.
- del Monte, A. A., Antonielli, E., De Tomasi, V., Luchetti, G., Papanozzi, E., & Gambacorta, G. (2018). Methods for source rock identification on seismic data: An example from the Tanezzuft Formation (Tunisia). *Marine and Petroleum Geology*, 91, 108-124.
- Ferreira, D. J. A., de Oliveira, G. M. B., Castro, T. M., Dias, R. M., & Lupinacci, W. M. (2021). Geostatistics assisted by machine learning for reservoir property modeling: A case study in presalt carbonates of Buzios Field, Brazil. *The Leading Edge*, 40(12), 876-885.
- Györfi, L., Kohler, M., Krzyżak, A., & Walk, H. (2002). *A distribution-free theory of nonparametric regression*. New York, NY: Springer New York.
- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, 20(8), 832-844.
- Lafferty, J., McCallum, A., & Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data.
- Lima, E. D. A. M., & Leite, J. F. (1978). Projeto estudo global dos recursos Minerais da Bacia sedimentar do Parnaíba. Integração geológico-metalogenética. CPRM.
- Løseth, H., Wensaas, L., Gading, M., Duffaut, K., & Springer, M. (2011). Can hydrocarbon source rocks be identified on seismic data?. *Geology*, 39(12), 1167-1170.
- Matheron G (1989). *Estimating and choosing*. Springer-Verlag, Berlin
- Mussa, A. (2020). O Potencial de geração de hidrocarbonetos na Formação Pimenteiras, Bacia do Parnaíba. Tese de Doutorado em Geociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.
- Quadfeul, S. A., & Aliouane, L. (2016). Total organic carbon estimation in shale-gas reservoirs using seismic genetic inversion with an example from the Barnett Shale. *The Leading Edge*, 35(9), 790-794.

- Peters, K. E., & Cassa, M. R. (1994). Applied source rock geochemistry.
- Rodrigues, R. (1995). A geoquímica orgânica na Bacia do Parnaíba. *Revista Brasileira de Geociências*, 25, 145-146.
- Sayers, C. M. (2013). The effect of kerogen on the AVO response of organic-rich shales. *The Leading Edge*, 32(12), 1514-1519.
- Thiéblemont, A., Castano-Urbe, C., Paternoster, B., Chaboureau, A. C., & Shipilova, E. (2022). Seismic characterization of source rocks: Why should it be assessed on a site-by-site basis?. *The Leading Edge*, 41(3), 177-183.
- Vernik, L., & Liu, X. (1997). Velocity anisotropy in shales: A petrophysical study. *Geophysics*, 62(2), 521-532.
- Vernik, L., & Nur, A. (1992). Ultrasonic velocity and anisotropy of hydrocarbon source rocks. *Geophysics*, 57(5), 727-735.
- Wackernagel, H. (2003). Ordinary kriging. In *Multivariate geostatistics: an introduction with applications* (pp. 79-88). Berlin, Heidelberg: Springer Berlin Heidelberg.