UNIVERSIDADE FEDERAL DA BAHIA
Curso de Pós-graduação em Geofísica

TESE DE DOUTORADO

AUTOMAÇÃO SÍSMICA HÍBRIDA: INTEGRAÇÃO DE INTELIGÊNCIA ARTIFICIAL E MÉTODOS DETERMINÍSTICOS NA ANÁLISE DE VELOCIDADES

MARCOS AUGUSTO DA LIMA LUZ

SALVADOR – BAHIA
SETEMBRO – 2025

Documento preparado com o sistema LaTeX.

Documento elaborado com os recursos gráficos e de informática do CPGG/UFBA

Automação sísmica híbrida: Integração de inteligência artificial e métodos determinísticos na análise de velocidades

por

Marcos Augusto da Lima Luz

Matemática (Universidade Federal do Pará – 2005)

Mestre em Geofísica (Universidade Federal do Rio Grande do Norte – 2012)

Orientador: Prof. Dr. Marcos Alberto Rodrigues Vasconcelos

Documento assinado digitalmente
gov.br MARCOS ALBERTO RODRIGUES VASCONCELOS
Data: 17/11/2025 10:15:35-0300
Verifique em https://validar.iti.gov.br

Documento assinado digitalmente
gov.br JUAREZ DOS SANTOS AZEVEDO
Data: 15/11/2025 10:01:41-0300
Verifique em https://validar.iti.gov.br

Documento assinado digitalmente
gov.br DIOGO LUIZ DE OLIVEIRA COELHO
Data: 15/11/2025 10:10:53-0300
Verifique em https://validar.iti.gov.br

Documento assinado digitalmente
gov.br SAULO POMPONET OLIVEIRA
Data: 14/11/2025 15:32:02-0300
Verifique em https://validar.iti.gov.br

Comissão Examinadora

Dr. Marcos Alberto R. Vasconcelos

Dr. Juarez dos Santos Azevedo

Dr. Diogo Luiz de Oliveira Coelho

Dr. Saulo Pomponet Oliveira

Dr. Hugo Esteban Poveda Nuñez

Aprovada em 30 de setembro de 2025

Ficha catalográfica elaborada pela Biblioteca Universitária de Ciências e Tecnologias Prof. Omar Catunda, SIBI – UFBA.

L979 Luz, Marcos Augusto da Lima

Automação sísmica híbrida: integração de inteligência artificial e métodos determinísticos na análise de velocidades / Marcos Augusto Lima da Luz. — Salvador – BA, 2025.

113 f. : il.

Orientador: Marcos Alberto Rodrigues Vasconcelos.

Tese (Doutorado) — Universidade Federal da Bahia, Instituto de Geociências, 2025.

1.Geofísica. 2. Inteligência Artificial. 3. Aprendizado Computacional. 4. K-means++. 5. Análise de Componentes Principais (PCA). 6. Equação de Dix. 7. Redes Neurais. 8. Campo de velocidades sísmicas. I. Vasconcelos, Marcos Alberto Rodrigues. II. Título.

CDU:550.34

“À memória de meu pai, José Augusto Pereira da Luz, cuja presença permanece viva em meus passos e cuja força e exemplo continuam a iluminar o meu caminho.”

Resumo

O presente trabalho apresenta uma metodologia de automação sísmica híbrida que integra técnicas de inteligência artificial e métodos determinísticos clássicos para a análise e determinação automática do campo de velocidades em dados sísmicos. O objetivo central é otimizar o processo de exploração de hidrocarbonetos, ampliando a precisão na modelagem e a segurança operacional durante a interpretação geológica. Tradicionalmente, a construção desse campo de velocidades depende de procedimentos manuais de *picking* em painéis de *semblance*, caracterizados por alta subjetividade e pela exigência de experiência técnica dos analistas, especialmente em ambientes com ruído e complexidade estrutural elevada. A metodologia proposta combina, de forma sequencial e integrada, abordagens estatísticas e de aprendizado de máquina. Inicialmente, aplica-se uma técnica de pré-agrupamento amostral, responsável pela estruturação prévia dos dados e pela determinação automática do número ótimo de agrupamentos. Em seguida, a associação do algoritmo K-means++ com a Análise de Componentes Principais (PCA) promove uma redução de dimensionalidade eficiente, realçando os padrões de coerência e representatividade sísmica. Na etapa determinística, a equação de Dix é empregada para a conversão das velocidades RMS em velocidades intervalares, formando o conjunto de treinamento para a rede neural MLP (Multilayer Perceptron). Essa rede supervisionada realiza o ajuste final do campo de velocidades, assegurando consistência física, suavidade e comportamento monotonicamente crescente. O caráter híbrido do método reside justamente na interação entre o rigor físico-matemático da formulação determinística e a capacidade preditiva adaptativa da inteligência artificial. A metodologia foi validada em modelos sintéticos e em dados reais do Golfo do México, evidenciando robustez, estabilidade e aplicabilidade em diferentes cenários geológicos. Os resultados demonstram que a automação sísmica híbrida proposta permite gerar modelos de velocidade mais realistas e de maior resolução espacial, reduzindo significativamente a interferência humana e contribuindo para a eficiência interpretativa em ambientes de exploração complexos.

Palavras-chave: Automação Sísmica Híbrida; Inteligência Artificial; K-means++; Análise de Componentes Principais (PCA); Equação de Dix; Rede Neural MLP; Campo de Velocidades; Sísmica de Reflexão.

Abstract

This research presents a hybrid seismic automation methodology that integrates artificial intelligence techniques with deterministic methods for the automatic analysis and estimation of seismic velocity fields. The main goal is to optimize hydrocarbon exploration by enhancing model accuracy and ensuring operational safety throughout geological interpretation. Traditionally, the construction of the velocity field relies on manual picking from semblance panels, a subjective and time-consuming procedure that demands expert interpretation, especially under noisy or geologically complex conditions. The proposed workflow combines statistical and machine learning approaches in a sequential and integrated manner. The process begins with a sample pre-clustering technique, responsible for the preliminary structuring of the data and for automatically determining the optimal number of clusters. Next, the joint application of the K-means++ algorithm and Principal Component Analysis (PCA) enables efficient dimensionality reduction, improving data coherence and representativeness. In the deterministic stage, the Dix equation is employed to convert RMS velocities into interval velocities, which serve as training data for a Multilayer Perceptron (MLP) neural network. This supervised model performs the final adjustment of the velocity field, ensuring physical consistency, smoothness, and monotonic behavior. The hybrid nature of the methodology arises from the synergistic integration between deterministic physical modeling and adaptive artificial intelligence prediction. The proposed approach was validated using both synthetic models and real seismic data from the Gulf of Mexico, demonstrating robustness, stability, and applicability across diverse geological scenarios. The results confirm that the hybrid seismic automation framework provides more realistic and continuous velocity models, substantially reducing human intervention and improving interpretive efficiency in complex exploration environments.

Keywords: Hybrid Seismic Automation; Artificial Intelligence; K-means++; Principal Component Analysis (PCA); Dix Equation; Multilayer Perceptron (MLP); Velocity Field; Reflection Seismics.

Índice

Índice de Tabelas

Índice de Figuras

Introdução

A estimativa manual de velocidades sísmicas constitui uma das etapas mais críticas e, simultaneamente, mais complexas do processamento de dados geofísicos. Esse procedimento exige do intérprete não apenas tempo considerável, mas também elevado grau de experiência para distinguir, com segurança, eventos primários de ruídos, múltiplos ou distorções impostas pela heterogeneidade geológica. Conforme destacado por Neidell e Taner (1971), a seleção manual no espectro de velocidades pode se tornar excessivamente demorada e sujeita a erros de julgamento, sobretudo em regiões onde a coerência dos refletores é limitada. Nessas condições, a negligência de anomalias, como zonas de sombra ou zonas de baixa velocidade, pode comprometer a qualidade da interpretação e ampliar os riscos exploratórios (Marfurt e Alves, 2015).

Com o intuito de mitigar tais limitações, a indústria passou a adotar estratégias baseadas em espectros de semelhança. Desde a proposta inicial de Taner e Koehler (1969), esse recurso consolidou-se como ferramenta essencial de apoio à escolha da velocidade de empilhamento, conferindo maior objetividade ao processo decisório. Ao longo do tempo, diferentes abordagens foram sugeridas para complementar essa técnica, entre elas o emprego de dados pré-empilhados (Zhang e Lu, 2016) e métodos híbridos que integram múltiplas fontes de informação, como espectros de semelhança, painéis pré-empilhados e seções empilhadas (Ferreira et al., 2020). Apesar desses avanços, a etapa de seleção ainda demanda significativa intervenção humana.

A complexidade do problema se intensifica quando se consideram diferentes ambientes de aquisição. Em levantamentos terrestres, ruídos incoerentes, originados por tráfego, vento ou atividades antrópicas, tendem a mascarar a continuidade dos refletores, dificultando a análise. Já em ambientes marinhos, embora tais ruídos sejam menos pronunciados, as reverberações de eventos múltiplos do fundo do mar sobrepõem eventos distintos, comprometendo a determinação confiável das velocidades. Assim, a análise de velocidades enfrenta desafios singulares em contextos onshore e offshore, o que reforça a necessidade de metodologias ajustadas às especificidades de cada cenário.

Nesse contexto, emerge a demanda pela automação do processo de seleção de velocidades. A presente pesquisa propõe uma abordagem inovadora que integra métodos estatísticos a técnicas de aprendizagem de máquina, visando maior eficiência na determinação do campo de velocidades. Inicialmente, os dados são organizados em classes e submetidos a algoritmos de aprendizagem não supervisionada, como o K-means++, capazes de identificar subgrupos representativos. Em seguida, aplica-se a Análise de Componentes Principais (PCA), que concentra a energia dos traços e facilita a identificação de eventos sísmicos relevantes.

Na etapa subsequente, incorporou-se aprendizado supervisionado por meio de redes neurais do tipo Perceptron Multicamadas (MLP). Esse modelo atua no ajuste dos valores de velocidade NMO estimados pela equação de Dix, assegurando suavidade e monotonicidade física aos perfis de velocidade. Tal ajuste mostra-se particularmente eficaz em áreas geologicamente complexas, como regiões salinas, onde as elevadas velocidades do halito distorcem os tempos de trânsito e dificultam a correta caracterização dos horizontes.

Dessa forma, a metodologia proposta busca não apenas reduzir a dependência da análise manual, mas também aprimorar a precisão e a confiabilidade na estimação do campo de velocidades sísmicas.

1

Fundamentos físicos e geométricos do método sísmico de reflexão

1.1 Princípios físicos e operacionais da sísmica de reflexão

A aquisição de dados sísmicos é uma etapa fundamental na exploração geofísica, envolvendo a geração e registro de ondas elásticas que se propagam e refletem nas interfaces das camadas subsuperficiais (Yilmaz e Doherty, 1987). Essas ondas são produzidas por fontes sísmicas, que podem ser de natureza explosiva, como dinamite, ou não explosiva, como sistemas vibratórios conhecidos como Vibroseis (Liner, 1999). As ondas geradas propagam-se pelo meio geológico e, ao encontrarem descontinuidades nas propriedades elásticas, são refletidas de volta à superfície, quando são detectadas por sensores especializados (Mayne, 1962).

Em ambientes marinhos, a aquisição sísmica é realizada utilizando fontes como canhões de ar comprimido (*air guns*) e receptores denominados hidrofones, estrategicamente posicionados para otimizar a qualidade dos dados coletados (Clark, 2016). As reflexões sísmicas registradas fornecem informações cruciais sobre a composição e estrutura geológica em profundidade, permitindo a construção de modelos detalhados da subsuperfície.

Os receptores, como geofones em ambientes terrestres ou hidrofones em ambientes marinhos, são essenciais para o registro das vibrações do solo ou variações de pressão na coluna d'água induzidas pelas ondas sísmicas. Esses dispositivos convertem a energia mecânica das ondas em sinais elétricos, que representam o deslocamento ou variação da pressão, resultantes das reflexões nas interfaces geológicas (Dobrin e Savit, 1976). Ao receber uma onda sísmica, o

receptor gera um sinal proporcional à amplitude da vibração, originando o que é conhecido como *traço sísmico*. Cada traço representa a resposta sísmica em um ponto específico ao longo do tempo e é registrado em um sismograma, que compila todos os traços obtidos em uma linha ou área de aquisição, fornecendo uma representação temporal e espacial das reflexões sísmicas (Yilmaz e Doherty, 1987).

Nesta pesquisa, o foco recai sobre a sísmica de reflexão marinha, amplamente empregada em programas de exploração e monitoramento offshore devido à sua elevada resolução e capacidade de imagear estruturas profundas sob o leito oceânico. A aquisição desses dados pode ocorrer por diferentes arranjos operacionais, destacando-se dois métodos principais: a sísmica com cabos flutuantes (*streamers*) e a sísmica com sensores de fundo oceânico (*Ocean Bottom Nodes* – OBN).

Na configuração streamer (Figura 1.1), um navio reboca canhões de ar comprimido (*air guns*) e cabos dotados de hidrofones, que registram as variações de pressão geradas pelas reflexões sísmicas ao longo de perfis paralelos de aquisição. Já no método OBN (Figura 1.2), sensores autônomos são posicionados diretamente sobre o fundo marinho por meio de veículos operados remotamente (ROVs), permitindo a obtenção de registros de alta fidelidade, com melhor relação sinal-ruído e repetibilidade temporal, características fundamentais em levantamentos 4D e em ambientes com infraestrutura submarina complexa, como sistemas de produção e risers de petróleo.

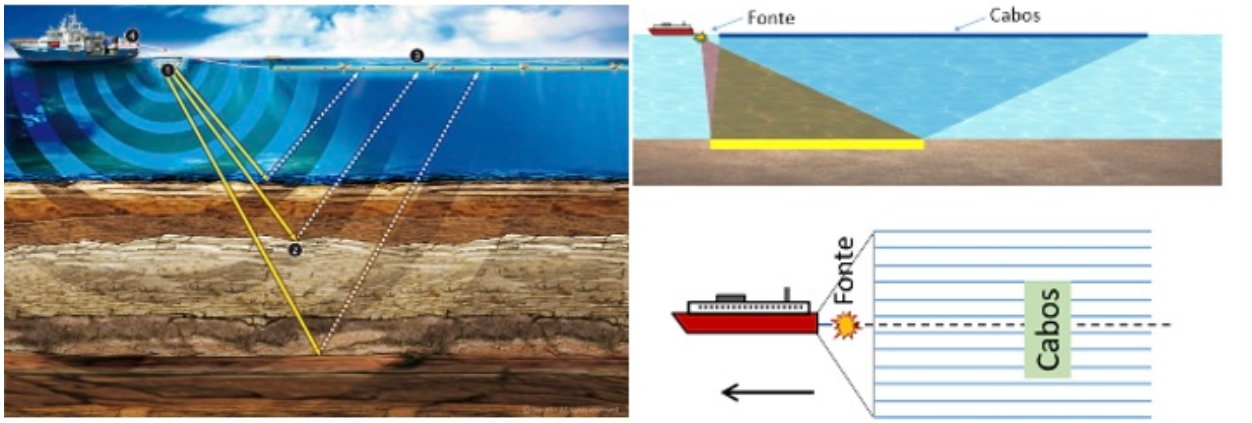


Figura 1.1: Configuração típica da sísmica marinha com cabos flutuantes (*streamers*), na qual o navio fonte reboca canhões de ar comprimido (*air guns*) e cabos equipados com hidrofones para registrar as reflexões sísmicas. Fonte: Adaptado de Petrobras (2025).

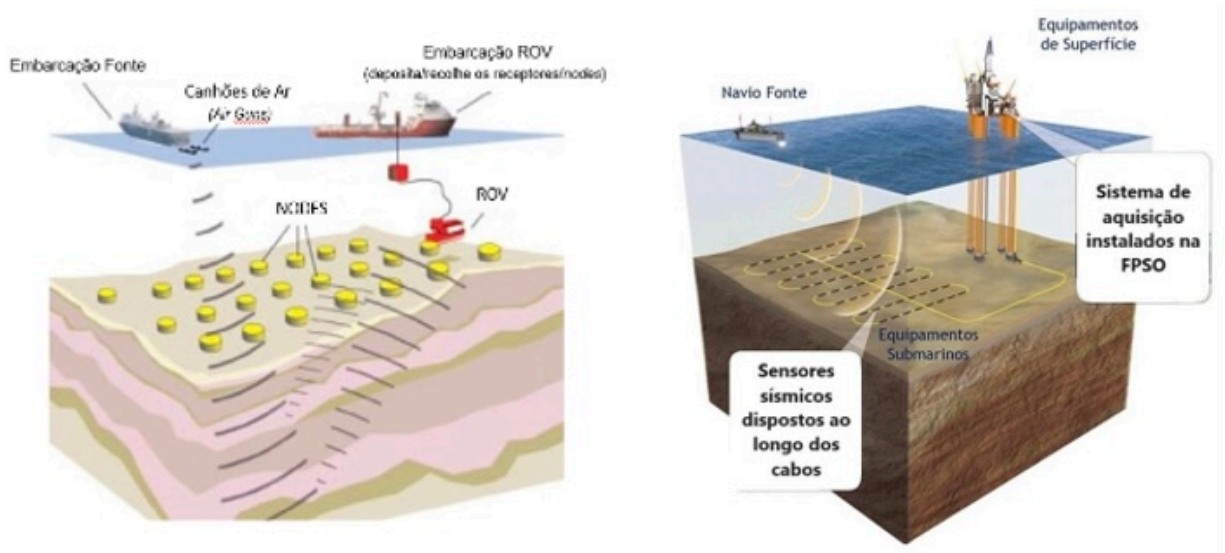


Figura 1.2: Representação do método *Ocean Bottom Nodes* (OBN), em que sensores sísmicos autônomos são posicionados no leito marinho por meio de veículos ROV, proporcionando registros de alta fidelidade e maior cobertura em áreas com infraestrutura submarina. Fonte: Adaptado de Petrobras (2025).

Apesar das diferenças tecnológicas e logísticas entre essas configurações, um aspecto permanece invariável: a necessidade de uma análise criteriosa do campo de velocidades. A determinação precisa das velocidades sísmicas constitui o núcleo interpretativo de qualquer processamento, pois controla as etapas de correção de sobretempo normal (NMO), empilhamento, migração e conversão tempo-profundidade, garantindo a coerência geométrica e física dos refletores (Yilmaz e Doherty, 1987; Sheriff e Geldart, 1995; Telford et al., 1990; Ashcroft, 2011). Assim, independentemente do tipo de aquisição ou do meio investigado, o estudo das velocidades representa uma etapa indispensável para a correta caracterização das propriedades elásticas da subsuperfície e a redução das incertezas associadas à imagem sísmica final.

A interpretação dos sismogramas gerados após as aquisições, permite inferir a existência das estruturas geológicas e, especialmente no caso de camadas, suas propriedades físicas e geométricas, tais como velocidade das ondas sísmicas, densidade e espessura das camadas, sendo fundamental para a identificação de estruturas potenciais de reservatórios de hidrocarbonetos. A precisão na aquisição e processamento dos dados sísmicos é determinante para a confiabilidade dos modelos geológicos gerados, impactando diretamente nas decisões exploratórias e de desenvolvimento de campos petrolíferos (Dobrin e Savit, 1976).

1.1.1 Arranjos de aquisição e organização dos dados

Os dados sísmicos são organizados em diferentes arranjos e domínios, dependendo da configuração das fontes e receptores. Em arranjos como o *split-spread*, as fontes são colocadas no

centro e os receptores se espalham simetricamente em ambos os lados, cobrindo amplamente a área de aquisição e proporcionando uma boa razão sinal-ruído, ideal para levantamentos detalhados Figura (1.3a). No arranjo *end-on* Figura (1.3b), os receptores são dispostos em linha com as fontes, ao longo da direção do tiro, o que simplifica a logística e é eficiente em termos de custo, embora possa limitar a resolução lateral (Telford et al., 1990; Knapp e Steeples, 1986).

Além disso, os dados são organizados nos domínios do tiro, CMP *(Common Midpoint)* e *offset* para otimizar a interpretação. No domínio do tiro, os dados de cada disparo são agrupados, facilitando a análise da propagação inicial das ondas. O domínio CMP agrega reflexões de diferentes tiros para o mesmo ponto da subsuperfície, fator essencial para realização da análise de velocidade das camadas. Já o domínio *offset* organiza os dados com base na distância entre fonte e receptor, permitindo empilhamento e aumento da razão sinal-ruído. Esse conjunto de arranjos e domínios proporciona uma base sólida para o processamento e interpretação dos dados, resultando em uma representação detalhada e precisa da subsuperfície (Moreira Neto et al., 2005; Porsani, 2000; de São Paulo, 2020).

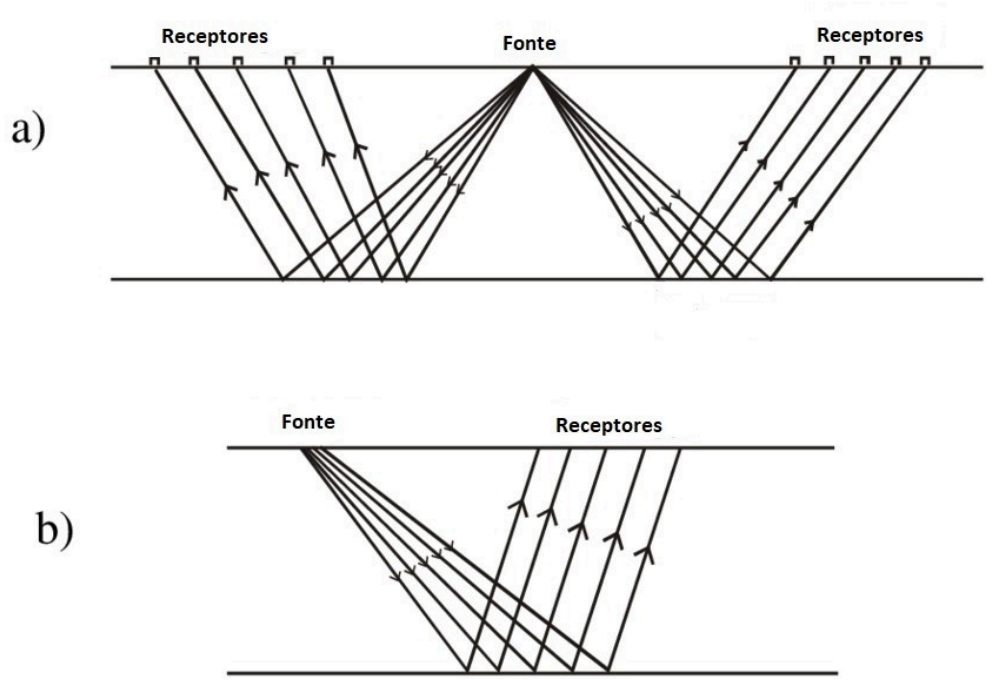


Figura 1.3: Geometria de aquisição sísmica multicanal. (a) Split-spread (b) End-on. Fonte: Adaptado de Ozegin (2012).

1.1.2 Técnica CDP e organização CMP

A organização CMP *(Common Midpoint)* é amplamente reconhecida como uma das abordagens mais eficazes para o imageamento das camadas geológicas em subsuperfície, devido à sua

capacidade de simplificar o processamento sísmico e melhorar a qualidade dos dados. Essa técnica se destaca, primeiramente, pela simplicidade matemática de seu desenvolvimento, especialmente em cenários em que as interfaces geológicas são horizontais. Além disso, o método CMP permite um aumento significativo na razão sinal-ruído, uma vez que o mesmo ponto em profundidade é amostrado diversas vezes, utilizando diferentes afastamentos entre fonte e receptor.

Quando as camadas da subsuperfície podem ser assumidas como plano-paralelas e não apresentam variação lateral de velocidade, os traços associados a um mesmo ponto médio comum estarão vinculados a um único ponto refletor em profundidade. Nesse caso específico, a organização CMP pode ser equivalente à organização CDP (Common Depth Point). Essa equivalência, entretanto, não se mantém quando os refletores são inclinados (Yilmaz, 2001). Para esses cenários, pontos distintos em profundidade são iluminados para cada tiro, tornando a equivalência entre CMP e CDP inválida, conforme ilustrado na Figura 1.4(a) e Figura 1.4(b).

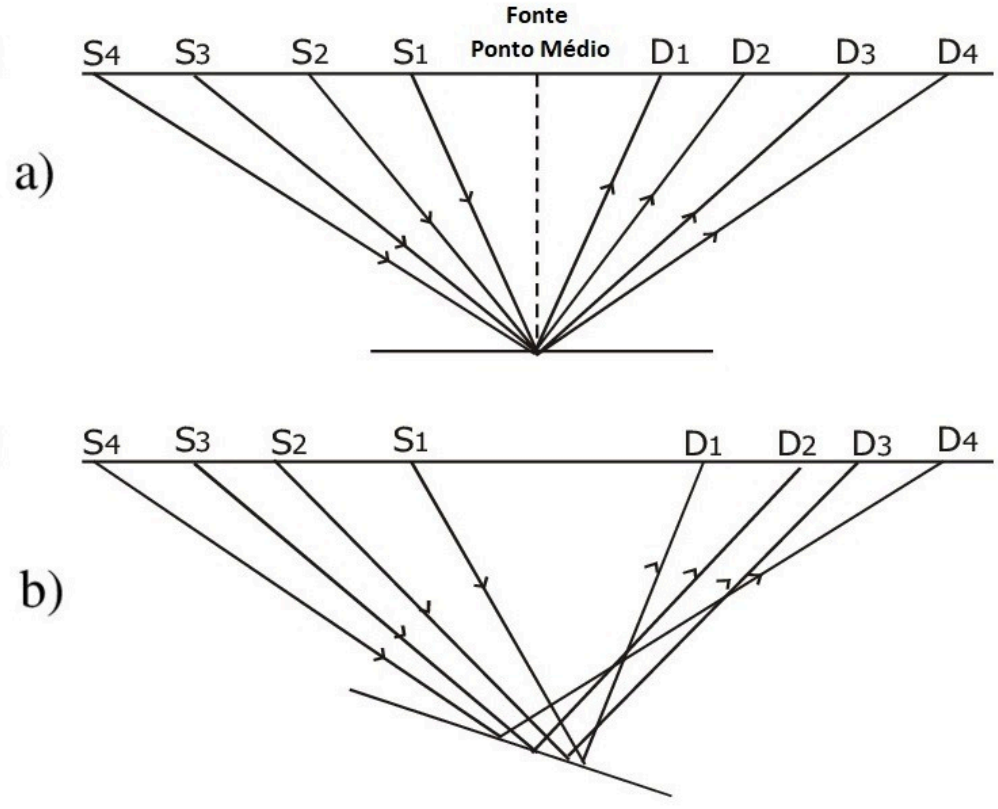


Figura 1.4: Perfil de reflexão de pontos em profundidade.(a) Ponto comum de reflexão em profundidade (CDP). (b) Refletor inclinado (Sem ponto comum de reflexão em profundidade). Fonte: Adaptado de Ozegin (2012).

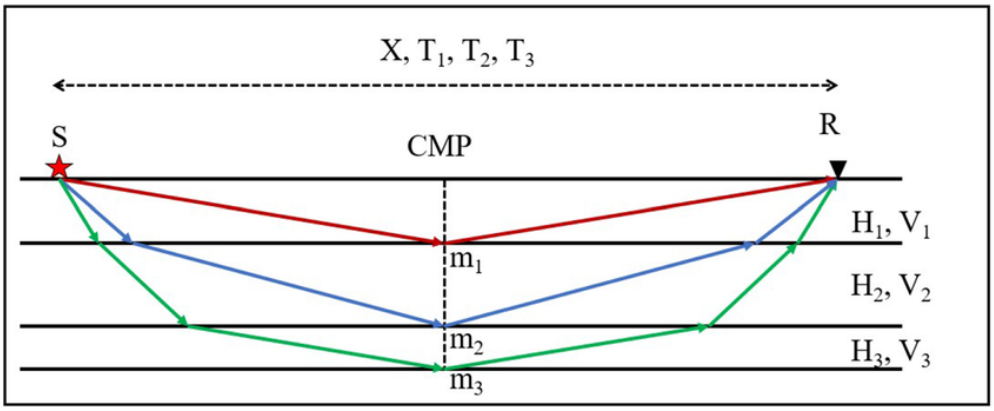


Figura 1.5: configuração CMP de fontes e receptores em um levantamento sísmico com camadas planas horizontais. Fonte: Alfuraidan et al. (2023).

Essa limitação destaca a importância de avaliar cuidadosamente as condições geológicas locais antes de aplicar o método CMP, garantindo que os pressupostos do modelo sejam adequados para a interpretação geofísica.

A configuração típica de uma aquisição sísmica utilizando um modelo simples de três camadas com representação CMP, é ilustrada na Figura 1.5 que inclui os seguintes elementos: **S** representa a fonte, **R** o receptor, e **X** a distância conhecida (*offset*) entre a fonte e o receptor. T_i denota o tempo de percurso de ida e volta da fonte ao receptor ao longo do raio refletido pela i-ésima camada (também conhecido), quando $i = 1, 2, 3$. A velocidade da onda sísmica na i-ésima camada, representada por V_i, é um parâmetro desconhecido, assim como a espessura da camada, indicada por H_i. Linhas pretas sólidas indicam as interfaces entre as camadas, enquanto setas coloridas representam os caminhos dos raios. Embora as ondas emitidas pela fonte se propaguem em várias direções, aquelas que atravessam as interfaces entre as camadas seguem a Lei de Snell equação (1.1). A velocidade das ondas sísmicas nas camadas subsuperficiais é responsável por determinar como essas ondas se propagam através das diferentes camadas:

$$\frac{\sin\theta_1}{V_1} = \frac{\sin\theta_2}{V_2} \tag{1.1}$$

Assim, θ_1 e θ_2 correspondem aos ângulos dos raios incidentes e transmitidos, medidos em relação à normal da interface. E as variáveis V_1 e V_2 representam as velocidades nas camadas em que os raios incidentes e transmitidos se propagam, respectivamente (Sheriff e Geldart, 1995).

A relação entre o tempo de percurso de ida e volta (t) de um raio refletido e o deslocamento horizontal (x) entre a fonte e o receptor pode ser descrita pela equação hiperbólica 1.2:

$$t^2(x) = t_0^2 + \frac{x^2}{v^2} \tag{1.2}$$

Nesta equação, $t_0 = \frac{2h}{v}$ é o tempo de ida e volta para deslocamento zero, sendo h a profundidade e v a velocidade de propagação da onda até a interface (Sheriff e Geldart, 1995). A camada mais próxima da superfície geralmente permite uma medição direta da velocidade e satisfaz a suposição hiperbólica da equação (1.2). No entanto, em camadas mais profundas, devido à curvatura dos raios, a equação pode não ser aplicável diretamente.

Apesar de a equação hiperbólica oferecer uma boa aproximação para a relação entre o tempo de percurso e o deslocamento em meios simples, sua aplicabilidade prática está condicionada a certas premissas que nem sempre são atendidas em ambientes geológicos reais. Especialmente em estruturas geologicamente complexas, como anticlinais, falhas ou zonas com forte anisotropia (Tsvankin e Gutierrez, 1996; Tsvankin, 1997; Grechka e Tsvankin, 1999).

À medida que se consideram camadas mais profundas, o desvio da trajetória dos raios sísmicos em relação à vertical tende a aumentar, acentuando os efeitos de curvatura e introduzindo erros sistemáticos na estimativa dos tempos de trânsito baseados no modelo hiperbólico. Nessas situações, a velocidade v da equação passa a representar uma média ponderada das velocidades, e não necessariamente a velocidade real de propagação no intervalo de interesse. Esse comportamento gera discrepâncias na correção do sobretempo normal (NMO), o que pode comprometer o alinhamento preciso dos refletores durante o empilhamento.

Para minimizar tais limitações, práticas modernas de processamento sísmico frequentemente incorporam análises detalhadas de velocidade, ajustes não hiperbólicos, e algoritmos de correção baseados em modelagem mais realista da propagação das ondas. Nessas abordagens, o uso de velocidades efetivas como a V_{NMO}, estimadas a partir de painéis de coerência, permite adaptar o modelo hiperbólico às condições observadas nos dados. Ainda assim, a validação dessas velocidades e a análise cuidadosa dos resíduos da correção são etapas fundamentais para garantir a consistência da imagem sísmica obtida.

Assim, embora a equação hiperbólica forneça a base teórica para a descrição do tempo de trânsito das ondas refletidas, e funcione adequadamente para camadas horizontais de velocidades constantes ou suavemente variantes, sua aplicação exige uma compreensão clara de suas limitações e das condições geológicas envolvidas, bem como a integração com métodos de análise mais refinados sempre que necessário.

1.1.3 Correção de sobretempo normal (NMO)

A correção de sobretempo normal (NMO - *Normal Moveout Correction*) é uma etapa crítica no processamento de dados sísmicos de reflexão. Essa correção busca compensar o aumento do tempo de trânsito das ondas refletidas causado pelo afastamento entre a fonte sísmica e o receptor (*offset*). Sem essa compensação, os refletores não se alinham horizontalmente, o que dificulta a interpretação geológica e compromete o empilhamento (*stacking*) (Yilmaz e Doherty, 1987).

A velocidade de correção de sobretempo normal, V_{NMO}, funciona como uma velocidade média eficaz utilizada para horizontalizar eventos hiperbólicos nos painéis CMP, mas ela nem sempre corresponde à velocidade intervalar real (V_{int}), especialmente em meios inclinados, heterogêneos ou anisotrópicos. Em contextos com anisotropia ou heterogeneidades verticais acentuadas, a curva de moveout deixa de ser exatamente hiperbólica e a velocidade estimada tende a representar um valor intermediário que otimiza o alinhamento dos refletores, ainda que não reflita com precisão as velocidades intervalares verdadeiras (Shah et al., n.d.; Grechka e Tsvankin, 2019).

A determinação de V_{NMO} é usualmente realizada por meio de *pickings* manuais em painéis de *semblance* que representam espectros de velocidades como medida de coerência para identificar eventos de reflexão primária. (Shah et al., n.d.; Taner e Koehler, 1969; Grechka e Tsvankin, 2019).

No entanto, a presença de ruído coerente e incoerente interfere diretamente no painel de *semblance* e pode comprometer tanto a qualidade da análise de velocidade quanto a confiabilidade do empilhamento resultante (Liu e Marfurt, 2015). Entre as limitações mais relevantes da NMO clássica está o pressuposto de refletor horizontal paralelo, o que não ocorre em estruturas curvadas e de maior complexidade, com ou em presença de anisotropia, condições nas quais surgem desvios não-hiperbólicos que dificultam a aplicação dessa correção sem ajustes adicionais (Grechka e Tsvankin, 2019).

Como resposta a essas limitações, as práticas avançadas de processamento sísmico têm adotado correções como DMO, empilhamento e migração pré-empilhamento, além de modelos de velocidade que incorporam efeitos anisotrópicos, buscando reduzir as imprecisões de tempo e fortalecer a robustez da imagem sísmica formada (Grechka e Tsvankin, 2019).

Para o caso de uma interface refletora horizontal em um meio isotrópico e homogêneo, o tempo de trânsito total da onda refletida segue uma relação hiperbólica, expressa pela equação:

$$t^2(x) = t_0^2 + \left(\frac{x^2}{V_{NMO}^2}\right) \tag{1.3}$$

Sendo, $t(x)$ o tempo de trânsito correspondente a um determinado afastamento x; t_0 o tempo de trânsito vertical, também conhecido como tempo zero-*offset*; e V_{NMO} a velocidade associada à correção de sobretempo normal.

Essa formulação permite determinar o tempo corrigido para cada traço sísmico em função do *offset*. A aplicação dessa correção tem como propósito a realocação dos refletores, originalmente curvos devido à geometria hiperbólica dos tempos de trânsito nos refletores, conforme pode ser visto na Figura 1.6 que exibe o mesmo sismograma antes e depois da correção NMO. Esse processo contribui diretamente para a eficácia da etapa subsequente de empilhamento comum de ponto médio (CMP *stacking*), aumentando a coerência dos sinais refletidos por meio do aprimoramento da razão sinal-ruído.

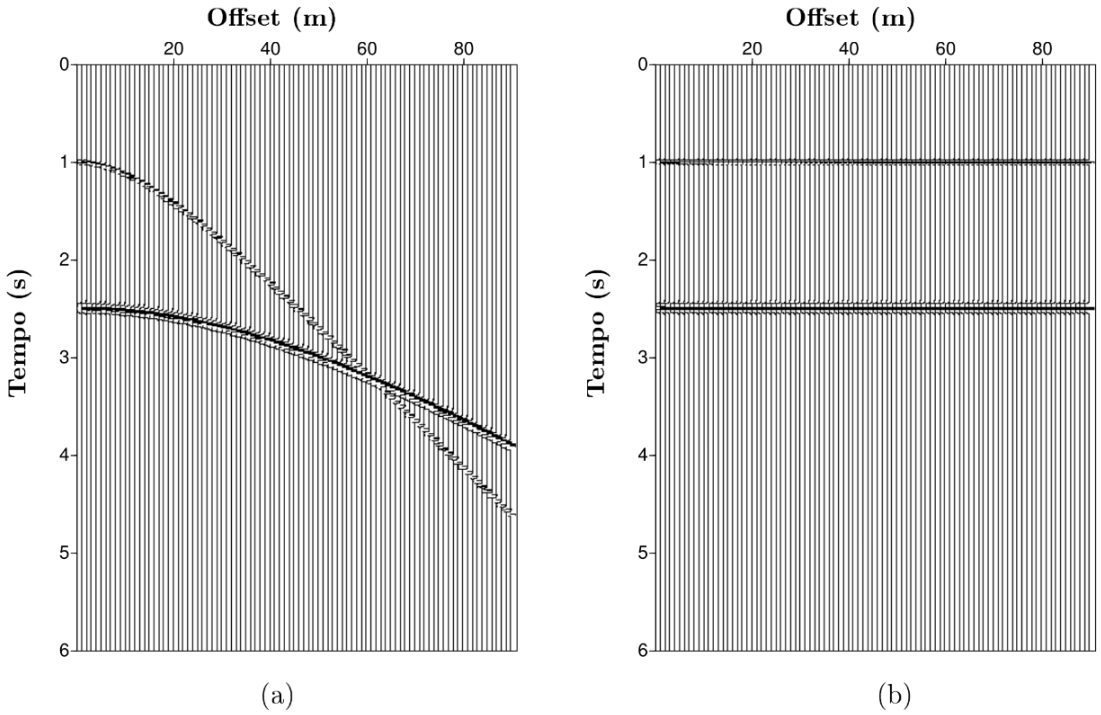


Figura 1.6: Sismograma para duas camadas: (a) Sismograma sem correção NMO. (b) Sismograma corrigido de NMO. Fonte: Adaptado de Souza (2014).

O processo de correção NMO envolve o mapeamento dos traços adquiridos para novos tempos de amostragem, de forma a alinhar os refletores horizontais. Esse procedimento exige a determinação do tempo de trânsito $t(x)$ original, a estimativa adequada de V_{NMO}, e a interpolação de amostras para reatribuição dos tempos corrigidos (Yilmaz, 2001; Sheriff e Geldart, 1995; Halliburton, 2009).

1.1.4 Estiramento de NMO (*NMO Stretching*)

Durante o processamento sísmico, um dos efeitos mais indesejáveis associados ao empilhamento é o chamado **estiramento de NMO** (*Normal Moveout Stretching*). Esse fenômeno ocorre quando, após a aplicação da correção de sobretempo normal (NMO), as amostras temporais são remapeadas de forma não uniforme, provocando uma dilatação dos sinais refletidos e, consequentemente, distorção da forma de onda (Yilmaz, 2001; Sheriff e Geldart, 1995; Ashcroft, 2011; Yang, 2014).

O problema torna-se mais pronunciado em eventos rasos (pequenos tempos de reflexão t_0) e em *offsets* longos, nos quais o deslocamento temporal $\Delta t = t(x) - t_0$ é mais significativo. Nesses casos, o remapeamento introduz um alongamento aparente do pulso sísmico, que reduz o conteúdo de altas frequências, altera a fase e compromete a coerência dos eventos (Dunkin e Levin, 1973; Chen et al., 2017; Khoshnavaz et al., 2021). Esse comportamento é descrito quantitativamente pelo **fator de estiramento**, definido como:

$$S = \frac{t(x) - t_0}{t_0}, \tag{1.4}$$

em que $t(x)$ é o tempo corrigido após o processo de NMO, t_0 é o tempo de reflexão vertical (*zero-offset*) e S representa o percentual de distorção temporal introduzida pela correção.

De acordo com Ashcroft (2011) e Yilmaz (2001), quando S ultrapassa valores entre **30% e 50%**, o evento refletido apresenta deformações significativas, o que leva à perda de coerência entre os traços. Essa perda é particularmente evidente nos eventos rasos, onde a diferença relativa $t(x) - t_0$ é mais expressiva, produzindo o chamado *pulso esticado*.

A Figura 1.7 ilustra de forma clara o efeito do estiramento. O painel (a) mostra os eventos originais com curvaturas hiperbólicas típicas, enquanto (b) exibe o resultado após a correção, onde se observa a dilatação e distorção das formas de onda em tempos rasos. O painel (c) apresenta a aplicação de uma janela de *muting*, utilizada para eliminar os segmentos mais afetados, e (d) mostra o empilhamento resultante após a mitigação do estiramento.

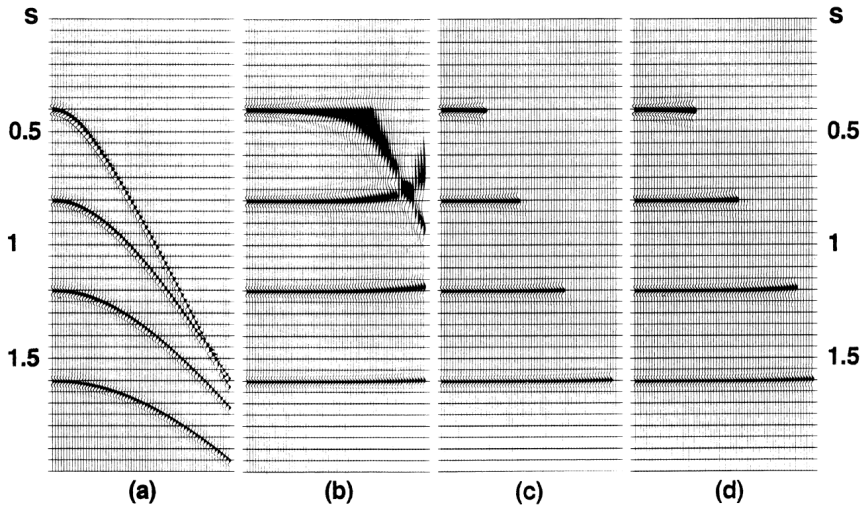


Figura 1.7: Efeito do estiramento de NMO (*NMO stretching*). (a) Seção original com curvaturas hiperbólicas; (b) distorção severa causada pelo estiramento em eventos rasos e grandes *offsets*; (c) aplicação de janela de *muting* para eliminar as zonas afetadas; (d) resultado empilhado após mitigação do efeito. O estiramento modifica a forma de onda e reduz a frequência efetiva do sinal. Fonte: Adaptado de Yilmaz (2001) e Ashcroft (2011).

O estiramento, segundo Sheriff e Geldart (1995), está diretamente relacionado à não linearidade do remapeamento temporal introduzido pela correção NMO. Como o processo não preserva a densidade amostral uniforme, o espaçamento entre as amostras varia ao longo do tempo, provocando compressão ou dilatação da forma de onda. Conforme observa Yang (2014), esse efeito acarreta perda de resolução vertical e deterioração da razão sinal-ruído, especialmente em camadas de baixa velocidade e em regiões caracterizadas por forte heterogeneidade lateral.

Para atenuar esse fenômeno, diversas estratégias têm sido aplicadas no processamento sísmico. Entre elas, destaca-se o uso de janelas de *muting* baseadas em *offset* ou profundidade, que permitem excluir as regiões mais afetadas pela distorção. Outro procedimento comum é a aplicação de técnicas de interpolação *spline* ou de reamostragem adaptativa com preservação de fase, voltadas à suavização do remapeamento temporal. Adicionalmente, a adoção de modelos de velocidade refinados contribui para reduzir o excesso de correção, minimizando a ocorrência de estiramento em eventos rasos (Chen et al., 2017; Khoshnavaz et al., 2021). Em trabalhos mais recentes, têm sido propostas abordagens modernas de NMO stretch-free, bem como correções realizadas no domínio τ–p, que buscam eliminar o estiramento sem comprometer a fidelidade do sinal (Khoshnavaz et al., 2021).

Em síntese, o controle do estiramento de NMO é essencial para garantir a integridade da forma de onda e preservar o conteúdo espectral dos dados empilhados. Sua mitigação constitui uma etapa indispensável em fluxos sísmicos modernos, assegurando a coerência dos refletores e a resolução temporal necessária para uma interpretação geológica precisa.

2

Fundamentação teórica: Equação de Dix e aprendizagem de máquina

2.1 Fundamentos físicos e avanços computacionais na análise de velocidade sísmica

No contexto da geofísica aplicada, é amplamente reconhecido que a velocidade de propagação das ondas sísmicas é uma variável crítica tanto no processamento quanto na interpretação dos dados de subsuperfície. Essa velocidade não é constante, pois depende diretamente de uma combinação de fatores físicos e químicos do meio rochoso, tais como a composição mineralógica, o teor de porosidade, o tipo de fluido presente nos poros, o grau de compactação, bem como a pressão de confinamento e a temperatura ambiente. Em particular, observa-se que a velocidade tende a aumentar com a profundidade em função da compactação progressiva, que reduz a porosidade e eleva a rigidez do material (Wyllie et al., 1956; Castagna et al., 1985; Wang e Nur, 1992; Schön, 2015).

Os primeiros estudos quantitativos sobre a relação entre a velocidade sísmica e as propriedades físicas das rochas remontam a Faust (1951), que propôs uma equação empírica associando a velocidade média ao tempo geológico, antecipando a variabilidade da velocidade com a profundidade e idade das formações (Faust, 1951). Posteriormente, a introdução da equação de Dix (1955) representou um marco para a geofísica, ao possibilitar a conversão das velocidades RMS em velocidades intervalares, permitindo inferências mais realistas sobre camadas individuais e não apenas sobre médias acumuladas. Esse avanço foi essencial para o desenvolvimento dos primeiros modelos de subsuperfície com maior fidelidade à estratigrafia real.

Nas décadas seguintes, os estudos em física de rochas evoluíram substancialmente, destacando o papel de propriedades como o conteúdo de argila, a saturação de fluidos e a microestrutura porosa na modulação das velocidades sísmicas. Trabalho como o de Castagna et al. (1985) refinou essa relação empírica, contribuindo para o desenvolvimento de equações de previsão baseadas em parâmetros petrofísicos.

Com a sísmica 3D e com métodos de migração mais sofisticados, a análise de velocidade passou a incorporar modelos anisotrópicos e técnicas baseadas em inversão sísmica, permitindo a construção de modelos dinâmicos que se adaptam à geometria real das camadas e à complexidade estrutural da subsuperfície (Mavko et al., 2009).

Atualmente, a análise de velocidade ocupa papel central em fluxos modernos de processamento, como migração pré-empilhamento e inversão *full-waveform*, reafirmando sua relevância na redução de incertezas e na acurácia das interpretações geológicas (Araújo, 2018).

Mais recentemente, metodologias baseadas em inteligência artificial têm revolucionado esse campo, introduzindo abordagens que aliam alta capacidade preditiva à eficiência computacional. Entre essas inovações, destaca-se o uso de redes neurais convolucionais (CNNs) e arquiteturas U-Net de alta resolução, que permitem aprimorar modelos de velocidade de baixa resolução e refinar estruturas geológicas complexas com elevada precisão (Kim et al., 2024). Outro avanço relevante é o modelo VelocityGPT, baseado em transformadores, que gera modelos de velocidade realistas de forma autoregressiva, condicionando camadas profundas às características estruturais das camadas rasas (Harsuko et al., 2025, 2024).

Além disso, abordagens baseadas em aprendizado profundo têm sido aplicadas com sucesso à inversão sísmica, tanto em dados sintéticos quanto reais, como no caso de estudos aplicados ao Golfo do México (Farris et al., 2023). Técnicas iterativas como o *Deep-Tomography* demonstram grande potencial na construção progressiva de modelos realistas, ao incorporar aprendizado profundo em ciclos sucessivos de refinamento (Muller et al., 2022). Complementando esse panorama, modelos de difusão condicionais vêm sendo explorados como ferramentas generativas para síntese controlada de modelos de velocidade, condicionados a dados geológicos, registros de poço ou imagens sísmicas, ampliando significativamente as possibilidades de modelagem inversa (Wang et al., 2024).

Esses avanços marcam uma nova era na análise de velocidade sísmica, na qual o conhecimento geológico tradicional é enriquecido por algoritmos de inteligência artificial, com implicações diretas na velocidade e qualidade da interpretação e na redução da ambiguidade dos modelos de subsuperfície.

Na prática da sísmica de reflexão, diferentes definições de velocidade são empregadas, de

acordo com o objetivo da análise. A **velocidade de empilhamento** (V_{stack}), obtida a partir da correção do sobretempo normal (NMO), corresponde àquela que melhor alinha os eventos refletidos em um conjunto de traços CMP, maximizando a coerência dos refletores. Essa velocidade, embora útil para fins de empilhamento, não representa diretamente a velocidade real das formações geológicas atravessadas pelas ondas.

2.1.1 Formulação e interpretação física da equação de Dix

Neste trabalho, para fins de modelagem e migração, a **velocidade intervalar** (V_{int}), que corresponde à velocidade média dentro de uma camada específica, e que é derivada das **velocidades RMS** (Root Mean Square), foi estimada por meio da equação de Dix (1955), a qual expressa a velocidade intervalar em função das diferenças quadráticas de tempo e velocidade entre camadas sucessivas:

$$V_{\text{int},n} = \sqrt{\frac{V_{\text{RMS},n}^2 \cdot t_n - V_{\text{RMS},n-1}^2 \cdot t_{n-1}}{t_n - t_{n-1}}} \tag{2.1}$$

sendo $V_{\text{int},n}$ a velocidade intervalar da camada n, $V_{\text{RMS},n}$ a velocidade RMS acumulada até a camada n, $V_{\text{RMS},n-1}$ a velocidade RMS até a camada anterior $(n-1)$, e t_n, t_{n-1} os tempos de dupla viagem (two-way travel time) acumulados até as interfaces n e $n-1$, respectivamente.

A equação de Dix (1955) parte do princípio de que a velocidade RMS até um tempo t_n representa uma média ponderada das velocidades intervalares das camadas anteriores. Portanto, ao subtrair a contribuição acumulada da camada anterior $(n-1)$, obtém-se a contribuição isolada da camada atual (n). A operação é realizada em termos quadráticos, pois as velocidades RMS e intervalares são definidas por médias quadráticas, refletindo a física da propagação das ondas em meios estratificados.

Do ponto de vista físico, a velocidade sísmica é também influenciada pelo estado de confinamento das rochas. Em regiões mais profundas, onde a pressão vertical é maior, a compactação reduz a porosidade e aumenta a rigidez do meio, elevando, consequentemente, a velocidade das ondas compressivas (ondas P). De forma complementar, variações bruscas nas propriedades elásticas, como mudanças de litologia ou presença de fluidos, introduzem contrastes de impedância que se refletem nas velocidades medidas.

2.2 Aprendizagem não supervisionada e k-means++: fundamentos e implementação

A aprendizagem não supervisionada constitui um ramo da inteligência artificial em que o objetivo é identificar estruturas subjacentes ou padrões em conjuntos de dados sem a necessidade de rótulos predefinidos (Jain, 2010). Dentro desse contexto, os algoritmos de agrupamento (*clustering*) desempenham papel central na descoberta de relações de similaridade, sendo o método k-means um dos mais amplamente utilizados pela sua simplicidade e eficiência (Xu e Wunsch, 2005; Lloyd, 1982).

O algoritmo k-means clássico foi introduzido por MacQueen et al. (1967) e posteriormente formalizado por Lloyd (1982), com o objetivo de particionar um conjunto de n observações em K grupos distintos, de modo que a variabilidade intracluster seja minimizada e a variabilidade intercluster maximizada. No entanto, sua eficácia depende fortemente da escolha inicial dos centróides, sendo esta uma das principais causas de convergência a mínimos locais e de instabilidade entre diferentes execuções (Arthur e Vassilvitskii, 2007; Fraley e Raftery, 2002).

Para superar essa limitação, Arthur e Vassilvitskii (2007) propuseram o algoritmo k-means++, uma modificação que aprimora a seleção inicial dos centróides por meio de uma estratégia probabilística ponderada pela distância ao quadrado, o que aumenta significativamente a probabilidade de convergência ao ótimo global. Estudos subsequentes confirmaram que o k-means++ reduz o erro médio de inicialização em até 90% em relação ao método tradicional (Bishop, 2006; Celebi et al., 2013; Bahmani et al., 2012). A Figura 2.1 ilustra, de forma esquemática, o processo de seleção dos centróides no k-means++, evidenciando como a probabilidade de escolha é influenciada pelas distâncias ao quadrado e como isso resulta em uma partição mais estável e representativa dos dados.

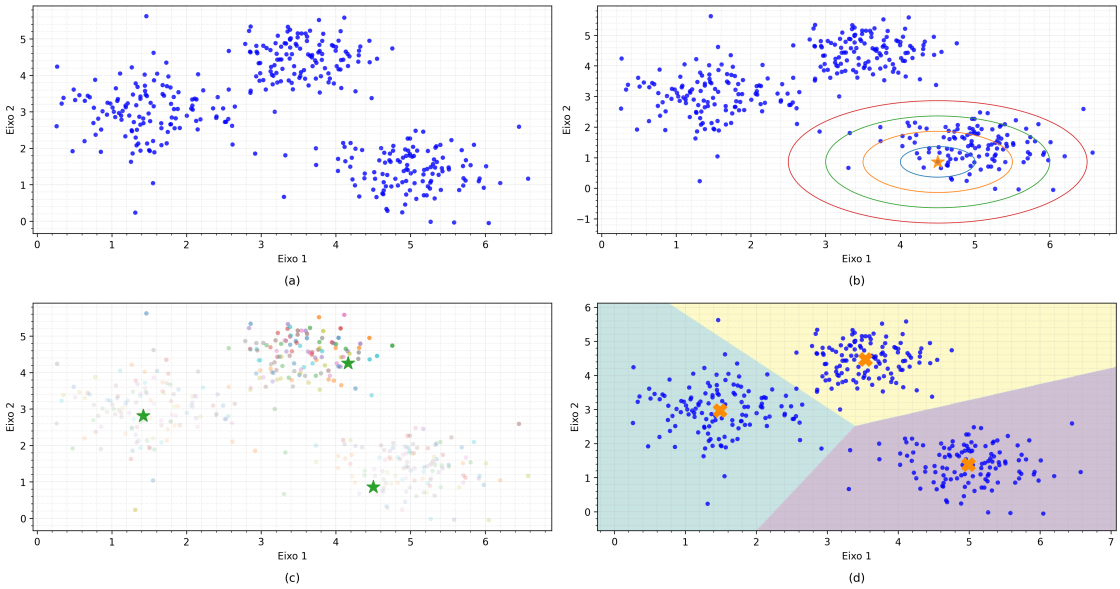


Figura 2.1: Exemplo K-means++: (a) Dados originais. (b) 1º centróide (aleatório) e D^2 preliminar. (c) Seleção probabilística $P(x_i) \propto D(x_i)^2$. (d) Partição final (k-means).

Assim, o k-means++ busca melhorar a qualidade da inicialização sem aumentar consideravelmente o custo computacional, mantendo o mesmo tempo de convergência assintótica, mas com desempenho estatisticamente mais consistente em grandes conjuntos de dados (Tan et al., 2019).

2.2.1 Inicialização dos centróides e cálculo das distâncias mínimas

A primeira etapa do k-means++ consiste na escolha aleatória do primeiro centróide μ_1 a partir de uma distribuição uniforme sobre o conjunto de dados $X = \{x_1, x_2, \ldots, x_n\}$, onde cada vetor $x_i \in \mathbb{R}^d$ representa um ponto em um espaço d-dimensional. Todos os pontos possuem igual probabilidade de serem escolhidos, de modo que:

$$P(\mu_1 = x_i) = \frac{1}{n}, \quad \forall x_i \in X. \tag{2.2}$$

Em seguida, calcula-se a distância mínima entre cada ponto e o conjunto de centróides já selecionados, representada por:

$$D(x_i)^2 = \min_{\mu_j \in M} \|x_i - \mu_j\|_2^2 = (x_i - \mu_j)^\top (x_i - \mu_j), \tag{2.3}$$

onde M é o conjunto dos centróides escolhidos até o momento. Essa métrica, baseada na norma euclidiana, quantifica a proximidade de cada ponto x_i em relação ao centróide mais próximo (Bishop, 2006).

Para otimizar o cálculo vetorial, pode-se representar as distâncias em forma matricial, considerando $X \in \mathbb{R}^{n \times d}$ e $M \in \mathbb{R}^{K \times d}$:

$$D^2 = \text{diag}(XX^\top)\mathbf{1}^\top - 2XM^\top + \mathbf{1}\text{diag}(MM^\top)^\top, \tag{2.4}$$

em que $\mathbf{1}$ representa um vetor coluna de uns, e $\text{diag}(A)$ denota o vetor da diagonal principal da matriz A. A menor distância para cada ponto é obtida por:

$$D_{\min,i} = \min_j D_{ij}^2, \quad \forall i \in \{1, \ldots, n\}. \tag{2.5}$$

Essa abordagem matricial é computacionalmente eficiente e apropriada para grandes volumes de dados (Hastie et al., 2009; Shindler et al., 2011).

2.2.2 Distribuição probabilística e seleção otimizada dos centróides

Após a definição do primeiro centróide, o algoritmo seleciona os centróides subsequentes de forma probabilística, atribuindo a cada ponto x_i uma probabilidade proporcional ao quadrado de sua distância ao centróide mais próximo:

$$P(x_i) = \frac{D(x_i)^2}{\sum_{j=1}^{n} D(x_j)^2}. \tag{2.6}$$

Essa formulação, proposta originalmente por Arthur e Vassilvitskii (2007), assegura que pontos distantes dos centróides existentes tenham maior probabilidade de serem escolhidos, promovendo uma dispersão inicial mais homogênea no espaço de dados.

Como apontam Steinley (2006) e Celebi et al. (2013), essa estratégia reduz o viés espacial de inicialização e melhora a estabilidade dos agrupamentos gerados, especialmente em conjuntos de dados não balanceados ou de alta dimensionalidade.

O processo é repetido até que sejam definidos os K centróides $\{\mu_1, \mu_2, \ldots, \mu_K\}$, garantindo que a configuração inicial esteja estatisticamente distribuída ao longo das principais regiões de densidade dos dados (Xu e Wunsch, 2008).

2.2.3 Iterações de agrupamento e convergência do algoritmo

Com os K centróides iniciais definidos, o algoritmo k-means++ prossegue com as iterações do k-means padrão. Cada ponto x_i é atribuído ao cluster C_k correspondente ao centróide mais próximo, conforme:

$$x_i \in C_k \quad \text{se} \quad \|x_i - \mu_k\|_2^2 = \min_{1 \leq j \leq K} \|x_i - \mu_j\|_2^2. \tag{2.7}$$

Após a atribuição, os centróides são recalculados como a média dos pontos em cada cluster:

$$\mu_k^{(t+1)} = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i. \tag{2.8}$$

A função objetivo, que mede a variabilidade intra-cluster e é minimizada iterativamente, é definida por:

$$W(C_k) = \sum_{x_i \in C_k} \|x_i - \mu_k\|_2^2, \tag{2.9}$$

ou, de forma global para todos os clusters:

$$\text{tot.withinss} = \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - \mu_k\|_2^2. \tag{2.10}$$

Quanto menor o valor de tot.withinss, maior a compactação dos clusters e mais eficiente o agrupamento (Kanungo et al., 2002). Essa otimização pode ser interpretada como uma minimização de energia, equivalente à decomposição de variância total do sistema em variância intracluster e intercluster, similar à análise de variância multivariada (ANOVA) (Milligan e Cooper, 1985).

Em algumas aplicações, a métrica euclidiana pode ser substituída por medidas de correlação para capturar dependências lineares entre as variáveis. A distância baseada na correlação de Pearson é expressa como:

$$d_{\text{cor}}(x, y) = 1 - \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{2.11}$$

onde $\bar{x}$ e $\bar{y}$ representam as médias das variáveis x e y. Essa medida é especialmente útil em dados sísmicos, pois considera correlações de fase e amplitude entre traços (Rajagopalan et al., 2019).

O processo iterativo continua até que o deslocamento dos centróides entre duas iterações sucessivas seja inferior a um limite δ, isto é:

$\|\mu_k^{(t+1)} - \mu_k^{(t)}\|_2 < \delta, \quad \forall k.$

Essa condição de parada garante a convergência do algoritmo, identificando padrões estáveis e fisicamente coerentes, reduzindo o erro de reconstrução e acelerando a convergência global (Bahmani et al., 2012; Tan et al., 2019; Pelleg e Moore, 2000; Ding et al., 2015).

No contexto da análise sísmica, essa técnica é particularmente relevante, pois permite segmentar grandes volumes de dados de amplitude em grupos coerentes com as variações de refletividade e estrutura das camadas subsuperficiais, servindo como etapa essencial na inferência automatizada de velocidades e padrões estruturais.

A abordagem de clusterização utilizada é apresentada no algoritmo abaixo:

Algorithm 1: Clusterização com k-means++

- **Passo 1: Escolher o primeiro centróide aleatoriamente**
  - Selecione o primeiro centróide aleatoriamente entre os pontos de dados.
- **Passo 2: Escolher os próximos centróides com base na distância**
  - Para cada ponto de dados, calcule a distância até o centróide mais próximo já selecionado.
- **Passo 3: Seleção probabilística dos centróides**
  - Selecione o próximo centróide com uma probabilidade proporcional ao quadrado da distância calculada no passo anterior.
- **Passo 4: Repetir até selecionar K centróides**
  - Continue o processo até que o número total de centróides desejado K seja atingido.
- **Passo 5: Executar o k-means++ padrão**
  - Agora que os K centróides iniciais foram selecionados, o algoritmo segue com o K-means padrão, realizando as seguintes etapas:
    - **Atribuir pontos aos centróides mais próximos**
      - Atribua cada ponto de dados ao centróide mais próximo, calculando a distância entre cada ponto e os centróides.
    - **Recalcular os centróides**
      - Após a atribuição de todos os pontos, recalculam-se os centróides como a média dos pontos pertencentes a cada cluster.
    - **Iterar até a convergência**
      - Continue a reatribuir os pontos e a recalcular os centróides, repetindo o processo até que não haja mais mudanças nas atribuições de pontos ou até atingir um critério de convergência.

2.2.4 Benefícios da variante k-means++

O k-means++ apresenta diversas vantagens em relação ao K-means tradicional, principalmente na fase de inicialização dos centróides.

1. Melhor desempenho: A escolha mais inteligente dos centróides iniciais reduz significativamente a probabilidade de uma má inicialização, que poderia levar a uma convergência lenta ou resultados de baixa qualidade. O K-means++ mitiga esse risco, melhorando a eficiência do processo.

2. Convergência mais rápida: Como os centróides iniciais estão melhor distribuídos, o algoritmo tende a convergir mais rapidamente. Isso ocorre porque a fase inicial de agrupamento já está otimizada, economizando iterações desnecessárias.

3. Menor sensibilidade à inicialização: O k-means padrão pode ser muito sensível à escolha inicial dos centróides, resultando em respostas variáveis entre execuções. O k-means++ diminui essa sensibilidade, tornando o processo mais robusto e os resultados mais consistentes.

Desse modo, a utilização do k-means++ melhora significativamente o k-means padrão ao selecionar centróides iniciais de forma mais eficaz. Utilizando uma estratégia probabilística baseada nas distâncias dos pontos de dados aos centróides já selecionados, o k-means++ garante uma melhor dispersão dos centróides e, consequentemente, uma maior eficiência no processo de agrupamento. A introdução do procedimento de pré-agrupamento amostral, em conjunto com o k-means++, assegura que os dados estejam devidamente organizados para maximizar a performance e a estabilidade do algoritmo.

2.3 Aprendizagem não supervisionada e Análise de Componentes Principais (PCA): fundamentos e aplicação

A *Principal Component Analysis* (PCA), ou Análise de Componentes Principais, é uma técnica estatística de aprendizagem não supervisionada amplamente utilizada para a redução de dimensionalidade em conjuntos de dados multivariados (Pearson, 1901; Hotelling, 1933). Seu princípio fundamental consiste em transformar um conjunto original de variáveis correlacionadas em um novo conjunto de variáveis ortogonais, denominadas *componentes principais*, que capturam a maior variância possível presente nos dados (Jolliffe, 2002; Jackson, 2005). A Figura 2.2 ilustra esse processo de transformação, mostrando como os eixos

originais são reorientados para alinhar-se às direções de maior variabilidade, resultando em uma representação mais compacta e informativa dos dados.

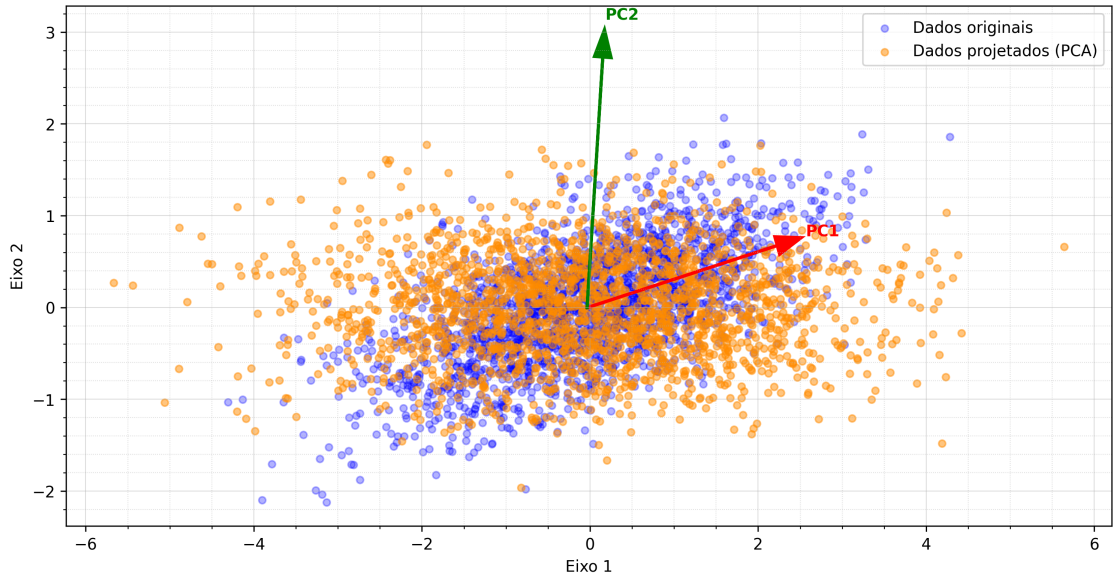


Figura 2.2: Exemplo ilustrativo do processo de redução de dimensionalidade por meio da PCA, no qual os dados originais são projetados sobre novos eixos correspondentes aos componentes principais.

A aplicação da PCA em dados sísmicos tem sido especialmente relevante em estudos que envolvem painéis de *Common Midpoint* (CMP), uma vez que esses dados contêm redundâncias significativas e forte correlação entre traços adjacentes (Bishop, 2006; Van der Maaten et al., 2009). Ao aplicar a PCA após a identificação de centróides via k-means++, é possível transformar o conjunto de traços sísmicos em um espaço de menor dimensionalidade, no qual as direções principais (autovetores) são ordenadas de acordo com a variância explicada (autovalores), ou seja, pela quantidade de informação retida em cada componente (Shlens, 2014; Vasconcelos, 2021).

Neste contexto, a PCA atua como um filtro estatístico que sintetiza as informações mais relevantes, eliminando ruídos e redundâncias e destacando as direções de máxima variação energética nos sinais (Wold et al., 1987). Assim, obtém-se um traço representativo que concentra a maior energia de todo o painel CMP, o que é essencial para a análise de coerência e para o cálculo de velocidades sísmicas mais estáveis (Ulrych et al., 2012).

2.3.1 Formulação matemática da PCA

Considerando um conjunto de dados sísmicos representado por uma matriz $X \in \mathbb{R}^{n \times p}$, onde cada linha x_i $(i = 1, 2, \ldots, n)$ corresponde a uma observação temporal (amostras de tempo) e cada coluna x_j $(j = 1, 2, \ldots, p)$ representa as amplitudes de um traço sísmico em instantes

de tempo distintos, o primeiro passo é centralizar os dados em torno da média:

$$X_c = X - \bar{X}, \tag{2.12}$$

em que $\bar{X}$ é o vetor das médias das colunas de X.

Em seguida, calcula-se a matriz de covariância $C \in \mathbb{R}^{p \times p}$, que descreve a relação de variabilidade entre as variáveis (amplitudes ao longo dos traços):

$$C = \frac{1}{n-1} X_c^T X_c. \tag{2.13}$$

Cada elemento C_{ij} da matriz C é definido como:

$$C_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j), \tag{2.14}$$

em que C_{ij} representa a covariância entre as variáveis i e j. Valores elevados de C_{ij} indicam que as variáveis variam de forma semelhante, refletindo alta correlação entre traços ou amostras temporais adjacentes (Anderson, 2003).

A determinação das direções principais é realizada por meio da decomposição espectral da matriz de covariância. Isso envolve a resolução do problema de autovalores e autovetores definido por:

$$Cv_i = \lambda_i v_i, \quad i = 1, 2, \ldots, p, \tag{2.15}$$

em que v_i é o autovetor associado ao autovalor λ_i. A equação característica correspondente é dada por:

$$(C - \lambda_i I) v_i = 0, \tag{2.16}$$

sendo I a matriz identidade de dimensão $p \times p$.

Os autovalores λ_i quantificam a variância explicada por cada direção principal, e os autovetores v_i definem as direções ortogonais ao longo das quais a variância é máxima (Bishop, 2006; Jolliffe, 2002). A ordenação dos autovalores em ordem decrescente ($\lambda_1 > \lambda_2 > \cdots > \lambda_p$) garante que o primeiro componente principal (PC_1) capture a maior parcela da variabilidade total do conjunto de dados.

2.3.2 Projeção no espaço dos componentes principais

Com os autovalores e autovetores determinados, o conjunto de dados original pode ser projetado em um novo espaço definido pelos autovetores dominantes, formando a matriz de componentes principais $Z \in \mathbb{R}^{n\times m}$:

$$Z = X_c V_m, \tag{2.17}$$

em que $V_m = [v_1, v_2, \ldots, v_m]$ contém os m autovetores correspondentes aos maiores autovalores. Cada coluna de Z representa uma componente principal Z_m, que captura uma fração específica da variância total dos dados.

O primeiro componente principal é dado por:

$$Z_1 = X_c v_1, \tag{2.18}$$

e reflete a direção de máxima variação, no caso sísmico, as principais flutuações de amplitude associadas aos refletores mais energéticos do painel CMP (Ulrych et al., 2012; Vassiliou et al., 2021).

A variância total do conjunto de dados é dada por:

$$\mathrm{Var}_{\mathrm{total}} = \sum_{j=1}^{p} \mathrm{Var}(X_j) = \sum_{j=1}^{p} \frac{1}{n} \sum_{i=1}^{n} x_{ij}^2, \tag{2.19}$$

enquanto a variância explicada pela m-ésima componente principal é:

$$\mathrm{Var}(Z_m) = \frac{1}{n} \sum_{i=1}^{n} z_{im}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{p} \phi_{jm} x_{ij} \right)^2, \tag{2.20}$$

sendo ϕ_{jm} o elemento da matriz de autovetores V_m. A proporção da variância explicada (PVE) pela m-ésima componente é dada por:

$$\mathrm{PVE}_m = \frac{\mathrm{Var}(Z_m)}{\mathrm{Var}_{\mathrm{total}}} = \frac{\sum_{i=1}^{n} z_{im}^2}{\sum_{j=1}^{p} \sum_{i=1}^{n} x_{ij}^2}. \tag{2.21}$$

Conforme Jolliffe (2002), as primeiras componentes geralmente explicam a maior parte da variância, enquanto as componentes de ordem superior contêm ruído e pequenas variações residuais. Em aplicações sísmicas, Moulik e Ekström (2014) e Sen et al. (2019) demonstram

que a retenção das duas ou três primeiras componentes já é suficiente para reconstruir padrões dominantes de reflexão com alta fidelidade.

2.3.3 Análise geométrica e aplicação sísmica

Geometricamente, a PCA corresponde a uma rotação do sistema de coordenadas original para um novo sistema alinhado às direções de máxima variação (Ringnér, 2008). Em termos práticos, a primeira componente principal define a direção de maior energia dos traços sísmicos, enquanto as subsequentes capturam variações secundárias ortogonais, frequentemente associadas a ruídos ou anisotropias locais (Shlens, 2014; Ulrych et al., 2012).

A aplicação da PCA em dados sísmicos permite identificar padrões coerentes entre traços, realçar feições estruturais e reduzir a redundância inerente a grandes volumes de dados de reflexão (Koch, 2007; Wang et al., 2019). Além disso, ao ser utilizada após a clusterização k-means++, a PCA atua como uma ferramenta de síntese em que cada componente representa uma combinação linear de traços com pesos determinados pelos autovetores correspondentes, o que garante que as variações mais significativas de amplitude e energia sejam preservadas (Vasconcelos, 2021; Bishop, 2006).

Portanto, ao ser aplicada sobre os traços sísmicos filtrados, a PCA reorganiza o conjunto de amplitudes ao longo das direções principais, destacando as variações mais relevantes e associando cada componente a propriedades físicas específicas do meio, como contraste de impedância e continuidade lateral dos refletores, pois, segundo Yilmaz (2001); Sheriff e Geldart (1995), as amplitudes registradas nos dados sísmicos resultam diretamente das diferenças de impedância acústica entre camadas adjacentes, que controlam a intensidade e a polaridade das reflexões. Essas variações de impedância estão intimamente relacionadas às propriedades litológicas e aos padrões de velocidade do meio, de modo que contrastes mais acentuados produzem reflexões mais fortes e bem definidas. Além disso, a continuidade lateral dos refletores influencia a coerência e a qualidade da análise de velocidade, uma vez que descontinuidades estruturais ou estratigráficas tendem a dispersar a energia refletida, dificultando a identificação de eventos consistentes em *gathers* CMP e comprometendo o delineamento preciso das curvas de *semblance*. Assim, a combinação entre contraste de impedância e continuidade lateral constitui um dos principais fatores que determinam a resolução, a estabilidade e a confiabilidade dos modelos de velocidade obtidos na interpretação sísmica.

Assim, o processo resulta em uma representação compacta e informativa dos dados, que reduz a dimensionalidade e preserva as principais características sísmicas, proporcionando análises mais robustas, rápidas e precisas.

2.4 Aprendizagem supervisionada Perceptron Multicamadas (MLP): Modelagem de perfis de velocidade

As redes neurais artificiais (RNAs) constituem modelos computacionais inspirados na estrutura e no funcionamento dos neurônios biológicos, capazes de aprender relações complexas entre variáveis por meio de processos iterativos de ajuste de parâmetros. Conforme Haykin (2009), uma RNA é um sistema adaptativo que modifica sua estrutura interna de modo a realizar tarefas específicas, como classificação, regressão e predição de padrões não lineares. A base conceitual desse modelo foi introduzida por McCulloch e Pitts (1943), que propuseram o primeiro neurônio artificial, formalizando o processo de somatório ponderado e ativação lógica como unidade fundamental de processamento.

O modelo *Perceptron* simples, formulado por Rosenblatt (1958), representa o primeiro modelo de aprendizagem supervisionada e consiste em uma única camada de pesos ajustáveis conectada a uma função de ativação. Embora eficiente para problemas linearmente separáveis, o perceptron simples apresenta limitações significativas ao lidar com funções não lineares, como demonstrado por Minsky e Papert (1969). Para contornar essas restrições, surgiu o modelo Perceptron Multicamadas (MLP – *Multilayer Perceptron*), que incorpora uma ou mais camadas ocultas e funções de ativação não lineares, permitindo a aproximação de qualquer função contínua com precisão arbitrária, segundo o teorema da aproximação universal proposto por Hornik et al. (1989).

2.5 Aprendizagem supervisionada Perceptron Multicamadas (MLP): Modelagem de perfis de velocidade

As redes neurais artificiais (RNAs) constituem modelos computacionais inspirados na estrutura e no funcionamento dos neurônios biológicos, capazes de aprender relações complexas entre variáveis por meio de processos iterativos de ajuste de parâmetros. Conforme Haykin (2009), uma RNA é um sistema adaptativo que modifica sua estrutura interna de modo a realizar tarefas específicas, como classificação, regressão e predição de padrões não lineares. A base conceitual desse modelo foi introduzida por McCulloch e Pitts (1943), que propuseram o primeiro neurônio artificial, formalizando o processo de somatório ponderado e ativação lógica como unidade fundamental de processamento.

O modelo *Perceptron* simples, formulado por Rosenblatt (1958), representa o primeiro modelo de aprendizagem supervisionada e consiste em uma única camada de pesos ajustáveis conectada a uma função de ativação. Embora eficiente para problemas linearmente sepa-

ráveis, o perceptron simples apresenta limitações significativas ao lidar com funções não lineares, como demonstrado por Minsky e Papert (1969). Para contornar essas restrições, surgiu o modelo Perceptron Multicamadas (MLP – *Multilayer Perceptron*), que incorpora uma ou mais camadas ocultas e funções de ativação não lineares, permitindo a aproximação de qualquer função contínua com precisão arbitrária, segundo o teorema da aproximação universal proposto por Hornik et al. (1989). A Figura 2.3 apresenta, de forma esquemática, a arquitetura de uma rede neural do tipo Perceptron Multicamadas (MLP), composta por uma camada de entrada, múltiplas camadas ocultas e uma camada de saída. Essa estrutura evidencia o fluxo de informações ao longo das conexões totalmente interligadas entre os neurônios, permitindo à rede modelar relações não lineares complexas entre as variáveis de entrada e os resultados previstos, característica fundamental que diferencia as arquiteturas multicamadas dos perceptrons simples.

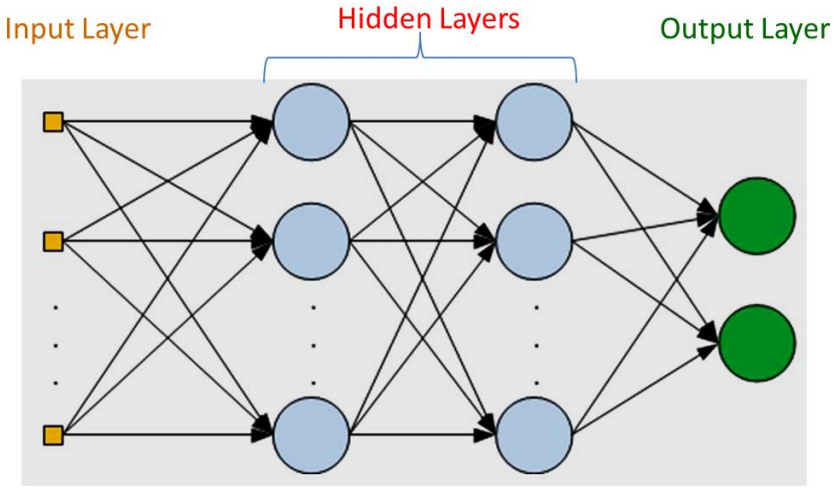


Figura 2.3: Estrutura esquemática de uma rede neural do tipo Perceptron Multicamadas (MLP), composta por uma camada de entrada (*Input Layer*), múltiplas camadas ocultas (*Hidden Layers*) e uma camada de saída (*Output Layer*). As conexões entre os neurônios são totalmente conectadas e ajustadas iterativamente durante o processo de treinamento supervisionado, com base no cálculo do erro e na retropropagação dos gradientes. Fonte: Adaptado de Book (2025).

A MLP utilizada nesta tese foi implementada para o ajuste e suavização dos perfis de velocidade $\hat{v}_{\mathrm{NMO}}(t_0)$ obtidos a partir da equação de Dix. Formalmente, o modelo pode ser expresso como uma função composta entre múltiplas camadas, descrita por:

$$\mathbf{y} = f^{(L)}\big(f^{(L-1)}\big(\cdots f^{(2)}\big(f^{(1)}(\mathbf{x}\mathbf{W}^{(1)} + \mathbf{b}^{(1)})\big)\cdots\big)\big)\,, \tag{2.22}$$

em que $\mathbf{x} \in \mathbb{R}^n$ representa o vetor de entrada, $\mathbf{W}^{(l)}$ e $\mathbf{b}^{(l)}$ são, respectivamente, a matriz de pesos e o vetor de *bias* da l-ésima camada, $f^{(l)}(\cdot)$ é a função de ativação da camada l, e L indica o número total de camadas (incluindo a de saída). Cada camada realiza uma

transformação linear seguida de uma não linearidade, sendo esta última essencial para a modelagem de relações complexas e não lineares (Bishop, 2006; Goodfellow et al., 2016).

A função de ativação empregada neste trabalho é a *logística sigmoide*, definida por:

$$f^{(l)}(z_i) = \frac{1}{1 + e^{-z_i}}, \tag{2.23}$$

onde z_i é a entrada ponderada do neurônio i na camada l. Essa função é amplamente utilizada em regressões sísmicas devido à sua suavidade e capacidade de mapear valores contínuos entre 0 e 1, o que facilita a normalização de amplitudes e velocidades (Zhang e Schuster, 2000; Liu et al., 2010).

O processo de aprendizagem da MLP baseia-se na minimização de uma função de custo $E(\mathbf{W})$, definida como o erro quadrático médio (*Mean Squared Error – MSE*), expressa por:

$$E(\mathbf{W}) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 , \tag{2.24}$$

em que y_i representa o valor alvo (velocidade verdadeira) e $\hat{y}_i$ é a saída estimada pela rede para a i-ésima amostra. O gradiente do erro em relação aos pesos é calculado por meio do algoritmo de *retropropagação do erro* (*backpropagation*), introduzido por Rumelhart et al. (1986), o qual aplica a regra da cadeia para propagar o erro das saídas em direção às camadas internas, ajustando iterativamente os pesos $\mathbf{W}$ para minimizar o erro global.

O ajuste dos pesos segue a regra de atualização:

$$\mathbf{W}^{(l)} \leftarrow \mathbf{W}^{(l)} - \eta \frac{\partial E}{\partial \mathbf{W}^{(l)}}, \tag{2.25}$$

em que η é a taxa de aprendizado (*learning rate*). Entretanto, em redes profundas ou dados ruidosos, o uso de gradiente descendente puro pode resultar em convergência lenta ou instável. Para mitigar esses efeitos, foi utilizado o otimizador Adam (*Adaptive Moment Estimation*), desenvolvido por Kingma e Ba (2015), que combina as vantagens dos métodos de momento e RMSProp, ajustando dinamicamente o tamanho do passo de aprendizado com base nas médias móveis dos gradientes de primeira e segunda ordem.

O cálculo das estimativas de momento e variância é realizado conforme:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla E_t, \qquad v_t = \beta_2 v_{t-1} + (1 - \beta_2)(\nabla E_t)^2, \tag{2.26}$$

em que m_t é o 1º momento (média, "momentum") e v_t é o 2º momento (média dos quadrados, "RMS"), já β_1 e β_2 são fatores de decaimento exponencial (tipicamente 0,9 e 0,999). Para corrigir o *bias* inicial nas estimativas, aplicam-se as expressões:

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \qquad \hat{v}_t = \frac{v_t}{1-\beta_2^t}, \tag{2.27}$$

resultando na atualização final dos parâmetros da rede:

$$\theta_t = \theta_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}, \tag{2.28}$$

em que θ_t representa os parâmetros ajustáveis da rede, e ϵ é um termo de regularização numérica para evitar divisões por zero.

A arquitetura MLP implementada nesta tese é composta por quatro camadas ocultas com 1024, 512, 256 e 128 neurônios, respectivamente, configuradas com ativação logística e regularização implícita via Adam. Essa configuração foi inspirada em estudos de redes profundas aplicadas à modelagem de perfis geofísicos (Zhang et al., 2019; Araya-Polo et al., 2020; Ma et al., 2021), nos quais observou-se que estruturas mais densas capturam melhor a variabilidade não linear das velocidades de propagação.

Após o treinamento, a saída da rede $\hat{v}_{\mathrm{NMO}}(t_0)$ é ajustada para obedecer ao comportamento físico de crescimento monotônico com o tempo duplo (t_0), conforme a função de acumulação máxima:

$$\hat{v}_{\mathrm{mon}}(t_i) = \max\{\hat{v}(t_1), \hat{v}(t_2), \ldots, \hat{v}(t_i)\}, \quad \forall i = 1, 2, \ldots, n, \tag{2.29}$$

assegurando que $\hat{v}_{\mathrm{mon}}(t_i) \geq \hat{v}_{\mathrm{mon}}(t_{i-1})$, o que preserva a coerência física da propagação sísmica, já que as velocidades empilhadas tendem a aumentar com a profundidade ou permanecer constantes (Yilmaz, 2001; Sen e Stoffa, 2019). Essa restrição, implementada após a predição, garante que as curvas de velocidade resultantes mantenham o comportamento esperado, eliminando oscilações espúrias decorrentes de ruído ou instabilidades numéricas.

Em síntese, o uso da MLP neste trabalho se justifica por sua capacidade de aprender relações complexas entre amplitudes, tempos de trânsito e velocidades derivadas, proporcionando suavização automática dos perfis e robustez frente a dados ruidosos. Como destacam Goodfellow et al. (2016) e LeCun et al. (2015), redes profundas com ativação não linear possuem elevada capacidade de generalização, tornando-as especialmente adequadas para modelagem de fenômenos físicos complexos, como a propagação de ondas sísmica em meios heterogêneos.

3

Metodologia híbrida para automação do campo de velocidades sísmicas

3.1 Modelagem e caracterização dos dados sísmicos

Com o objetivo de validar os procedimentos metodológicos propostos e avaliar sua eficácia em um ambiente controlado, foi gerado um modelo sintético de dados sísmicos. Esse tipo de modelagem constitui uma etapa fundamental em estudos de processamento, pois permite isolar variáveis, introduzir condições específicas de aquisição e analisar a resposta do sistema antes da aplicação em dados reais.

O modelo sintético adotado, ilustrado na Figura 3.1, foi construído no software de processamento sísmico *Seismic Unix* (**SU**), uma ferramenta de código aberto desenvolvida e mantida pelo *Center for Wave Phenomena* (CWP) da *Colorado School of Mines* (CSM). Trata-se de um cenário de camadas planas inclinadas, idealizado para reproduzir feições típicas da subsuperfície e incluir efeitos de múltiplas ordens, de modo a representar um desafio próximo ao encontrado em situações reais de aquisição.

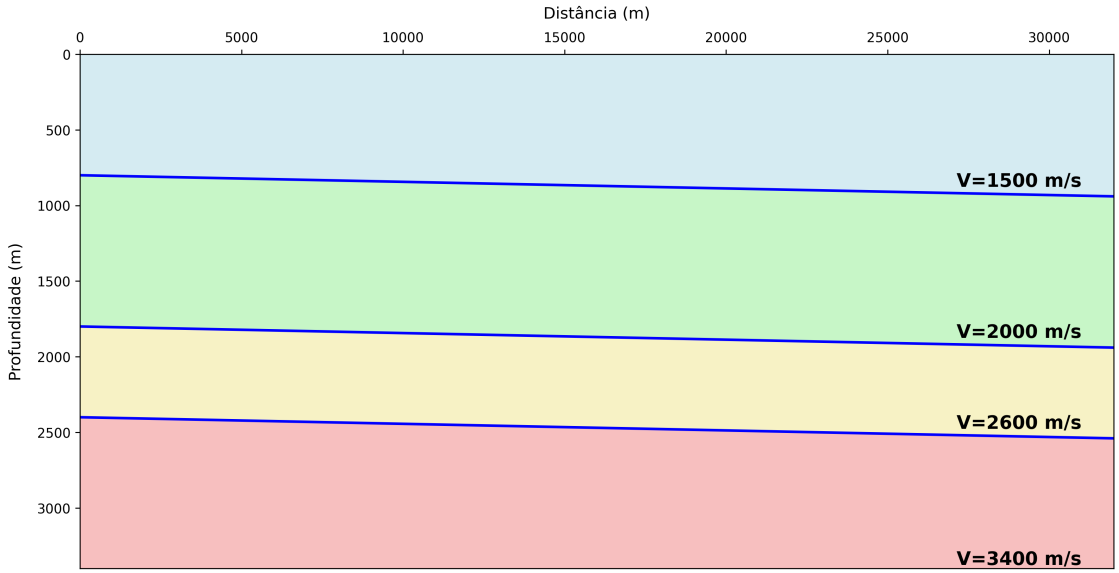


Figura 3.1: Modelo de camadas planas inclinadas.

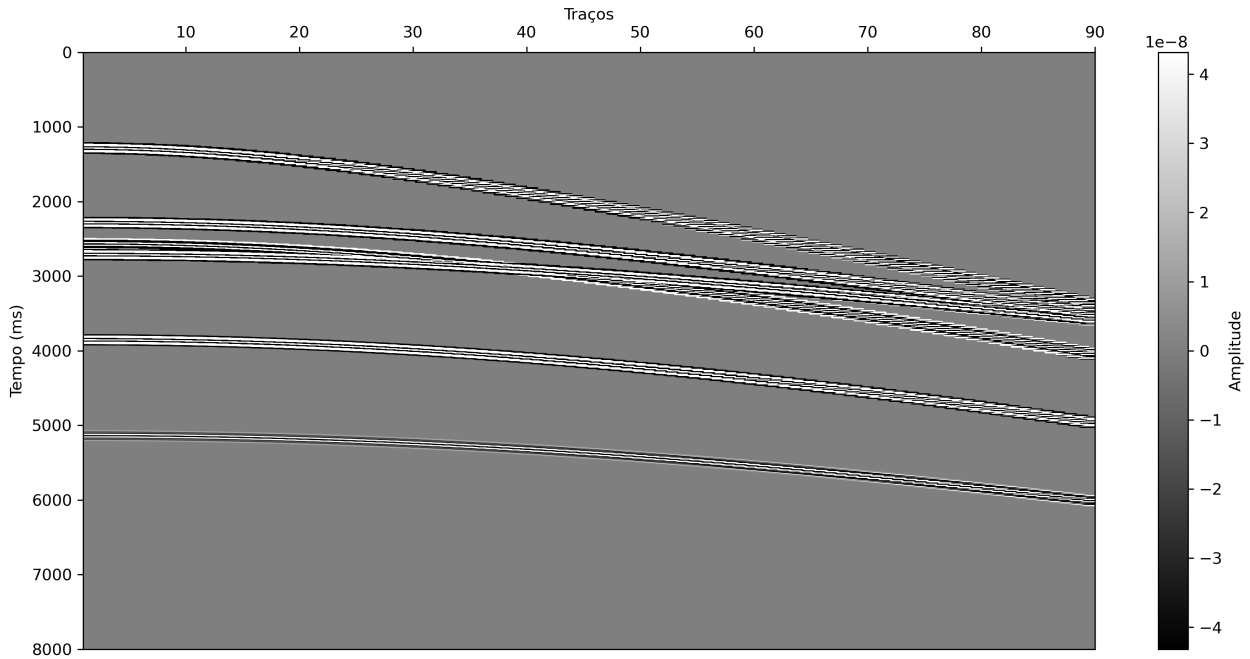


Figura 3.2: CMP central com eventos múltiplos de primeira, segunda e terceira ordens.

Na Figura 3.2, apresenta-se a seção CMP central, onde se destacam os eventos primários e múltiplos de primeira, segunda e terceira ordens. A escolha dessa seção se deve à sua representatividade, permitindo observar tanto a continuidade das reflexões quanto a estabilidade dos eventos sísmicos ao longo do tempo, servindo como referência para as análises subsequentes.

Os parâmetros utilizados para este estudo incluem um intervalo de 26 metros entre as estações

e entre os receptores, e um intervalo fixo de 52 metros entre a fonte e o receptor. Desse modo, a configuração do CMP fica com intervalos de 13 metros. Foram realizados 500 tiros e capturados dados através de 180 canais. O intervalo de amostragem de 4 ms foi escolhido para capturar as nuances das reflexões sísmicas ao longo do tempo de registro de 8 segundos, resultando em um total de 2001 amostras por traço.

Com esta configuração, a seção sísmica gerada oferece uma base sólida para análise e aplicação do procedimento denominado de pré-agrupamento amostral, possibilitando identificar reflexões, avaliar a qualidade dos sinais e verificar a presença de variabilidades dentro dos parâmetros de controle definidos.

Já o dado real utilizado foi uma linha sísmica 2D do Golfo do México que, assim como o dado sintético, foi processado, quando necessário, utilizando as rotinas do **SU**. Essa linha sísmica corresponde a aproximadamente 40 km de extensão, e foi obtida por meio do arranjo tipo end-on.

A aquisição foi realizada com intervalo entre estações, receptores e fonte-receptor de 87,5 pés. O intervalo entre pontos médios comuns (CMP) foi de 13,33 pés. Foram adquiridos 1001 tiros, cada um registrado em 180 canais. O sinal foi amostrado em intervalos de 4 ms, totalizando 1501 amostras por traço, o que corresponde a um tempo total de registro de 4s. A geometria de aquisição contemplou afastamentos mínimos de -330 pés e máximos de -15.993 pés.

Assim, o conjunto de parâmetros de aquisição adotado fornece uma base consistente para o processamento e posterior estimação das velocidades, assegurando equilíbrio entre resolução temporal, amostragem espacial e alcance de profundidade.

3.2 Estrutura híbrida do fluxo de automação sísmica

O processamento sísmico proposto neste trabalho foi estruturado de forma sequencial e automatizada, com o objetivo de reduzir a subjetividade inerente às etapas tradicionais de análise de velocidades e, ao mesmo tempo, assegurar maior consistência física nos resultados. Para alcançar esse propósito, foi elaborado um fluxo híbrido que combina organização geométrica, técnicas de agrupamento, redução de dimensionalidade e métodos de aprendizado supervisionado.

A Figura 3.3 apresenta uma visão esquemática desse fluxo de automação. Nela, observam-se as etapas que partem dos dados originais no formato **SEG-Y** e avançam até a obtenção do campo de velocidades suavizado, passando por processos intermediários como pré-

agrupamento amostral, aplicação do k-means++, extração de traços representativos via PCA, associação com a geometria real, cálculo das velocidades intervalares pela equação de Dix e ajuste final por redes neurais MLP.

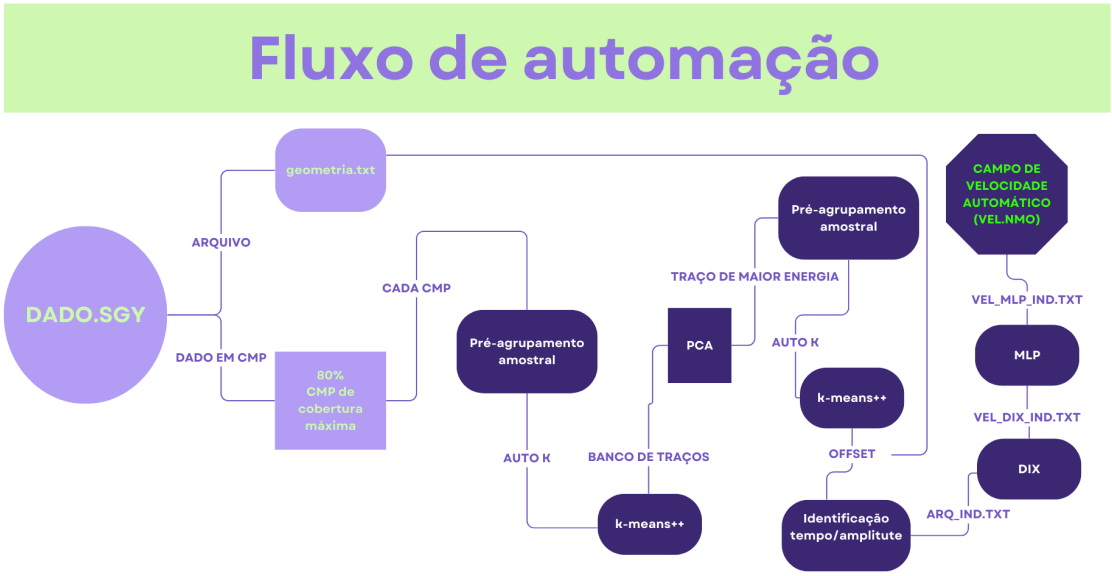


Figura 3.3: Fluxo de integração de técnicas para obtenção automática do campo de velocidade sísmica.

Esse encadeamento metodológico permite que cada módulo desempenhe uma função complementar: a clusterização auxilia na identificação de padrões estáveis, a PCA sintetiza informações relevantes, a equação de Dix fornece as estimativas iniciais de velocidade e a MLP refina os resultados, garantindo um perfil mais coerente. Assim, o fluxo delineado proporciona uma integração entre métodos estatísticos e físicos, gerando saídas automatizadas que podem ser aplicadas em diferentes cenários de análise sísmica.

Embora a relação proposta por Dix (1955) seja válida apenas sob o pressuposto de camadas horizontais, isotrópicas e com interfaces planas, condições que muitas vezes não se verificam em ambientes geologicamente complexos, neste trabalho essa limitação não compromete os resultados. Isso ocorre porque a equação de Dix (1955) não é empregada como solução final para a estimativa das velocidades, mas sim como um processo intermediário, cujo objetivo é fornecer um campo de velocidades **preliminar**. Esses valores preliminares constituem os dados de entrada para o treinamento de uma rede neural do tipo MLP (Multilayer Perceptron). Somente após o aprendizado da rede, os dados de saída passam a representar as velocidades reais, mais consistentes com a complexidade geológica do meio investigado.

Dessa forma, a integração de uma formulação clássica, já consolidada na sísmica de reflexão, com técnicas modernas de aprendizagem de máquina, possibilita a construção de modelos de velocidade mais fidedignos da subsuperfície. Tais modelos servem de base para etapas cruciais

do processamento sísmico, como migração e inversão, permitindo refinar a interpretação geológica e reduzir incertezas na identificação de alvos em profundidade, sobretudo em regiões marcadas por forte heterogeneidade lateral ou anisotropia.

Por fim, nesta tese, além da utilização de métodos de aprendizagem não supervisionada como o k-means++ e PCA, optou-se também por uma abordagem supervisionada com redes MLP, aplicada à suavização e estabilização dos valores de V_{NMO} obtidos após a aplicação da equação de Dix (1955). Como resultado, foram gerados perfis de velocidade mais realistas e compatíveis com cenários geologicamente complexos.

3.3 Construção da geometria e seleção dos CMPs de cobertura máxima

Antes da execução dos algoritmos, foi elaborado um arquivo contendo a geometria associada a cada traço dos painéis CMP. Esse arquivo serviu como referência indispensável para a associação entre tempo de trânsito, *offset* e amplitudes. Tal recurso foi utilizado nas etapas seguintes, quando se buscou relacionar os centroides derivados da análise estatística ao comportamento físico registrado nos dados originais.

Já para seleção dos CMPs de cobertura máxima, o processo foi implementado de forma automática por meio de rotinas em *Shell Script* integradas ao pacote *Seismic Unix* (**SU**). Essa etapa é fundamental para assegurar que apenas os painéis mais representativos da aquisição sísmica sejam analisados, reduzindo a influência de regiões com baixa amostragem e garantindo maior confiabilidade nos cálculos de velocidades.

O procedimento inicia-se com a ordenação dos traços por CMP e *offset*, permitindo a visualização da distribuição de cobertura ao longo do perfil. Em seguida, o script aplica comandos que identificam o valor mínimo e máximo de CMP presentes no conjunto de dados, corrigindo inconsistências como registros nulos ou negativos. A partir desses limites, é calculada uma faixa percentual intermediária, no caso, 20% do intervalo total que delimita os CMPs de maior cobertura, evitando tanto bordas com dados escassos quanto redundâncias de painéis pouco representativos.

Uma vez estabelecido esse intervalo, o script verifica a existência real dos CMPs selecionados e ajusta automaticamente os valores caso algum índice não esteja presente no conjunto. Para cada CMP dentro do intervalo definido, são criados arquivos independentes contendo os traços correspondentes. Essa divisão garante que os painéis de cobertura máxima sejam isolados de forma sistemática, eliminando a necessidade de seleção manual e assegurando a

padronização do processo.

Além disso, o procedimento inclui uma etapa de validação, em que arquivos vazios ou inconsistentes são descartados, assegurando que apenas painéis com dados válidos sejam mantidos para análise. Por fim, os CMPs selecionados são organizados em diretórios específicos e convertidos para formatos **SU** e **SEG-Y**, o que viabiliza sua utilização nas etapas subsequentes do fluxo de automação.

Dessa maneira, a seleção dos CMPs de cobertura máxima é conduzida de forma criteriosa e automatizada, garantindo que os algoritmos de clusterização e redução de dimensionalidade operem sobre um conjunto de dados otimizado, estatisticamente representativo e fisicamente consistente.

3.4 Pré-agrupamento amostral

Em estudos geofísicos, a análise cuidadosa da variabilidade dos dados amostrados é essencial para interpretar corretamente as condições das estruturas em subsuperfície, especialmente em aplicações como a aquisição de dados sísmicos. No entanto, esses dados frequentemente contêm interferências, ou ruídos, que podem ser causados por fatores ambientais, equipamentos ou outras fontes. O ruído coerente, por exemplo, caracteriza-se por um padrão espacial ou temporal específico, sendo gerado por fenômenos físicos e geológicos, como ondas superficiais, no caso de aquisição terrestre, e múltiplas no caso de aquisição marinha, que muitas vezes mascaram os sinais de interesse (Ebadi, 2017; Schimmel e Paulssen, 1997). Por outro lado, o ruído incoerente é aleatório e sem padrão definido, originado de fontes ambientais, como tráfego, vento e ondas do mar, ou interferências eletrônicas, podendo se sobrepor de maneira imprevisível ao sinal desejado (Chen e Simaan, 1991; Hanna e Simaan, 1987).

Esses ruídos geram *outlier* e dificultam o processamento dos dados e a interpretação das feições sísmicas, tornando necessário o uso de métodos que ajudem a distinguir as variações no processo de aquisição desses dados. Para lidar com essas interferências, são utilizados métodos de processamento avançados, como filtros espectrais e técnicas de atenuação, que ajudam a melhorar a razão sinal-ruído e permitem uma melhor análise das estruturas geológicas.

A aplicação de técnicas adequadas para mitigar esses diferentes tipos de ruído é fundamental para garantir maior precisão dos dados, reforçando a importância de metodologias que priorizem a qualidade e a confiabilidade dos sinais amostrados. Desse modo, para atenuar efeitos como esses e fornecer informações iniciais de agrupamento, foram realizados procedimentos estatísticos que lidam adequadamente com volumes de dados dessa complexidade.

A distribuição em classes de dados amostrais é uma técnica recomendada para conjuntos de dados com uma quantidade substancial de valores, permitindo uma estrutura mais organizada e simplificada para a análise. Segundo Freedman et al. (2007), a classificação em intervalos ajuda a reduzir a complexidade dos cálculos e facilita a identificação de padrões gerais nos dados. Moore et al. (2012) também ressaltam que essa abordagem é essencial ao lidar com grandes volumes de dados, pois auxilia na atenuação dos efeitos de flutuações extremas e valores discrepantes, resultando em uma análise estatística mais robusta e confiável. Johnson e Wichern (2007) afirmam que a organização dos dados em classes é particularmente benéfica para grandes volumes de dados, onde a estrutura hierárquica das classes contribui para a interpretação otimizada dos resultados. Assim, ao organizar os dados sísmicos correspondentes a cada painel de CMP de maior cobertura em classes, é possível identificar a quantidade de subgrupos em que os dados podem ser dispostos, mas contabilizar como classes relevantes apenas aquelas que possuem valores associados à sua frequência absoluta diferentes de 0 e 1, por estas serem consideradas de baixa relevância estatística. Essa etapa inicial é necessária para identificar o número de clusters a ser utilizado na técnica de agrupamento k-means++ que compõe parte da metodologia proposta.

A análise estatística dos valores de amplitudes amostrados começa com sua organização em classes, um passo fundamental para lidar com grandes volumes de informações frequentemente contaminadas por ruídos. Esse processo simplifica os cálculos, facilita a identificação de padrões e auxilia na atenuação da dispersão, garantindo maior robustez à análise. Inicialmente, a amplitude amostral dos dados é calculada para determinar os limites gerais do conjunto. Com base nisso, a Regra de Sturges (1926) é utilizada para definir o número ideal de classes, equilibrando a precisão das amostras. A amplitude de cada classe é então calculada, e os dados são distribuídos em intervalos correspondentes, permitindo o cálculo das frequências absolutas. Nesse processo, o somatório do número de classes relevantes será o valor de k a ser utilizado dentro da técnica k-means++, o que configura uma nova forma de determinar automaticamente o número de clusters k. Essa abordagem assegura uma estrutura organizada e confiável, indispensável para minimizar os impactos dos ruídos e maximizar a qualidade das inferências.

Primeiramente, calcula-se a amplitude amostral (A_a) dos dados, conforme a equação (3.1), que representa a diferença entre o limite superior L_{sup} e o limite inferior l_{inf} da amostra:

$$A_a = L_{sup} - l_{inf} \tag{3.1}$$

Com (A_a) calculada, utiliza-se a regra de Sturges (1926) para determinar o número de classes (K) em função do número total de valores n da variável analisada, conforme a equação (3.2):

$$K = 1 + 3.32 \log 10(n) \tag{3.2}$$

Em seguida, calcula-se a amplitude de classe (h) por meio da equação (3.3), que representa a razão entre (A_a) e (K):

$$h = \frac{A_a}{K} \tag{3.3}$$

A próxima etapa envolve o cálculo da frequência absoluta f_i de cada classe, representando o número de amostras contidas em cada intervalo. Nesse estágio, são identificadas, mas não contabilizadas, as classes com frequências nulas ou unitárias, por representarem valores irrelevantes para a análise dos dados. Somente as classes que apresentam uma quantidade significativa de dados são contabilizadas, fornecendo uma base confiável para a determinação automática do número de clusters que será utilizado na aplicação do k-means++. Vale ressaltar que, por se tratar de uma análise monocanal, n corresponde à quantidade de valores de amplitude existentes em cada traço sísmico analisado. Esse procedimento é denominado pré-agrupamento amostral e está representado na Tabela 3.1. Para garantir a efetividade da técnica, foi realizada a análise de variabilidade e controle estatístico desses dados, conforme descrito no Apêndice A.

	Classes	f_i	f_{ri}(%)	F_i	F_{ri} (%)	x_i	x_i^2	$f_i \cdot x_i$	$f_i \cdot x_i^2$	y_i	y_i^2	$f_i \cdot y_i$	$f_i \cdot y_i^2$
0	[-681.4014 ⊢ -563.0735]	1	0.09	1	0.09	-622.237	387179.405	-622.237	387179.405	-5	25	-5	25
1	[-563.0735 ⊢ -444.7456]	0	0.00	1	0.09	-503.910	253924.799	-0.000	0.000	-4	16	-0	0
2	[-444.7456 ⊢ -326.4177]	2	0.18	3	0.27	-385.582	148673.179	-771.163	297346.357	-3	9	-6	18
3	[-326.4177 ⊢ -208.0898]	15	1.33	18	1.60	-267.254	71424.544	-4008.806	1071368.161	-2	4	-30	60
4	[-208.0898 ⊢ -89.7619]	64	5.68	82	7.28	-148.926	22178.895	-9531.251	1419449.279	-1	1	-64	64
5	[-89.7619 ⊢ 28.5661]	831	73.74	913	81.01	-30.598	936.231	-25426.855	778008.364	0	0	0	0
6	[28.5661 ⊢ 146.8940]	179	15.88	1092	96.89	87.730	7696.554	15703.671	1377683.083	1	1	179	179
7	[146.8940 ⊢ 265.2219]	18	1.60	1110	98.49	206.058	42459.861	3709.042	764277.501	2	4	36	72
8	[265.2219 ⊢ 383.5498]	12	1.06	1122	99.56	324.386	105226.154	3892.630	1262713.852	3	9	36	108
9	[383.5498 ⊢ 501.8777]	2	0.18	1124	99.73	442.714	195995.433	885.427	391990.866	4	16	8	32
10	[501.8777 ⊢ 620.2056]	3	0.27	1127	100.00	561.042	314767.697	1683.125	944303.092	5	25	15	75
11	Σ	1127	100.00	1127	100.00	-336.577	1550462.752	-14486.418	8694319.959	0	110	169	633

Tabela 3.1: Distribuição em classes dos valores de amplitudes de um traço sísmico.

3.5 Agrupamento não supervisionado com o algoritmo k-means++

O procedimento inicial consistiu na aplicação de um pré-agrupamento amostral, etapa essencial para garantir maior estabilidade ao algoritmo k-means++, que, conforme discutido

por James et al. (2013), tende a apresentar limitações ligadas à escolha inicial de centroides. O aprendizado não supervisionado, por sua natureza exploratória, envolve certo grau de subjetividade, o que pode gerar inconsistências na análise. Como apontado por James et al. (2013), os métodos de clusterização buscam identificar subgrupos dentro de um conjunto de dados, mas o desempenho do k-means++ depende diretamente da qualidade da preparação. Nesse sentido, o pré-agrupamento forneceu um arranjo mais homogêneo, reduzindo desvios e aumentando a eficácia da técnica.

3.6 Associação entre centroides e amplitudes reais

Com os centroides obtidos em cada CMP, iniciou-se a etapa de associação às amostras originais de cada traço. Essa relação possibilitou comparar valores médios de agrupamentos com amplitudes reais, gerando uma matriz de distâncias que indicava a proximidade de cada ponto em relação a cada centro. A métrica utilizada foi a distância euclidiana, descrita pela equação 3.4.

$$d_{ik} = \sqrt{\sum_{j=1}^{n} (x_{ij} - \mu_{kj})^2} \tag{3.4}$$

em que x_{ij} corresponde às coordenadas do ponto x_i e μ_{kj} às do centróide μ_k, considerando n dimensões. Os índices de mínima distância foram obtidos pela expressão 3.5, de forma a selecionar o traço que melhor representava cada centróide.

$$i_{\min} = \arg\min_{i}(d_{ik}) \tag{3.5}$$

Assim, tornou-se possível vincular as informações estatísticas à geometria real, assegurando que os traços de maior representatividade fossem identificados e utilizados nos cálculos posteriores, pois essa ação ajuda a formar um banco de traços mais coerente para cada painel CMP. Na Figura 3.4, temos uma ilustração em que μ representa os centróides calculados nos traços, e a busca e identificação por valores iguais ou similares em todo o conjunto de traços do painel CMP. Já a Figura 3.5 mostra os traços identificados e selecionados do painel CMP 70280 do dado do Golfo do México. Esse processo foi repetido para todos os CMPs de cobertura máxima, consolidando a ligação entre os resultados de agrupamento e os registros originais.

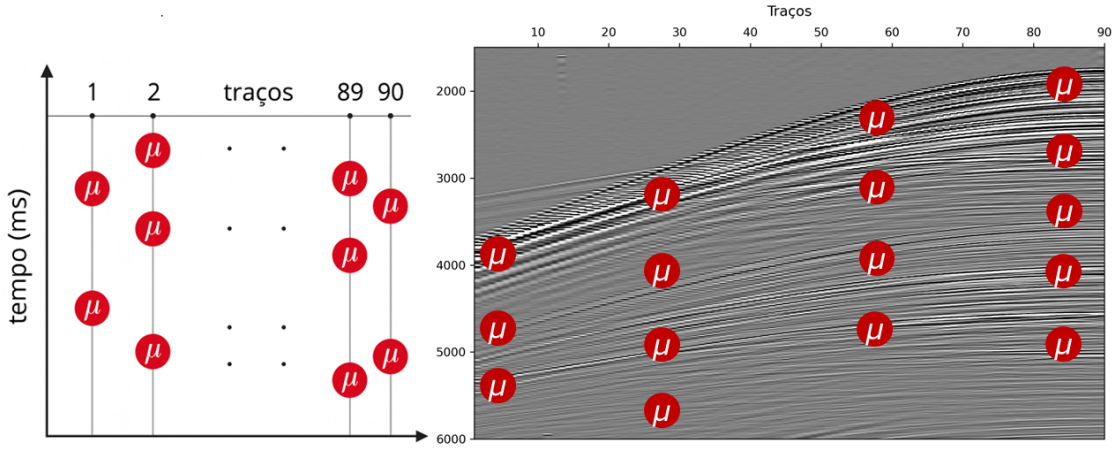


Figura 3.4: Cálculo de centróides, busca e identificação nos traços de cada painel CMP para formação do banco de traços de maior representatividade.

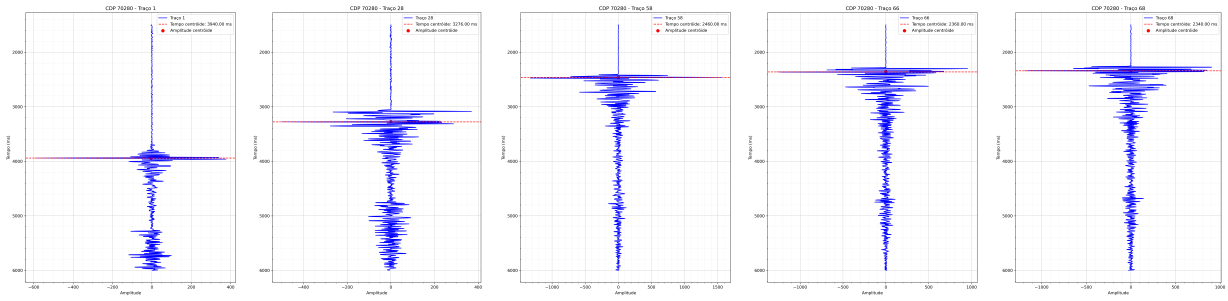


Figura 3.5: Registro de traços selecionados com marcações de tempo de amplitude. Traços identificados: 1, 28, 58, 66 e 68.

3.7 Redução de dimensionalidade e síntese de traços via PCA

A associação entre k-means++ e PCA é amplamente reconhecida como complementar, sendo útil para lidar com grandes volumes de informação. Enquanto o k-means++ organiza padrões e estruturas em grupos bem definidos, a PCA reduz a dimensionalidade e reorganiza os dados de modo a ressaltar apenas as variáveis mais significativas, simplificando a análise de sinais complexos.

O k-means++ aprimora o tradicional k-means ao implementar um mecanismo probabilístico para inicialização dos centroides, aumentando a eficiência da distribuição inicial e diminuindo a quantidade de iterações até a convergência (Arthur e Vassilvitskii, 2007). Por sua vez, a PCA transforma variáveis correlacionadas em componentes principais não correlacionados, hierarquizados segundo a variabilidade que explicam. Esse processo, segundo Vasconcelos (2021), elimina redundâncias e reestrutura dados de alta dimensionalidade, tendo como base

a decomposição da matriz de covariância em autovalores e autovetores (Anton e Rorres, 2004). Aplicações típicas incluem compressão, reconhecimento de padrões e processamento de imagens (Gonzalez e Woods, 2000).

Desse modo, a etapa seguinte envolveu a aplicação da PCA para sintetizar os traços mais significativos em termos de energia, seguida do pré-agrupamento amostral e do k-means++ novamente, agora para identificar os tempos associados aos valores de amplitude do traço de maior energia. A combinação integrada dessas técnicas no fluxo de trabalho mostrou-se robusta, pois o k-means++ identifica grupos consistentes e a PCA sintetiza a estrutura essencial dos dados, conforme pode ser visualizado na Figura 3.6. Esse encadeamento fortalece a interpretação sísmica, reduz a instabilidade inerente e torna o processamento adequado para análises que envolvem grandes volumes de informações.

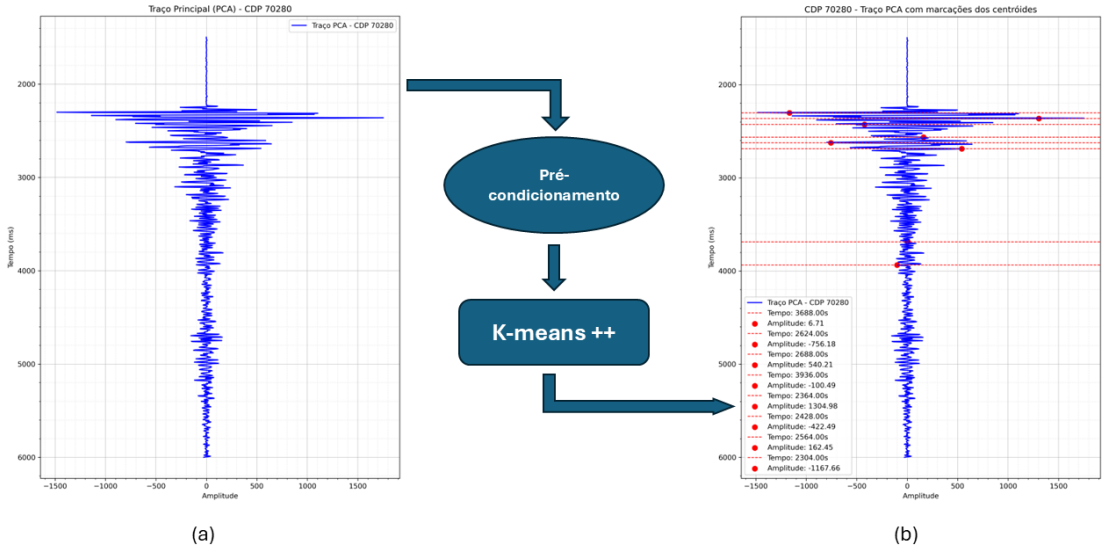


Figura 3.6: PCA aplicada aos traços identificados. (a) Traço de maior energia. (b) Traço de maior energia com registro de tempo de amplitude.

A organização gráfica desse resultado posicionou o tempo no eixo vertical e as amplitudes no horizontal, com a inversão do eixo y para manter a escala sísmica convencional. Esse traço sintetizado destacou os eventos de maior relevância, preparando a base para um novo ciclo de agrupamento automático que alimentou o k-means++. Combinado ao arquivo de geometria, esse processo identificou os traços que mantiveram energia estável, tornando-os aptos para o cálculo de velocidades, pois cada valor de amplitude identificado carrega a geometria do seu traço correspondente.

3.8 Obtenção de velocidades via equação de Dix

Após a etapa de preparação, os dados sísmicos foram utilizados no cálculo das velocidades de propagação das camadas subsuperficiais por meio da aplicação da equação de Dix. A determinação das velocidades intervalares é fundamentada na análise dos tempos de trânsito obtidos para cada refletor e das velocidades quadráticas médias (*Root Mean Square* – V_{RMS}), estimadas a partir da relação hiperbólica de NMO.

Nos dados reais, a velocidade associada ao primeiro intervalo pode ser obtida diretamente a partir dos tempos médios de trânsito e deslocamentos correspondentes, caracterizando a propagação até a primeira interface refletora. Para camadas mais profundas, entretanto, torna-se necessário considerar a contribuição acumulada das camadas superiores. Nesse contexto, a equação de Dix fornece um meio robusto de extrair as velocidades intervalares (V_{int}) a partir das velocidades RMS sucessivas e dos tempos associados a cada horizonte refletor. Esse procedimento resultou nas velocidades intervalares V_{int}, posteriormente refinadas por meio de aprendizado supervisionado.

3.9 Ajuste de perfis com redes neurais MLP

Para atenuar ruídos e assegurar consistência física, as velocidades derivadas foram ajustadas utilizando uma rede neural MLP. Perfis de velocidade derivados apenas da equação de Dix podem conter segmentos não monotônicos e instabilidades em zonas complexas, como aquelas com presença de sal, que introduz efeitos de *pull-up*, zonas de sombra e múltiplos trajetos de onda. A rede MLP, estruturada com quatro camadas ocultas de 1024, 512, 256 e 128 neurônios, ativação logística e otimização pelo algoritmo Adam, foi capaz de aprender relações não lineares entre tempo *zero-offset* e velocidade V_{NMO}. Essa abordagem suavizou o perfil e impôs crescimento monotônico, preservando as tendências globais e reduzindo dispersões locais. A utilização da MLP mostrou-se vantajosa por dispensar parametrizações rígidas, diferindo de ajustes polinomiais tradicionais. Como resultado, foram obtidas curvas de velocidade mais fiéis ao comportamento físico da propagação sísmica, fornecendo suporte consistente para etapas de migração e interpretação estrutural.

4

Resultados e discussões

4.1 Resultados em dado sintético com eventos múltiplos de fundo do mar

Para a etapa de validação inicial da metodologia proposta, foi construído um modelo sintético representativo de um ambiente marinho com três camadas planas, no qual foram incluídos eventos múltiplos de fundo do mar de até terceira ordem. A introdução desses múltiplos teve como propósito simular um cenário realista de reverberações na lâmina d'água, uma das principais fontes de ruído coerente em dados *offshore*. Tais eventos, embora não correspondam a interfaces geológicas verdadeiras, reproduzem trajetórias recorrentes de energia que tendem a mascarar as reflexões primárias e, portanto, representam um teste rigoroso à robustez do fluxo automatizado desenvolvido.

A partir do desenvolvimento e validação do modelo sintético, o processamento seguiu a sequência integrada descrita na metodologia, iniciando-se pela construção da geometria completa dos painéis CMP e pela organização dos traços que compõem o conjunto de dados. O pré-agrupamento amostral foi aplicado para estimar o número ótimo de clusters, seguido da execução do algoritmo k-means++, responsável por identificar agrupamentos coerentes de amplitudes e seus centróides representativos. Em cada painel CMP, os traços mais próximos dos centróides foram extraídos e utilizados como amostras de maior relevância.

Na sequência, os conjuntos de traços selecionados foram submetidos à Análise de Componentes Principais (PCA), que possibilitou a síntese de traços de maior energia e a eliminação de redundâncias. Sobre esses dados reduzidos, os procedimentos de pré-agrupamento e clusterização foram novamente aplicados, gerando centróides refinados e valores medianos de tempo

e amplitude para as zonas de maior coerência sísmica. Essa etapa garantiu a preservação das relações temporais e espaciais entre os eventos refletidos.

Com a associação entre os centróides e as amplitudes correspondentes geometricamente estabelecida, os tempos médios obtidos em cada CMP foram utilizados como base para o cálculo das velocidades preliminares por meio da equação de Dix (1955). Os resultados foram integrados em um banco de dados que representa o campo de velocidades preliminar, posteriormente empregado como conjunto de treinamento para a rede neural MLP. Esse modelo supervisionado foi então responsável por refinar os perfis e impor o comportamento fisicamente monotônico das velocidades em profundidade.

Nas Figuras 4.1 e 4.2 tem-se a comparação entre o modelo empilhado utilizando o campo de velocidade obtido de forma automática e o modelo empilhado utilizando o campo de velocidade obtido via *picking*. Já na Figura 4.3 tem-se a diferença entre os dados sintéticos empilhados. Os resultados obtidos por meio da aplicação do procedimento desenvolvido demonstraram alto desempenho na obtenção automática do campo de velocidade do dado sintético desenvolvido.

A avaliação comparativa entre a abordagem tradicional (manual) e a estratégia automática proposta revela diferenças significativas na seleção dos traços sísmicos mais representativos. A Figura 4.4, que apresenta a comparação da amplitude média dos traços selecionados, evidencia que o método automático preserva maior conteúdo energético em relação à abordagem manual, demonstrando que o valor de amplitude média obtido pelo ajuste automático representa um incremento de aproximadamente 80,13% em relação ao valor obtido pelo ajuste manual. Essa característica é particularmente importante, pois está diretamente associada à preservação dos eventos primários e à caracterização de componentes ruidosos, como as múltiplas do fundo do mar. A maior amplitude média dos traços selecionados automaticamente reforça a capacidade do método em identificar zonas com maior coerência na impedância acústica e relevância geológica.

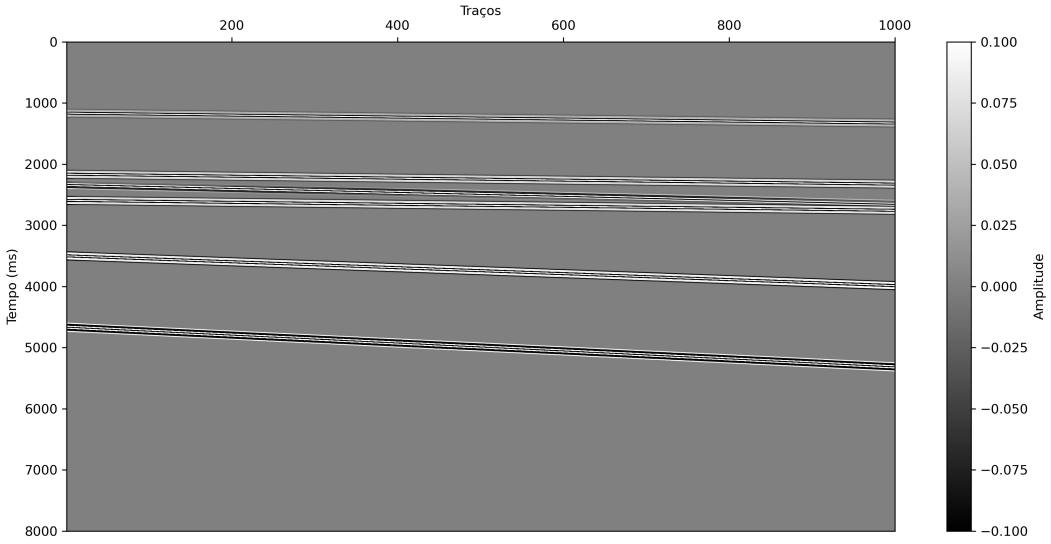


Figura 4.1: Dado sintético empilhado com campo de velocidade obtido de forma automática.

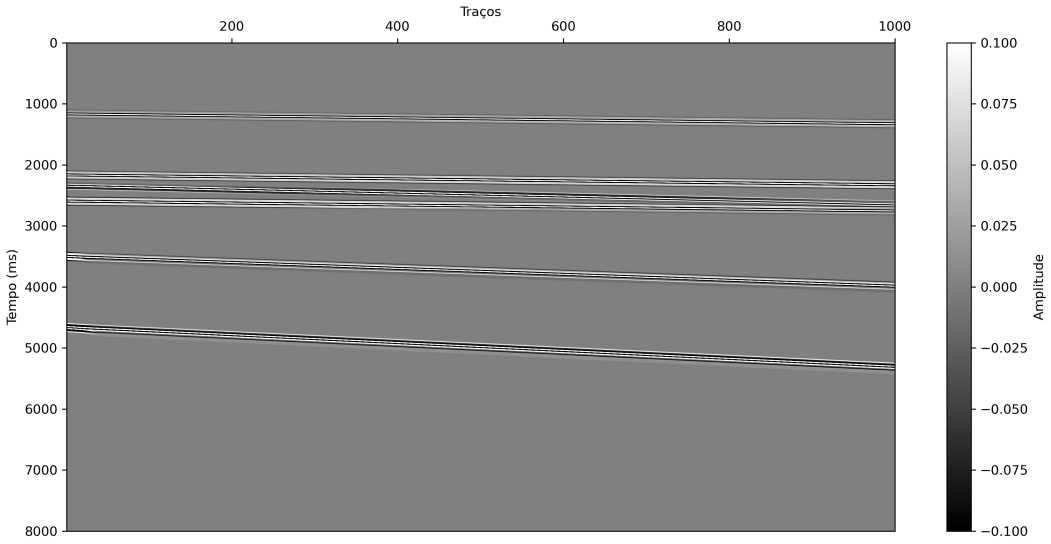


Figura 4.2: Dado sintético empilhado com campo de velocidade obtido via *picking*.

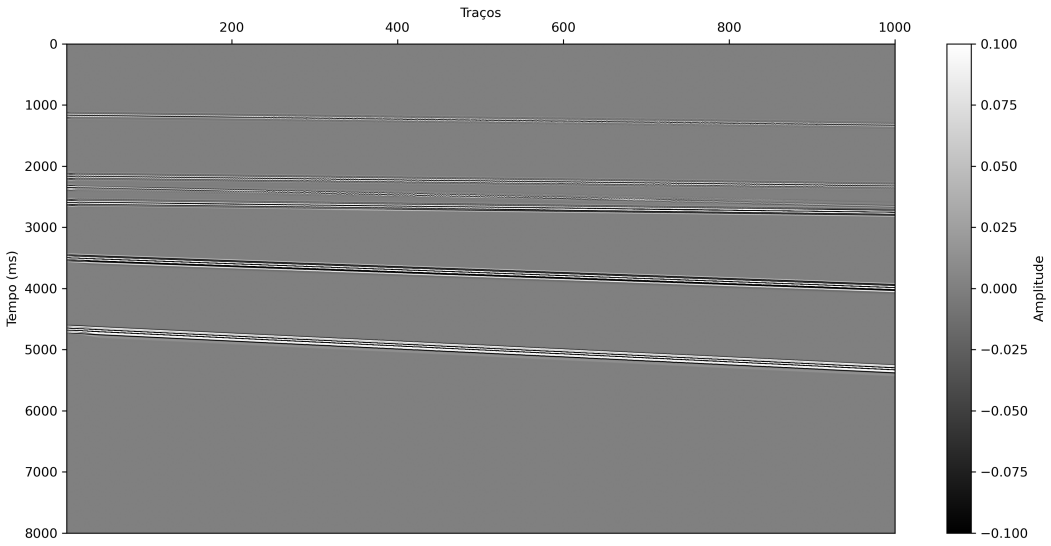


Figura 4.3: Diferença entre os dados sintéticos empilhados com campo de velocidade automático e com campo de velocidade obtido via *picking*.

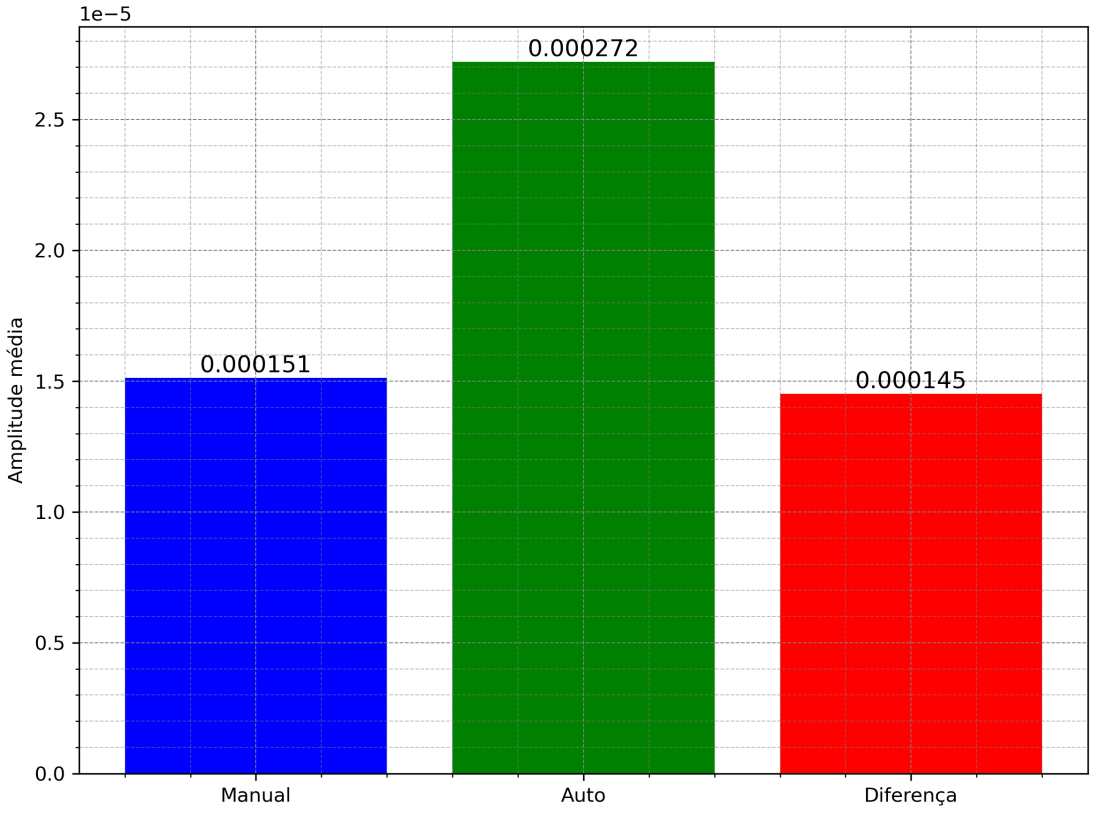


Figura 4.4: Comparação da amplitude média entre os dados sintéticos empilhados.

A Figura 4.5, por sua vez, mostra a diferença de amplitude por traço entre os métodos. Observa-se uma tendência de flutuação constante até o traço 1000. Essa descontinuidade pode estar relacionada à atuação do algoritmo de agrupamento, que identifica agrupamentos

distintos de traços com características acústicas diferentes. O comportamento oscilatório observado evidencia que os traços selecionados automaticamente apresentam maior variabilidade, o que pode indicar uma cobertura mais ampla das informações geológicas relevantes.

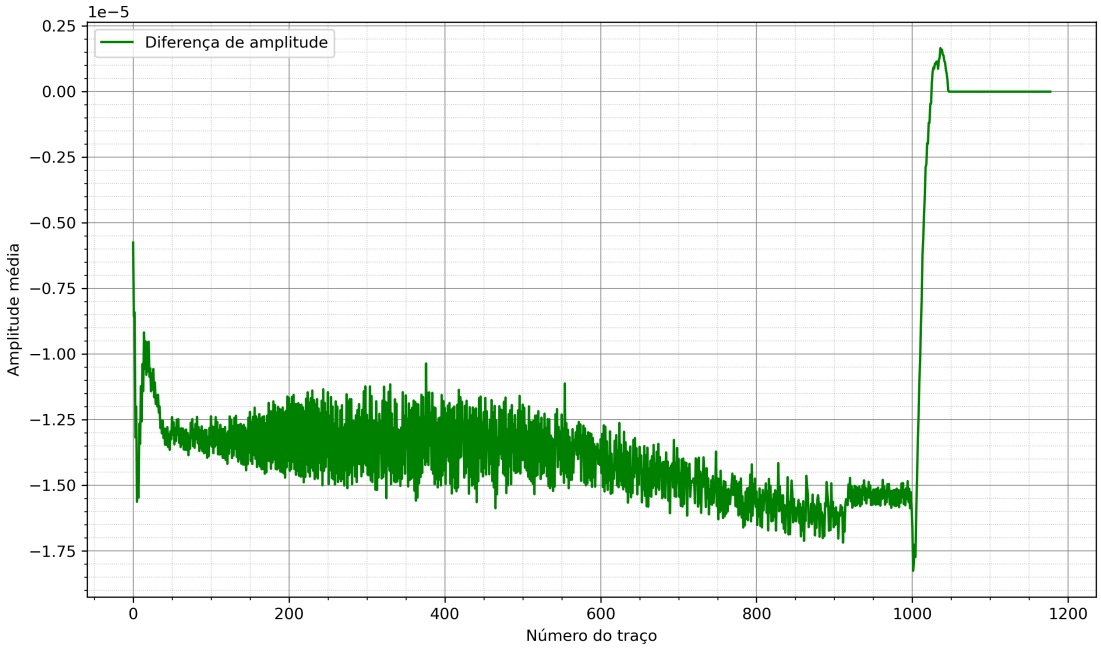


Figura 4.5: Diferença de amplitude por traço.

Na Figura 4.6, é analisada a diferença de energia acumulada entre os dois métodos. Nota-se uma predominância de valores negativos, indicando que a energia total dos traços escolhidos pelo modelo automático é, na maioria dos casos, superior à do modelo manual. Esse resultado confirma que o novo método consegue preservar melhor a coerência e a integralidade do sinal ao longo da seção, o que é essencial para ambientes com forte presença de ruídos, como os dados sintéticos com eventos múltiplos de alta ordem utilizados neste estudo. Os picos locais de energia apontam para regiões onde o modelo manual pode ter perdido traços relevantes ou incluído ruídos, enquanto a técnica automática demonstrou maior seletividade e precisão.

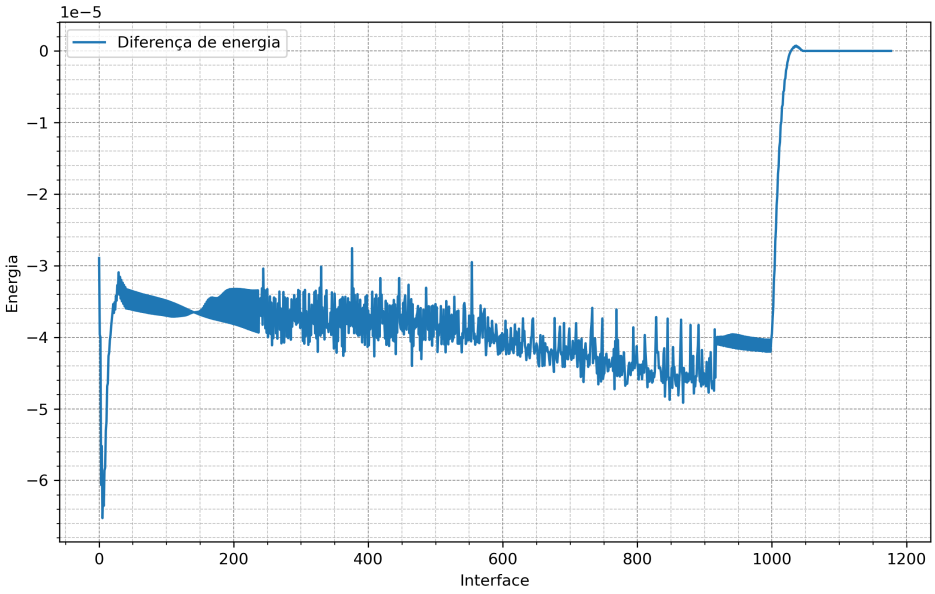


Figura 4.6: Diferença de energia acumulada.

Por fim, a Figura 4.7 traz uma análise espectral comparativa entre os dois métodos, revelando que o espectro de frequências dos traços do modelo automático cobre uma faixa mais ampla, especialmente entre 30 e 45 Hz, e com maior amplitude. Isso indica uma maior riqueza de detalhes e melhor resolução vertical, características essenciais para a interpretação de interfaces complexas. A curva de diferença no espectro reforça que, embora existam zonas de sobreposição, há um ganho consistente de frequência nos traços selecionados automaticamente, o que pode contribuir para uma melhor definição de refletores geológicos mesmo em regiões mais profundas.

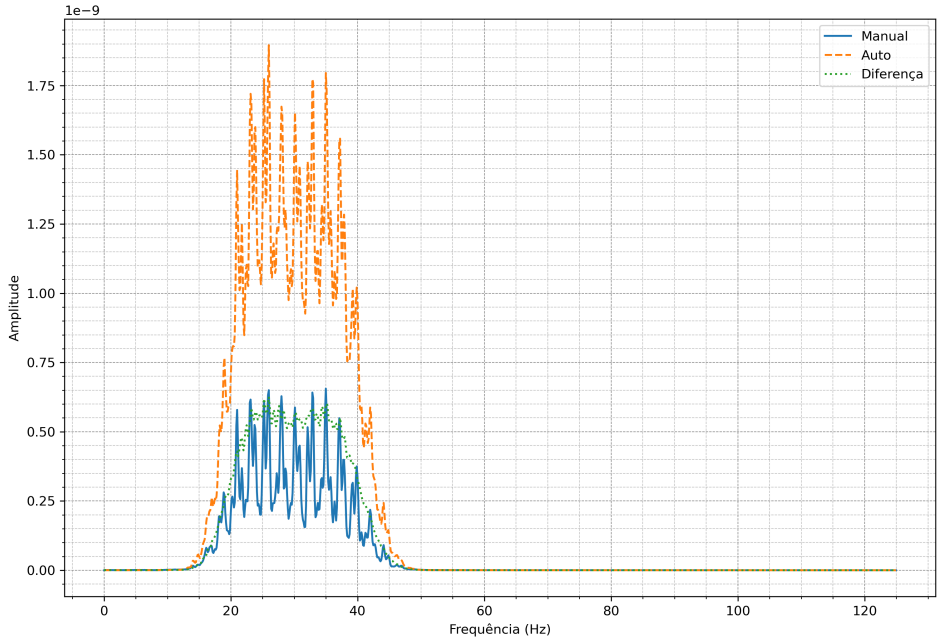


Figura 4.7: Análise espectral comparativa.

Dessa forma, os resultados obtidos a partir dos dados sintéticos, mostram que o método automático proposto não apenas mantém a qualidade dos sinais sísmicos, como também aumenta significativamente sua riqueza informacional, energia e resolução espectral. Esse desempenho superior se destaca diante da complexidade dos dados utilizados, marcados por eventos múltiplos do fundo do mar de até terceira ordem, que podem mascarar eventos primários se não forem adequadamente atenuados.

4.2 Resultados dados reais: Golfo do México

Dada a importância de se obter um campo de velocidade capaz de horizontalizar de forma consistente os refletores sísmicos, torna-se evidente a relevância do uso de técnicas automáticas em relação ao método manual baseado em *picking*. A Figura 4.8 do painel CMP(17612) do dado do Golfo do México mostra que, inicialmente, o painel CMP (à esquerda) apresenta os eventos refletidos com curvaturas hiperbólicas, evidenciando a ausência de correção de movimento normal (NMO). No centro, observa-se o painel de semblance, que indica as regiões de maior coerência em função da velocidade e do tempo duplo de trânsito; entretanto, esse painel não apresenta concentrações de energia bem definidas, o que dificulta a análise do intérprete e a escolha precisa da curva de velocidade. Por fim, o painel CMP corrigido (à direita), obtido a partir do campo de velocidades extraído manualmente via picking, mostra

que os refletores superficiais foram parcialmente horizontalizados, mas a partir do tempo de 3.5 (ms) observa-se que a correção NMO não foi satisfatória, resultando em eventos mal alinhados e com perda de coerência lateral.

Enquanto a abordagem manual conduz a resultados dependentes da subjetividade do intérprete, o procedimento automático, guiado pela análise estatística, demonstra maior estabilidade. Observa-se, na seção corrigida por NMO (Figura 4.9), que o campo de velocidade obtido de forma automática produz refletores bem alinhados e com maior continuidade lateral, além de minimizar o mascaramento causado por múltiplas, ruídos residuais e zonas de baixa velocidades. Assim, a técnica automática não apenas reduz o tempo de processamento e a intervenção humana, mas também assegura maior confiabilidade dos resultados.

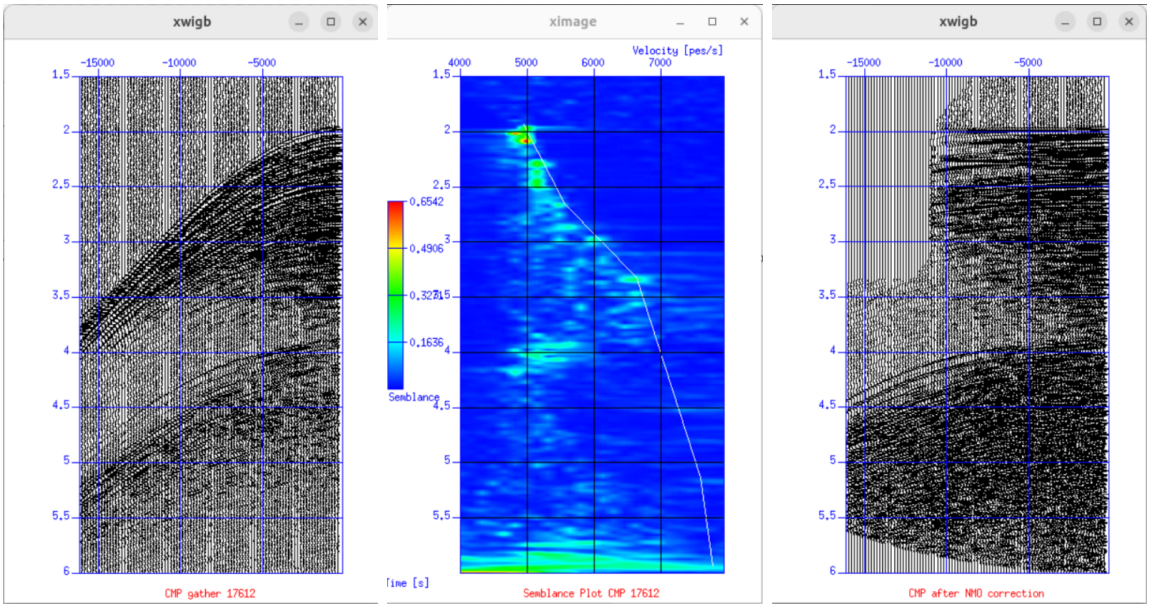


Figura 4.8: Análise do painel CMP(17612) do dado do Golfo do México na obtenção do campo de velocidade via *picking*.

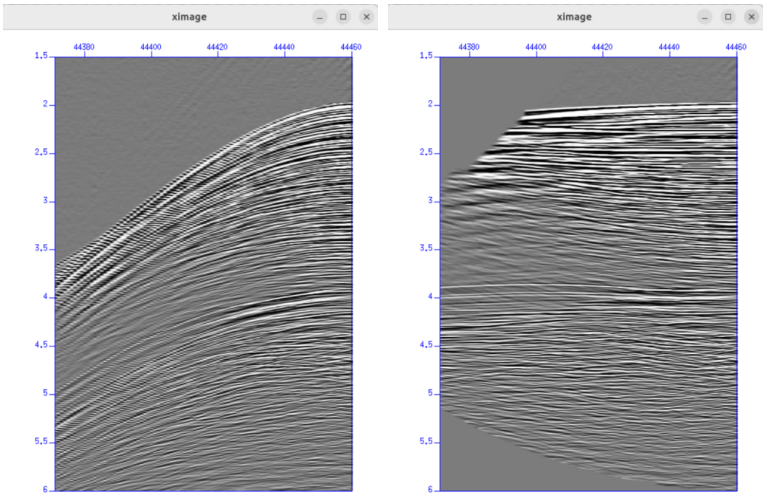


Figura 4.9: Análise do painel CMP(17612) do dado do Golfo do México na obtenção do campo de velocidade da forma automática.

Agora, seguindo o mesmo fluxo de tratamento do dado para obtenção do campo de velocidade de forma automática, aplicado ao dado sintético, tem-se no Anexo II, as imagens dos traços selecionados para os casos de CMPs de cobertura mínima, média e máxima do dado real, além das imagens dos dados após a aplicação da PCA aos conjuntos de traços selecionados dos referidos CMPs.

Nas Figuras 4.10 e 4.11 tem-se a comparação entre o modelo empilhado utilizando o campo de velocidade obtido de forma automática e o modelo empilhado utilizando o campo de velocidade obtido via *picking*, agora do dado real. Já na Figura 4.12 tem-se a diferença entre os dados.

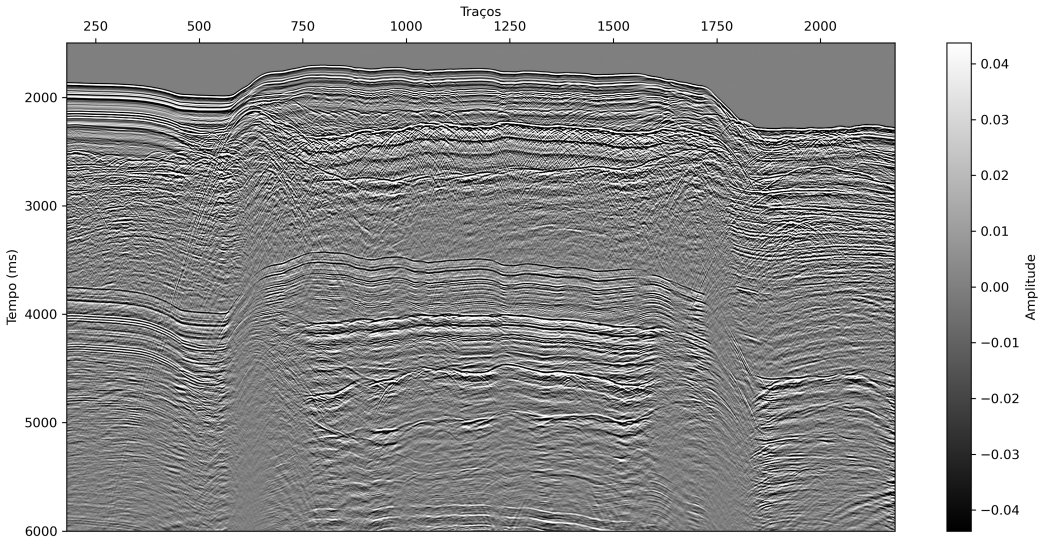


Figura 4.10: Seção sísmica do Golfo do México empilhada com campo de velocidade obtido de forma automática.

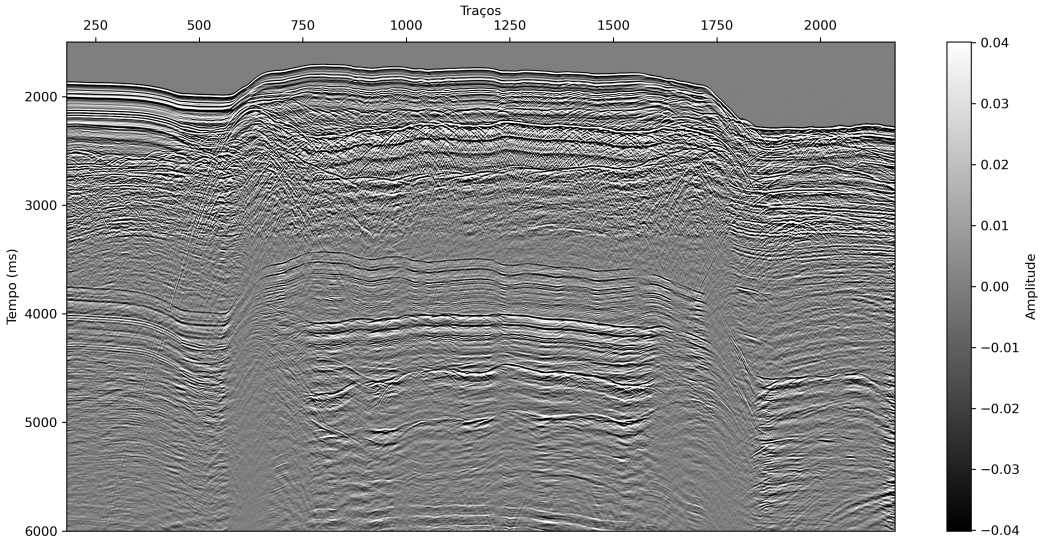


Figura 4.11: Seção sísmica do Golfo do México empilhada com campo de velocidade obtido via *picking*.

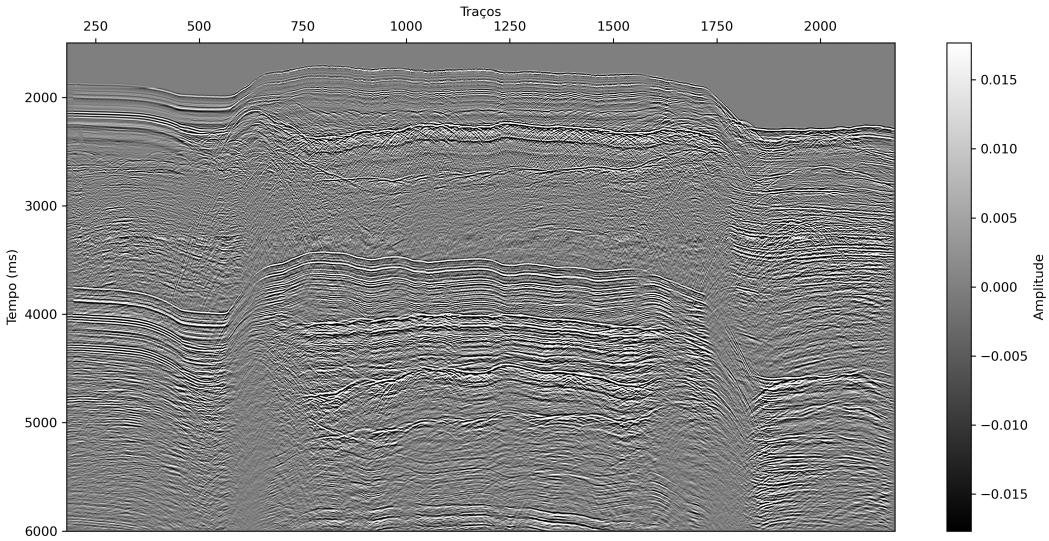


Figura 4.12: Diferença entre os dados empilhados com campo de velocidade automático e com campo de velocidade obtido via *picking*, para os dados sísmicos do Golfo do México.

A comparação entre os métodos de obtenção de velocidades no dado do Golfo do México revela diferenças significativas, tanto em termos de amplitude, quanto de energia e conteúdo espectral. A Figura 4.13 mostra que a amplitude média dos traços obtidos pelo método automatizado (0.182702) foi ligeiramente superior à do método manual (0.168381), o que indica uma resposta sísmica mais intensa no resultado automático. A diferença média de amplitude entre os dois métodos (0.073889), embora relativamente pequena e correspondendo a um incremento percentual de aproximadamente 8,5% na amplitude média, sugere que o algoritmo automático teve uma leve tendência a realçar a energia dos sinais.

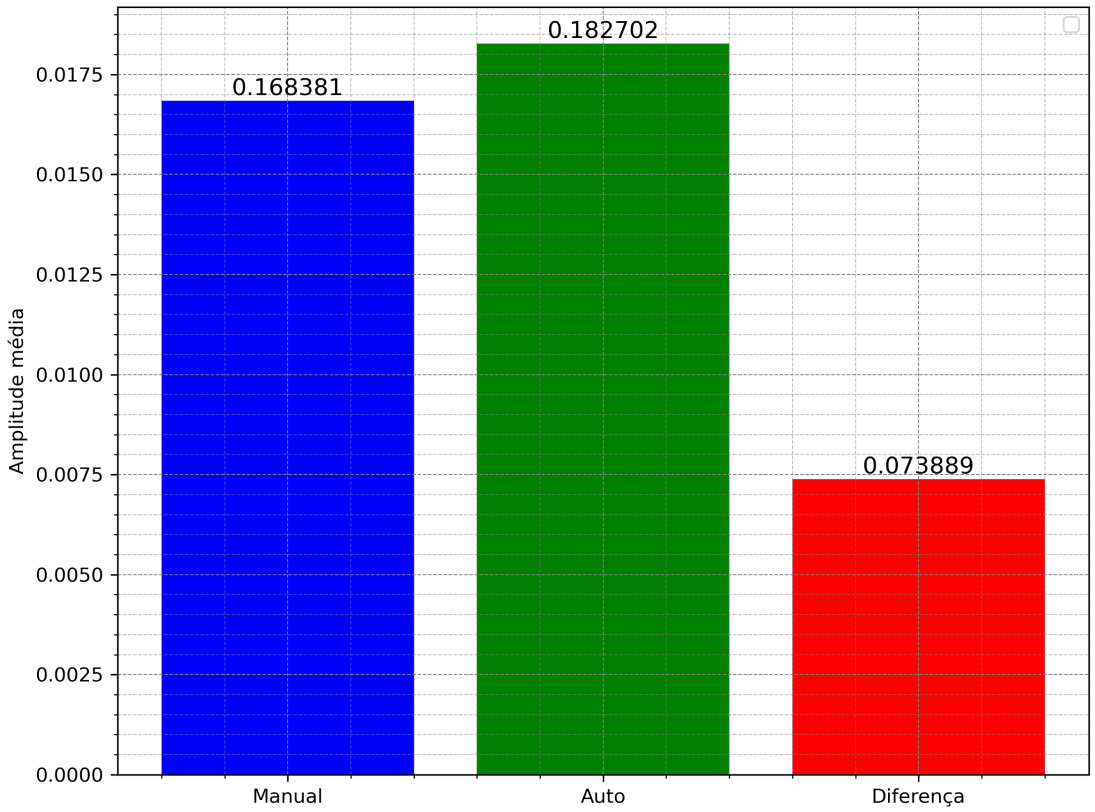


Figura 4.13: Comparação da amplitude média entre os dados empilhados, para os dados sísmicos do Golfo do México.

Essa observação é reforçada pela Figura 4.14, que ilustra a diferença da amplitude média por traço, demonstrando variações pontuais ao longo do painel sísmico. É notável a presença de oscilações de amplitude em vários trechos, o que pode estar relacionado tanto à variabilidade natural do dado quanto à sensibilidade do modelo automático às pequenas mudanças estratigráficas. Essas variações, apesar de oscilarem em torno de zero, revelam uma flutuação não desprezível, que deve ser interpretada com cautela em ambientes com ruídos, como eventos múltiplos do fundo do mar.

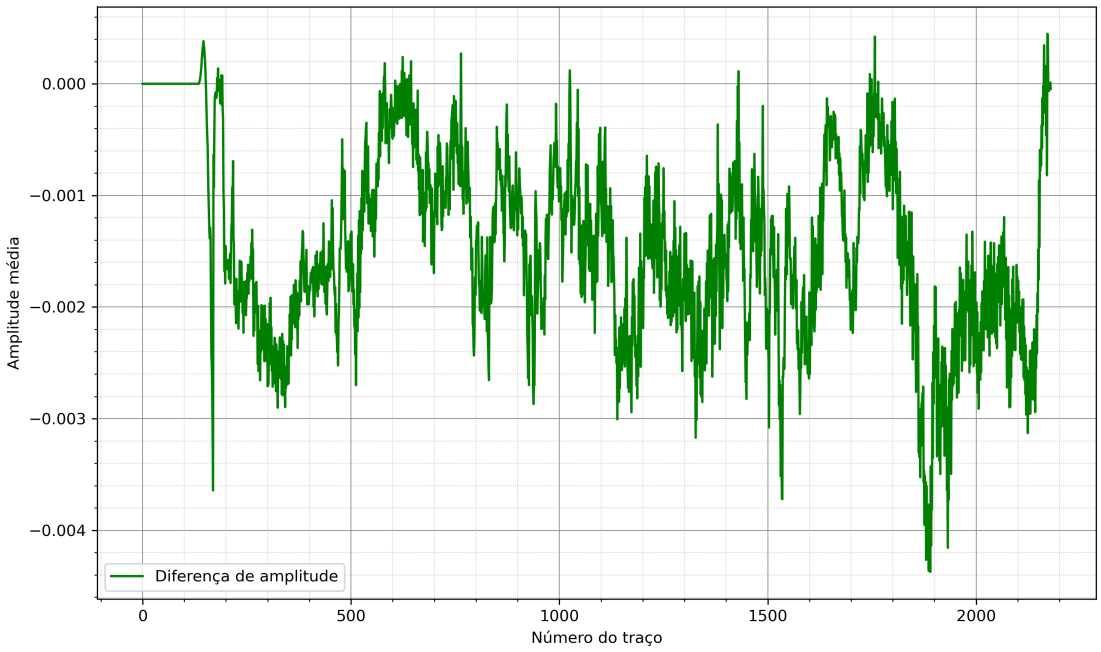


Figura 4.14: Diferença de amplitude por traço, para os dados sísmicos do Golfo do México.

A Figura 4.15 reforça essa análise ao exibir a diferença de energia entre os dois métodos ao longo das interfaces. A presença de picos de energia acima de 1.5 em certos pontos do painel indica que o método automático pode estar atribuindo mais energia a determinados eventos, o que poderia representar uma amplificação de reflexões múltiplas ou difrações. A variabilidade energética também pode sugerir um ajuste de velocidade que, embora mais dinâmico, introduz contrastes mais intensos ao longo da seção.

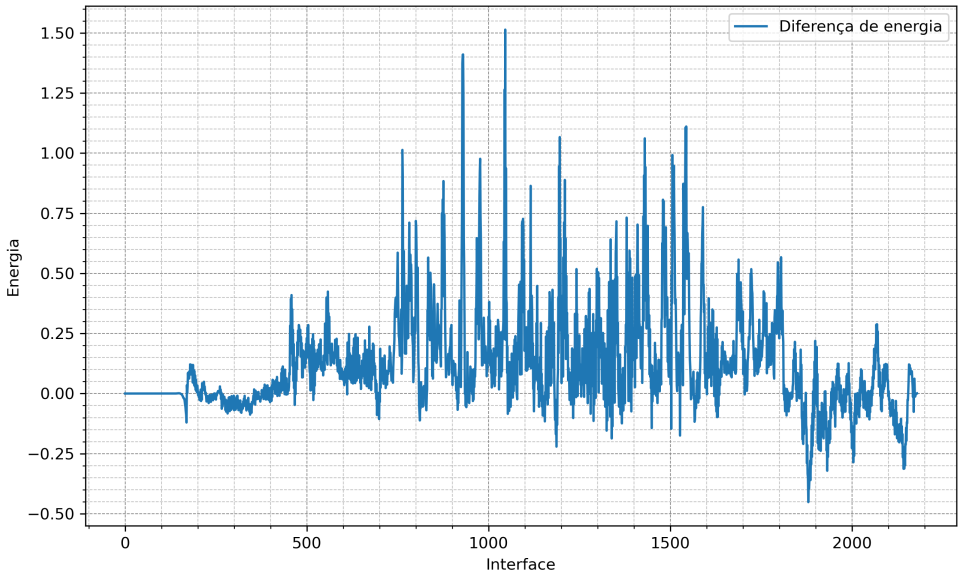


Figura 4.15: Diferença de energia acumulada, para os dados sísmicos do Golfo do México.

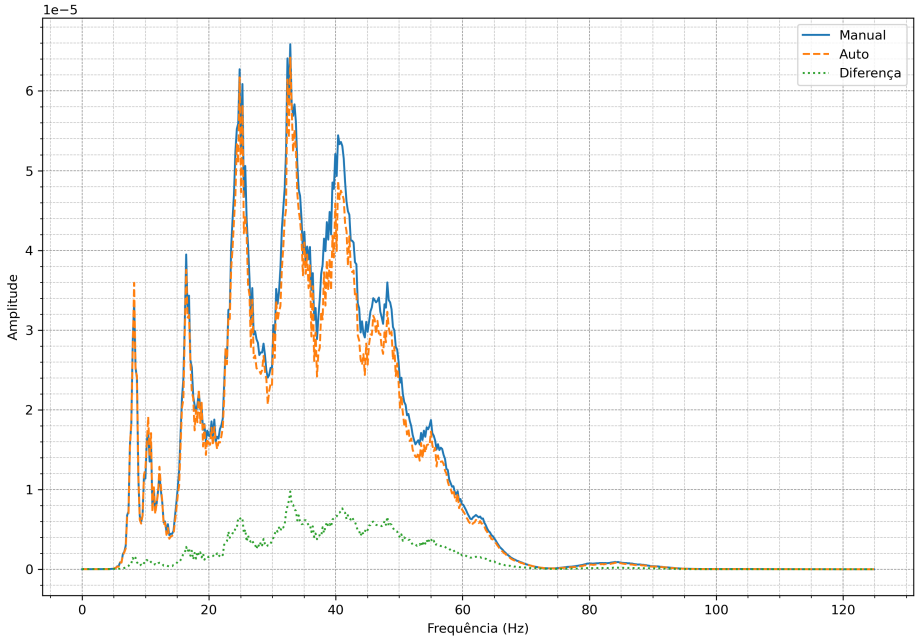


Figura 4.16: Análise espectral comparativa, para os dados sísmicos do Golfo do México.

Por fim, a Figura 4.16, com a análise do espectro de frequência, evidencia a boa correspondência entre os métodos nas bandas dominantes (20–60 Hz), embora o método automático apresente ligeiras diferenças em amplitudes específicas. O espectro da diferença confirma que

os dois métodos são coerentes na maior parte da faixa útil de frequência, mas apresentam padrões distintos na região de 20 a 40 Hz, o que pode ser indicativo de uma sensibilidade maior do algoritmo automático aos eventos múltiplos ou a pequenas variações estruturais.

Assim, a metodologia proposta demonstra ser uma ferramenta promissora para obtenção robusta do campo de velocidades em ambientes sísmicos marinhos complexos, superando limitações dos métodos convencionais, proporcionando agilidade e confiabilidade aos resultados obtidos.

5

Conclusões

As conclusões do presente estudo evidenciam o êxito na implementação e integração de técnicas de aprendizagem de máquina para a determinação automática do campo de velocidades em dados sísmicos. Este trabalho não apenas respondeu às questões formuladas inicialmente, mas também trouxe contribuições significativas para a área de processamento de dados geofísicos.

A aplicação da metodologia baseada no pré-agrupamento amostral mostrou-se fundamental para estruturar os dados antes da execução do algoritmo K-means++. Esse procedimento inicial possibilitou a eliminação de classes estatisticamente irrelevantes, otimizando o agrupamento subsequente e aumentando a estabilidade dos resultados. A integração do K-means++ com a Análise de Componentes Principais (PCA) revelou-se particularmente eficaz, pois reduziu a dimensionalidade dos dados sem comprometer a precisão e a consistência nos cálculos de velocidade. Além disso, essa abordagem assegurou uma distribuição mais equilibrada dos centróides, minimizando a influência de ruídos e valores extremos.

O estudo também ressaltou a eficiência do pré-agrupamento em grandes volumes de dados amostrais, proporcionando ganhos computacionais e interpretativos. Ao reduzir a variabilidade e aumentar a homogeneidade, foi possível aprimorar a confiabilidade dos modelos gerados. A análise estatística detalhada dos resultados confirmou a consistência e a robustez das técnicas implementadas, oferecendo uma abordagem replicável em diferentes cenários e contribuindo para o avanço tecnológico na indústria de óleo e gás.

O uso da equação de Dix, possibilitou a criação de um banco de dados com valores de velocidade preliminares. Esses valores foram empregados como dados de entrada para o treinamento de uma rede neural MLP, cuja aplicação desempenhou papel essencial na su-

avização e estabilização dos perfis de velocidade. Essa estratégia não apenas viabilizou a construção de modelos mais robustos e eficientes, reduzindo o tempo e o esforço humano na interpretação sísmica, como também assegurou que os valores resultantes fossem fisicamente consistentes, respeitando o crescimento com a profundidade e minimizando artefatos oriundos de variações temporais e de eventos ruidosos.

Os resultados obtidos nas aplicações da metodologia tanto em dado sintético quanto no dado sísmico do Golfo do México, demonstraram que ela é capaz de identificar com precisão as velocidades subsuperficiais em diferentes camadas geológicas.

Ao atingir os objetivos estabelecidos, este estudo contribuiu para avanços metodológicos relevantes, introduzindo soluções inovadoras para desafios existentes na sísmica de reflexão. A combinação das técnicas apresentadas mostrou-se eficiente na otimização dos processos exploratórios, aumentando a precisão na caracterização geológica e reduzindo riscos ambientais e operacionais associados à exploração de hidrocarbonetos.

Referências Bibliográficas

Alfuraidan, M.; Al-Shuhail, A.; Hanafy, S. e Sarumi, I. (2023) Approximation of seismic velocities from the spectrum of weighted graphs, GEM - International Journal on Geomathematics, **14**.

Anderson, T. W. (2003) An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, Hoboken, NJ, 3º edic..

Anton, H. e Rorres, C. (2004) Álgebra Linear com Aplicações, Bookman, Porto Alegre.

Araya-Polo, S.; Blanchard, T. M.; Giraldo, F. F. e Brossier, R. (2020) Neural network architectures for seismic velocity model building, IEEE Transactions on Geoscience and Remote Sensing, **58**(6):4387–4400.

Araújo, K. (2018) Plataforma interativa para análise sísmica: uma proposta metodológica, `https://repositorio.ufpa.br/jspui/handle/2011/10556`, Dissertação de Mestrado, Universidade Federal do Pará.

Arthur, D. e Vassilvitskii, S. (2007) k-means++: The advantages of careful seeding, In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SIAM.

Ashcroft, W. (2011) A Petroleum Geologist's Guide to Seismic Reflection, Wiley-Blackwell, Oxford.

Bahmani, B.; Moseley, B.; Vattani, A.; Kumar, R. e Vassilvitskii, S. (2012) Scalable k-means++, Proceedings of the VLDB Endowment, **5**(7):622–633.

Bishop, C. M. (2006) Pattern Recognition and Machine Learning, Springer, New York, ISBN 978-0-387-31073-2.

Book, D. L. (2025) As principais arquiteturas de redes neurais, `https://www.deeplearningbook.com.br/as-principais-arquiteturas-de-redes-neurais/`, Acesso em: 4 set. 2025.

Castagna, J.; Batzle, M. e Eastwood, R. (1985) Relationship between compressional-wave and shear-wave velocities in clastic silicate rocks, Geophysics, **50**(4):571–581.

Celebi, M. E.; Kingravi, H. A. e Vela, P. A. (2013) A comparative study of efficient initialization methods for the k-means clustering algorithm, Expert Systems with Applications, **40**(1):200–210.

Chen, C.-M. e Simaan, M. A. (1991) Velocity filters for multiple interference attenuation in two-dimensional geophysical data, IEEE Transactions on Geoscience and Remote Sensing, **29**(4):563–570.

Chen, S.; Jin, S.; Li, X.-Y. e Yang, W. (2017) Nonstretching nmo correction using a dynamic time warping algorithm, SEG Annual Meeting Expanded Abstracts.

Clark, R. (2016) Seismic acquisition techniques in marine environments, Journal of Marine Geophysics, **45**(3):215–230.

Ding, C.; He, X. e Simon, H. D. (2015) On the equivalence of nonnegative matrix factorization and k-means, spectral clustering, and pca, Pattern Recognition, **46**(1):284–296.

Dix, C. (1955) Seismic velocities from surface measurements, Geophysics, **20**(1):68–86.

Dobrin, M. B. e Savit, C. H. (1976) Introduction to geophysical prospecting, McGraw-Hill.

Dunkin, J. W. e Levin, F. K. (1973) Effect of normal moveout on a seismic pulse, Geophysics, **38**(4):635–645.

Ebadi, M. R. (2017) Coherent and incoherent seismic noise attenuation using parabolic radon transform and its application in environmental geophysics, Modeling Earth Systems and Environment, **3**(18).

Farris, S.; Clapp, R. e Araya-Polo, M. (2023) Learning-based seismic velocity inversion with synthetic and field data, Sensors, **23**(19).

Faust, L. (1951) Seismic velocity as a function of depth and geologic time, Geophysics, **16**(2):192–206.

Ferreira, R. S.; Oliveira, D. A.; Semin, D. G. e Zaytsev, S. (2020) Automatic velocity analysis using a hybrid regression approach with convolutional neural networks, IEEE Transactions on Geoscience and Remote Sensing, **59**(5):4464–4470.

Fraley, C. e Raftery, A. E. (2002) Model-based clustering, discriminant analysis, and density estimation, Journal of the American Statistical Association, **97**(458):611–631.

Freedman, D.; Pisani, R. e Purves, R. (2007) Statistics, W. W. Norton & Company, New York, 4º edic., ISBN 9780393929720.

Gonzalez, R. C. e Woods, R. E. (2000) Processamento de Imagens Digitais, Edgard Blucher, São Paulo.

Goodfellow, I.; Bengio, Y. e Courville, A. (2016) Deep Learning, MIT Press, Cambridge, MA, ISBN 978-0262035613.

Grechka, V. e Tsvankin, I. (1999) 3-d description of normal moveout in anisotropic inhomogeneous media, Geophysics, **64**(2):652–663.

Grechka, V. e Tsvankin, I. (2019) Nmo-velocity surfaces and dix-type formulae in anisotropic heterogeneous media, arXiv preprint.

Halliburton (2009) Promax/seisspace technical reference - nmo correction, `https://www.landmark.solutions`.

Hanna, M. T. e Simaan, M. A. (1987) Design and implementation of velocity filters using multichannel array processing techniques, IEEE Transactions on Acoustics, Speech, and Signal Processing, **35**(6):864–877.

Harsuko, R.; Cheng, S. e Alkhalifah, T. (2024) Propagating the prior from shallow to deep with a pre-trained velocity-model generative transformer, Journal of Geophysical Research.

Harsuko, R.; Cheng, S. e Alkhalifah, T. (2025) Synthesizing realistic-scale seismic velocity models using a spatially-aware generative model, **2025**(1):1–5.

Hastie, T.; Tibshirani, R. e Friedman, J. (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, New York, 2º edic..

Haykin, S. (2009) Neural Networks and Learning Machines, Pearson Education, Upper Saddle River, NJ, 3º edic., ISBN 978-0131471399.

Hornik, K.; Stinchcombe, M. e White, H. (1989) Multilayer feedforward networks are universal approximators, Neural Networks, **2**(5):359–366.

Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components, Journal of Educational Psychology, **24**(6):417–441.

Jackson, J. E. (2005) A User's Guide to Principal Components, John Wiley & Sons, Hoboken, NJ.

Jain, A. K. (2010) Data clustering: 50 years beyond k-means, Pattern Recognition Letters, **31**(8):651–666.

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

Johnson, R. A. e Wichern, D. W. (2007) Applied Multivariate Statistical Analysis, Pearson, Upper Saddle River, NJ, 6º edic., ISBN 9780131877153.

Jolliffe, I. T. (2002) Principal Component Analysis, Springer, New York, 2º edic..

Kanungo, T.; Mount, D. M.; Netanyahu, N. S.; Piatko, C. D.; Silverman, R. e Wu, A. Y. (2002) An efficient k-means clustering algorithm: Analysis and implementation, IEEE Transactions on Pattern Analysis and Machine Intelligence, **24**(7):881–892.

Khoshnavaz, H.; Tavakoli, A. e Ghaffari, H. (2021) Stretch-free nmo correction in cmp gathers, Journal of Geophysics and Engineering, **18**(5):701–713.

Kim, S.; Cho, Y. e Jun, H. (2024) Resolution enhancement for a seismic velocity model using machine learning, Geophysical Journal International, **238**(2):681–699.

Kingma, D. P. e Ba, J. (2015) Adam: A method for stochastic optimization, International Conference on Learning Representations (ICLR).

Knapp, R. W. e Steeples, D. W. (1986) High-resolution common-depth-point seismic reflection profiling: Instrumentation, Geophysics, **51**(2):276–282.

Koch, K. R. (2007) Principal component analysis of seismic data: theory and applications, Geophysical Prospecting, **55**(3):413–429.

LeCun, Y.; Bengio, Y. e Hinton, G. (2015) Deep learning, Nature, **521**(7553):436–444.

Liner, C. L. (1999) Elements of 3D Seismology, PennWell Books.

Liu, W. e Marfurt, K. (2015) Elimination of acquisition footprint in 3d land seismic data using dip-steered median filter, Geophysical Prospecting, **63**(2):340–354.

Liu, Z.; Wong, H. K. e Jun, M. B. G. (2010) A multilayer perceptron neural network approach to seismic velocity inversion, Journal of Applied Geophysics, **70**(2):90–98.

Lloyd, S. (1982) Least squares quantization in pcm, IEEE Transactions on Information Theory, **28**(2):129–137.

Ma, Y.; Tang, J.; Liu, K. e Zhang, S. (2021) Seismic velocity inversion using deep learning with physics constraints, Geophysics, **86**(3):R233–R247.

Van der Maaten, L.; Postma, E. e Van den Herik, J. (2009) Dimensionality reduction: A comparative review, Journal of Machine Learning Research, **10**:66–71.

MacQueen, J. et al. (1967) Some methods for classification and analysis of multivariate observations, In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 281–297, Oakland, CA, USA.

Marfurt, K. J. e Alves, T. M. (2015) Pitfalls and limitations in seismic attribute interpretation of tectonic features, Interpretation, **3**(1):A5–A15.

Mavko, G.; Mukerji, T. e Dvorkin, J. (2009) The Rock Physics Handbook: Tools for Seismic Analysis of Porous Media, Cambridge University Press, 2º edic..

Mayne, W. H. (1962) Common reflection point horizontal data stacking techniques, Geophysics, **27**(6):927–938.

McCulloch, W. S. e Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity, The Bulletin of Mathematical Biophysics, **5**:115–133.

Milligan, G. W. e Cooper, M. C. (1985) An examination of procedures for determining the number of clusters in a data set, Psychometrika, **50**(2):159–179.

Minsky, M. e Papert, S. (1969) Perceptrons: An Introduction to Computational Geometry, MIT Press, Cambridge, MA, ISBN 978-0262631112.

Moore, D. S.; McCabe, G. P. e Craig, B. A. (2012) Introduction to the Practice of Statistics, W. H. Freeman, New York, 7º edic., ISBN 9781429274340.

Moreira Neto, C. A.; Pestana, R. C. e Aldunate, G. C. (2005) Migração pré-empilhamento em profundidade no domínio da frequência de seções de ondas planas, Revista Brasileira de Geofísica, **23**(4):361–372.

Moulik, P. e Ekström, G. (2014) The relationship between large-scale variations in shear velocity, density, and compressional velocity in the earth's mantle, Journal of Geophysical Research: Solid Earth, **119**(7):5510–5527.

Muller, A.; Bom, C.; Costa, J.; Klatt, M.; Faria, E.; Silva, B. e Albuquerque, M. (2022) Deep-tomography: iterative velocity model building with deep learning, arXiv preprint arXiv:2209.12804.

Neidell, N. S. e Taner, M. T. (1971) Semblance and other coherency measures for multichannel data, Geophysics, **36**(3):482–497.

Ozegin, K. (2012) FUNDAMENTALS OF ACTIVE METHODS OF GEOPHYSICAL PROSPECTING, ISBN 978-978-900-913-8.

Pearson, K. (1901) Liii. on lines and planes of closest fit to systems of points in space, Philosophical Magazine, **2**(11):559–572.

Pelleg, D. e Moore, A. W. (2000) X-means: Extending k-means with efficient estimation of the number of clusters, In: *Proceedings of the 17th International Conference on Machine Learning (ICML)*, pp. 727–734, Morgan Kaufmann.

Porsani, M. J. (2000) Imageamento multifoco de refletores sísmicos, Dissert. de Mestrado, Universidade Federal do Pará.

Rajagopalan, S.; Ghosh, S. e Rao, S. S. (2019) Unsupervised learning for seismic facies classification using clustering and correlation analysis, Interpretation, **7**(4):T813–T829.

Ringnér, M. (2008) What is principal component analysis?, Nature Biotechnology, **26**(3):303–304.

Rosenblatt, F. (1958) The perceptron: A probabilistic model for information storage and organization in the brain, Psychological Review, **65**(6):386–408.

Rumelhart, D. E.; Hinton, G. E. e Williams, R. J. (1986) Learning representations by back-propagating errors, Nature, **323**(6088):533–536.

Schimmel, M. e Paulssen, H. (1997) Noise reduction and detection of weak, coherent signals through phase-weighted stacks, Geophysical Journal International, **130**(2):497–505.

Schön, J. (2015) Compressional-wave seismic velocity, bulk density, and their variation with mineralogy, porosity, pressure, temperature and fracturing, USGS Report SIR2023–5061.

Sen, M.; Khan, S. D. e Vernik, L. (2019) Principal component analysis (pca) of seismic attributes for reservoir characterization, Interpretation, **7**(2):T293–T304.

Sen, M. K. e Stoffa, P. (2019) Seismic inversion methods: A practical approach, Cambridge University Press.

Shah; Levin et al. (n.d.) An analysis of stacking, rms, average, and interval velocities over a horizontally layered ground, ResearchGate Archive.

Sheriff, R. E. e Geldart, L. P. (1995) Exploration Seismology, Cambridge University Press, Cambridge, 2º edic..

Shindler, M.; Wong, A. e Meyerson, A. (2011) Fast and accurate k-means for large datasets, In: *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 24, pp. 2375–2383, Curran Associates, Inc.

Shlens, J. (2014) A tutorial on principal component analysis, arXiv preprint arXiv:1404.1100.

Souza, M. S. d. (2014) Determinação automática da velocidade de empilhamento e obtenção da seção zero-offset, Orientador: Prof. Dr. Milton J. Porsani.

Steinley, D. (2006) K-means clustering: A half-century synthesis, British Journal of Mathematical and Statistical Psychology, **59**(1):1–34.

Sturges, H. A. (1926) The choice of a class interval, Journal of the American Statistical Association, **21**(153):65–66.

de São Paulo, U. (2020) Introdução aos métodos sísmicos, Disponível em: `https://edisciplinas.usp.br/pluginfile.php/5606827/mod_resource/content/1/agg1162020-aula3-sismica.pdf`.

Tan, P.-N.; Steinbach, M.; Karpatne, A. e Kumar, V. (2019) Introduction to Data Mining, Pearson, Boston, 2º edic..

Taner, M. e Koehler, F. (1969) Velocity spectra—digital computer derivation applications of velocity functions, Geophysics, **34**(6):859–881.

Telford, W. M.; Geldart, L. P. e Sheriff, R. E. (1990) Applied geophysics, Cambridge University Press, pp. 38–43.

Tsvankin, I. (1997) Anisotropic parameters and p-wave velocity for orthorhombic media, Geophysics, **62**(4):1292–1309.

Tsvankin, I. e Gutierrez, M. A. (1996) Nonhyperbolic reflection moveout in anisotropic media, In: *SEG Technical Program Expanded Abstracts 1996*, pp. 1179–1182, Society of Exploration Geophysicists.

Ulrych, T. J.; Sacchi, M. D. e Woodbury, A. D. (2012) Principles of Geophysical Data Inversion, Cambridge University Press, Cambridge.

Vasconcelos, S. (2021) Análise de Componentes Principais (PCA), Documento técnico.

Vassiliou, A. A.; Maravelakis, P. e Kotsiantis, S. (2021) Signal processing and pattern recognition with pca: A comprehensive review, Pattern Recognition Letters, **152**:12–28.

Wang, F.; Huang, X. e Alkhalifah, T. (2024) Controllable seismic velocity synthesis using generative diffusion models, arXiv preprint arXiv:2403.11852.

Wang, Z. e Nur, A. (1992) Effects of porosity and clay content on acoustic velocities in sandstones, Geophysics, **57**(5):573–582.

Wang, Z.; Li, X. e Zhao, J. (2019) Principal component analysis of seismic attributes based on multi-scale wavelet decomposition, Geophysical Journal International, **217**(2):1103–1117.

Wold, S.; Esbensen, K. e Geladi, P. (1987) Principal Component Analysis, vol. 2.

Wyllie, M.; Gregory, A. e Gardner, L. (1956) Elastic wave velocities in heterogeneous and porous media, Geophysics, **21**(1):41–70.

Xu, R. e Wunsch, D. (2005) Survey of clustering algorithms, IEEE Transactions on Neural Networks, **16**(3):645–678.

Xu, R. e Wunsch, D. C. (2008) Clustering, Wiley-IEEE Press, Hoboken, NJ, ISBN 9780470290011.

Yang, W. (2014) Reflection Seismology: Theory, Data Processing and Interpretation, Elsevier, Amsterdam.

Yilmaz, O. e Doherty, S. M. (1987) Seismic data processing, Society of Exploration Geophysicists.

Yilmaz, (2001) Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data, Society of Exploration Geophysicists (SEG), Tulsa, OK, ISBN 978-1-56080-094-1.

Zhang, P. e Lu, W. (2016) Automatic time-domain velocity estimation based on an accelerated clustering method, Geophysics, **81**(4):U13–U23.

Zhang, R. e Schuster, S. G. (2000) Neural network implementation of velocity analysis, Geophysics, **65**(5):1512–1521.

Zhang, R.; Chen, Z.; Chen, T. e Zhang, J. (2019) Deepseis: Deep learning for seismic horizon tracking, Geophysics, **84**(6):IM35–IM45.

Agradecimentos

Aos meus pais, Maria Secundina Lima da Luz e José Augusto Pereira da Luz (in memoriam), meus grandes apoiadores em toda a minha trajetória acadêmica. À minha mãe, pelo amor, força e incentivo constantes; e ao meu pai, cuja memória permanece viva em mim e segue sendo uma fonte de inspiração e dedicação.

À minha esposa, Rosana Maria do Nascimento Luz, pelo apoio incondicional e pela dedicação integral a esta jornada. Desde o mestrado até o doutorado, em muitos momentos, assumiu sozinha todas as responsabilidades familiares para que eu pudesse me concentrar nos estudos. Este título também lhe pertence, pois nada disso seria possível sem seu amor, dedicação, paciência e companheirismo.

Aos meus filhos, João Marcos, Marcele Lorena e Monique Dayane, pela compreensão nos momentos de ausência, pelo carinho diário e pela motivação. Vocês são a razão do meu esforço e a inspiração para acreditar que sempre é possível ir além.

Aos meus irmãos, Márcia Cristina e Márcio Rodrigo, pelo companheirismo e apoio em todos os momentos, compartilhando não apenas os laços familiares, mas também os sonhos e conquistas ao longo da vida.

Ao meu orientador, Prof. Dr. Marcos Alberto Rodrigues Vasconcelos, pela disponibilidade e generosidade em aceitar o desafio de me orientar em um momento em que eu estava sem direção acadêmica. Sua confiança, dedicação, paciência e incentivo foram fundamentais para a conclusão desta pesquisa. Mais do que um orientador, foi um verdadeiro parceiro de caminhada científica, cuja contribuição jamais esquecerei.

Aos meus colegas do Doutorado, pela convivência enriquecedora, pela troca de experiências, pelas discussões científicas e pelo apoio mútuo ao longo dessa jornada. A caminhada foi mais leve e produtiva graças à colaboração, à amizade e ao espírito de companheirismo construído nesse período.

À minha querida amiga Nilda de Araújo Lima, que por tantos anos foi a alma acolhedora da secretaria do Programa de Pós-Graduação em Geodinâmica e Geofísica da UFRN. Hoje

já aposentada, Nilda continua sendo uma presença marcante e muito estimada em minha vida e na de todos que tiveram o privilégio de conviver com ela. Sempre solícita, atenciosa e generosa, foi e segue sendo um exemplo de dedicação e humanidade. Mesmo após a mudança de cidade e de estado, mantivemos o contato e a amizade sincera construída ao longo do tempo, que permanece como uma das mais belas lembranças e conquistas do meu percurso acadêmico e pessoal.

À Universidade Federal Rural da Amazônia (UFRA), pelo afastamento das minhas atividades docentes, medida essencial que possibilitou minha dedicação integral ao programa de Doutorado.

Aos professores do Programa de Pós-Graduação em Geofísica da Universidade Federal da Bahia (UFBA), pela excelência na formação, pelo rigor científico e pela disponibilidade em compartilhar conhecimentos e orientações fundamentais para meu desenvolvimento acadêmico e profissional. Em especial, ao professor Hédison Kiuity Sato, com quem tive a oportunidade de conviver durante esta trajetória. Foram inúmeros finais de semana passados na universidade, não apenas dedicados ao trabalho, mas também em conversas sobre os mais diversos temas, que extrapolavam o ambiente acadêmico. Essa convivência resultou não apenas em aprendizado, mas também na construção de uma amizade verdadeira. Mais do que um professor, o professor Sato tornou-se um grande amigo que levarei para toda a vida.

Ao Programa de Pós-Graduação em Geofísica do Instituto de Geociências da UFBA e ao Centro de Pesquisa em Geofísica e Geologia (CPGG), pela infraestrutura e pelo ambiente de pesquisa estimulante, que proporcionaram as condições ideais para o avanço desta investigação.

À CAPES, pelo financiamento e suporte, que viabilizaram esta etapa tão importante da minha formação acadêmica.

Apêndice A

Análise de eficiência

A modelagem estatística é aplicada para verificar a eficiência do processo de divisão em classes dos dados sísmicos amostrados sem eventos múltiplos. O principal objetivo é avaliar a qualidade do processo, utilizando o coeficiente de variação (CV) e a análise gráfica de variabilidade como métricas de homogeneidade dos dados distribuídos em classes, em comparação com dados amostrais sem agrupamento.

Com os intervalos de classe e frequências devidamente calculados, procede-se ao cálculo do ponto médio de classe, em que l_{inf} e L_{sup} passam a representar os limites inferior e superior da classe, respectivamente. Dessa forma, o ponto médio de classe (x_i) é calculado conforme a equação (A.1):

$$x_i = \frac{(l_{inf}) + (L_{sup})}{2} \tag{A.1}$$

Para evitar o uso de valores elevados e obter ganhos computacionais significativos, foi realizada uma mudança de variável por meio da equação (A.2):

$$y_i = \frac{x_i - x_0}{h} \tag{A.2}$$

Desse modo, a média (μ) passa a ser obtida pela equação (A.3), em que f_i representa a frequência absoluta, y_i são os valores transformados da variável, x_0 é o valor de referência e h a amplitude da classe:

$$\mu = x_0 + h\frac{\sum_{i=1}^{n}(f_i y_i)}{n} \tag{A.3}$$

A moda para dados agrupados representa o valor ou intervalo de valores que aparece com mais frequência em uma distribuição. Sua fórmula leva em consideração a classe com a maior frequência simples, além das frequências das classes adjacentes. Sendo calculada nesse caso, pela equação (A.4):

$$M_o = l_i + \left(\frac{f - f_{(\text{ant})}}{2f - f_{(\text{ant})} - f_{(\text{pos})}}\right) h \tag{A.4}$$

Nessa fórmula, l_i é o limite inferior da classe modal, o termo f representa a frequência da classe modal, enquanto $f_{(\text{ant})}$ e $f_{(\text{pos})}$ correspondem, respectivamente, às frequências das classes imediatamente anterior e posterior à classe modal. Já h representa a amplitude da classe modal. Esse cálculo considera a influência das classes adjacentes para fornecer uma estimativa ajustada da moda, evitando uma aproximação rígida apenas ao ponto médio da classe modal, oferecendo estimativas precisas e permitindo uma interpretação mais robusta da distribuição dos dados em contextos onde o agrupamento em classes é necessário.

Já a mediana é a medida que representa o valor central de uma distribuição, dividindo-a em duas partes iguais. Em uma distribuição de dados agrupados, a mediana é localizada dentro da classe cuja frequência acumulada atinge ou ultrapassa a metade do total das observações. Diferente da média, que considera todos os valores, a mediana se concentra apenas no ponto que separa os 50% menores dos 50% maiores valores da distribuição, o que a torna uma medida resistente a valores extremos. Para calcular a mediana em dados agrupados, utiliza-se a fórmula:

$$M_d = l_i + \left(\frac{\frac{\sum_{i=1}^{n} f_i}{2} - F_{(\text{ant})}}{f}\right) h \tag{A.5}$$

Nesta expressão, l_i é o limite inferior da classe que contém a mediana. A posição da mediana é obtida dividindo a soma total das frequências por dois, representada pelo termo $\sum_{i=1}^{n} f_i/2$. Esse valor indica a posição central da distribuição e permite identificar a classe mediana. O termo $F_{(\text{ant})}$ corresponde à frequência acumulada das classes anteriores à classe mediana,

enquanto f é a frequência absoluta da própria classe mediana. Já h é a amplitude da classe. A fórmula ajusta a posição da mediana dentro do intervalo de sua classe, considerando a proporção da frequência acumulada até essa classe, sendo assim, uma medida importante para entender o comportamento dos dados sem a interferência de valores atípicos.

Seguindo para a análise de dispersão dos dados, temos o cálculo da variância para dados agrupados (σ^2), esta mede a dispersão ou variabilidade dos dados em torno da média, permitindo avaliar como os dados se distribuem em relação ao seu valor médio. Quando os dados são agrupados, a variância é calculada com uma fórmula que considera a amplitude de classe (h), a frequência absoluta de classe (f_i), e a posição relativa dos pontos de classe em relação à média da distribuição (y_i). O cálculo também leva em conta o número total de observações (n). A fórmula para a variância agrupada é:

$$\sigma^2 = h^2 \left(\frac{\sum_{i=1}^{n} \left(f_i y_i^2 \right)}{n-1} - \left(\frac{\sum_{i=1}^{n} \left(f_i y_i \right)}{n-1} \right)^2 \right) \tag{A.6}$$

Utilizar a variância para dados agrupados apresenta a vantagem de simplificar o cálculo da variabilidade em grandes conjuntos de dados e por preservar a estrutura da distribuição. Esse método é especialmente útil para distribuições que apresentam padrões consistentes dentro de intervalos, possibilitando uma interpretação mais robusta da estabilidade e variabilidade dos dados.

Já o desvio padrão (σ) é calculado pela equação (A.7), que é obtida extraindo a raiz quadrada da variância:

$$\sigma = h \sqrt{\frac{\sum_{i=1}^{n} \left(f_i y_i^2 \right)}{n-1} - \left(\frac{\sum_{i=1}^{n} \left(f_i y_i \right)}{n-1} \right)^2} \tag{A.7}$$

A eficiência do agrupamento dos dados em classes é analisada através do coeficiente de variação (CV), uma métrica que expressa a variação relativa dos dados em relação à média. Quanto menor o CV, mais homogêneos são os dados em cada classe. O coeficiente de variação é calculado pela equação (A.8), onde o desvio padrão (σ) é dividido pela média (μ) e multiplicado por 100:

$$CV = \frac{\sigma}{\mu}100 \tag{A.8}$$

A.1 Análise da variabilidade do dados: Gráfico de controle estatístico.

O Gráfico de controle estatístico é uma ferramenta eficaz nesse contexto, por permitir monitorar a estabilidade e a variabilidade dos dados sísmicos ao longo do tempo ou da profundidade. Esse gráfico ajuda a identificar se as variações presentes nos dados são parte de um comportamento esperado (causas comuns) ou se indicam interferências ou problemas específicos (causas especiais). Para construir esse gráfico, utilizam-se cálculos de média e desvio padrão, que servem como parâmetros para definir limites de controle. Esses limites representam os intervalos dentro dos quais a variabilidade dos dados é considerada aceitável. Esse procedimento é aplicado tanto a dados agrupados (organizados em classes) quanto a dados não agrupados (brutos), para que seja possível visualizar qual procedimento melhor centraliza a distribuição das amostras com menor dispersão.

Os limites de controle superior e inferior para os dados agrupados são determinados da seguinte forma:

$$\text{Limite Superior Agrupado } (LSA) = \mu + 2\sigma$$

$$\text{Limite Inferior Agrupado } (LIA) = \mu - 2\sigma$$

Esses limites ajudam a identificar se os dados agrupados apresentam uma variabilidade consistente e esperada em torno da média.

Para os dados não agrupados, a média e o desvio padrão são calculados diretamente a partir dos valores individuais da amostra, sem a necessidade de divisão em classes. A média não agrupada, $\bar{x}_{\text{não agrupada}}$, é dada pela soma de todos os valores de amplitude x_i dividida pelo número total de observações n, conforme a equação (A.9):

$$\bar{x}_{\text{não agrupada}} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{A.9}$$

O desvio padrão não agrupado, $\sigma_{\text{não agrupado}}$, mede a dispersão dos valores em torno da média e é calculado pela raiz quadrada da variância dos dados não agrupados, segundo a equação (A.10):

$$\sigma_{\text{não agrupado}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_{\text{não agrupada}})^2} \tag{A.10}$$

Os limites de controle para dados não agrupados são determinados da seguinte forma:

$$\text{Limite Superior Não Agrupado}(LSNA) = \bar{x}_{\text{não agrupada}} + 2\sigma_{\text{não agrupado}}$$

$$\text{Limite Inferior Não Agrupado } (LINA) = \bar{x}_{\text{não agrupada}} - 2\sigma_{\text{não agrupado}}$$

No Gráfico de controle de processo, os limites de controle definidos por essas médias e desvios padrões (tanto para dados agrupados quanto para dados não agrupados) permitem monitorar a variabilidade e identificar possíveis anomalias. Para os dados agrupados, os limites de controle proporcionam uma análise mais estável e filtrada, o que facilita a observação de tendências. Por outro lado, os limites para dados não agrupados capturam uma variabilidade mais ampla.

Complementando essa análise, o coeficiente de variação (CV) fornece uma medida relativa da dispersão dos dados em relação à média, expressa em percentual. Esse coeficiente é especialmente importante na utilização da técnica k-means++, pois um CV menor indica que os dados estão mais concentrados ao redor da média, proporcionando maior precisão na interpretação e consequentemente, na identificação do tempo de reflexão que está relacionado com as maiores concentrações de valores de amplitude, relação esta que será fundamental para o cálculo dos valores de velocidade por meio da equação de (Dix, 1955) explorada no próximo capítulo.

A.2 Resultados

A.2.1 Análise do traço 1

As métricas calculadas do traço 1 para dados agrupados e não agrupados são apresentadas na Tabela A.1. É possível identificar que a média e a mediana dos dados agrupados estão ligeiramente mais altas em comparação com os valores não agrupados, refletindo uma maior estabilidade dos dados após o agrupamento. Além disso, o coeficiente de variação para os dados agrupados é positivo e razoavelmente baixo, indicando uma distribuição mais consistente em torno da média, enquanto o coeficiente de variação dos dados não agrupados

exibe um valor extremo e negativo, apontando para uma dispersão elevada e a presença de variabilidade não controlada.

A análise do traço 1 (Figura A.1) revela uma série de características importantes sobre a variabilidade e o controle estatístico dos dados analisados, bem como os benefícios do processo de agrupamento. O gráfico de controle estatístico (Figura A.2), permite observar a distribuição das amplitudes ao longo do tempo (ou profundidade) para o traço 1, com dados não agrupados exibidos na linha azul e os limites de controle definidos para dados agrupados e não agrupados. As linhas verdes representam a média e os limites de controle para dados agrupados, enquanto as linhas vermelhas correspondem aos mesmos parâmetros para os dados não agrupados. A comparação entre os limites agrupados e não agrupados destaca a vantagem de suavização proporcionada pelo agrupamento, que reduz a influência de outliers e permite uma interpretação mais clara da estabilidade do processo.

Tabela A.1: Medidas estatísticas descritivas para o traço 1: comparação entre dados agrupados e não agrupados

MÉTRICAS	**VALORES**
Média agrupada	0,2863
Média calc_padrão	-0,0002
Mediana agrupada	0,2983
Mediana calc_padrão	0,0000
Moda agrupada	0,3022
Moda calc_padrão	0,0000 (Unimodal)
Variância agrupada	0,1158
Variância calc_padrão	0,1110
Desvio padrão agrupado	0,3403
Desvio padrão calc_padrão	0,3332
Coeficiente de variação agrupado	118,8617%
Coeficiente de variação calc_padrão	-178965,1664%

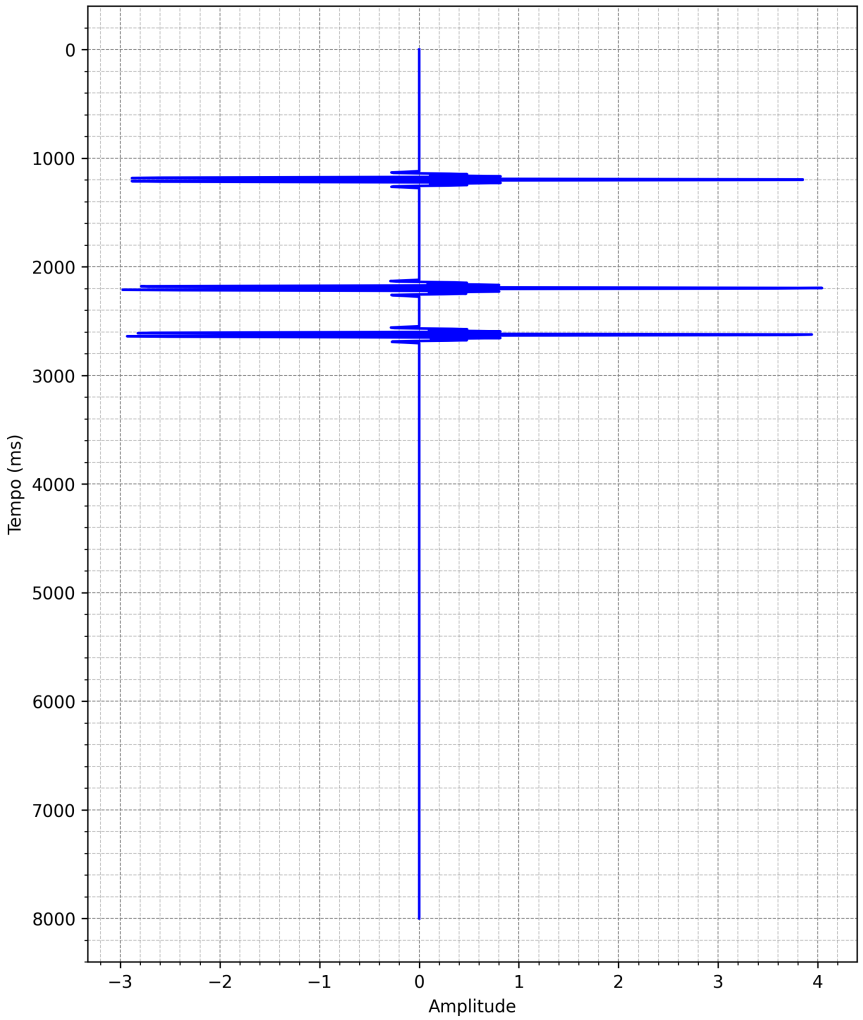


Figura A.1: Traço sísmico 1.

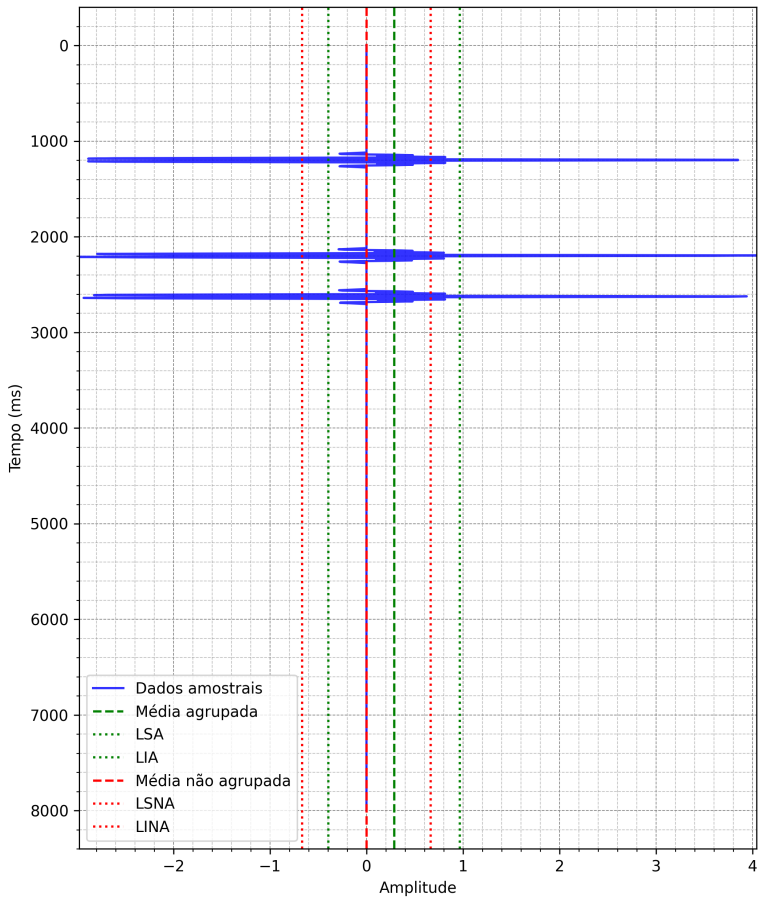


Figura A.2: Análise gráfica de variabilidade para o traço sísmico 1.

A.2.2 Análise do traço 45

Na Tabela A.2, os valores estatísticos calculados para o traço 45 (Figura A.3) são apresentados de forma detalhada, confirmando uma tendência semelhante à observada no traço 1. A média, mediana e moda dos dados agrupados estão alinhadas de forma mais robusta em torno da média central, enquanto as mesmas métricas para os dados não agrupados apresentam uma dispersão mais elevada. Notavelmente, o coeficiente de variação para os dados agrupados está mais estabilizado em comparação com o valor negativo extremo dos dados não agrupados. Essa estabilidade reflete uma maior homogeneidade na distribuição dos dados agrupados, o que é essencial para interpretações mais confiáveis em análises sísmicas. A análise do traço 45 (Figura A.3) também segue a abordagem de agrupamento, com uma avaliação detalhada de suas métricas e do gráfico de controle. A Figura A.4 apresenta o gráfico de controle do traço 45, destacando as mesmas configurações de média e limites de controle para dados agrupados e não agrupados. Aqui, a distinção entre os limites de controle verde e vermelho é igualmente evidente, com os dados agrupados mostrando uma maior contenção das variações dentro dos limites de controle, enquanto os dados não agrupados exibem uma amplitude maior e, consequentemente, uma variabilidade mais ampla.

Tabela A.2: Medidas estatísticas descritivas para o traço 45: comparação entre dados agrupados e não agrupados

MÉTRICAS	**VALORES**
Média agrupada	0,0300
Média calc_padrão	-0,0002
Mediana agrupada	0,0314
Mediana calc_padrão	0,0000
Moda agrupada	0,0315
Moda calc_padrão	0,0000 (Unimodal)
Variância agrupada	0,1134
Variância calc_padrão	0,1110
Desvio padrão agrupado	0,3367
Desvio padrão calc_padrão	0,3332
Coeficiente de variação agrupado	1121,3143%
Coeficiente de variação calc_padrão	-177966,8455%

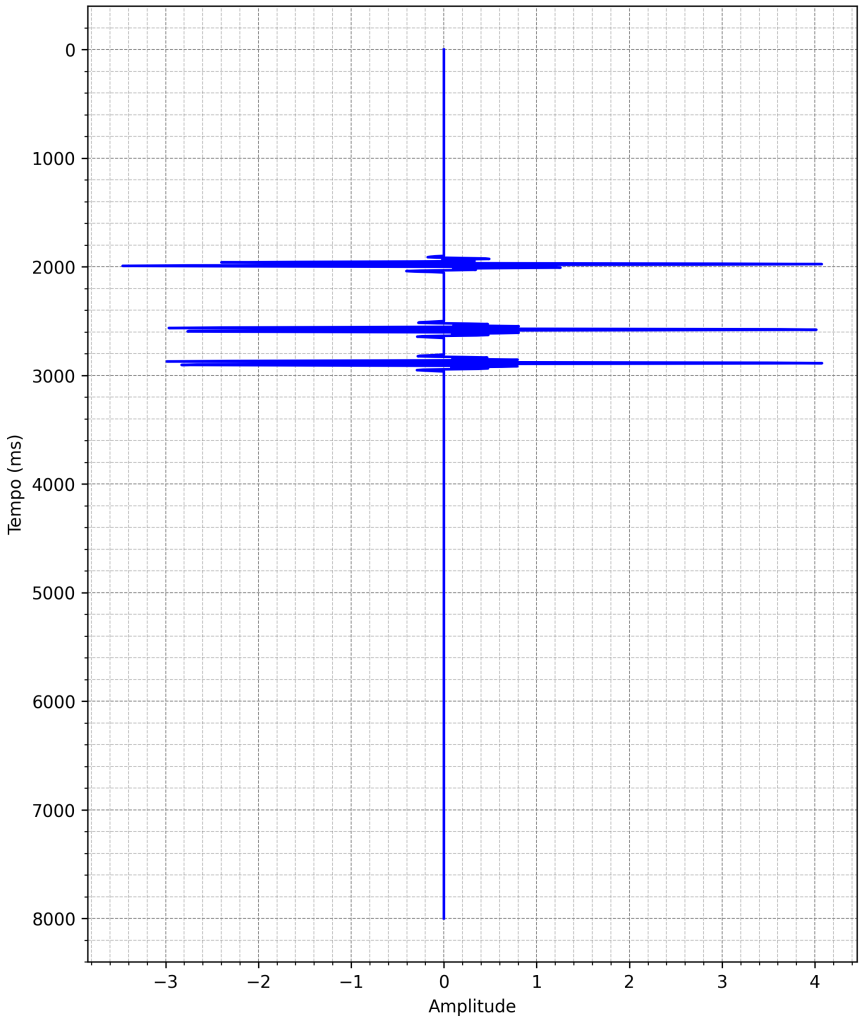


Figura A.3: Traço sísmico 45.

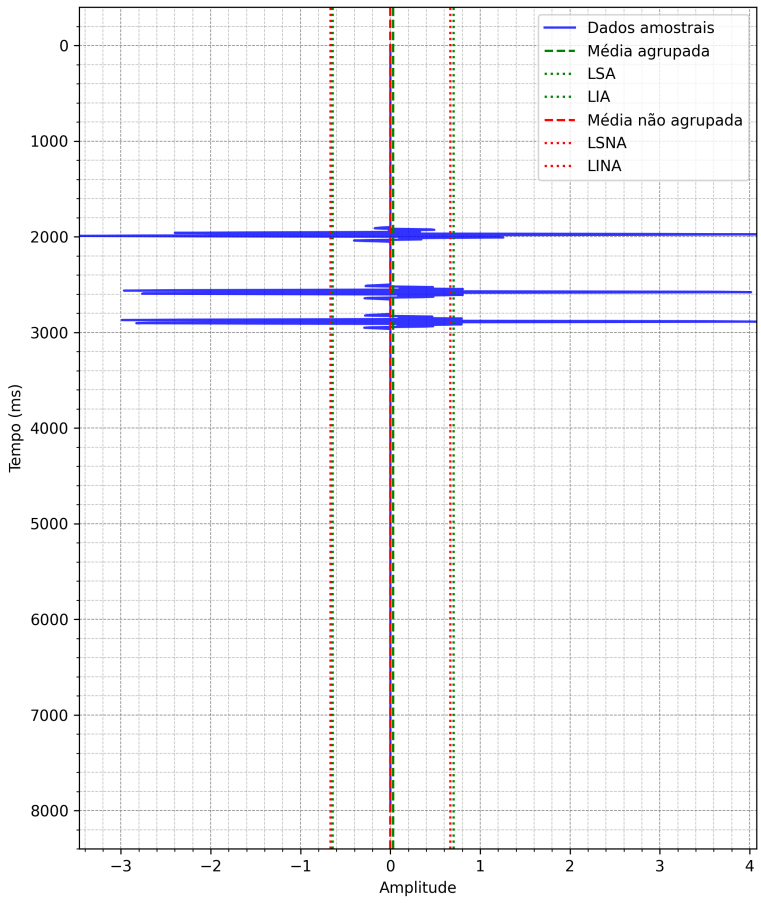


Figura A.4: Análise gráfica de variabilidade para o traço sísmico 45.

A.2.3 Análise do traço 90

A Tabela A.3 resume as métricas para o traço 90 (Figura A.5), onde se observa que os valores de média e mediana dos dados agrupados são mais homogêneos em comparação aos valores não agrupados. O coeficiente de variação para os dados agrupados é moderado e negativo, indicando uma variabilidade controlada e dentro de expectativas, enquanto o coeficiente para os dados não agrupados é de mesmo sinal, mas extremamente elevado, sinalizando um processo menos estável.

Para o traço 90 (Figura A.5), a análise de controle estatístico e variabilidade segue os mesmos princípios, com a Figura A.6 exibindo o gráfico de controle que compara dados agrupados e não agrupados. Observa-se que os dados agrupados permanecem consistentemente dentro dos limites de controle, enquanto os dados não agrupados têm uma maior amplitude de variabilidade, evidenciando o benefício do agrupamento em suavizar as flutuações de amplitude. Essa suavização é crucial nesse tipo de análise, por permitir a interpretação do comportamento do sinal sem o impacto excessivo de ruídos ou variações anômalas e por indicar os pontos de maior concentração dos valores de amplitudes, procedimento essencial para que a identificação de clusters, por meio da técnica de agrupamento k-means++, seja mais rápida e assertiva.

Tabela A.3: Medidas estatísticas descritivas para o traço 90: comparação entre dados agrupados e não agrupados

MÉTRICAS	**VALORES**
Média agrupada	-0,3380
Média calc_padrão	-0,0002
Mediana agrupada	-0,3473
Mediana calc_padrão	0,0000
Moda agrupada	-0,3504
Moda calc_padrão	0,0000 (Unimodal)
Variância agrupada	0,0458
Variância calc_padrão	0,0416
Desvio padrão agrupado	0,2139
Desvio padrão calc_padrão	0,2041
Coeficiente de variação agrupado	-63,2816%
Coeficiente de variação calc_padrão	-116600,7791%

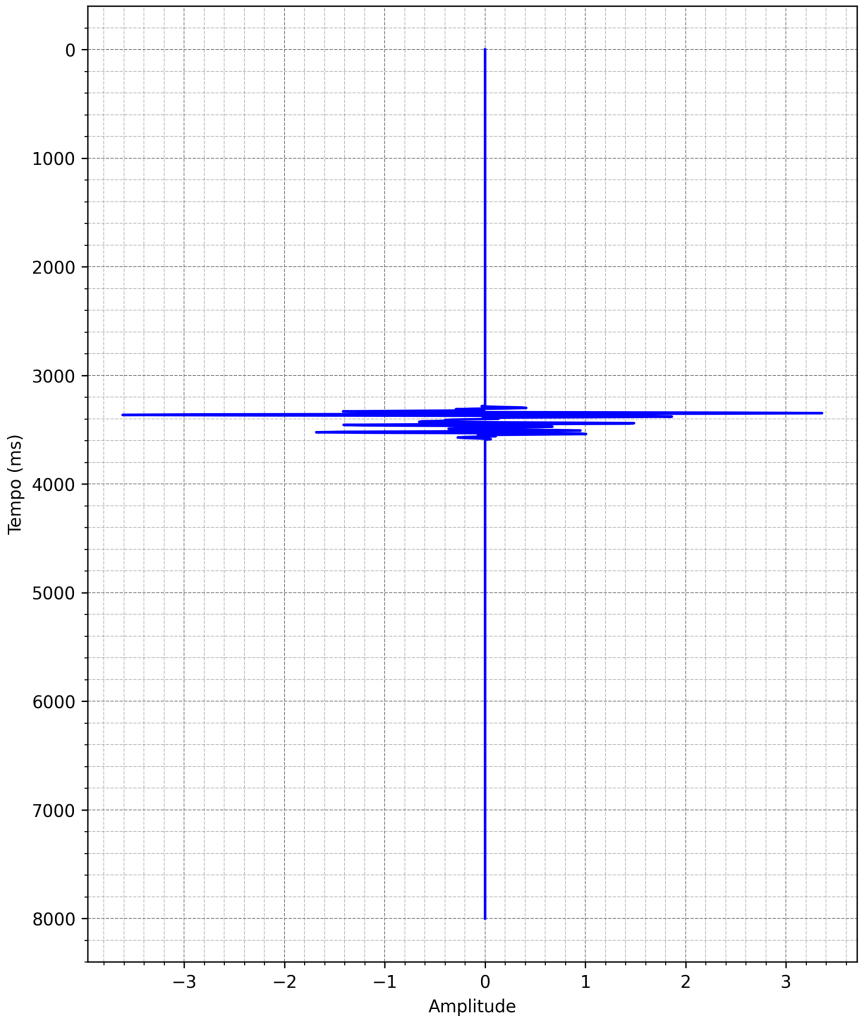


Figura A.5: Traço sísmico 90.

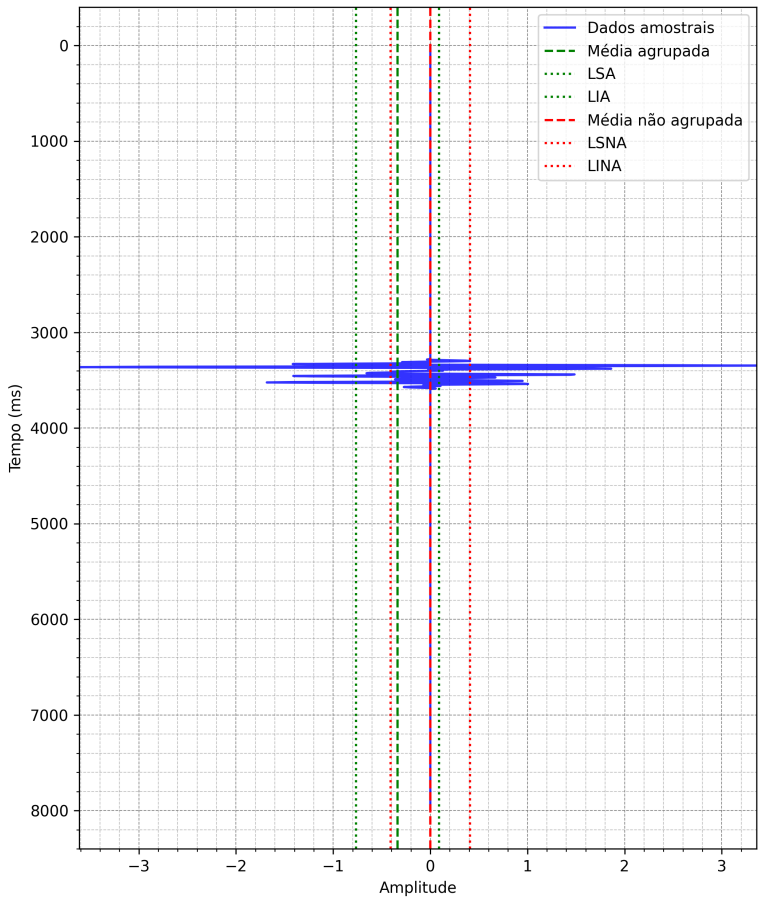


Figura A.6: Análise gráfica de variabilidade para o traço sísmico 90.

O processo de agrupamento aplicado neste estudo de modelagem estatística mostrou-se fundamental para garantir a qualidade e a estabilidade dos dados. Essa abordagem também permitiu a detecção de tendências centrais e variações com maior clareza, minimizando a influência de valores extremos.

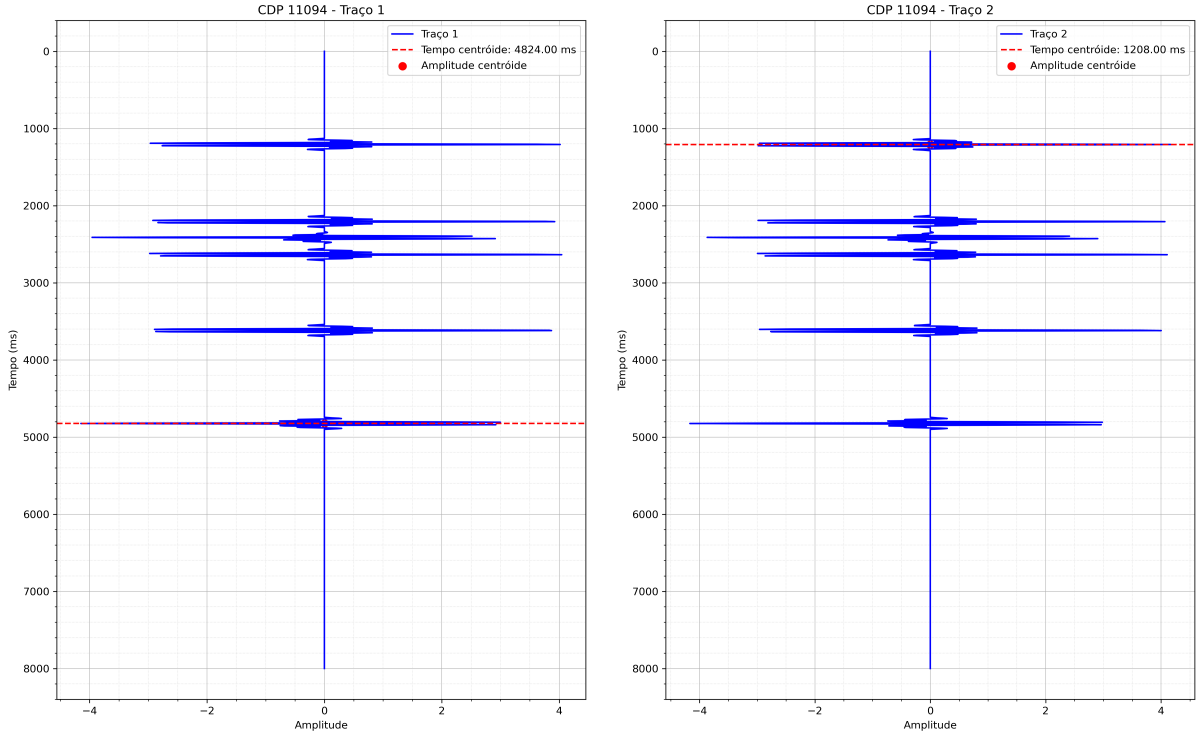
As análises gráficas de variabilidade para os traços 1, 45 e 90 (Figuras A.2, A.4 e A.6) destacaram os benefícios dessa aplicação. É possível visualizar que os dados agrupados permaneceram dentro dos limites de controle, indicando uma variabilidade esperada e controlada, enquanto os dados não agrupados apresentaram maior amplitude de variação, refletindo a forte influência de ruídos. A estabilidade proporcionada pela técnica é crucial em estudos sísmicos, pois a identificação de padrões consistentes é essencial para interpretações precisas e confiáveis de subsuperfície.

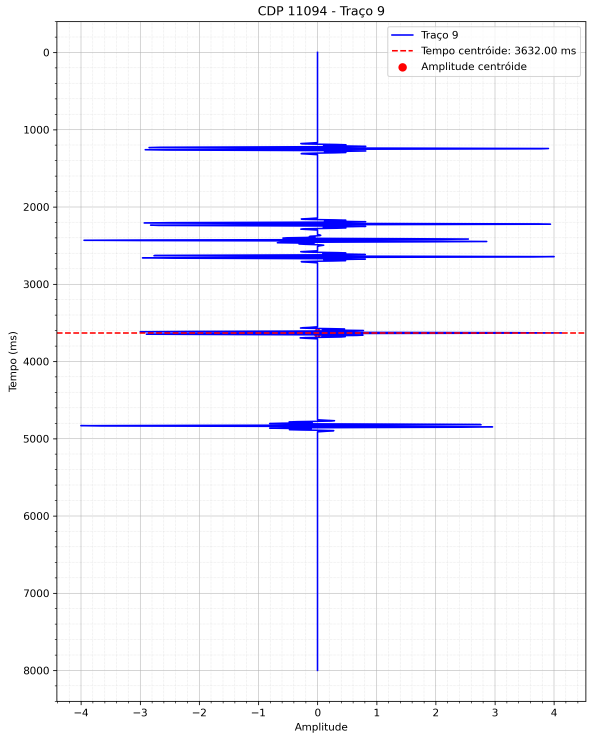
Portanto, a técnica de pré-agrupamento amostral mostrou-se eficaz na otimização dos dados a serem submetidos às técnicas k-means++ e PCA, utilizadas para obtenção automática do campo de velocidade. Esse processo garantiu resultados mais robustos e confiáveis, essenciais para aplicações nas diversas fases do processamento de dados sísmicos.

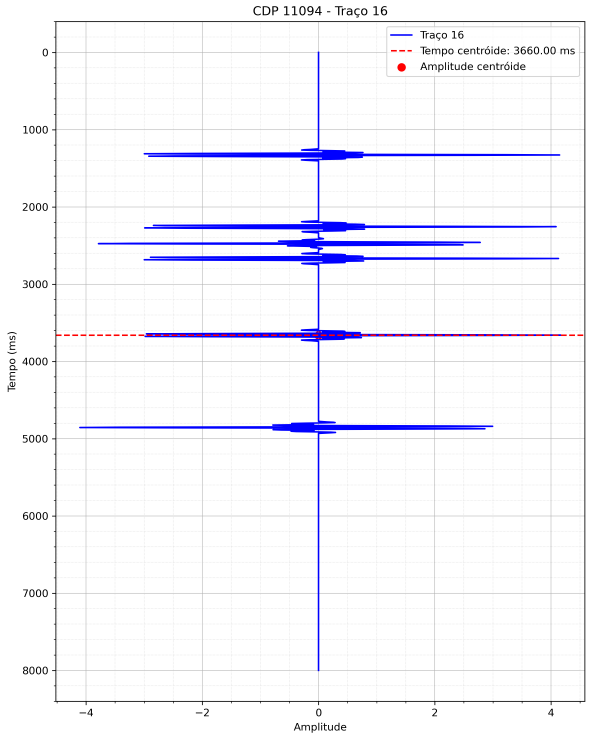
Anexo I

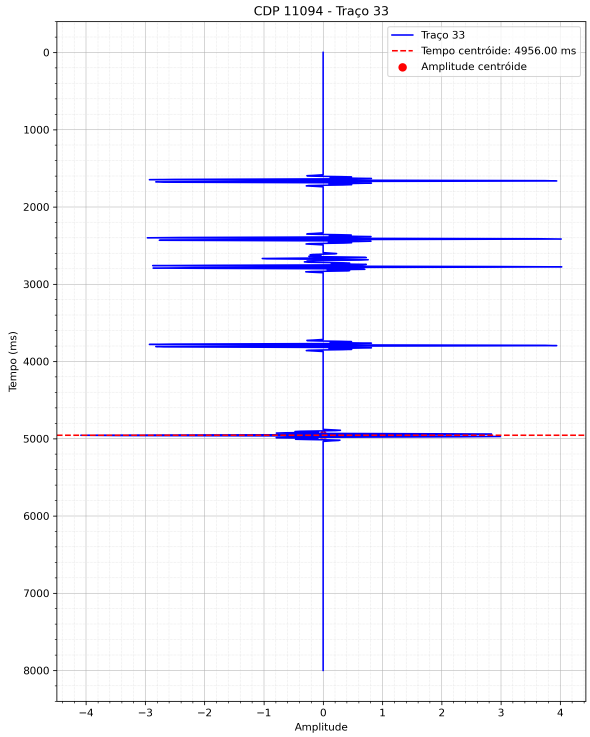
Traços representativos do modelo sintético com eventos múltiplos

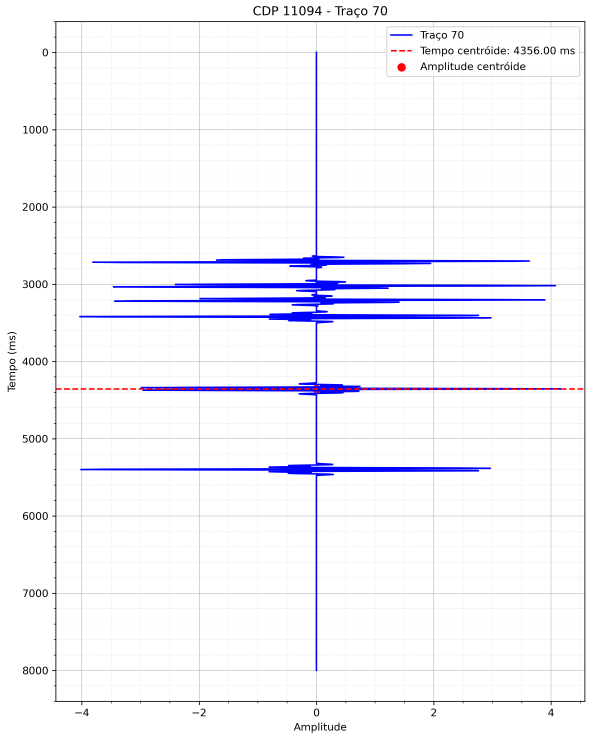
I.1 Traços identificados: CMP (11094) inicial de cobertura máxima

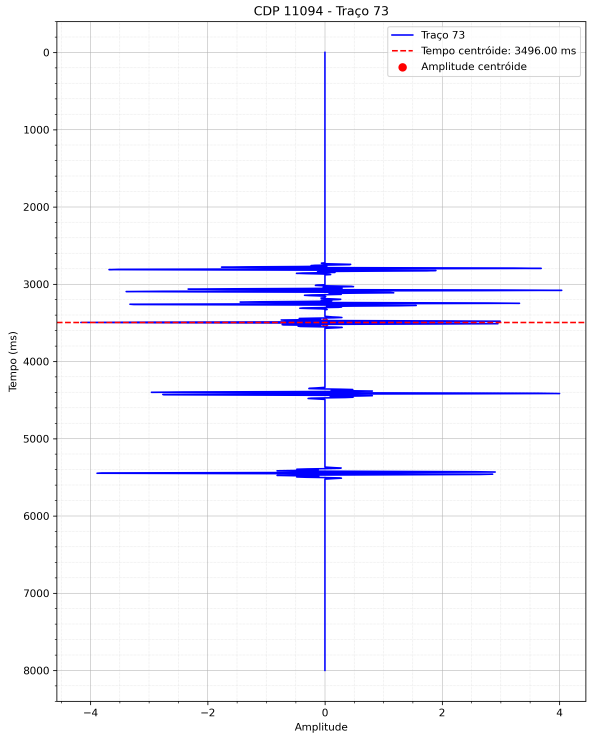


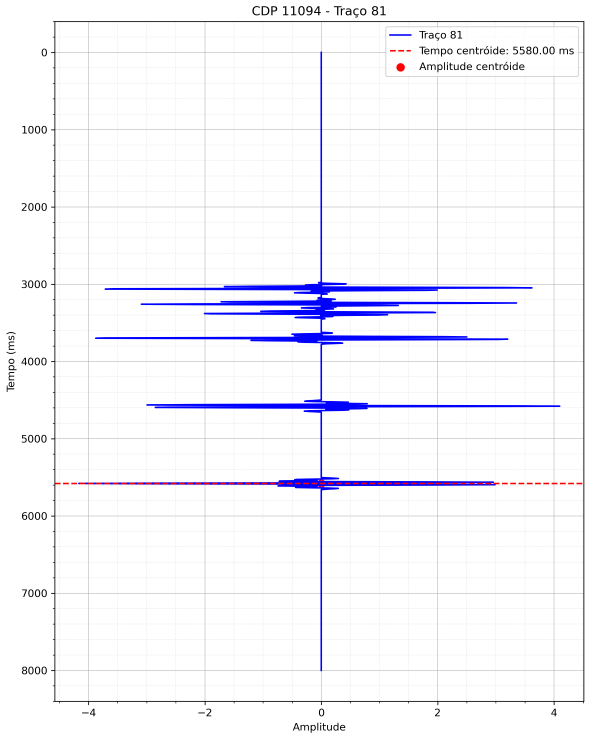

CDP 11094 - Traço 9
Traço 9
Tempo centróide: 3632.00 ms
Amplitude centróide
Tempo (ms)
Amplitude


CDP 11094 - Traço 16
Traço 16
Tempo centróide: 3660.00 ms
Amplitude centróide
Tempo (ms)
Amplitude

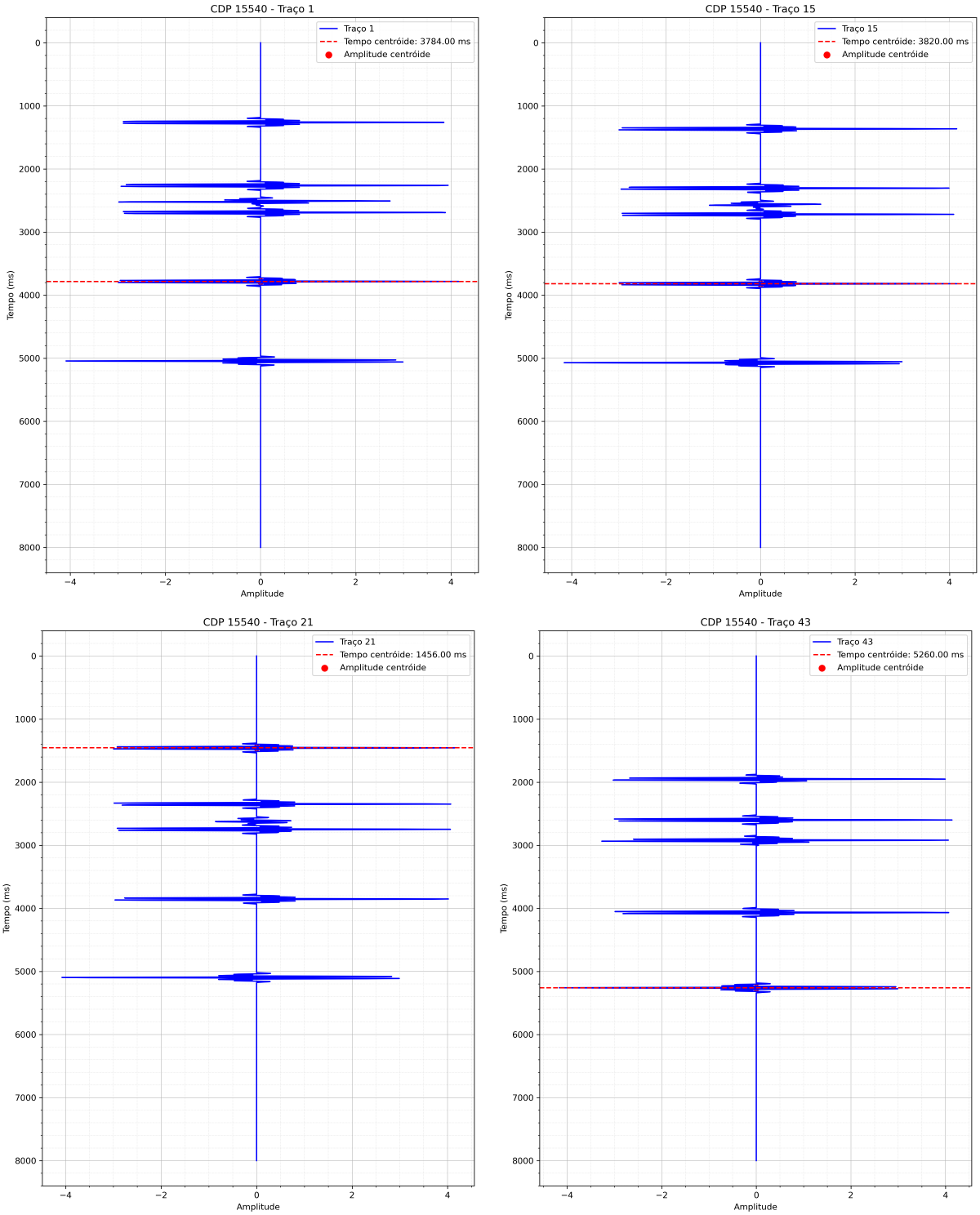

CDP 11094 - Traço 33
Traço 33
Tempo centróide: 4956.00 ms
Amplitude centróide
Tempo (ms)
Amplitude

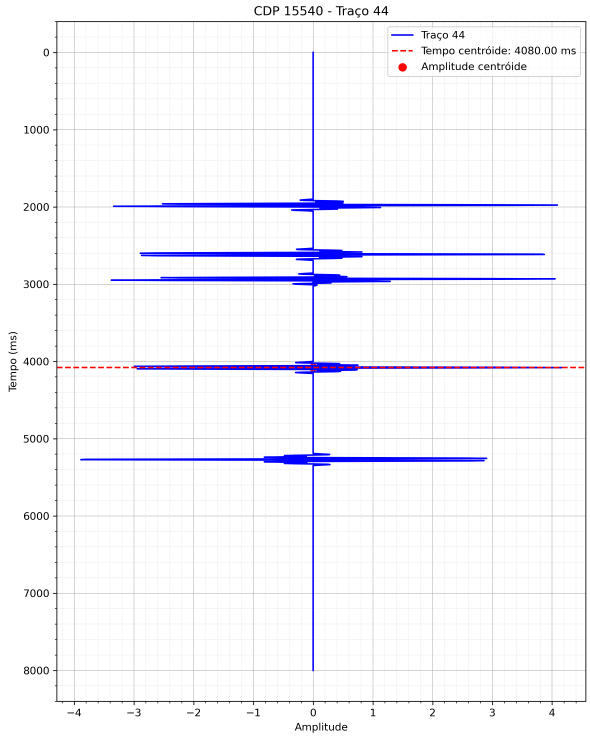

CDP 11094 - Traço 70
Traço 70
Tempo centróide: 4356.00 ms
Amplitude centróide
Tempo (ms)
Amplitude

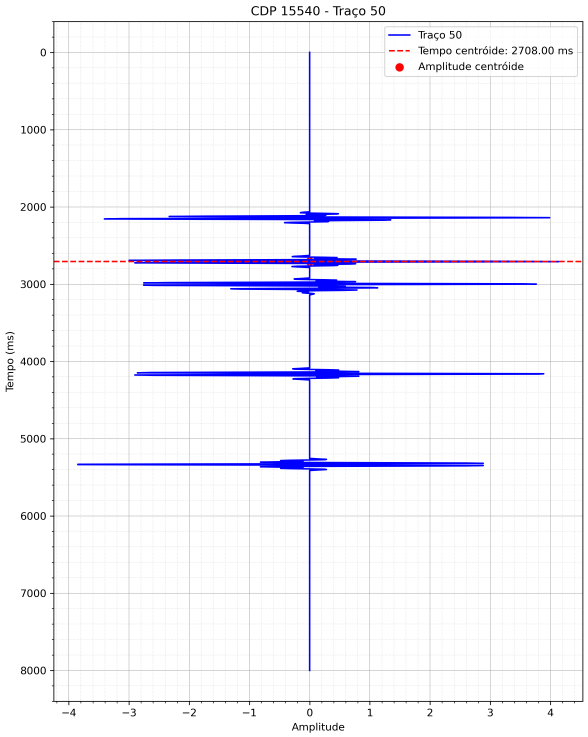

CDP 11094 - Traço 73
Traço 73
Tempo centróide: 3496.00 ms
Amplitude centróide
Tempo (ms)
Amplitude
0
1000
2000
3000
4000
5000
6000
7000
8000
−4
−2
0
2
4

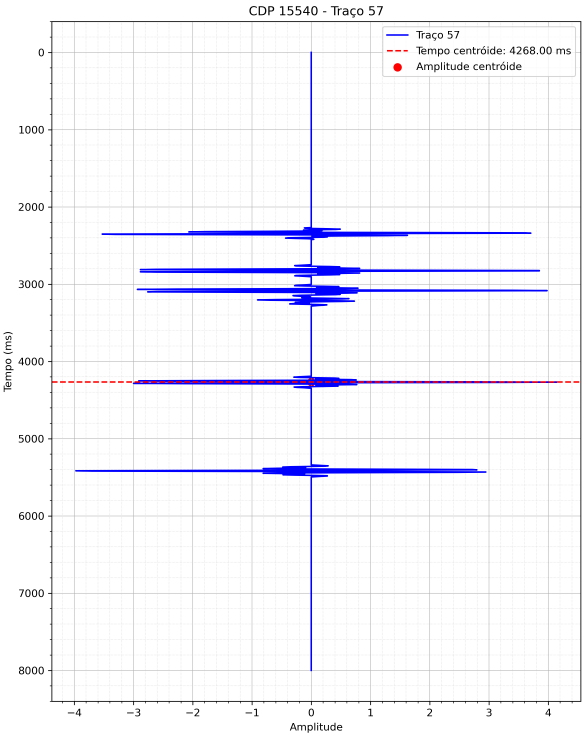

CDP 11094 - Traço 81
Traço 81
Tempo centróide: 5580.00 ms
Amplitude centróide
Tempo (ms)
Amplitude
0
1000
2000
3000
4000
5000
6000
7000
8000
−4
−2
0
2
4

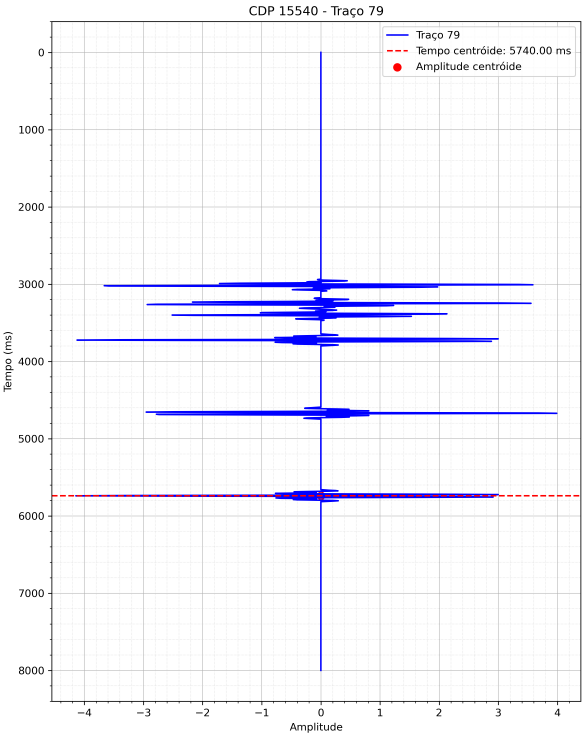
I.2 Traços identificados: CMP (15540) médio de cobertura máxima



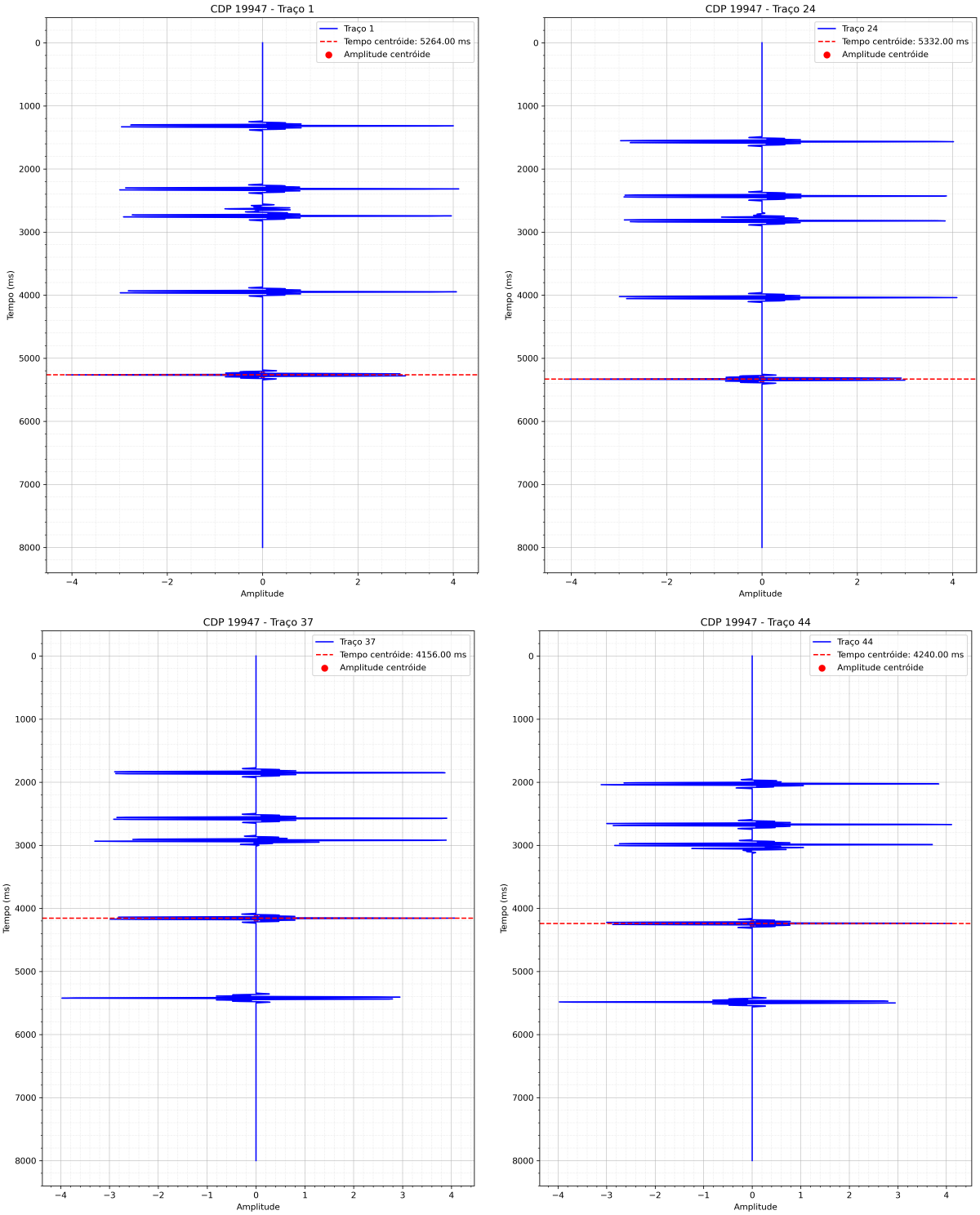

CDP 15540 - Traço 44
Traço 44
Tempo centróide: 4080.00 ms
Amplitude centróide
Tempo (ms)
Amplitude

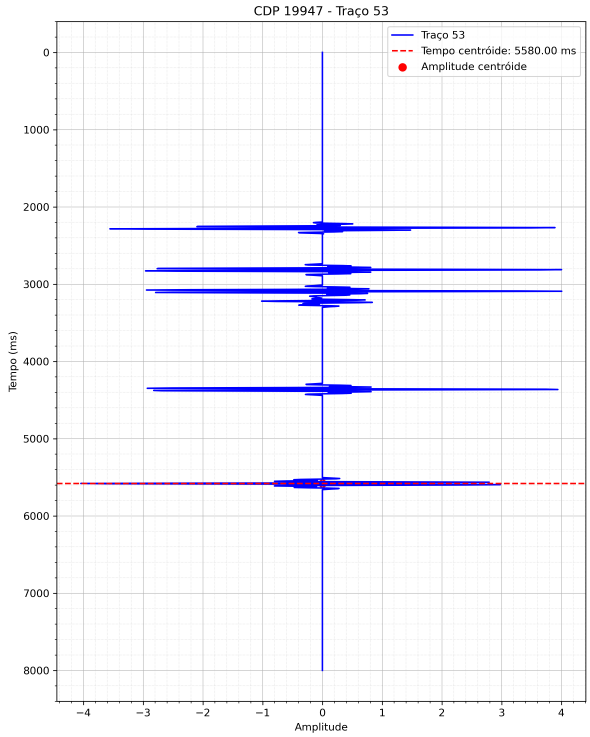

CDP 15540 - Traço 50
Traço 50
Tempo centróide: 2708.00 ms
Amplitude centróide
Tempo (ms)
Amplitude

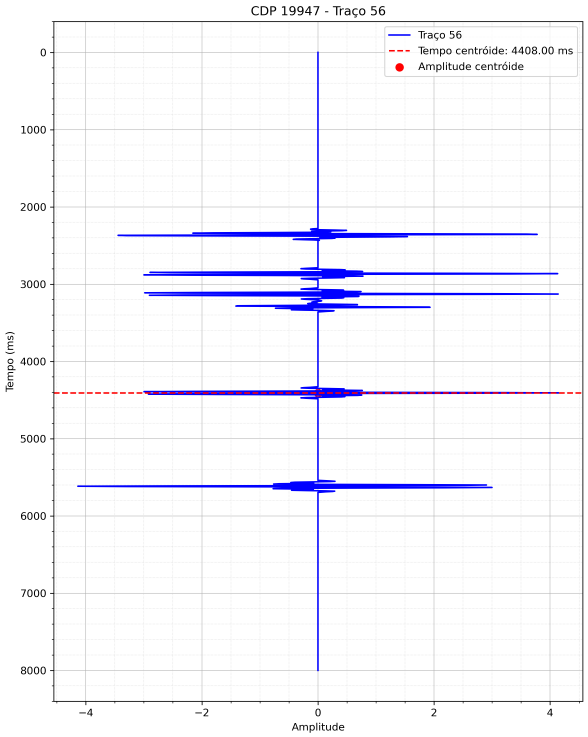

CDP 15540 - Traço 57
Traço 57
Tempo centróide: 4268.00 ms
Amplitude centróide
Tempo (ms)
Amplitude

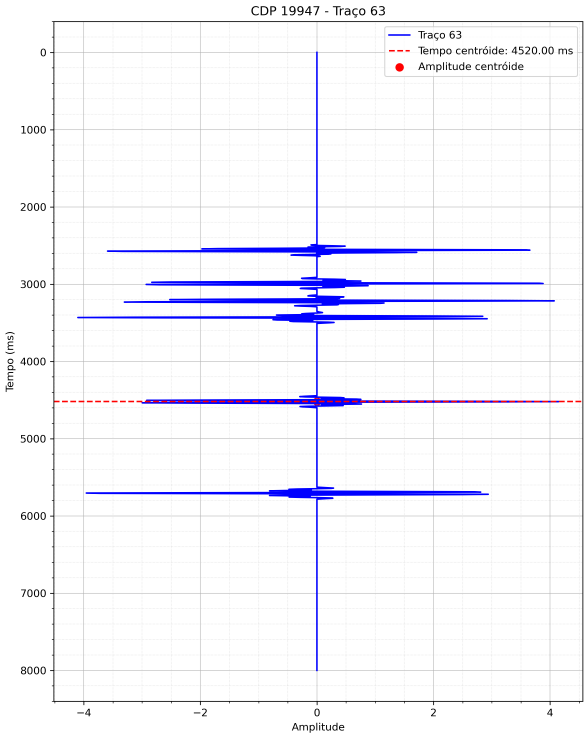

CDP 15540 - Traço 79
Traço 79
Tempo centróide: 5740.00 ms
Amplitude centróide
Tempo (ms)
Amplitude

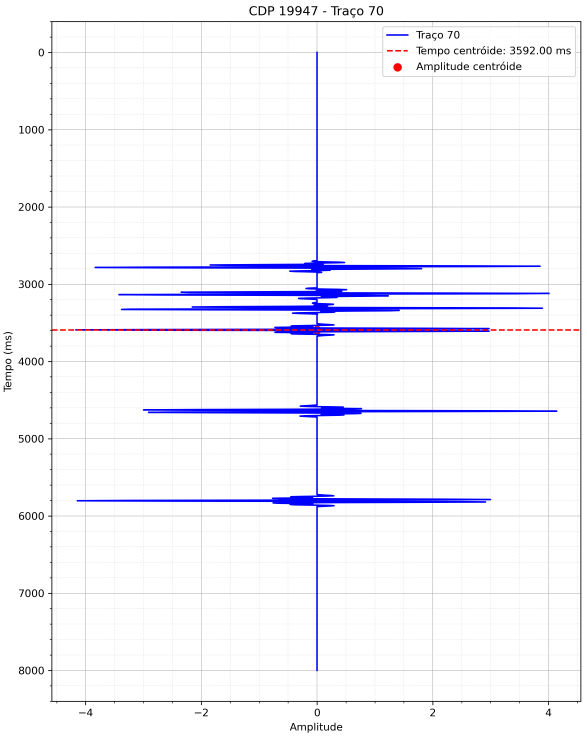
I.3 Traços identificados: CMP (19947) máximo de cobertura máxima



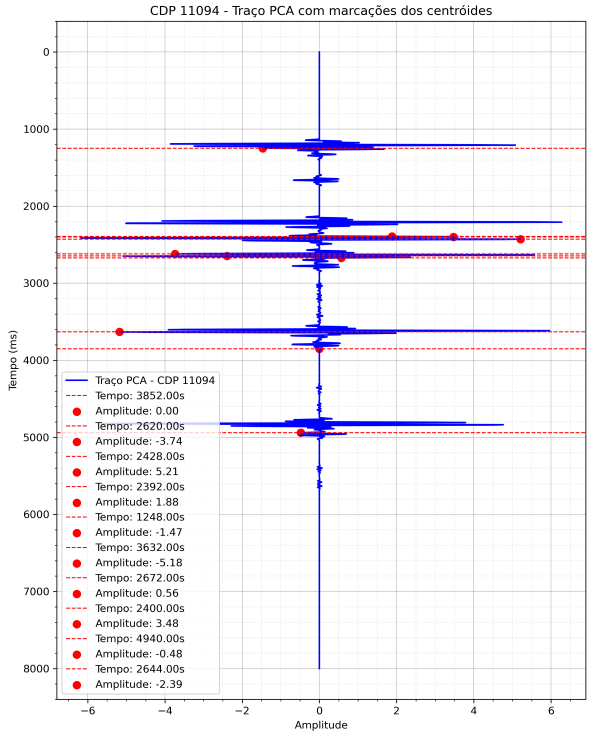

CDP 19947 - Traço 53
Traço 53
Tempo centróide: 5580.00 ms
Amplitude centróide
Tempo (ms)
Amplitude

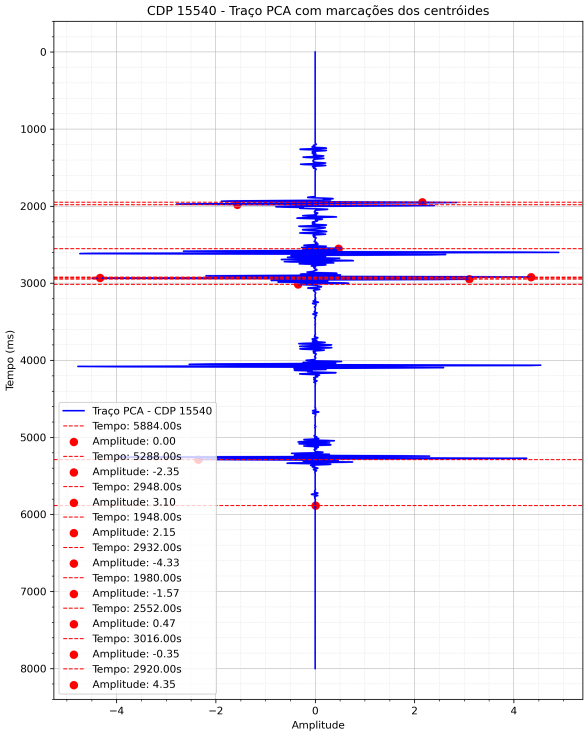

CDP 19947 - Traço 56
Traço 56
Tempo centróide: 4408.00 ms
Amplitude centróide
Tempo (ms)
Amplitude

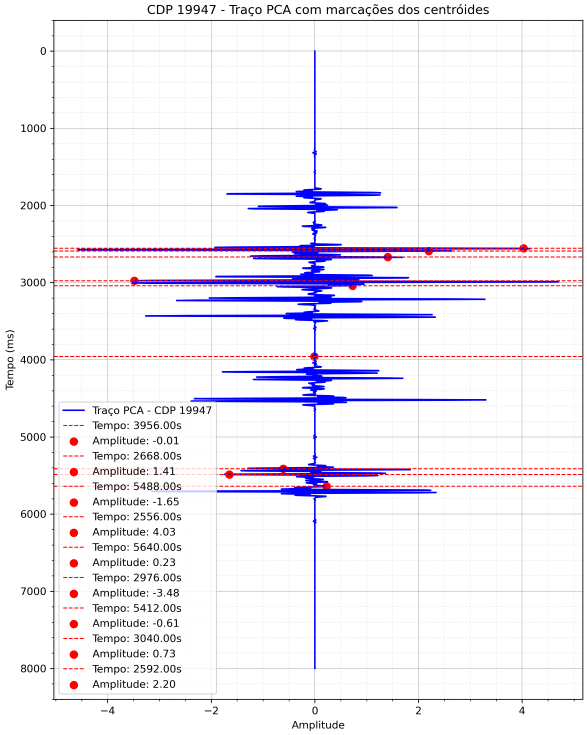

CDP 19947 - Traço 63
Traço 63
Tempo centróide: 4520.00 ms
Amplitude centróide
Tempo (ms)
Amplitude


CDP 19947 - Traço 70
Traço 70
Tempo centróide: 3592.00 ms
Amplitude centróide
Tempo (ms)
Amplitude

I.4 Traços de maior energia formados a partir da aplicação da PCA



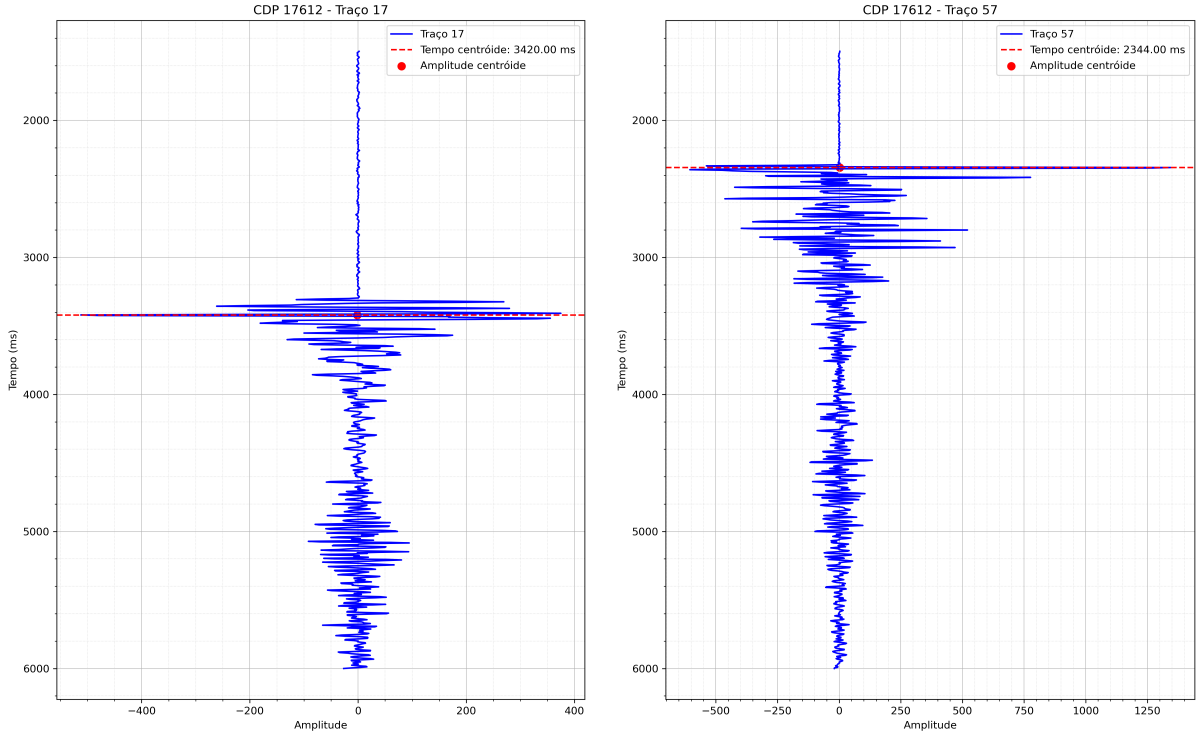


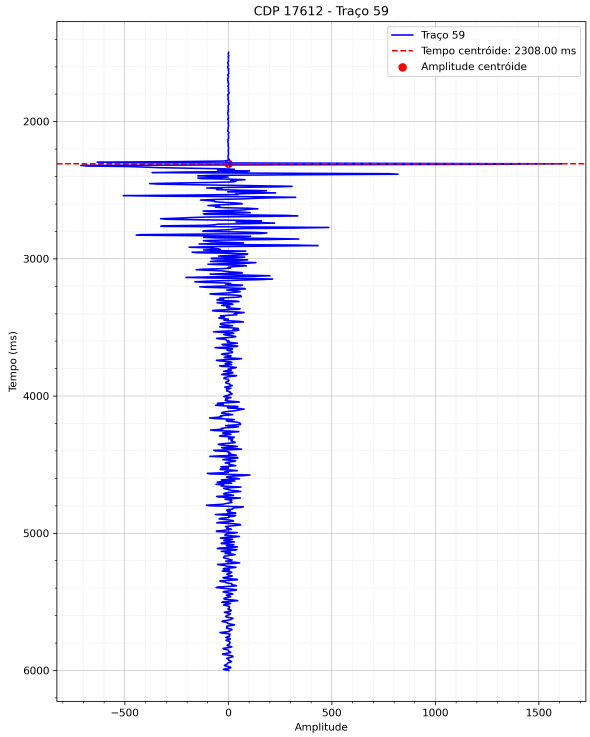


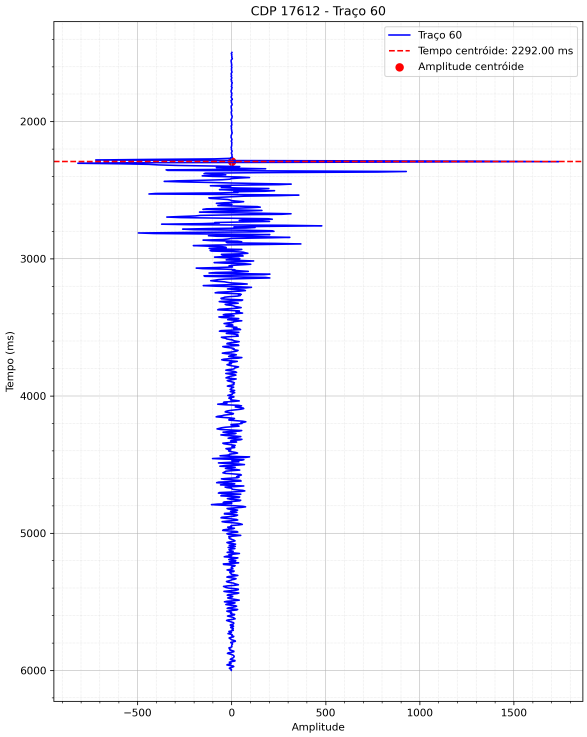
Anexo II

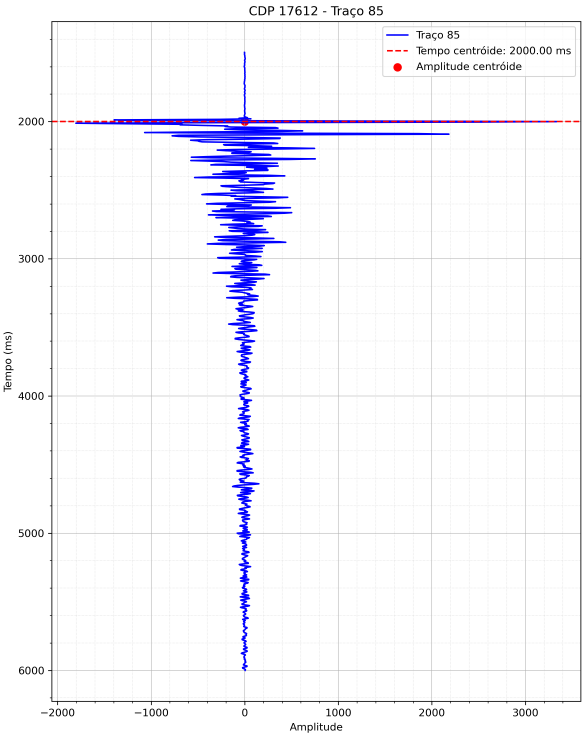
Traços representativos do dado real do Golfo do México

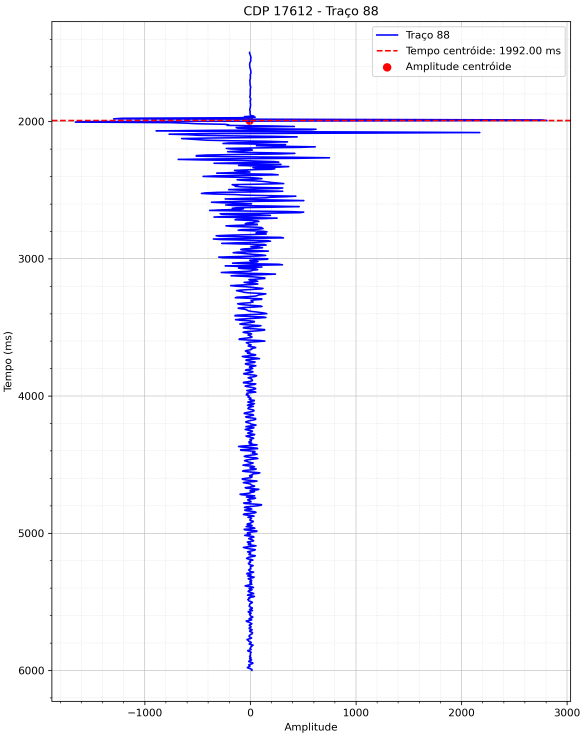
II.1 Traços identificados: CMP (17612) inicial de cobertura máxima



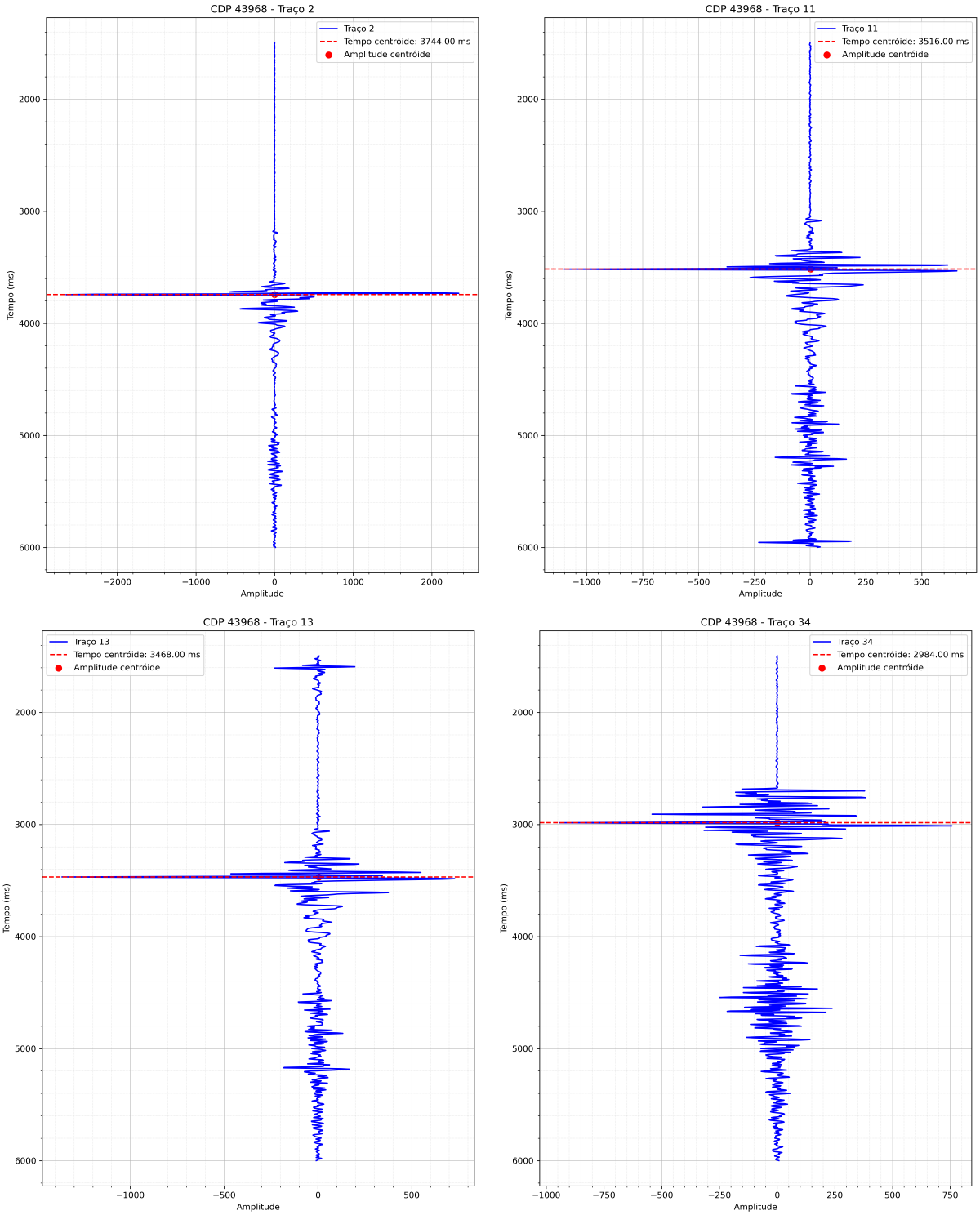

CDP 17612 - Traço 59
Traço 59
Tempo centróide: 2308.00 ms
Amplitude centróide
Tempo (ms)
Amplitude

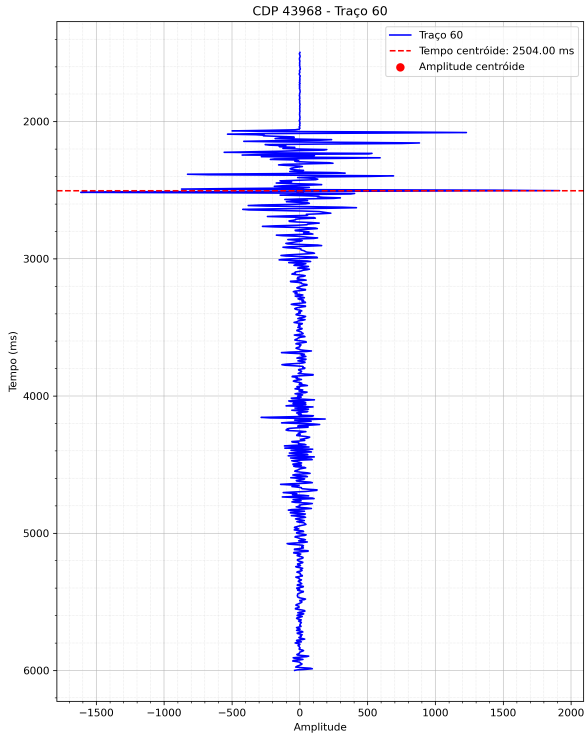

CDP 17612 - Traço 60
Traço 60
Tempo centróide: 2292.00 ms
Amplitude centróide
Tempo (ms)
Amplitude

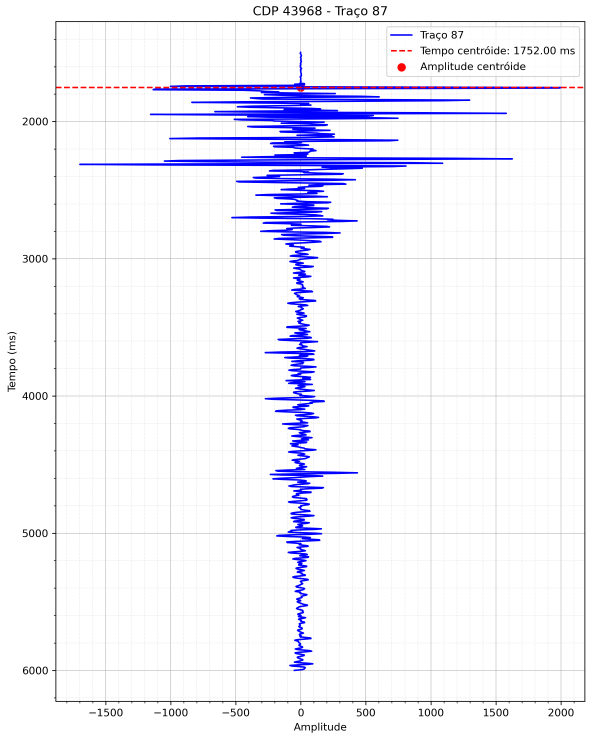

CDP 17612 - Traço 85
Traço 85
Tempo centróide: 2000.00 ms
Amplitude centróide
Tempo (ms)
Amplitude


CDP 17612 - Traço 88
Traço 88
Tempo centróide: 1992.00 ms
Amplitude centróide
Tempo (ms)
Amplitude

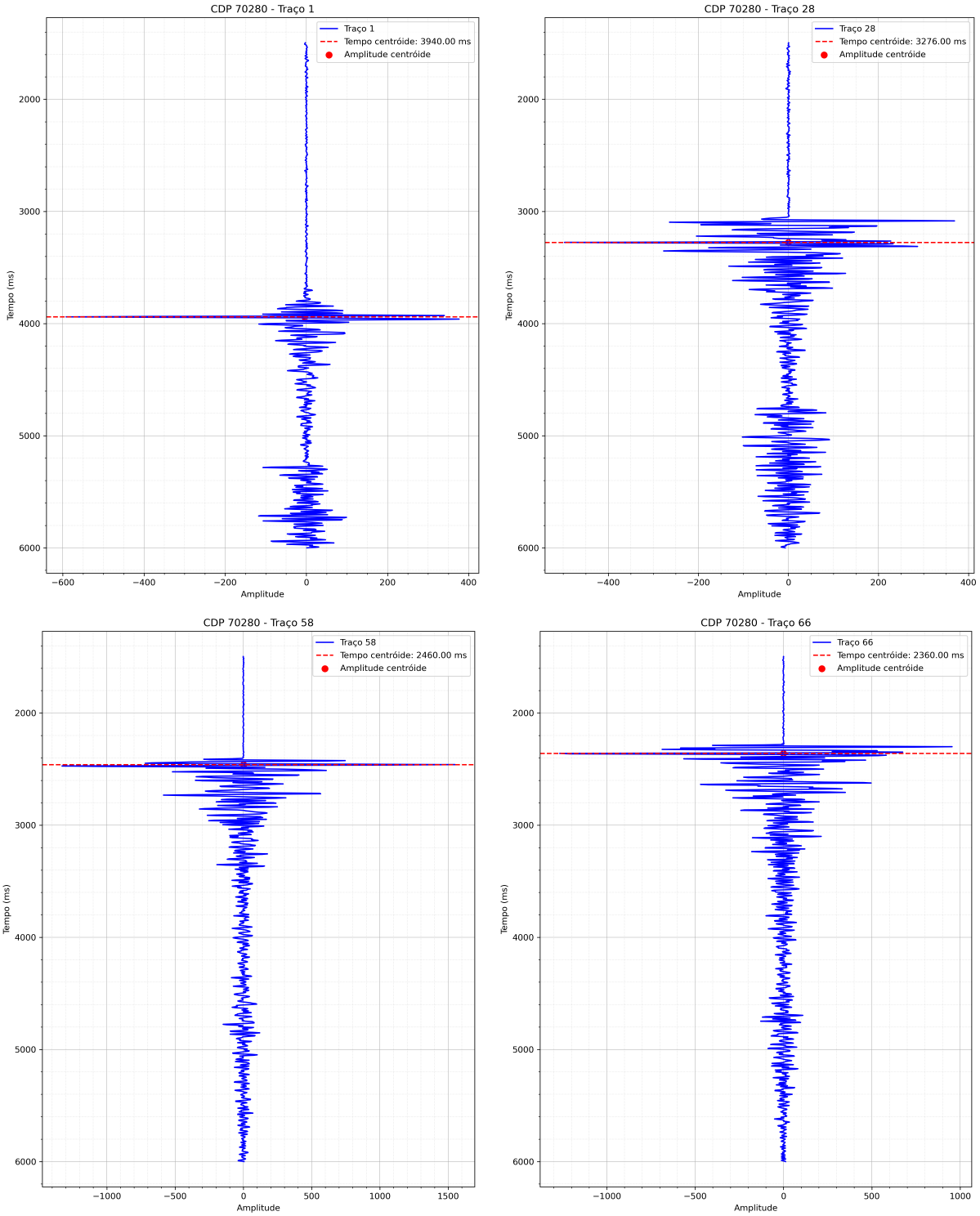
II.2 Traços identificados: CMP (43968) médio de cobertura máxima

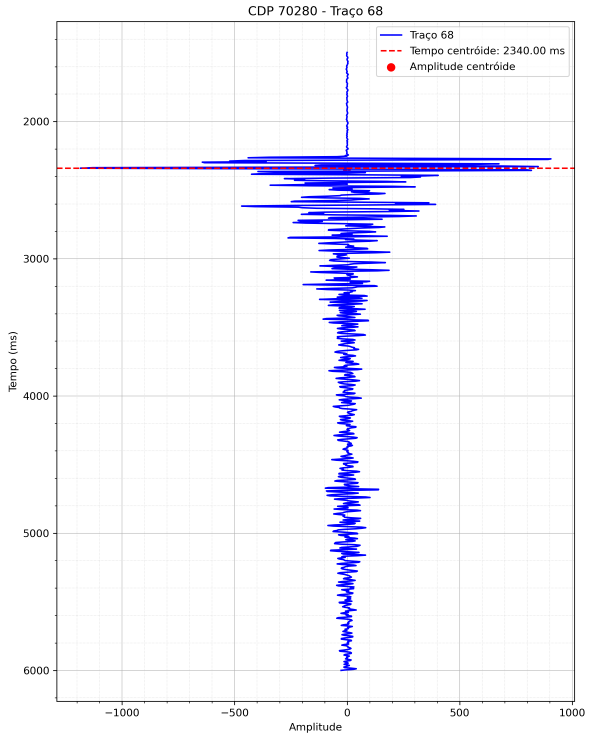


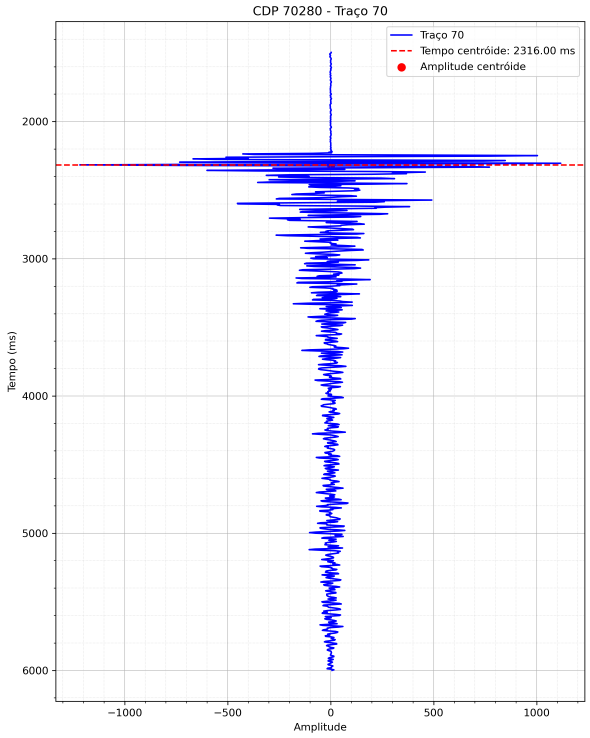

CDP 43968 - Traço 60
Traço 60
Tempo centróide: 2504.00 ms
Amplitude centróide
Tempo (ms)
Amplitude
2000
3000
4000
5000
6000
−1500
−1000
−500
0
500
1000
1500
2000


CDP 43968 - Traço 87
Traço 87
Tempo centróide: 1752.00 ms
Amplitude centróide
Tempo (ms)
Amplitude
2000
3000
4000
5000
6000
−1500
−1000
−500
0
500
1000
1500
2000

II.3 Traços identificados: CMP (70280) máximo de cobertura máxima




CDP 70280 - Traço 68
Traço 68
Tempo centróide: 2340.00 ms
Amplitude centróide
Tempo (ms)
Amplitude
2000
3000
4000
5000
6000
−1000
−500
0
500
1000


CDP 70280 - Traço 70
Traço 70
Tempo centróide: 2316.00 ms
Amplitude centróide
Tempo (ms)
Amplitude
2000
3000
4000
5000
6000
−1000
−500
0
500
1000

II.4 Traços de maior energia formados a partir da aplicação da PCA