# Geração de Perfis Sônicos Sintéticos para Poços Perfurados na Bacia de Sergipe: Uma Análise de Linearidade Usando Regressão

**José Gabriel da S. Carvalho**[1], **Vitor Hugo Simon**[2], **Rodolfo B. de B. Garcia**[1]

[1] Departamento de Computação – Universidade Federal de Sergipe (UFS)
Av. Marcelo Deda Chagas, s/n – Bairro Rosa Elze – 49107-230
São Cristóvão – SE – Brasil

[2]Departamento de Engenharia de Petróleo – Universidade Federal de Sergipe (UFS)

jose.carvalho@dcomp.ufs.br, rodolfo.botto@dcomp.ufs.br,

vitor.h.simon@gmail.com

***Abstract.*** *The prediction of sonic logs is essential in the oil industry, as it allows the estimation of geological properties without direct measurements, which are often costly or unfeasible. This work applies linear regression to generate synthetic sonic logs based on other geophysical logs from the ANP database for the Sergipe Onshore Basin. The methodology uses Pearson correlation to select relevant variables for the prediction. The model achieved an* $r^2$ *of 0.77, indicating its viability to reduce costs by replacing direct measurements with statistical predictions.*

***Resumo.*** *A predição de perfis sônicos é essencial na indústria petrolífera, pois permite estimar propriedades geológicas sem medições diretas, muitas vezes caras ou inviáveis. Este trabalho aplica regressão linear para gerar perfis sônicos sintéticos com base em outros perfis geofísicos da base de dados da ANP na Bacia Sergipe Terrestre. A metodologia utiliza correlação de Pearson para selecionar variáveis relevantes à predição. O modelo obteve* $r^2$ *de 0,77, indicando sua viabilidade para reduzir custos ao substituir medições diretas por predições estatísticas.*

## 1. Introdução

O petróleo vem sendo, ao longo dos séculos, a mais significativa fonte de energia global [Alhelfia et al. 2021], e é uma das principais matérias-primas para uma ampla gama de produtos industriais no mundo contemporâneo. No entanto, sua exploração é um processo complexo, pois seus reservatórios estão localizados no subsolo, tornando impossível a observação direta [Cao et al. 2017]. Para compreender a composição e as propriedades desses reservatórios, é essencial realizar uma caracterização detalhada das formações geológicas por meio da análise de perfis geofísicos, que fornecem informações sobre as propriedades petrofísicas das rochas ao longo do poço perfurado [Archie 1950, Augusto and Martins 2009]. Entre os perfis mais comuns estão os de raios gama, potencial espontâneo, resistividade e sônico [Cranganu and Breaban 2013].

O perfil sônico é fundamental na caracterização dos reservatórios, pois permite a determinação de propriedades essenciais, como porosidade, litologia e propriedades elásticas [Nero et al. 2023]. Por outro lado, esse perfil nem sempre está disponível, seja por falhas instrumentais, más condições do poço, perda de dados durante o armazenamento, erros de calibração ou por uma decisão operacional de não adquiri-lo em poços considerados menos relevantes, especialmente quando o perfil de densidade já está programado, visto que o custo de obtenção, embora não muito elevado, ainda pode ser um fator limitante [Nero et al. 2023, Ogbamikhumi et al. 2020, Pratikna et al. 2022].

A predição de perfis sônicos por inteligência artificial tem se consolidado na indústria petrolífera como solução eficiente para substituir medições ausentes, utilizando perfis acessíveis já disponíveis na documentação dos poços. Alinhada à digitalização do setor e ao quarto paradigma da ciência de dados [Alhelfia et al. 2021, Gressling 2020], essa abordagem reduz custos e tempo, otimiza a caracterização de formações rochosas, melhora a interpretação sísmica e contribui para a redução de impactos ambientais decorrentes de operações adicionais, maximizando o aproveitamento dos dados já coletados.

Segundo [Ellis and Singer 2007] e comprovado neste trabalho, algumas curvas de perfis comumente referenciados em documentação de poços indicam relações lineares com o perfil sônico. Ainda mais, fortes correlações entre esses perfis e o tempo de trânsito podem ser presenciadas, tornando a predição do perfil sônico viável a partir de modelos lineares. Embora o uso de aprendizado de máquina para a geração de perfis sônicos tenha crescido nos últimos anos, há uma escassez de estudos que empregam a regressão linear, especialmente em bacias sedimentares brasileiras. Um exemplo relevante aplicado fora do país é o estudo de [Akinyemi et al. 2023], que avaliou diversos algoritmos de aprendizado de máquina na previsão do perfil sônico na Bacia do Delta do Níger, demonstrando a eficácia de diferentes abordagens mas, por outro lado, a regressão linear foi o pior método.

Diante desse contexto, este trabalho tem como objetivo gerar perfis sônicos sintéticos utilizando o algoritmo de regressão linear, a partir de dados de perfis geofísicos mais acessíveis da Bacia de Sergipe. A proposta parte do indício de que a relação entre os perfis geofísicos relevantes e o tempo de trânsito sônico é linear. Por fim, este trabalho busca contribuir para o avanço das técnicas de predição de perfis sônicos e na identificação geofísica de rochas no cenário brasileiro.

Os resultados obtidos indicaram fortes indícios de linearidade entre curvas de perfis tradicionais, como o NPHI, RHOB e GR, e o sinal sônico, que resultaram em desempenho satisfatório de $R^2$ médio em 0,77, com baios índices de erros e cujo melhor resultado foi de $R^2 = 0{,}88$, superior a resultados de trabalhos prévios. Isso evidencia o potencial da abordagem proposta em contextos onde o perfil sônico não está disponível.

O restante deste artigo está organizado da seguinte forma: a Seção 2 apresenta a fundamentação teórica necessária para o desenvolvimento do trabalho. A Seção 3 descreve a metodologia adotada, incluindo o tratamento dos dados e os procedimentos de modelagem. A Seção 4 apresenta e discute os principais resultados obtidos e a discussão sobre eles. Por fim, a Seção 5 reúne a conclusão parcial do estudo e propõe direções para trabalhos futuros.

## 2. Fundamentação Teórica

Esta sessão é destinada à explicação de termos pontuais abordados neste trabalho, cujo objetivo principal é gerar perfis sônicos sintéticos utilizando o algoritmo de regressão linear, a partir de dados de perfis geofísicos mais acessíveis na Bacia de Sergipe.

### 2.1. Perfis Geofísicos e o Perfil Sônico

As rochas-reservatório são essenciais na exploração de petróleo por apresentarem porosidade e permeabilidade, características que permitem o armazenamento e o fluxo de hidrocarbonetos [Selley 1998, Tissot and Welte 2013]. A identificação dessas rochas é feita por meio de perfis geofísicos, que fornecem informações detalhadas sobre as propriedades das formações em profundidade. Por exemplo, os raios gama fornecem informações sobre a composição mineral das rochas; a densidade estima a porosidade e a composição litológica das rochas que, juntamente com a porosidade de nêutrons, ajudam também na detecção de gás em reservatórios.

O perfil sônico (DT) registra o tempo de trânsito das ondas compressionais no meio rochoso (ondas P), refletindo propriedades elásticas e densidade da formação, além de estimar a porosidade [Ahammod et al. 2014, Ellis and Singer 2007]. Por todos os fatores que dificultam a aquisição de seus dados, outros perfis comumente utilizados durante a perfuração de um poço, como, por exemplo, os raios gama, a densidade, a porosidade de nêutrons, o cáliper, a indução e a própria profundidade, podem servir como descritores de um perfil sônico sintético.

### 2.2. Regressão Linear

Segundo [Montgomery et al. 2021], a regressão linear é uma técnica estatística amplamente utilizada para problemas de predição numérica, especialmente nas áreas de inteligência artificial e aprendizado de máquina. O objetivo principal dessa abordagem é modelar a relação entre uma variável dependente (ou resposta) e uma ou mais variáveis independentes (ou preditoras), assumindo que essa relação seja linear. Ou seja, espera-se que as variações na variável dependente possam ser explicadas como combinações lineares proporcionais das variações das variáveis preditoras.

Matematicamente, o modelo de regressão linear é representado pela Equação 1, onde $y$ é o valor previsto da variável dependente, $\omega_0$ é o intercepto, $\omega_1, \omega_2, \ldots, \omega_n$ são os coeficientes que indicam a influência de cada variável preditora $x_1, x_2, \ldots, x_n$, e $\epsilon$ é o termo de erro que captura a variação não explicada pelo modelo:

$$y = \omega_0 + \omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_n x_n + \epsilon \tag{1}$$

## 3. Metodologia

Nesta seção, apresenta-se a metodologia adotada neste trabalho, estruturada nas subseções a seguir: aquisição e pré-processamento dos dados de poços da Bacia de Sergipe; treinamento, validação e testes do modelo linear.

### 3.1. Aquisição dos Dados

Para a realização deste trabalho, foi utilizado o banco de dados da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), disponível em `https://reate.`

cprm.gov.br/anp/TERRESTRE. A partir desse banco, foram selecionados os arquivos AGP e DLIS de 8 poços da Bacia de Sergipe.

As profundidades máximas variam entre os diferentes poços, com valores entre 654,9 metros e 765,5 metros, refletindo a extensão das medições realizadas em cada local. Por causa da grande quantidade de informação nos arquivos dos poços, a dimensão final das bases será tratada na seção a seguir.

### 3.2. Pré-processamento dos Dados

O pré-processamento dos dados, assim como a escolha dos poços, foi baseado no trabalho de [SOUSA 2024]. Nele, juntamente com a consultoria de especialistas na área da Geologia e Engenharia de Petróleo e Gás, foi realizada a extração dos dados de perfis geofísicos, e etapas como remoção de amostras completas contendo valores nulos ou inválidos (como NaN) foram realizadas para assegurar maior qualidade e confiabilidade nos dados.

Dos dados gerados por [SOUSA 2024], este trabalho manteve apenas as informações sobre curvas de perfis e profundidades. Desta forma, o conjunto de características analisado foi composto por: Profundidade, DCAL, GR, Log10_RESD, DT, RHOB, DRHO, NPHI e PE.

Após o pré-processamento e a exclusão de registros inválidos, a soma de amostras válidas em todos os poços foi de 28.507 amostras contendo as curvas selecionadas e seus respectivos valores de profundidade.

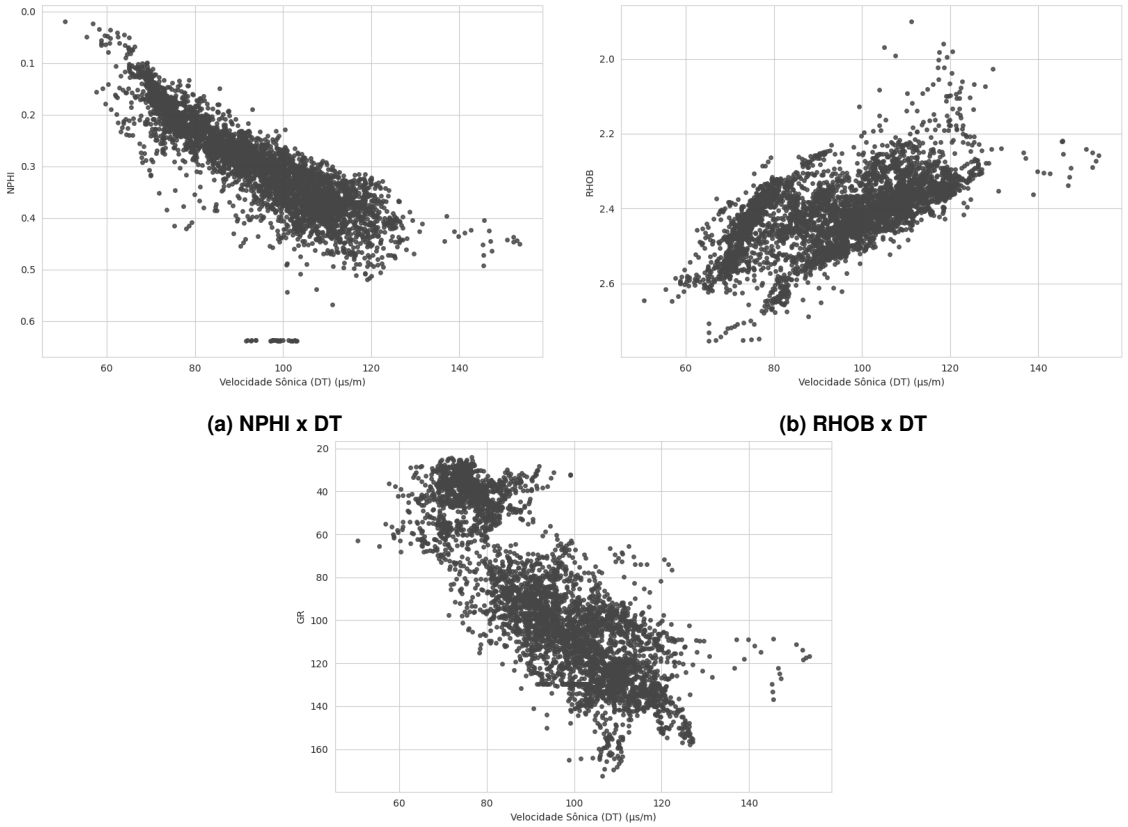
#### 3.2.1. Correlação dos Perfis

**Tabela 1. Correlações de Pearson e Spearman**

| Método | Prof. | DCAL | GR | Log10_RESD | PE | RHOB | DRHO | NPHI |
|---|---|---|---|---|---|---|---|---|
| Pearson | -0.28 | 0.05 | 0.60 | -0.58 | -0.34 | -0.78 | 0.19 | 0.88 |
| Spearman | -0.24 | 0.14 | 0.61 | -0.56 | -0.26 | -0.72 | 0.20 | 0.88 |

Para a análise do banco de dados, foi realizada inicialmente uma avaliação da correlação entre os perfis geofísicos disponíveis. Esse processo permite identificar quais perfis apresentam maior proximidade com o perfil sônico (DT), com a intenção inicial de manter apenas os dados mais relevantes e excluir perfis que pudessem introduzir vieses nas amostras, comprometendo a qualidade dos resultados. Foram aplicados dois métodos estatísticos de correlação (Tabela 1): o coeficiente de correlação de Pearson, que mede relações lineares, e a correlação de Spearman, que captura associações não necessariamente lineares. O uso combinado dessas abordagens permite uma avaliação mais abrangente, fornecendo uma compreensão mais precisa das interdependências entre os perfis e suas relações com o DT.

Ao analisar a Tabela 1 é possível observar a proximidade entre os valores dos coeficientes com relação ao perfil DT, indiciando que a relação dos perfis e da profundidade com o DT seja próxima de tipos lineares. As únicas curvas que obtiveram correlações discrepantes com o perfil sônico foram relacionadas com o DCAL e o PE, demonstrando relações não lineares com o DT.




**(a) NPHI x DT** **(b) RHOB x DT**

**(c) GR x DT**

**Figura 1. Relação linear entre perfis ao DT**

Ainda na Tabela 1, observa-se que os perfis NPHI, RHOB, GR e Log10_RESD apresentam as maiores correlações com o perfil DT e estão na lista de perfis com relação linear. Essa informação é fundamental para orientar o treinamento do modelo, pois indica que esses perfis são os mais relevantes e têm maior potencial para contribuir para um resultado mais preciso no modelo final.

Como forma de confirmação dessas informações, a Figura 1a mostra a distribuição das três maiores correlações com o DT (PHI na Figura 1a, o RHOB na Figura 1b e o GR na Figura 1c). Apesar de não serem linhas retas perfeitas, talvez pela presença de ruídos durante a obtenção dos perfis, ele se aproxima de um formato de reta.

### 3.3. Geração de Perfil Sônico Sintético

Os experimentos aqui executados foram baseados na documentação oficial da biblioteca *Scikit-learn* para Python, que apresenta uma descrição detalhada sobre a partição das bases de dados entre treino, validação e teste, e sobre as métricas estatísticas, incluindo suas definições matemáticas e aplicações. A documentação completa pode ser acessada em: `https://scikit-learn.org/stable/modules/model_evaluation.html`.

Para garantir uma avaliação mais robusta e generalizável do modelo, foi adotada a técnica de Leave-One-Out Cross-Validation (LOOCV). Nesse método, dos 8 poços disponíveis, em cada uma das 8 iterações, um poço é reservado exclusivamente para teste,

enquanto os 7 restantes são utilizados para treinamento. Ao final, os resultados obtidos em cada iteração são agregados por meio da média das métricas de desempenho. Essa abordagem permite que todos os poços participem tanto da fase de treinamento quanto da de teste, promovendo uma avaliação mais imparcial e abrangente da capacidade preditiva do modelo.

Em cada iteração foi aplicada a inclusão progressiva de preditores. A sequência de prioridades dos preditores foi baseada na correlação de Pearson calculada no pré-processamento dos dados (Tabela 1). A sequência foi composta, da maior correlação à menor, por: NPHI, RHOB, GR, Log10_RESD, PE, Profundidade, DRHO e DCAL. Essa priorização considera a proximidade da relação, independentemente de ser direta ou inversamente proporcional.

### 3.3.1. Métricas Estatísticas

As métricas MAE (Equação 2), MSE (Equação 3), RMSE (Equação 4) e $R^2$ (Equação 5) são amplamente utilizadas para avaliar o desempenho de modelos de regressão. O Erro Absoluto Médio (MAE) calcula a média dos desvios absolutos entre valores preditos e reais, sendo de fácil interpretação por estar na mesma unidade da variável de saída. O Erro Quadrático Médio (MSE) utiliza os desvios ao quadrado, penalizando mais fortemente erros maiores e capturando melhor a variabilidade. A Raiz do Erro Quadrático Médio (RMSE) é a raiz do MSE, combinando a sensibilidade a grandes erros com a interpretação facilitada por manter a unidade original. Por fim, o coeficiente de determinação ($R^2$) quantifica a proporção da variância explicada pelo modelo, indicando o quão bem os dados se ajustam à regressão. Quanto menores os valores de MAE, MSE e RMSE, e mais próximo de 1 for o $R^2$, melhor o desempenho do modelo.

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} |y_i - \hat{y}_i| \tag{2}$$

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2 \tag{3}$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2} \tag{4}$$

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{5}$$

## 4. Resultados e Discussões

Os resultados obtidos após as 8 iterações do LOOCV, apresentados na Tabela 2, demonstram que a inclusão progressiva dos preditores selecionados — sejam curvas de perfil ou informações de profundidade — melhorou continuamente o desempenho do modelo na geração do perfil sônico sintético. Isso se refletiu em ganhos nas métricas estatísticas, indicando maior precisão e qualidade na predição.

**Tabela 2. Resultados dos Testes**

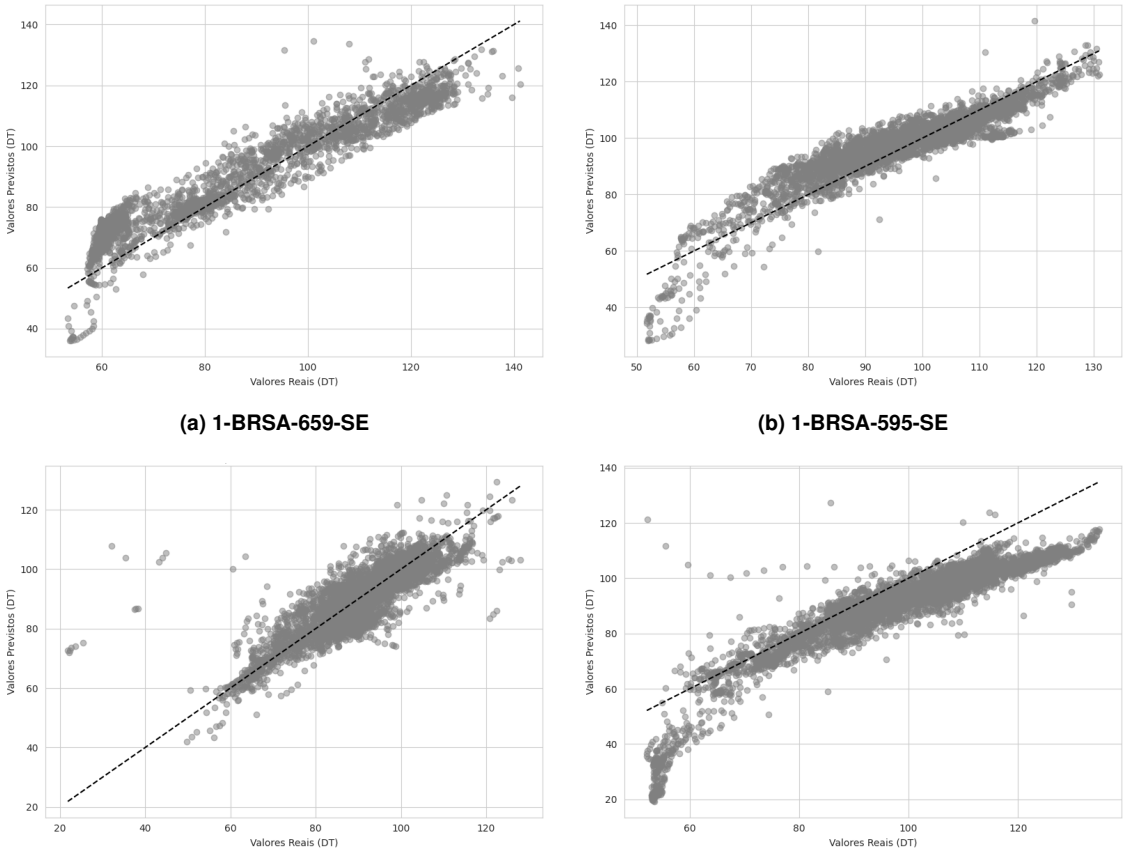
| Perfis Utilizados | $R^2$ | RMSE | MSE | MAE |
|---|---|---|---|---|
| NPHI | 0.69 | 9.24 | 89.55 | 7.29 |
| NPHI + RHOB | 0.72 | 8.72 | 78.74 | 6.87 |
| NPHI + RHOB + GR | 0.74 | 8.83 | 82.70 | 6.98 |
| NPHI + RHOB + GR + Log10_RESD | 0.73 | 8.93 | 84.96 | 7.03 |
| NPHI + RHOB + GR + Log10_RESD + PE | 0.73 | 8.96 | 85.39 | 7.07 |
| NPHI + RHOB + GR + Log10_RESD + PE + Profundidade | 0.76 | 8.13 | 68.71 | 6.40 |
| NPHI + RHOB + GR + Log10_RESD + PE + Profundidade + DRHO | 0.77 | 8.07 | 67.42 | 6.33 |
| NPHI + RHOB + GR + Log10_RESD + PE + Profundidade + DRHO + DCAL | 0.77 | 8.01 | 67.22 | 6.29 |

**Tabela 3. Resultados Individuais dos Poços**

| Poço | $R^2$ | RMSE | MSE | MAE |
|---|---|---|---|---|
| 1-BRSA-551-SE | 0.76 | 7.73 | 59.82 | 5.94 |
| 1-BRSA-574-SE | 0.63 | 11.27 | 126.91 | 8.99 |
| 1-BRSA-595-SE | 0.84 | 5.70 | 32.54 | 4.34 |
| 1-BRSA-605-SE | 0.78 | 6.57 | 43.14 | 5.40 |
| 1-BRSA-659-SE | 0.88 | 7.79 | 60.71 | 6.35 |
| 1-BRSA-689-SE | 0.82 | 10.23 | 104.62 | 8.47 |
| 1-BRSA-696-SE | 0.76 | 7.73 | 59.82 | 5.94 |
| 1-BRSA-698-SE | 0.68 | 7.09 | 50.24 | 4.90 |

Com apenas um preditor, o coeficiente de determinação ($R^2$) médio inicial foi de 0,69, aumentando gradualmente até atingir 0,77 com a utilização de todos os perfis disponíveis. Esses valores já são maiores que de trabalhos prévios, como o de [Akinyemi et al. 2023], em que os valores da regressão linear ficaram próximos a 0,65. Outras observações relacionadas ao $R^2$ são: (a) o grande ganho quando incluídos o RHOB e o GR, mostrando a relação linear entre eles e o DT; (b) a estagnação quando colocados a resistividade e o PE, o que pode mostrar redundância entre eles; (c) o grande ganho ao incluir a profundidade, mostrando que pode ser um dado dependente de outra curva anteriormente analisada; (d) o pequeno ganho das últimas iterações mostram uma relação menos linear, mas que pode agregar qualidade se forem tratadas de forma adequada.

Tabela 3, os resultados obtidos ao longo das oito iterações mostram que o poço 1-BRSA-659-SE obteve $R^2$ médio de 0,88, muito próximo do melhor resultado obtido por [Akinyemi et al. 2023] com o CatBoost. Em geral, foi obtido desempenho satisfatório, com $R^2$ médio de 0,77, além de valores médios de MAE = 6,29, MSE = 67,22 e RMSE = 8,01. A performance variou entre os poços, com desvios mais expressivos concentrados em faixas específicas de profundidade, como nas seções iniciais ou finais, conforme o poço analisado. Nos gráficos de dispersão (Figura 2c), a proximidade dos pontos em relação à linha de referência indica boa correspondência entre os valores reais e os preditos. Já nos gráficos de linha (Figura 3a), a linha cinza representa o perfil real e a linha




**(a) 1-BRSA-659-SE** **(b) 1-BRSA-595-SE**

**(c) 1-BRSA-698-SE** **(d) 1-BRSA-574-SE**

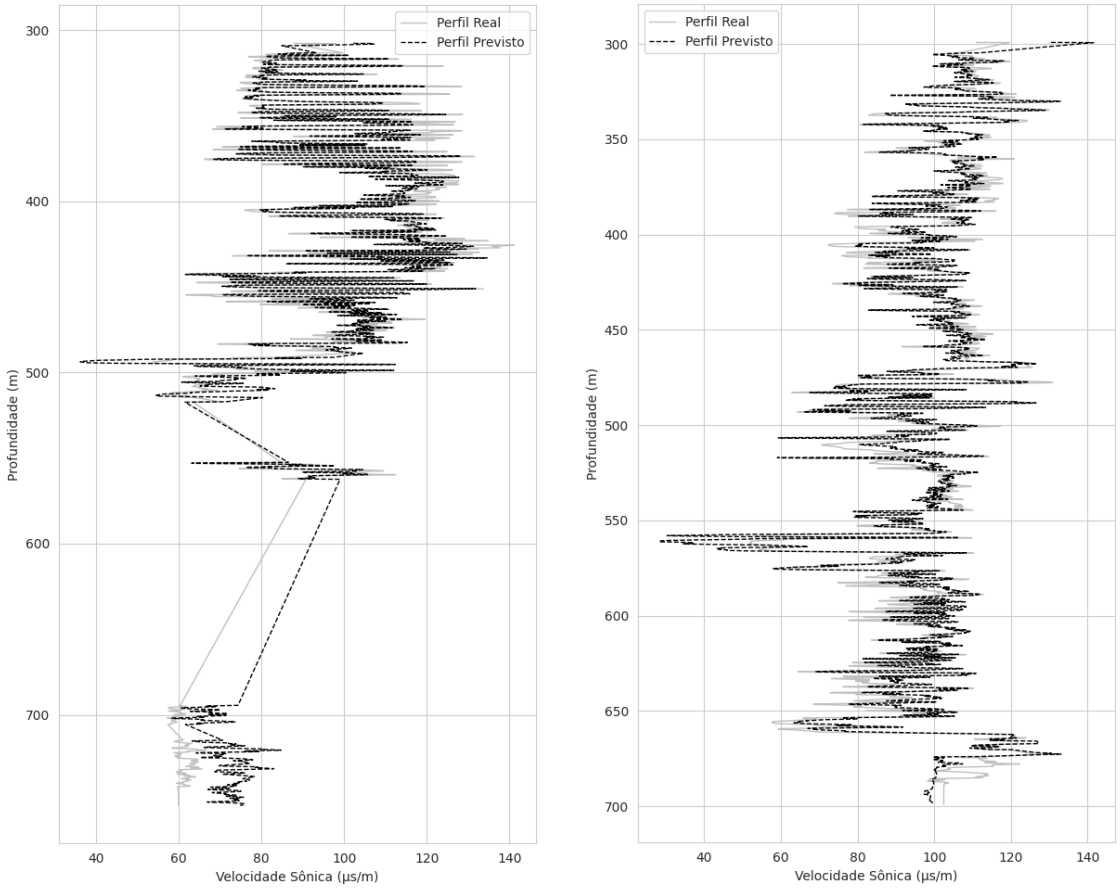
**Figura 2. Previsão dos Poços**

preta pontilhada indica o valor predito — quanto maior a sobreposição entre elas, maior a acurácia da predição.

Ainda na Tabela 2, observa-se que a inclusão dos perfis Log10_RESD e PE resultou em uma leve queda nas métricas médias de desempenho. No entanto, essa redução foi majoritariamente causada pelo poço 1-BRSA-574-SE, que apresentou uma deterioração significativa nas métricas. Enquanto os demais poços mantiveram desempenho estável, o $R^2$ desse poço caiu de 0,68, no conjunto anterior (NPHI, RHOB e GR), para 0,59 com os novos perfis. Esse comportamento sugere que o poço 1-BRSA-574-SE pode ser um outlier em relação aos demais, afetando negativamente a média geral das métricas. Essa discrepância é evidenciada na Tabela 2, onde os resultados desse poço destoam do padrão observado nos demais.

## 5. Conclusão Parcial e Trabalhos Futuros

Este estudo avaliou o uso da regressão linear na predição do perfil sônico (DT) em poços da Bacia de Sergipe, obtendo desempenho médio satisfatório ($R^2 = 0{,}77$) e com $R^2$ obtido pelo poço 1-BRSA-659-SE de 0,88. Os bons resultados mostram a relação linear entre alguns perfis e o DT, especialmente os NPHI, RHOB e GR. Como visto na Figura 1a, as distribuições aparentam ser prejudicadas por ruídos que merecem ser melhor pesquisado, buscando se aproximar ainda mais da distribuíção em forma de linha reta. Outra




**(a) 1-BRSA-659-SE** **(b) 1-BRSA-C595-SE**

**Figura 3. Curva Real x Prevista**

observação que vale a pena ser investigada futuramente é a relação entre a profundidade e a curva DT, assim como relações multivariadas entre os preditores, a fim de garantir modelos mais estáveis e interpretáveis. Embora o modelo apresente boa interpretabilidade, sua capacidade preditiva é limitada diante de relações não lineares, o que sugere que outras abordagens não lineares, como as redes neurais, possam tratar as curvas de perfis não lineares, e poder oferecer ganhos expressivos na geração das curvas sônicas.

Por fim, verificou-se também que o desempenho do modelo está mais associado à escolha adequada dos perfis preditores do que à sua quantidade, ressaltando a importância de uma seleção criteriosa das variáveis de entrada e do auxílio de especialistas da área da Geologia e Engenharia de Petróleo e Gás, que podem selecionar os melhores poços para análises e esclarecer a ocorrência de eventos peculiares. Além disso, recomenda-se o uso de dados brutos, possibilitando maior controle sobre o tratamento das variáveis e inclusão de atributos categóricos (como litologia), que podem contribuir para reduzir erros em faixas críticas de profundidade.

## Referências

Ahammod, S., Hai, M. A., Islam, M. R., and Abu, S. (2014). Petro-physical analysis of reservoir rock of fenchuganj gas field (well# 03) using wireline log. *American Journal of Engineering Research (AJER)*, 3(8):37–48.

Akinyemi, O. D., Elsaadany, M., Siddiqui, N. A., Elkurdy, S., Olutoki, J. O., and Islam, M. M. (2023). Machine learning application for prediction of sonic wave transit time-a case of niger delta basin. *Results in Engineering*, 20:101528.

Alhelfia, L. M., Ali, H. M., and Ahmed, S. H. (2021). P-wave sonic log predictive modeling with optimal artificial neural networks topology. *Journal of Al-Qadisiyah for computer science and mathematics*, 13(3):Page–142.

Archie, G. E. (1950). Introduction to petrophysics of reservoir rocks. *AAPG bulletin*, 34(5):943–961.

Augusto, F. d. O. A. and Martins, J. L. (2009). A well-log regression analysis for p-wave velocity prediction in the namorado oil field, campos basin. *Revista Brasileira de Geofisica*, 27:595–608.

Cao, J., Shi, Y., Wang, D., and Zhang, X. (2017). Acoustic log prediction on the basis of kernel extreme learning machine for wells in gjh survey, erdos basin. *Journal of Electrical and Computer Engineering*, 2017(1):3824086.

Cranganu, C. and Breaban, M. (2013). Using support vector regression to estimate sonic log distributions: a case study from the anadarko basin, oklahoma. *Journal of Petroleum Science and Engineering*, 103:1–13.

Ellis, D. V. and Singer, J. M. (2007). *Well logging for earth scientists*, volume 692. Springer.

Gressling, T. (2020). *Data Science in chemistry: artificial intelligence, big data, chemometrics and quantum computing with jupyter*. Walter de Gruyter GmbH & Co KG.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.

Nero, C., Aning, A. A., Danuor, S. K., and Mensah, V. (2023). Prediction of compressional sonic log in the western (tano) sedimentary basin of ghana, west africa using supervised machine learning algorithms. *Heliyon*, 9(9).

Ogbamikhumi, A., Salami, S., and Uwadiae, W. (2020). Neural network prediction of p-wave log for reservoir characterization in the niger delta basin. *Nigerian Journal of Technological Development*, 17(1):28–32.

Pratikna, K., Rahman, M., Torabi, A., and Mondol, N. (2022). Machine learning application for compressional wave velocity log prediction in sleipner co2 storage, offshore norway. In *83rd EAGE Annual Conference & Exhibition*, volume 2022, pages 1–5. European Association of Geoscientists & Engineers.

Selley, R. C. (1998). *Elements of petroleum geology*. Gulf Professional Publishing.

SOUSA, A. (2024). Validação externa dos agrupamentos de eletrofácies com aprendizado não supervisionado utilizando litologias interpretadas em poços de petróleo. 107 páginas.

Tissot, B. P. and Welte, D. H. (2013). *Petroleum formation and occurrence*. Springer Science & Business Media.