

TESE DE DOUTORADO  
Nº 613

APLICAÇÃO DE TÉCNICAS DE APRENDIZADO  
DE MÁQUINA NA DETERMINAÇÃO DE  
PROPRIEDADES PETROFÍSICAS E  
GEOQUÍMICAS EM ROCHAS RESERVATÓRIOS  
E GERADORAS DA BACIA DO PARNAÍBA

Julio César Ardito

Profa. Dra. Liliane Janikian Paes de Almeida  
Orientadora







UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE GEOCIÊNCIAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM GEOTECTÔNICA E GEOQUÍMICA

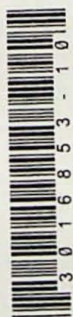
**Aplicação de Técnicas de Aprendizado de Máquina na Determinação de Propriedades Petrofísicas e Geoquímicas em Rochas Reservatórios e Geradoras da Bacia do Parnaíba**

JULIO CÉSAR ARDITO

Tese apresentada Instituto de Geociências da Universidade de São Paulo para obtenção do título de Doutor em Ciências pelo Programa de Pós-graduação em Geologia Sedimentar. Área de Concentração: Petrofísica e Aprendizagem de máquina.

**Orientadora: Profa. Dra. Liliane Janikian Paes de Almeida**

IGc-Instituto de Geociências



3 0 1 6 8 5 3 - 1 0 1 1

Aplicação de técnicas de aprendizado de máquina na determinação de propriedades petrofísicas

T6168

São Paulo  
2019



T  
AG76  
JC.a

Serviço de Biblioteca e Documentação do IGc/USP

Ficha catalográfica gerada automaticamente com dados fornecidos pelo(a) autor(a) via programa desenvolvido pela Seção Técnica de Informática do ICMC/USP

Bibliotecários responsáveis pela estrutura de catalogação da publicação:

Sonia Regina Yole Guerra - CRB-8/4208 | Anderson de Santana - CRB-8/6658

Ardito, Julio

Aplicação de Técnicas de Aprendizado de Máquina na Determinação de Propriedades Petrofísicas e Geoquímicas em Rochas Reservatórios e Geradoras da Bacia do Parnaíba / Julio Ardito; orientadora Liliane Janikian Paes de Almeida. -- São Paulo, 2019. 183 p.

Tese (Doutorado - Programa de Pós-Graduação em Geoquímica e Geotectônica) -- Instituto de Geociências, Universidade de São Paulo, 2019.

1. Petrofísica de rochas reservatório. 2. Aprendizado de Máquina. 3. Porosidade e Permeabilidade. I. Janikian Paes de Almeida, Liliane, orient. II. Título.



UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE GEOCIÊNCIAS

**APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE  
MÁQUINA NA DETERMINAÇÃO DE PROPRIEDADES  
PETROFÍSICAS E GEOQUÍMICAS EM ROCHAS  
RESERVATÓRIOS E GERADORAS DA BACIA DO  
PARNAÍBA**

**JULIO CÉSAR ARDITO**

**Orientadora: Profa. Dra. Liliane Janikian Paes de Almeida**

Tese de Doutorado

**Nº 613**

COMISSÃO JULGADORA

Dra. Liliane Janikian Paes de Almeida

Dr. Humberto Luis Siqueira Reis

Dr. André Luiz Vizine Pereira

Dr. André Oliveira Sawakuchi

Dr. Ricardo Cabral de Azevedo

Dr. Vinicius Hector Abud Louro

SÃO PAULO

2019

*Aos meus filhos, Renato, Ricardo e Raphael, a minha esposa Cláudia e aos meus pais, César e Alice, que sempre acreditaram em mim, muito mais do que eu mesmo.*



## Agradecimentos

Agradeço a Deus, que me deu a força e a energia que precisei ao longo desta trajetória.

Não posso deixar de agradecer minha família, que são meu porto seguro, especialmente meus filhos Renato, Ricardo e Raphael com quem deixei inúmeras vezes de brincar e curtir em função da pesquisa. Agradeço minha esposa Cláudia e meus pais César e Alice por me apoiarem incondicionalmente.

Agradeço minha orientadora Liliane Janikian Paes de Almeida, pela amizade, pelos conselhos e orientação. Ao Renato Paes de Almeida pela amizade, pelas aulas e ajuda durante todo este período.

Agradeço aos amigos do IGC que me ajudaram demais, além de terem feito minha vida mais alegre, Larissa, Cristiano, Maria, Ariel, Carlos, Nazaré, Dorília, Marcelo, Carlos, Grega. Aos amigos Bruno e Marcelo pela amizade.

A Senergy por ceder à licença do Interative Petrophysics.

E a Agência Nacional do Petróleo e Biocombustíveis (ANP) por ter cedido os dados para este projeto.

Ao CNPq pela bolsa concedida, que tanto me ajudou e a FAPESP.

*Aos funcionários do IGc, especialmente Alexandre, Katerine e Leonardo Capucci sempre tão atenciosos e prestativos.*

*E a USP, que tem sido minha segunda casa desde 2005.*



## Resumo

Ardito, Júlio César. DETERMINAÇÃO DE PROPRIEDADES PETROFÍSICAS E GEOQUÍMICAS DA BACIA DO PARNAÍBA POR MEIO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA. 2019. Tese (Doutorado em Ciências) – Instituto de Geociências, Universidade de São Paulo, São Paulo, 2019.

Propriedades de rocha como porosidade, permeabilidade e saturação de fluídos são de fundamental importância na compreensão de um reservatório e no auxílio à tomada de decisões sobre os rumos da exploração de um sistema petrolífero. Estas propriedades normalmente são calculadas com base em dados de perfilagem em um processo conhecido como avaliação petrofísica e que requer o uso de softwares pagos e da expertise de um profissional altamente experiente. Além disso, é um processo que eventualmente pode demorar dias, implicando em altos custos de *stand by* de sonda, desmobilização ou mobilização de equipes para realização de testes adicionais como extração de testemunhos, teste de formação e fraturamento hidráulico. Desta forma, este trabalho propõe um fluxo de trabalho alternativo capaz de estimar parâmetros como porosidade, permeabilidade e espessura do intervalo portador de gás por meio de uma série de algoritmos baseados em aprendizagem de máquina. Técnicas como *Gradient Boosting*, *Redes Neurais Artificiais (RNA)*, *K-means* e *Máquinas de Vetores de Suporte (SVM)* foram utilizadas para construir modelos preditivos de porosidade, permeabilidade e também para classificar e agrupar as rochas-reservatório possibilitando estimar os principais intervalos portadores de gás. Além disso, o uso destas técnicas possibilitou recuperar com sucesso perfis de densidade em poços antigos, perfilados em um período anterior ao advento das ferramentas de medição de densidade. Para isso, foram utilizados registros de perfilagem de poço, resultados de análises laboratoriais de porosidade, permeabilidade, teor de carbono orgânico e dados de cromatografia gasosa. Outro objetivo importante foi estudar e compreender as características petrofísicas e geoquímicas dos sistemas petrolíferos da Bacia do Parnaíba com uso de dados de 54 poços perfilados na bacia.

Palavras-chave: Propriedades petrofísicas, Porosidade, Permeabilidade, Aprendizado de máquina.



## Abstract

Ardito, Júlio César. DETERMINATION OF PETROPHYSICAL AND GEOCHEMICAL PROPERTIES OF THE PARNAIBA BASIN BY MACHINE LEARNING TECHNIQUES. 2019. Thesis – Institute of Geoscience, University of São Paulo, São Paulo, 2019.

Sedimentary rock properties such as porosity, permeability and fluid saturation are of fundamental importance in understanding a reservoir and in helping to make decisions of direction of exploration of petroleum system.

These properties are usually calculated based on profiling data in a process known as petrophysical appraisal that requires the use of paid software and the expertise of a highly experienced professional.

In addition, it may take a few days to complete, resulting in high costs of stand by in drilling operation, demobilization or team mobilization for additional testing such as core extraction, formation testing and hydraulic fracturing. Thus, this paper proposes an alternative workflow capable of estimating parameters such as porosity, permeability and *Netpay* thickness by a series of machine learning based algorithms. Techniques such as Gradient Boosting, Artificial Neural Networks (RNA), K-means and Support Vector Machines (SVM) were used to built predictive models of porosity, permeability and also to classify and group the reservoir, allowing to estimate the main netpay zones. In addition, the use of these techniques has made it possible to successfully recover density profiles in old wells profiled in a period prior to the advent of density measurement tools. For this purpose, well logging records, results of laboratory analysis of porosity, permeability, organic carbon content and gas chromatography data were used.

Another important objective of this work was to study and understand the petrophysical and geochemical characteristics of the Parnaíba basin petroleum systems. For this purpose data from 54 wells profiled in the basin were used.

Key words: Petrophysical Properties, Porosity, Permeability, Machine Learning.



## Lista de figuras

Figura 1: Mapa do relevo da bacia do Parnaíba com limites da bacia em vermelho (Julio Ardito, 2019). .....	24
Figura 2: Mapa Geológico da Bacia do Parnaíba realçando o Grupo Serra (Siluriano) na borda leste (Modificado de Bizzi et al.,2003).....	25
Figura 3: Representação esquemática da revisão litoestratigrafica do Grupo Serra Grande. (Modificado de Quadros, 1982). .....	27
Figura 4: Diagrama litoestratigrafico da Bacia do Parnaíba. (Modificado de Vaz et al., 2007). .....	28
Figura 5: Esquemático de comunicação entre neurônios (modificado de Haykin, 2001). .....	39
Figura 6: Esquemático do funcionamento de uma rede neural artificial. ....	40
Figura 7: Arquitetura de uma rede neural artificial com apenas uma camada. ....	41
Figura 8: Exemplo esquemático da arquitetura de uma rede neural artificial multicamada.....	41
Figura 9: Gráfico da função tangente hiperbólica.....	42
Figura 10: Padrão de conexões entre os neurônios da camada de entrada e um neurônio da camada intermediária, passando pela função soma e pela função de ativação. As células demarcadas com a letra “x” representam o dado de entrada e as setas demarcadas com a letra “w” são os pesos....	43
Figura 11: Fluxo de processamento e treinamento da rede neural artificial. ....	44
Figura 12: Exemplo de árvore de decisão (modificado de Ahmed et al., 2018). ....	45
Figura 13: Exemplo de uma árvore de decisão. (a) partição de entradas associada (b) aos nós da árvore. (modificado de Kang et al., 2016).....	46
Figura 14: Hiperplanos de separação (em azul) de um conjunto de dados com duas classes. A reta vermelha representa o hiperplano ideal (Julio Ardito, 2019). .....	51
Figura 15: Exemplo de envoltória convexa. (Modificado de Hughes e Correl, 2016). .....	52
Figura 16: Fluxograma de trabalho convencional utilizado neste trabalho para avaliação das Formações Poti, Cabeças e Jaicós. ....	56



Figura 17: Perfil Gama com valores de limites (clean e clay) para determinação da curva de argilosidade (à esquerda). A área verde hachurada representa o volume de argila na rocha. ....	57
Figura 18: Aplicação do método <i>Pickett Plot</i> para determinação do valor de <i>m</i> (coeficiente de cimentação), por meio da comparação de valores de porosidade e resistividade.....	58
Figura 19: Avaliação petrofísica final da Formação Cabeças nos poços 1, 2, 3 e 4. A coluna 6 contém o perfil de permeabilidade, a coluna 7 mostra a de saturação de água e a 8 mostra porosidade e as regiões portadoras de gás. ....	60
Figura 20: Perfis da avaliação petrofísica final da Formação Cabeças nos poços 5, 6 e 7. A coluna 6 contém o perfil permeabilidade, a coluna 7 mostra a saturação de água, a porosidade e as regiões portadoras de gás (netpay). A coluna 8, mostra em outra escala a porosidade e as zonas com gás....	61
Figura 21: Perfis da avaliação petrofísica final da Formação Jaicós nos poços 11, 19 e 25. A coluna 6 contém o perfil de saturação de água, a coluna 7 mostra a permeabilidade e a 8 mostra porosidade e as regiões portadoras de gás. ....	62
Figura 22: Perfis da avaliação petrofísica final da Formação Poti nos poços 22, 23, 24 e 25. A coluna 6 contém o perfil de saturação de água, a coluna 7 mostra a permeabilidade e a 8 mostra porosidade e as regiões portadoras de gás. ....	63
Figura 23: Perfis da avaliação petrofísica com destaque para a última coluna que contem as curvas de permeabilidade calculada (em azul) e medida por ressonância magnética nuclear (em vermelho) da Formação Cabeças no poços 7.....	64
Figura 24: Diagrama de dispersão da densidade pelo perfil sônico de um dos poços da Bacia do Parnaíba. Em A as cores representam a variação Gama. Em B, as cores indicam a formação, sem rosa a Formação Poti e Marrom a Formação Cabeças. ....	64
Figura 25: Perfis de porosidade, saturação de água e volume de argila. Os intervalos em verde mostram o reservatório efetivo e em vermelho os intervalos portadores de gás do poço 22 na Formação Poti.....	65



Figura 26: Fluxo de trabalho alternativo desenvolvido neste trabalho para avaliação de sistema petrolífero. ....	66
Figura 27: Diagrama de dispersão da porosidade preditiva gerado por Redes Neurais Artificiais. A figura mostra a correlação entre a porosidade calculada e a porosidade predita no intervalo Cabeças dos poços 5, 6, 7 e 9. ....	68
Figura 28: Diagrama de dispersão com a correlação entre a porosidade calculada e a porosidade predita por Redes Neurais Artificiais no intervalo Cabeças do poço 5. ....	70
Figura 29: Diagrama de dispersão com a correlação entre a porosidade calculada e a porosidade predita por Redes Neurais Artificiais no intervalo Cabeças do poço 7. ....	70
Figura 30: Perfis de densidade. Em azul as densidades preditas pela rede neural artificial e em vermelho as curvas de densidade medidas na perfilagem dos poços 5 (A) e 7 (B). ....	71
Figura 31: Diagrama de dispersão com a Correlação entre a densidade calculada e a predita por regressão linear com múltiplas variáveis no poço 5. ....	72
Figura 32: Diagrama de dispersão com a Correlação entre a densidade calculada e a predita por regressão linear com múltiplas variáveis no poço 7. ....	72
Figura 33: Perfis de densidade. Em azul as densidades preditas pela rede neural artificial e em vermelho as curvas de densidade medidas na perfilagem dos poços 5 e 7. ....	73
Figura 34: Gráfico do erro padrão em função do número de iterações para o conjunto de dados de testes e para o conjunto de dados de treino. ....	75
Figura 35: Diagrama de dispersão da densidade calculada pela predita por Gradiente Boosting para os dados do poço 5. A figura mostra o coeficiente de correlação R de 0,85. ....	77
Figura 36: Diagrama de dispersão da densidade calculada pela predita por Gradiente Boosting para os dados do poço 7. A figura mostra o coeficiente de correlação R de 0,87. ....	77
Figura 37: Curvas de densidade pela profundidade dos poços 5 (A) e 7 (B). Em azul as curvas preditas com o uso do método Gradient Boosting e em vermelho as densidades medidas na perfilagem. ....	78
Figura 38: Diagramas de dispersão com as correlações entre as densidades medidas e preditas pelo modelo nos poços 1 (em A), 2 (em B), 3 (em C) e	



4 (em D). Em azul é mostrada a reta de regressão linear dos dados de cada poço e a linha vermelha representa a reta expectativa, onde  $x=y$ ....79

Figura 39: Diagramas de dispersão com as correlações entre as densidades medidas e preditas pelo modelo nos poços 5 (em A), 6 (em B), 7 (em C) e 8 (em D). Em azul é mostrada a reta de regressão linear dos dados de cada poço e a linha vermelha representa a reta expectativa, onde $x=y$ ....	80
Figura 40: Perfis com as densidades preditas (em azul) e as densidades medidas (em vermelho) dos poços 1 (em A), 2 (em B), 3 (em C) e 4 (em D) da Formação Cabeças.....	81
Figura 41: Perfis com as densidades preditas (em azul) e as densidades medidas (em vermelho) dos poços 5 (em A), 6 (em B), 7 (em C) e 8 (em D) da Formação Cabeças.....	81
Figura 42: Porosidades calculadas por meio do perfil densidade (em azul) e por meio da densidade predita (em violeta) com o uso do método Gradient Boosting dos poços 1, 2, 3 e 4.....	82
Figura 43: Porosidades calculadas por meio do perfil densidade (em azul) e por meio da densidade predita (em violeta) com o uso do método <i>Gradient Boosting</i> dos poços 5, 6 e 7.....	83
Figura 44: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e preditas pelo modelo no poço 5.....	86
Figura 45: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e preditas pelo modelo no poço 6.....	86
Figura 46: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e preditas pelo modelo no poço 7.....	87
Figura 47: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e preditas pelo modelo no poço 8.....	87
Figura 48: Perfis com as porosidades calculadas (linhas azuis) e preditas (linhas vermelhas) pelo modelo para os poços 5 (em A) e 6 (em B).....	89
Figura 49: Diagrama de dispersão com a correlação linear entre o conjunto de dados de teste e o conjunto de dados de treino. As porosidades foram preditas com uso de ressonância magnética nuclear.....	91
Figura 50: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e as porosidades preditas pelo modelo no poço 22 do intervalo Poti.....	92



Figura 51: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e as porosidades previstas pelo modelo no poço 23 do intervalo Poti.....	92
Figura 52: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e as porosidades previstas pelo modelo no poço 24 do intervalo Poti.....	93
Figura 53: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e as porosidades previstas pelo modelo no poço 25 do intervalo Poti.....	93
Figura 54: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e as porosidades previstas pelo modelo no poço 26 do intervalo Poti.....	94
Figura 55: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e as porosidades previstas pelo modelo no poço 27 do intervalo Poti.....	94
Figura 56: Diagrama de dispersão com a correlação linear entre a permeabilidade medida por ressonância magnética nuclear e permeabilidade prevista no intervalo da formação Cabeças usando a primeira abordagem. ....	100
Figura 57: Gráfico da permeabilidade medida com o perfil ressonância magnética pela permeabilidade determinada pelo modelo no poço 2 da formação Cabeças.....	100
Figura 58: Gráfico da permeabilidade medida com o perfil ressonância magnética pela permeabilidade determinada pelo modelo no poço 3 da formação Cabeças.....	101
Figura 59: Gráfico da permeabilidade medida com o perfil ressonância magnética pela permeabilidade determinada pelo modelo no poço 4 da formação Cabeças.....	101
Figura 60: Diagrama de dispersão com dados de permeabilidade medida com o perfil ressonância magnética em função da permeabilidade preditiva no intervalo da Formação Cabeças. A linha azul mostra o coeficiente de correlação R. ....	102
Figura 61: Gráfico da permeabilidade medida com o perfil ressonância magnética pela permeabilidade determinada pelo modelo no intervalo da formação Poti.....	103



Figura 62: Gráfico do fator de importância de cada perfil (atributo de entrada) considerado pelo modelo no processo de treinamento.....	104
Figura 63: Diagramas de dispersão construídos com dados de 3 poços da Formação Cabeças. Os diagramas contemplam todos os atributos envolvidos, dois a dois e foram divididos em dois grupos (amarelo e azul), por semelhança de características, pelo algoritmo K-means. ....	106
Figura 64: Diagramas de dispersão construídos com dados de 3 poços do reservatório Poti. Os diagramas contemplam todos os atributos envolvidos, dois a dois e foram divididos em dois grupos (amarelo e azul), por semelhança de características, pelo algoritmo K-means. ....	107
Figura 65: Zonas de Net pay em perfis do poço 1 no intervalo do reservatório Cabeças do modelo A e do modelo B. A ausência de net pay aqui é interpretada como net gross.....	108
Figura 66: Espessuras das zonas portadoras de gás (Net pay) em perfis do poço 2 no intervalo do reservatório Cabeças estimadas por dois modelos distintos, A e B. O valor 1,0 significa portador de gás e 0,0 representa ausência de gás. ....	110
Figura 67: Espessuras das zonas portadoras de gás (Net pay) em perfis do poço 3 no intervalo do reservatório Cabeças estimadas por dois modelos distintos, A e B. O valor 1,0 significa portador de gás e 0,0 representa ausência de gás. ....	111
Figura 68: Espessuras das zonas portadoras de gás (Net pay) em perfis do poço 1 no intervalo do reservatório Poti estimadas por dois modelos distintos, A e B. O valor 1,0 significa portador de gás e 0,0 representa ausência de gás. ....	113
Figura 69: Espessuras das zonas portadoras de gás (Net pay) em perfis do poço 3 no intervalo do reservatório Poti estimadas por dois modelos distintos, A e B. O valor 1,0 significa portador de gás e 0,0 representa ausência de gás. ....	115
Figura 70: Mapa com a distribuição de porosidades médias da Formação Poti na Bacia do Parnaíba.....	117
Figura 71: Mapa com a distribuição de porosidades médias da Formação Cabeças na Bacia do Parnaíba.....	118



Figura 72: Mapa com a distribuição de permeabilidades médias da Formação Poti (em A) e da Formação Cabeças (em B) da bacia do Parnaíba.....	119
Figura 73: Mapa com a distribuição das espessuras médias portadoras de gás pela Bacia do PArnaíba. Em A são representadas as espessuras da da Formação Poti e em B da Formação Cabeças (em B).....	120
Figura 74: Diagrama de dispersão com a correlação linear (reta em azul) do teor de carbono orgânico medido pelo TOC predito com uso do método <i>Gradient Boosting</i> . A reta em vermelho mostra a função expectativa onde $x=y$ . A figura ilustra os resultados da comparação entre o conjunto de treino com o conjunto de dados de testes na razão 80 / 20.....	122
Figura 75: Diagrama de dispersão com a correlação linear (reta em azul) do teor de carbono orgânico medido pelo TOC predito com uso do método <i>Gradient Boosting</i> . A reta em vermelho mostra a função expectativa onde $x=y$ . A figura ilustra os resultados da comparação entre o conjunto de treino com o conjunto de dados de testes na razão 60 / 40.....	123
Figura 76: Variância do erro das predições em função do número de iterações realizadas pelo modelo durante o processo de treinamento. ....	123
Figura 77: Diagrama de dispersão com a correlação linear (reta em azul) do teor de carbono orgânico medido pelo TOC predito com uso do método <i>Gradient Boosting</i> . ....	124
Figura 78: Diagrama de dispersão com a correlação linear (reta em azul) do teor de carbono orgânico medido pelo TOC predito com uso do método <i>Gradient Boosting</i> . ....	125
Figura 79: Gráfico de barras do fator de importância normalizado dos atributos de entrada encontrados pelo modelo para predição de TMAX. ....	126
Figura 80: Diagrama de dispersão do perfil “pseudo-sônico” (DlogR) em função da resistividade, com destaque para a linha base (em amarelo) que representa a gênese do <i>log</i> pseudo-sônico. ....	127
Figura 81: Análises do teor de carbono orgânico (TOC) pelos perfis gama, sônico e “pseudo-sônico”. A coluna 4 mostra os valores de TOC calculados em laboratório e a coluna 6 mostra a comparação do perfil sônico com o DlogR, onde as áreas hachuradas em azul mostram as regiões com maiores teores de carbono orgânico. ....	127



Figura 82: Mapas com isópacas das formações geradoras da Bacia do Parnaíba. A) Formação Longá, B) Soleira da formação Longá, C) Formação Pimenteiras, D) Soleira da formação Pimenteiras, E) Formação Tianguá. .....	129
Figura 83: Gráficos dos valores de carbono orgânico, S2, índice de hidrogênio, refletância de vitrinita calculada e do índice de oxigênio da formação Longá. Os valores mostrados foram retirados de resultados dos ensaios de pirólise em amostras. ....	130
Figura 84: Diagrama de Van Krevelen para análise do tipo de matéria orgânica. O diagrama mostra os valores do índice de hidrogênio em função di índice de oxigênio da formação Longá. ....	131
Figura 85: Diagrama de carbono orgânico total pelo potencial remanescente de geração (S2) da formação Longá. O diagrama é usado para classificar e avaliar a qualidade do Querogênio da formação geradora. ....	132
Figura 86: Diagrama de Índice de hidrogênio pelo TMAX para análise e classificação do tipo de querogênio do intervalo gerador Longá. ....	132
Figura 87: Isólitais da formação Longá, onde: A) Razão dos índices de hidrogênio pelos índices de oxigênios, B) interpolação dos valores de TMAX nos poços estudados, C) Distribuição dos valores de TOC e D) Índice de Hidrogênio. ....	133
Figura 88: Gráficos dos valores de carbono orgânico, S2, índice de hidrogênio, do índice de oxigênio da formação Pimenteiras. Os valores mostrados foram retirados de resultados dos ensaios de pirólise em amostras. ....	134
Figura 89: Diagrama de Van Krevelen para análise do tipo de matéria orgânica. O diagrama mostra os valores do índice de hidrogênio em função di índice de oxigênio da Formação Pimenteiras. ....	135
Figura 90: Isólitais dos folhelhos radioativos (em A) e do teor de carbono orgânico (TOC) (em B) do tipo C. (modificado de Rodrigues, 95). ....	136
Figura 91: Diagrama do carbono orgânico total em função do potencial remanescente de geração (S2) da Formação Pimenteiras. O diagrama é usado para classificar e avaliar a qualidade do Querogênio da formação geradora. ....	137
Figura 92: Diagrama do Índice de hidrogênio pelo TMAX para análise e classificação do tipo de querogênio do intervalo gerador Pimenteiras. As formas e cores	



dos pontos correspondem a dados de diversos poços espalhados pela bacia.....	138
Figura 93: Isólitais da formação Pimenteiras onde: A) Razão dos índices de hidrogênio pelos índices de oxigênios, B) interpolação dos valores de TMAX nos poços estudados, C) Distribuição dos valores de TOC e D) Índice de Hidrogênio. ....	139
Figura 94: Gráficos dos valores de carbono orgânico, S <sub>2</sub> , índice de hidrogênio, refletância de vitrinita calculada e do índice de oxigênio da formação Tianguá. Os valores mostrados foram retirados de resultados dos ensaios de pirólise em amostras. ....	140
Figura 95: Diagrama de Van Krevelen para análise do tipo de matéria orgânica. O diagrama mostra os valores do índice de hidrogênio em função do índice de oxigênio da Formação Tianguá. ....	141
Figura 96: Isólitais da formação Tianguá onde: A) Razão dos índices de hidrogênio pelos índices de oxigênios, B) interpolação dos valores de TMAX nos poços estudados, C) Distribuição dos valores de TOC e D) Índice de Hidrogênio. ....	142



## Lista de tabelas

Tabela 1: Funções “perda” e gradientes (modificado de Hastie et al., 2001). .....	50
Tabela 2: Quantidade de registros de análises geoquímicas de Pirólise de cada parâmetro. ....	54
Tabela 3: Parâmetros utilizados para treinamento da rede neural artificial. ....	71
Tabela 4: Métricas e parâmetros utilizados para treinamento da rede neural artificial. ....	74
Tabela 5: Métricas e parâmetros utilizados para treinamento do modelo. ....	75
Tabela 6: Métricas e parâmetros utilizados para treinamento do modelo. ....	84
Tabela 7: Métricas e parâmetros utilizados para treinamento do modelo. ....	85
Tabela 8: Classificação prévia de porosidade de amostras de calha. ....	95
Tabela 9: Parâmetros utilizados no treinamento da SVM. ....	96
Tabela 10: Matriz de confusão do modelo treinado com SVM. ....	96
Tabela 11: Estimadores estatísticos da matriz de confusão. ....	97
Tabela 12: Dado de perfis de poço utilizados no treinamento do modelo A. ....	98
Tabela 13: Dado de perfis de poço utilizados no treinamento do modelo B. ....	98
Tabela 14: Fator de importância dado pelo modelo a cada perfil utilizado. ....	104
Tabela 15: Atributos de entrada utilizados no treinamento do modelo A. ....	107
Tabela 16: Atributos de entrada utilizados no treinamento do modelo B. ....	108
Tabela 17: Parâmetros da matriz de confusão utilizados no treinamento dos modelos A e B. ....	109
Tabela 18: Métricas da comparação dos resultados de cada modelo com o cálculo tradicional de NetPay. ....	109
Tabela 19: Matriz de confusão dos modelos A e B. ....	109
Tabela 20: Métricas para avaliação dos modelos A e B. ....	110
Tabela 21: Parâmetros da matriz de confusão utilizados no treinamento dos modelos A e B. ....	112
Tabela 22: Métricas da comparação dos resultados de cada modelo com o cálculo tradicional dos intervalos portadores de gás. ....	112
Tabela 23: Matriz de confusão utilizados no treinamento dos modelos A e B. ....	112
Tabela 24: Métricas dos modelos A e B em relação na comparação com os intervalos de NetPay calculados. ....	112



Tabela 25: Parâmetros da matriz de confusão utilizados no treinamento dos modelos A e B. ....	114
Tabela 26: Métricas da comparação dos resultados de cada modelo com o cálculo tradicional. ....	114
Tabela 27: Métricas da comparação dos resultados de cada modelo com o cálculo tradicional de NetPay. ....	114
Tabela 28: Métricas da comparação dos resultados de cada modelo com o cálculo tradicional de NetPay. ....	115



## Sumário

<b>1. Introdução .....</b>	<b>20</b>
<b>2. Revisão Bibliográfica .....</b>	<b>23</b>
2.1 <i>Contexto Geológico</i> .....	23
2.1.1 Grupo Serra Grande .....	24
2.1.2 Grupo Canindé .....	28
2.1.3 Grupo Balsas .....	30
2.2 <i>Propriedades de Rocha e Métodos de Medição</i> .....	30
2.2.1 Caliper .....	31
2.2.2 Raios Gamma .....	31
2.2.3 Densidade .....	32
2.2.4 Neutrão .....	32
2.2.5 Microresistividade .....	32
2.2.6 Porosidade .....	33
2.2.7 Permeabilidade .....	35
2.2.8 Volume de Argila .....	35
2.2.9 Saturação de Água .....	36
<b>3. METODOLOGIA .....</b>	<b>38</b>
3.1 <i>Técnicas de Aprendizagem de Máquina</i> .....	38
3.1.1 Regressão Linear .....	38
3.1.2 Redes Neurais Artificiais .....	39
3.1.3 Árvores de decisão .....	44
3.1.4 Máquinas de vetores de suporte (SVM) .....	50
<b>4. Processamento e análise de resultados .....</b>	<b>55</b>
4.1 <i>Análises petrofísicas convencionais</i> .....	56
4.2 <i>Fluxo de trabalho alternativo</i> .....	66
4.2.1 Recuperação do perfil densidade .....	67
4.2.2 Redes Neurais Artificiais .....	69
4.2.3 Regressão Linear com Múltiplas Variáveis .....	72
4.2.4 Gradient Boosting .....	75
4.3 <i>Modelo de classificação do reservatório</i> .....	95



4.4	<i>Implementação e treinamento de um modelo preditivo de permeabilidades com uso do algoritmo Gradient Boosting</i> .....	97
4.5	<i>Delimitação do reservatório e de zonas portadoras de gás (netpay) por agrupamento com uso do método K-means</i> .....	105
4.6	<i>Predição do teor de carbono orgânico (TOC) e do parâmetro T<sub>max</sub></i> .....	121
4.6.1	Abordagem 1 .....	122
4.6.2	Abordagem 2 .....	124
4.6.3	Abordagem 3 .....	126
4.6.4	Análises do potencial de geração da bacia .....	128
<b>5.</b>	<b>Conclusões</b> .....	<b>143</b>
	<b>Referências</b> .....	<b>146</b>
	<b>Anexo A – Trabalhos de Campo e Estudo Faciológico</b> .....	<b>152</b>
	Petrofísica Básica.....	162
	<b>Anexo B – Artigo Submetido: Variabilidade da Arquitetura Fluvial do Grupo Serra Grande</b> .....	<b>166</b>
	<b>Abstract</b> .....	<b>166</b>
	<b>Introduction</b> .....	<b>167</b>
	<b>Study sites and methods</b> .....	<b>168</b>
	<b>Fluvial styles and their spatial distribution</b> .....	<b>169</b>
	<b>Discussion and conclusions</b> .....	<b>171</b>



## 1. Introdução

As estimativas de propriedades físicas de rochas como porosidade, permeabilidade e saturação de fluidos de um reservatório são de fundamental importância para a continuidade da exploração de hidrocarbonetos em uma bacia, pois permitem subsidiar estudos de viabilidade econômica, volumetria e produção (ELKATATNY et al., 2018). Estas propriedades podem ser conhecidas por meio de uma avaliação petrofísica, que consiste na utilização de registros de perfilagem, dados de laboratório e outras informações relevantes em uma série de cálculos e processos. Embora este tipo de avaliação seja usualmente realizado por meio de um processo tradicional e normalmente eficaz, este tipo de procedimento demanda de um intervalo de tempo que pode levar horas ou em ou mesmo dias. Desta forma, este tempo entre o término da perfilagem, no campo e os resultados da avaliação petrofísica, no escritório, pode gerar uma série de gastos significativos com a sonda parada (*stand by*) e principalmente com espera para mobilização ou demora na desmobilização de equipes até que uma decisão sobre o rumo da exploração no *play* seja tomada. Esta decisão ocorre justamente em função dos resultados da avaliação petrofísica e pode implicar em testes adicionais como testes de formação, fraturamento hidráulico, retiradas de mais amostras laterais, testemunhos ou mesmo abandono do poço. Além disso, toda avaliação petrofísica requer o uso de um *software* específico, o que envolve usualmente altos custos e demanda da expertise de um profissional experiente e qualificado.

Um dos caminhos alternativos que tem sido explorado para calcular parâmetros petrofísicos de reservatório é o uso de estatística multivariada (SAFTIĆ et al., 2001; KURODA, 2012; RAFIK; KAMEL, 2017; SINHA et al., 2016) em conjunto com técnicas de inteligência artificial (VERMA et al., 2012; ANIFOWOSE; LABADIN; ABDULRAHEEM, 2013; ELKATATNY et al., 2018 e KUMAR; SINHA, 2013). Ao invés de utilizar equações genéricas, as técnicas de *machine* e *deep learning* utilizam algoritmos autônomos capazes de ler uma grande quantidade de dados, identificar padrões, desenvolver e ajustar um modelo preditivo adaptado às peculiaridades locais do reservatório estudado. Kuroda et al. (2012) usaram métodos não supervisionados para minimizar incertezas na caracterização de reservatórios.



Roy et. al, 2010 utilizaram métodos não supervisionados para identificar e classificar automaticamente fácies sísmicas.

Além de estudar e caracterizar o sistema petrolífero da bacia do Parnaíba, este projeto propõe estabelecer um fluxo de trabalho capaz de minimizar o intervalo de tempo de processamento petrofísico, produzindo estimativas primárias confiáveis acerca dos parâmetros permo-porosos do reservatório, além de minimizar custos com softwares e mão de obra específica, uma vez que o processo pode ser realizado em software livre e executado por qualquer profissional da área tecnologia da informação ou engenharia.

As etapas deste projeto envolvem o uso de estatística multivariada juntamente com método algoritmos supervisionados, como *Gradient Boosting* e não supervisionados, como redes neurais artificiais e *K-means*, executados em softwares livres, não necessariamente por um petrofísico.

Para isso, foram utilizados dados de poços da Bacia do Parnaíba, fornecidos em dois lotes pela Agência Nacional do Petróleo (ANP). A maior parte deles contém dados de perfilagem, porém raros são registros de ensaios laboratoriais como *Rock-Eval* e análises petrofísicas de rotina básicas e avançadas.

A bacia estudada, outrora conhecida como Bacia do Maranhão possui idade paleozóica e é classificada por diversos autores como intracratônica ou sinéclise (Caputo et al., 2005). Ela está localizada no nordeste ocidental brasileiro e abrange parte dos Estados do Maranhão, Piauí, Ceará, Tocantins, Pará e Bahia, ocupando uma área de cerca de 600km<sup>2</sup>. No depocentro da bacia são encontradas sucessões sedimentares com espessuras da ordem de 3500m (Milani e Zalán, 1999). Os limites com a Bacia de Barreirinhas, ao norte, e com a Bacia do São Francisco, ao sul, são respectivamente, o arco Ferre Urbano-Santos e o arco do São Francisco.

Segundo a agência (ANP, 2015), os principais reservatórios da bacia encontram-se entre os sedimentos devonianos do Grupo Canindé, sendo os arenitos porosos da Formação Cabeças e da Formação Poti, os principais reservatórios, e os folhelhos escuros ricos em matéria orgânica da Formação Pimenteiras compõe o principal gerador. Um terceiro sistema petrolífero seria um sistema não convencional na própria Formação Pimenteiras (Miranda, 2014). Além disso, há ainda indícios de um quarto sistema, muito menos estudado, no Grupo Serra Grande, de idade Siluriana.



Embora a estratigrafia e o arcabouço estrutural da bacia sejam bem conhecidos, são raríssimos os trabalhos a respeito dos elementos que compõem o sistema petrolífero, como por exemplo, Miranda (2014) e inexistem estudos detalhados a respeito das propriedades físicas das rochas que compõe os Grupos Canindé e Serra Grande. Além disso, poucos são os estudos sobre o gradiente geotérmico da bacia e suas implicações para o modelo de subsidência da bacia.

Desta forma, o presente trabalho propõe um fluxo completo e alternativo aos métodos convencionais para estimar as características permo-porosas dos reservatórios Poti e Cabeças com o uso de dados de perfilagem, dados de laboratório e a aplicação de diversas técnicas de *machine* e *deep learning*. Além disso, pretende-se caracterizar as rochas geradoras da bacia com o uso de *logs* de perfilagem e a reinterpretação de análises geoquímicas de laboratório, visando o entendimento das relações com os análogos de afloramento, e a avaliação do potencial gerador da bacia.

Nos anexos são apresentados dois estudos, que adicionalmente a este projeto levou a caracterização de fácies sedimentares do Grupo Serra Grande. Ambos os estudos, foram desenvolvidos durante o período de doutorado, a partir de duas viagens de campo. Um destes estudos trata de um modelo preliminar de fácies e análises de amostras de afloramentos do Grupo Serra Grande (anexo A) e o outro é um artigo submetido (JANIKIAN *et al.*, 2019) sobre a variabilidade da arquitetura fluvial do Grupo Serra Grande (anexo B).



## 2. Revisão Bibliográfica

### 2.1 Contexto Geológico

A bacia paleozoica do Parnaíba, localizada no nordeste brasileiro, é classificada pela maior parte dos autores como intracratônica e ocupa uma de cerca de 600km, conforme mostra o mapa de relevo apresentado na Figura 1.

A sedimentação siluriana na bacia desenvolveu-se sobre embasamento continental durante a estabilização da plataforma Sul-americana (Almeida e Carneiro, 1998). O lineamento transbrasiliano, a leste e sul, exerceu forte influência sobre o arcabouço estrutural, contribuindo na formação de *rifts* mais antigos. A origem da subsidência desta bacia ainda é muito controversa e alguns autores defendem que ela esteja relacionada à deformação *sin* e *pós* orogênica do Ciclo Brasileiro (Milani, 2000; Zalán, 1990). Outros trabalhos consideram a origem termal (Galushkin, 1997). Considera-se que o mais provável seja uma combinação de uma subsidência mecânica devido a um *rift* inicial com posterior predomínio de uma subsidência termal (Milani e Zalán, 1999; Castro et al., 2016). A exploração de hidrocarbonetos na Bacia do Parnaíba teve início na década de 50 pelo Conselho Nacional do Petróleo (CNP) e posteriormente pela Petrobrás. Após a década de 70 outras empresas participaram de novas fases da pesquisa na bacia. Ao longo de toda a campanha exploratória foram perfurados 89 poços, adquiridos cerca de 25500km lineares de seções sísmicas 2D e 294km<sup>2</sup> de sísmica 3D. Atualmente existem 3 campos de gás em produção na região centro oeste da bacia, os campos de Gavião Azul, Gavião Real e Gavião Branco, que produzem cerca de 4,7 milhões de m<sup>3</sup> de gás (ANP, 2015).

Sotopostos às unidades da Bacia do Parnaíba, ocorrem sedimentos pelíticos de idade proterozóica, denominados como unidade Riachão, considerada por alguns autores parte do embasamento (Caputo et al., 2005).

Por sua vez, o Grupo Jaibaras, de idade Neoproterozóica a Ordoviciano, também considerado também como uma bacia distinta por alguns autores (de Castro et al., 2016) é identificado em seções sísmicas nas áreas leste e nordeste, que segundo Vaz et al (2007) estaria relacionado a gênese da bacia.



## Bacia do Parnaíba



Figura 1: Mapa do relevo da bacia do Parnaíba com limites da bacia em vermelho (Julio Ardito, 2019).

A bacia do Parnaíba é constituída por três grandes grupos, sendo o Serra Grande, de idade Siluriana, sobreposto pelo grupo Canindé, Devoniano e o grupo Balsas de idade Carbonífera-Triássica (Figura 2).

### 2.1.1 Grupo Serra Grande

O Grupo Serra Grande ocupa a porção basal da Bacia do Parnaíba e aflora nas proximidades das bordas nordeste, leste e sudeste (Figura 2).

A "seção" Serra Grande foi primeiramente estudada por Small (1914) e então, classificada por ele como uma "série", sendo posteriormente reclassificada por Campbell et al.(1949) como uma "formação" constituída por arenitos Carboníferos (Pensilvaniano). Mesner e Woodrigde (1964) atribuíram idade siluriana à formação e



a dividiram em Membro Inferior constituído por arenitos e conglomerados e Membro Superior composto por siltitos e folhelhos.

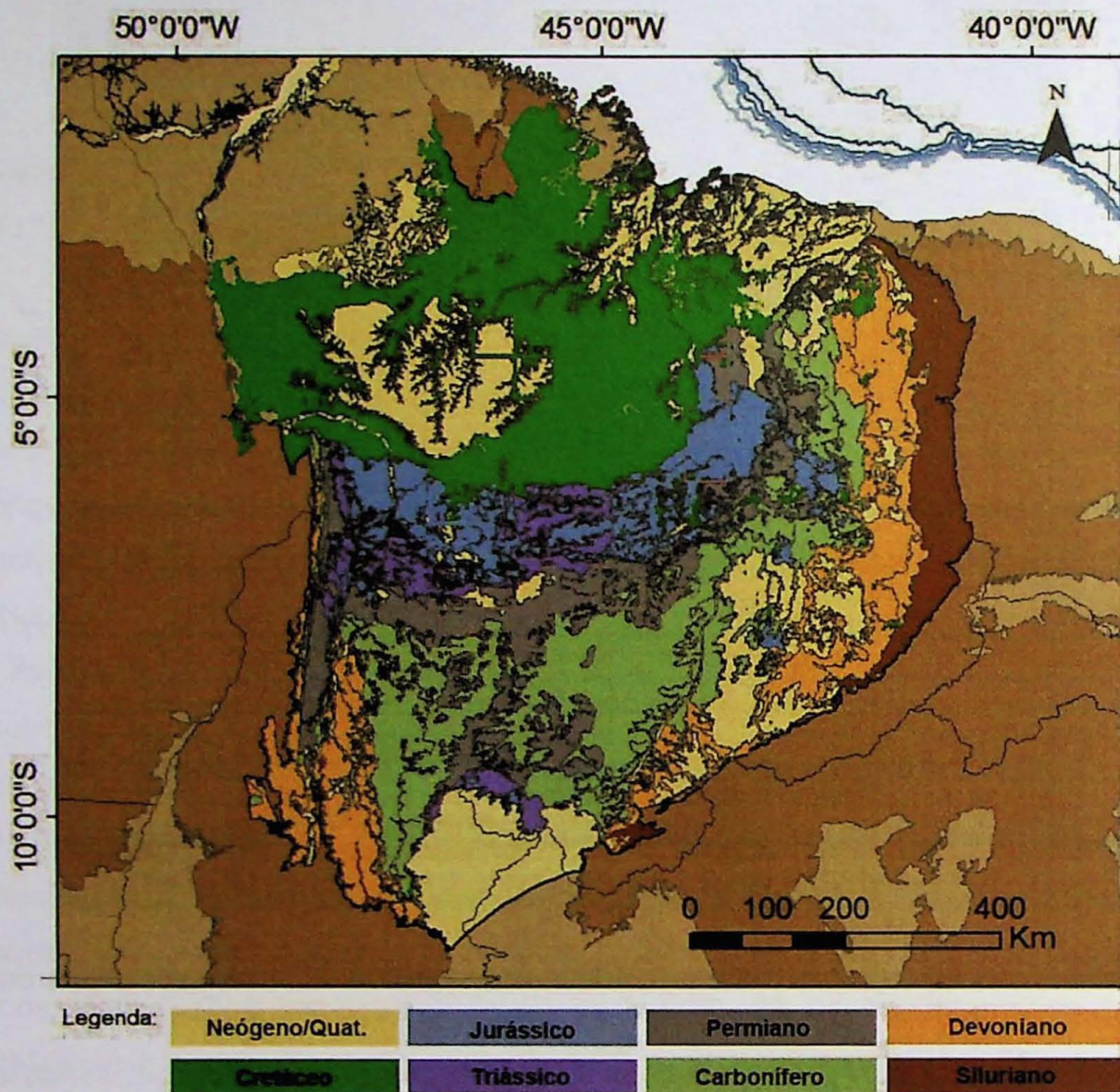


Figura 2: Mapa Geológico da Bacia do Parnaíba realçando o Grupo Serra (Siluriano) na borda leste (Modificado de Bizzi et al., 2003).

Posteriormente, Carozzi et al. (1975) promoveram esta unidade à categoria de "grupo", subdividindo-o em quatro unidades litológicas silu-devonianas mapeáveis em toda a bacia a partir da base, as formações Ipú, Tianguá, Itaim e Mirador. Caputo e Lima (1984) alteraram os limites da formação Mirador considerando apenas Ipu (basal), Tianguá (intermediária) e Jaicós (superior) como formações pertencentes ao grupo Serra Grande. Um contato discordante separa a base do grupo Serra Grande



do embasamento metamórfico e da sequência molássica constituída pelas formações Riachão e Mirador (Góes e Feijó, 1994).

Quadros (1982) sintetizou a evolução histórica das diferentes nomenclaturas, descrições, idades, unidades e colunas estratigráficas usadas por cada autor na Bacia do Parnaíba (Figura 3).

Uma das primeiras descrições de que se tem registro sobre a denominada "camada" Ipú foi feita por Campbell et al.(1949), quando descreveu arenitos arcózios de granulação grossa e conglomerados basais em sua composição. No trabalho citado os autores sugeriram que esses sedimentos foram depositados em ambiente deltaico. Caputo e Lima (1984) fizeram uma descrição mais detalhada e registraram conglomerados, arenitos de granulometria fina a grossa com estratificação cruzada depositados em ambiente fluvial entrelaçado e diamectitos peri-glaciais, ocorrendo principalmente na região nordeste da bacia, onde a espessura máxima é de cerca de 300m.

O Membro Tianguá, assim proposto inicialmente por Rodrigues (1967) para designar uma unidade litoestratigráfica intermediária composta por arenitos finos, siltitos e folhelhos foi posteriormente elevado à categoria de "formação" por Carozzi et al.(1975). Esta formação também é inicialmente encontrada na borda nordeste da bacia e apresenta espessuras variando entre 25m e 270m. Além disso, são descritos como sedimentos gerados em ambiente marinho raso, com idades obtidas por microfósseis na Formação Tianguá compatíveis às bacias silurianas (Caputo e Lima, 1984).

O termo Jaicós foi inicialmente proposto por Plummer (1946) para designar os arenitos e conglomerados do topo da Serra Grande. Carozzi et al.(1975) classificaram esta seção como "formação".

Caputo e Lima (1984) a descreveram como constituída por arenito médio a conglomerático, cinza esbranquiçado, com estratificação cruzada, mal selecionado, depositado em leques aluviais e fandelts. O quadro da Figura 3 mostra a evolução cronológica na interpretação do Grupo Serra Grande.

Góes e Feijó (1994) acrescentaram que estes sedimentos foram depositados por sistemas fluviais entrelaçados de idade neo-siluriana. Embora houvesse uma divergência entre alguns autores a respeito das idades para a Formação Jaicós, GRAHN e CAPUTO (1992) identificaram microfósseis de idade siluriana e esta é hoje a idade mais aceita.



Santos et al. (2004) descreveram esta formação, estudando diversos afloramentos nas bordas leste e sudeste da bacia e notaram a presença de uma camada basal composta por arenitos finos a médios com laminações cruzadas (*ripples*) e estratificação cruzada planar sobreposta por uma camada intermediária de conglomerados glaciais sustentados por clastos com estratificação cruzada acanalada e logo acima arenitos grossos com níveis cascalhosos e estratificação cruzada acanalada. Com base nessas descrições, eles propuseram um modelo faciológico caracterizado por uma rede de canais instáveis onde se desenvolvem formas de leito de cascalho de sistemas fluviais associados a leques aluviais glaciais.

Embora haja, atualmente, um consenso sobre a idade siluriana dos sedimentos do Grupo Serra Grande, alguns estudos chegaram a contestar essa idade (Petri e Fulfaro, 1983; Quadros, 1982). Por sua vez, Caputo e Lima (1984) estudaram a distribuição biocrológica dos quitinozoários na formação Tianguá e revelaram uma concentração de espécies de idade eo-siluriana.

A “sequencia” siluriana corresponde ao Grupo Serra Grande e consiste de um ciclo transgressivo-regressivo completo representativo da primeira incursão marinha na bacia (GÓES; FEIJÓ, 1994). A Figura 4 apresenta o diagrama litoestratigráfico da bacia do Parnaíba.

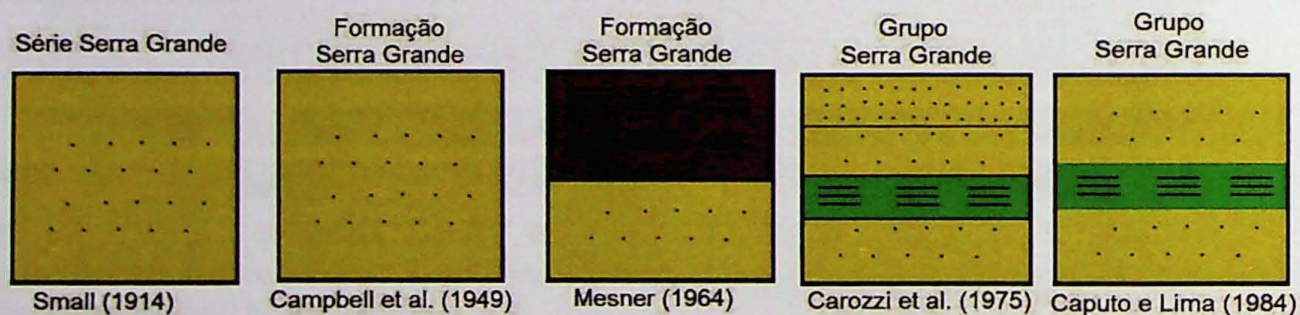


Figura 3: Representação esquemática da revisão litoestratigráfica do Grupo Serra Grande. (Modificado de Quadros, 1982).



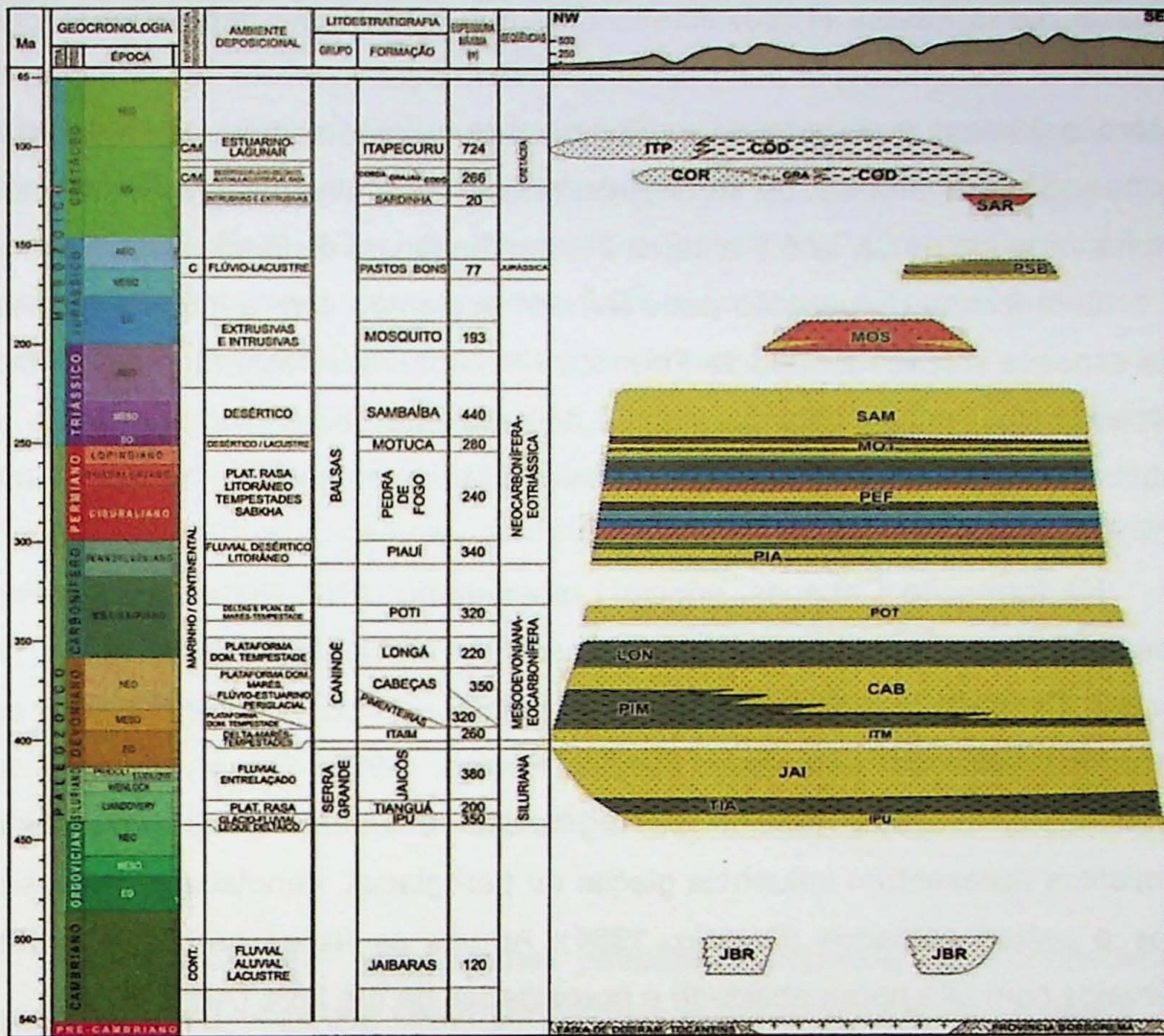


Figura 4: Diagrama litoestratigráfico da Bacia do Parnaíba. (Modificado de Vaz et al., 2007).

### 2.1.2 Grupo Canindé

O grupo denominado Canindé por Rodrigues (1967) continha inicialmente as formações Pimenteiras, Cabeças e Longá. As formações Itaim e Poti foram posteriormente incluídas no grupo por Caputo e Lima (1984) e Góes (1992) respectivamente. Os trabalhos de Góes e Feijó (1994) e Caputo (1992) sintetizaram a descrição litoestratigráfica e os ambientes deposicionais de cada formação.

A Formação Itaim, proposta por Kegel (1953), foi descrita por Caputo (1994) como arenitos finos, esbranquiçados, bem selecionados, com alta esfericidade, além de folhelhos cinza de ambientes deltaicos e plataformais dominados por correntes induzidas por processos de marés e de tempestades (Vaz et al. 2007).

A Formação Pimenteiras estudada inicialmente por Small (1914) foi descrita e classificada por diversos autores como uma camada predominantemente pelítica, apresentando folhelhos cinza a pretos representando a principal ingressão marinha



na bacia. Della Fávera (1990) acrescentou que o ambiente deposicional foi nerítico plataformal dominado por tempestades. A formação é muito radioativa e rica em matéria orgânica, apresentando teores de carbono orgânico que variam entre 0,24 e 4,41% nos poços analisados. Rodrigues, (1995) determinou valores de refletância de vitrinita variando de cerca 0,7 a cerca 2% em função da distância até a soleira.

O processo de geração pode ter sido acelerado com a intrusão pervasiva de uma espessa soleira na base da Formação Pimenteiras. Desta forma, o calor gerado pela soleira contribuiu para o aumento do gradiente geotérmico, induzindo assim o craqueamento do querogênio e alcançando a janela de geração de gás. (Rodrigues, 2007; Miranda, 2014).

Na formação Cabeças, principal reservatório (ANP, 2015), predominam dois tipos de fácies, arenitos médios a grossos de ambiente fluvial (Caputo, 1994) e arenitos finos, bem selecionados de ambiente nerítico plataformal com ação de correntes induzidas por marés (Della Fávera, 1982, Goes e Feijo, 1994) e estratificação cruzada tabular ou sigmoidal. Eventuais fácies compostas por diamictitos apresentam influência glacial ou peri glacial, denotada pela presença de tilitos e seixos estriados (Caputo, 1984). Ambas as fácies areníticas apresentam intervalos com alta permeabilidade e porosidades de até 26% (ANP, 2017).

Albuquerque e Dequech (1946) descreveram siltitos e folhelhos brancos a cinzas de ambiente nerítico plataformal dominado por tempestades na formação Longá. Lima e Leite (1978) descreveram folhelhos cinza escuros a pretos, bem laminados e bioturbados com um pacote de arenitos e siltitos na porção média.

A partir da década de 50 a Formação Poti foi aceita como o intervalo de sedimentos sobrepostos à formação Longá e sotopostos à formação Piauí (GÓES, 1995). Estes depósitos são compostos por arenitos de boa porosidade com intercalações de folhelhos e siltitos. Além de ser um excelente reservatório, a formação Poti é um importante intervalo produtor de gás na porção central da bacia. O ambiente deposicional desses depósitos foi interpretado por Paiva (1937) como deltáicos e de planícies de maré, eventualmente influenciados por tempestades.

Embora a estratigrafia da bacia seja bem conhecido, são raríssimos os trabalhos a respeito do sistema petrolífero, e inexistem estudos detalhados a respeito das propriedades físicas das rochas que compõe os Grupos Canindé e Serra Grande. Além disso, poucos são os estudos sobre o gradiente geotérmico da bacia e suas implicações para o modelo de subsidência da bacia.



### 2.1.3 Grupo Balsas

Piauí é a formação mais basal do Grupo Balsas. Lima e Leite (1978) descreveram arenitos médios com estratificação cruzada e intercalação de folhelhos de ambiente fluvial com eventuais incursões marinhas e contribuições eólicas.

Gois e Feijó (1994) descreveram a Formação Pedra de Fogo como calcários oolíticos intercalados com arenitos finos a médios, siltitos e folhelhos de idade permiana depositados em ambiente marinho raso a litorâneo.

A Formação Motuca, segundo Lima e Leite (1978) e Góis e Feijó (1994) é composta por siltitos vermelhos a marrom, arenitos finos a médios e eventualmente são encontrados folhelhos, anidrita e calcários de ambiente desértico com associação de lagos.

## 2.2 Propriedades de Rocha e Métodos de Medição

A compreensão do sistema petrolífero depende do conhecimento das propriedades físicas que controlam as rochas associadas a este sistema. Dentre as principais propriedades destacam-se porosidade, permeabilidade, resistividade, volume de argila, saturação de água e de fluidos.

Nesta seção será apresentada uma breve descrição das principais propriedades petrofísicas da rocha, bem como as ferramentas de medição e em seguida serão descritos os métodos para processamento, análise e predição utilizados neste trabalho.

A perfilagem é o processo de registro de propriedades físicas, químicas e elétricas da rocha ao longo do poço. Os perfis ou *logs* de perfilagem gerados são os registros contínuos realizados a partir da superfície do poço por meio de sensores elétricos, radioativos ou sônicos. A partir dos dados obtidos são realizadas correções e conversões apropriadas e o perfil é usado para inferir determinadas propriedades petrofísicas.

Diversos trabalhos descrevem os equipamentos e formas de aquisição dos dados e interpretações (Schlumberger, Log Interpretation Chartbook, 1995; Crain, 2000),



Nesta seção será feita uma breve descrição dos principais logs obtidos na perfilagem e um breve resumo dos principais registros.

### 2.2.1 Caliper

O perfil caliper (cal) é o registro do levantamento de sensores de distância (“patins”) que detectam a forma das paredes do poço. Por meio dele pode-se determinar o diâmetro em qualquer profundidade do poço e desta forma pode auxiliar no controle de qualidade de outros perfis, já que pode revelar alterações na parede do poço como possíveis irregularidades ou até zonas de arrombamento/reboco, onde parte da parede do poço pode ter sofrido desmoronamento ou preenchimento. O perfil caliper também é utilizado como base para o desenvolvimento do programa de cimentação e revestimento do poço, além de servir como base para a interpretação de dados adquiridos por equipamentos que utilizam patins que correm na parede do poço e que necessitam que haja a integridade deste, como o equipamento de microresistividade.

### 2.2.2 Raios Gamma

O perfil gama (GR) é obtido com auxílio de uma ferramenta composta por uma fonte gama e um detector Geiger que medem a radiação natural das rochas, desta forma, rochas com mineralogia composta por maior teor de urânio, potássio e tório, tendem a apresentar os maiores índices gama. Isso ocorre em rochas como arenitos micáceos, feldspáticos e argilosos ou em rochas argilosas (folhelhos). Uma das vantagens desta ferramenta é a capacidade de detectar a radioatividade natural de rochas, mesmo se aplicada em poços com as paredes revestidas com aço ou cimento.

O método foi desenvolvido e utilizado a primeira vez pela empresa Lane Wells em 1936 e foi aprimorado e padronizado para registro em unidades API após 1960 com o conhecimento prévio de litologias de um poço em Houston.

As principais aplicações deste perfil incluem a determinação da litologia, do volume de argila, a correlação poço a poço e a correlação com profundidade, para auxiliar a calibração de outras ferramentas de perfilagem.



### 2.2.3 Densidade

A ferramenta usada para medir densidade da rocha foi desenvolvida em 1957. Ela se baseia na absorção de raios gama por efeito Compton e para isso a ferramenta conta com uma fonte química de raios gama na base e um contador (cintilômetro) instalado acima da fonte. A maior parte dos raios gama emitidos pela fonte são absorvidos pela rocha, porém a parte restante é captada pelo contador e é inversamente proporcional a densidade eletrônica da rocha em análise.

Os resultados do *log* de densidade são calibrados para o arcabouço da rocha e usados para determinar a porosidade da rocha, conforme mostra a equação 2.1, uma vez que a densidade é proporcional à porosidade.

$$\phi_D = \frac{\rho - k_d}{1 - k_d} \quad (2.1)$$

Onde o fator  $K_d$  varia conforme a litologia e  $\rho$  é a densidade. Os valores de 2,65 g/cm<sup>3</sup> são assumidos para arenitos, 2,71g/cm<sup>3</sup> para carbonatos e 2,87 g/cm<sup>3</sup> para dolomitos.

Além do parâmetro densidade, a ferramenta é capaz de registrar a energia de retorno dos raios, que pode ser usada como função do fator ou efeito fotoelétrico (PE). O Efeito fotoelétrico é um perfil utilizado como indicativo para determinação da mineralogia.

### 2.2.4 Neutrão

A ferramenta com uma fonte de nêutrons na base e um contador instalado mais acima foi desenvolvida em 1945, pela empresa *Lane Wells*. Quando usada durante a perfuração de um poço ocorre um bombardeamento de nêutrons na rocha e o sensor registra aqueles elementos que retornam após colidir contra núcleos de átomos, na grande maioria hidrogênio. O hidrogênio normalmente encontra-se presente nos poros, sob a forma de água, óleo ou gás e assim é possível estabelecer uma relação entre os nêutrons identificados e a porosidade da rocha.

### 2.2.5 Microresistividade

O ensaio denominado micro resistividade é uma versão dos métodos geofísicos eletroresistivos adaptada para medir a resistividade das rochas em 1D na parede de um poço. Ela se baseia nas leis de Ohm, na qual um par de eletrodos injeta uma corrente constante e outro par mede a diferença de potencial (ddp). A



relação entre a resistividade da rocha pode ser determinada por meio das equações 3.2 e 3.3.

$$R = \frac{V}{I} \quad (3.2)$$

e

$$\sigma = R * \frac{L}{A} \quad (3.3)$$

Onde  $R$  é a resistência em ( $\Omega$ ),  $V$  a tensão em (volts),  $I$  é a corrente injetada em (Amperes),  $L$  é o comprimento percorrido pela corrente em metros,  $A$  é a área da seção em metros quadrados e  $\rho$  é a resistividade da formação em ohm.m.

Desta forma é possível calcular a resistividade da rocha em até 1,5 polegada da parede do poço e então elabora-se um perfil conhecido como microlog ou microlaterallog, que abrange a variação da resistividade, em ohm.m, ao longo do poço.

Usualmente é usado na indústria um segundo par de eletrodos de potencial mais afastado capaz de medir a resistividade da rocha um pouco mais profunda, entre 3 e 6 polegadas da parede do poço.

#### 2.2.6 Porosidade

Porosidade é o principal parâmetro na avaliação da qualidade de um reservatório, principalmente pela contribuição em quantificar reserva (Elkatatny et. al, 2018) e pode ser definida como a razão entre o volume de espaços vazios e o volume total de uma rocha. Este parâmetro pode ser classificado em porosidade primária, ou seja, os poros que foram formados durante o processo de deposição sedimentar e a secundária, ou seja, os poros gerados por processos pós deposicionais. Além disso, a porosidade também é classificada em relação a sua abrangência, neste caso em porosidade total (equação 2.4) e porosidade efetiva (equação 2.5). A primeira é atribuída ao volume total de poros e a segunda ao volume de poros interconectados, ambas divididas pelo volume total da rocha ou volume de Bulk.

$$\Phi_{total} = \frac{\text{Volume de poros}}{\text{Volume total}} \quad (2.4)$$

e

$$\Phi_{efetiva} = \frac{\text{Volume dos poros interconectados}}{\text{Volume total}} \quad (2.5)$$



Existem diversos tipos de porosidades em arenitos, alguma delas mais frequentemente associadas à porosidade primária, concomitante à sedimentação e preservação durante os processos de compactação. Outras geralmente associadas à porosidade secundária, geradas por fraturas ou dissolução, podendo ocorrer sob a forma intergranular ou cristalina. Além disso, ocorrem também porosidades em reservatórios carbonáticos, como porosidade intra e intergranular, móldica, vugular, fenestral e caverna (Choquette & Pray, 1970), porém essas não serão discutidas neste trabalho.

De maneira geral, os fatores que controlam a porosidade primária em rochas siliciclásticas são grau de seleção, esfericidade, empacotamento, tamanho de grãos e trama. Para uma má seleção granulométrica dos grãos, a tendência é uma diminuição da porosidade. Por outro lado, uma maior granulometria e uniformidade no tamanho dos grãos tende a produzir porosidades com gargantas de poros maiores.

No caso da porosidade secundária, os fatores de controle são principalmente cimentação, compactação, dissolução de grãos, fraturas e conteúdo de argilominerais. A porosidade, em arenitos, tende a ser menor em rochas mais antigas e profundas devido às tensões de sobrecarga na rocha que geram um maior empacotamento, além da cimentação.

Para estimar a porosidade de uma rocha através de perfis, pode-se usar os registros de densidade (equação 2.6), neutrão e sônico, ou a combinação deles com eventuais correções. Já em laboratório os principais métodos para a determinação da porosidade de rocha envolvem equipamentos como o porosímetro a gás ou a mercúrio, o microtomógrafo e a ressonância magnética nuclear (RMN).

$$\phi = \frac{\rho_{ma} - \rho_{bulk}}{\rho_{ma} - \rho_{fl}} \quad (2.6)$$

onde  $\rho_{ma}$  é a densidade de matriz,  $\rho_{fl}$  é a densidade de fluidos e  $\rho_{bulk}$  é a densidade de bulk.



### 2.2.7 Permeabilidade

A capacidade de percolação de um fluido através da rocha, meio ou substância é chamada de permeabilidade e usualmente é medida em mili Darcy (mD). Ela representa o grau de conectividade entre poros, relaciona a taxa de fluxo de um fluido com o gradiente de pressão e viscosidade do fluido e é controlada principalmente pelo tamanho médio e distribuição das gargantas de poros (Schon, 2011).

A permeabilidade absoluta é uma característica física da rocha e é representativa do fluxo de um único fluido. Já a permeabilidade relativa de uma rocha para óleo, gás ou água é uma função da permeabilidade absoluta e da viscosidade do fluido e geralmente está relacionada à medida do fluxo de um fluido na presença de outro fluido e pode ser medida em amostras de rochas no laboratório, ou em um poço.

A permeabilidade está relacionada ao volume do fluxo, a viscosidade e ao gradiente de pressão por meio da equação de Darcy:

$$Q = \frac{k.A.\Delta p}{\mu.L} \quad (2.7)$$

onde:

$k$  é a permeabilidade em mD,  $\Delta p/L$  representa o gradiente de pressão (ATM),  $\mu$  é a viscosidade (cp),  $A$  é a área de seção ( $\text{cm}^2$ ) e  $Q$  é o volume do fluxo ( $\text{cm}^2/\text{s}$ ).

A permeabilidade pode ser estimada por diversos métodos, *in situ*, ou em laboratório, dependendo da escala, sendo os métodos diretos mais comuns: medidas com o permeômetro portátil, análises em permeômetro de precisão e teste de poço como teste de formação e de pressão por isolamento. Por meio do uso de logs de perfilagem é possível e muito usual a obtenção da permeabilidade de maneira indireta.

### 2.2.8 Volume de Argila

O volume de argila, também conhecido como  $V_{\text{clay}}$  ( $V_{\text{cl}}$ ) ou  $V_{\text{shale}}$  ( $V_{\text{sh}}$ ) é um dos fatores que têm grande influência sobre a porosidade e a permeabilidade e desta forma também está diretamente relacionado à qualidade do reservatório. Este parâmetro é normalmente expresso em porcentagens ou frações decimais e em



reservatórios porosos pode ser calculado pelo registro do perfil gamma (GR) como mostra a equação abaixo:

$$V_{sh} = \frac{GR_{log} - GR_{min}}{GR_{max} - GR_{min}} \quad (2.8)$$

Um reservatório com grande volume de argila tem sua produção suprimida, porém uma pequena quantidade de argila disseminada nos poros pode contribuir para o trapeamento da água intersticial, estimulando a produção em reservatórios com alta saturação de água (Asquith, 1992).

A argila contida na rocha pode ser classificada como estrutural, laminar ou dispersa. As duas últimas afetam significativamente a porosidade do reservatório, fechando as gargantas de poros e reduzindo a permeabilidade no sistema.

Correções de porosidade devem ser aplicadas para estimar a porosidade efetiva em arenitos com fração de argilominerais. O volume de argila ( $V_{clay}$ ) pode ser calculado por meio do log de gama (GR), ou por crossplots do perfil de densidade (RHOB) e porosidade neutrônica (NPHI) (Schlumberger, Log Interpretation Chartbook, 1995; Interactive Petrophysics IP help manual, 2011).

### 2.2.9 Saturação de Água

A saturação de água, usualmente denominada  $S_w$  pode ser definida pela razão do volume de água ou do fluido pelo volume total de poros, conforme mostra a equação 2.9. Da mesma forma pode-se determinar o volume de outro fluido existente no reservatório.

$$S_w = \frac{\text{Volume do fluido}}{\text{Volume de poros}} \quad (2.9)$$

Sendo que a soma da saturação de todos os fluidos na rocha é 1, ou seja,  $S_w + S_{gas} + S_{oleo} = 1$ .

Grandes partes dos modelos para determinação da saturação provêm dos trabalhos de Archie (1942), nos quais, em laboratório foi determinado que, a resistividade específica da rocha saturada com água  $R_0$  é proporcional à resistividade elétrica específica da água  $R_w$  (equação 2.10).

Desta forma:

$$F = \frac{R_0}{R_w} \quad (2.10)$$

onde  $F$  é o fator de formação.



Na situação onde a água ou a salmoura nos poros é o único condutor é possível relacionar o fator de formação com a porosidade (equação 2.11).

$$F = \frac{R_0}{R_w} = \frac{a}{\phi^m} \quad (2.11)$$

onde  $m$  é o expoente de cimentação

Para rocha livre de argilosidade pode se calcular o índice do fator de formação por:

$$I = \frac{R_t}{R_0} \quad (2.12)$$

e desta forma estimar o expoente de saturação ( $n$ ):

$$S_w = \left(\frac{1}{I}\right)^{\frac{1}{n}} \quad (2.13)$$

Rearranjando as últimas equações chega-se a uma forma mais usual para a equação de Archie para "água":

$$S_{wa} = \left(\frac{A.R_w}{\phi^m.ResDeep}\right)^{\frac{1}{n}} \quad (2.14)$$



### 3. METODOLOGIA

Neste capítulo serão apresentados os métodos e técnicas utilizados neste trabalho para o processamento de dados, análise, modelagem e predição dos parâmetros dos reservatórios e geradores da bacia do Parnaíba.

Todas as técnicas alternativas propostas neste trabalho utilizam software livre. Em geral, nesta tese foi adotada a linguagem de programação *Python*, na versão 3.7, com uso dos módulos *Pandas*, *SciKitLearn*, *Matplotlib*, *Numpy* e *Seaborn*.

#### 3.1 Técnicas de Aprendizagem de Máquina

##### 3.1.1 Regressão Linear

A regressão linear é geralmente a base dos métodos mais comuns de análise preditiva. Trata-se de um método de análise estatística capaz de estimar o valor de uma variável dependente (resposta) como função de outras variáveis independentes (preditoras) (Montgomery, et al, 2012). A expressão abaixo ilustra a relação:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_n x_{in} + e_i \quad (3.1)$$

sendo  $y_i$  a variável dependente,  $x_{i1}, x_{i2}, x_{in}$  as variáveis independentes,  $b_0$  o *intercepto* da regressão, ou seja, o valor esperado de  $y$  quando todos os preditores tem valor = 0,  $\beta_1, \beta_2, \dots$ , são os coeficientes estimados da regressão e *epsilon* representa o erro, ou seja, a diferença entre os valores esperados e os valores observados de  $y_i$ .

A regressão indica o grau de significância entre as variáveis dependentes e independentes e o impacto que as variáveis independentes têm sobre a dependente.

A estimativa dos coeficientes é usualmente realizada pelo método dos *mínimos quadrados*. Esta técnica consiste no processo de otimização que busca o melhor ajuste do modelo através da minimização dos quadrados do erro da regressão.

Desta forma, utilizando  $\beta_0$  e  $\beta_1$  para representar os coeficientes da reta  $r$  e  $\varepsilon$  como variável de erro. Assim, a reta de ajuste pode ser escrita como:

$$y_i = \beta_0 + \beta_1 * x_i + \varepsilon_i \quad (3.2)$$

onde  $\beta_0$ ,  $\beta_1$  e  $e_i$  são constantes.



A soma de quadrados pode ser escrita como:

$$SQ(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 * x_i)^2 = \sum_{i=1}^n e_i^2 \quad (3.3)$$

Derivando a expressão acima em relação a  $\beta_0$  e  $\beta_1$  ficamos as expressões:

$$b_0 = \frac{(\sum y) * (\sum x^2) - (\sum x) * (\sum x * y)}{n * (\sum x^2) - (\sum x)^2} \quad (3.4)$$

$$b_1 = \frac{n * (\sum x * y) - (\sum x) * (\sum y)}{n * (\sum x^2) - (\sum x)^2} \quad (3.5)$$

A equação da regressão linear pode ser estimada por:

$$\hat{y} = b_0 + b_1 * x \quad (3.6)$$

### 3.1.2 Redes Neurais Artificiais

As redes neurais artificiais (RNA) consistem em um método preditivo e supervisionado com uma alta diversidade de aplicações, como por exemplo, descoberta de medicamentos, entendimento de linguagem natural, desenvolvimento de carros autônomos e reconhecimento facial.

As redes neurais artificiais funcionam como dispositivos de processamento de informação caracterizados pela interconexão de unidades elementares de processamento (neurônios artificiais) (HAYKIN, 2008). Em geral, o método RNA é ideal para situações em que se têm uma enorme quantidade de dados e problemas complexos.

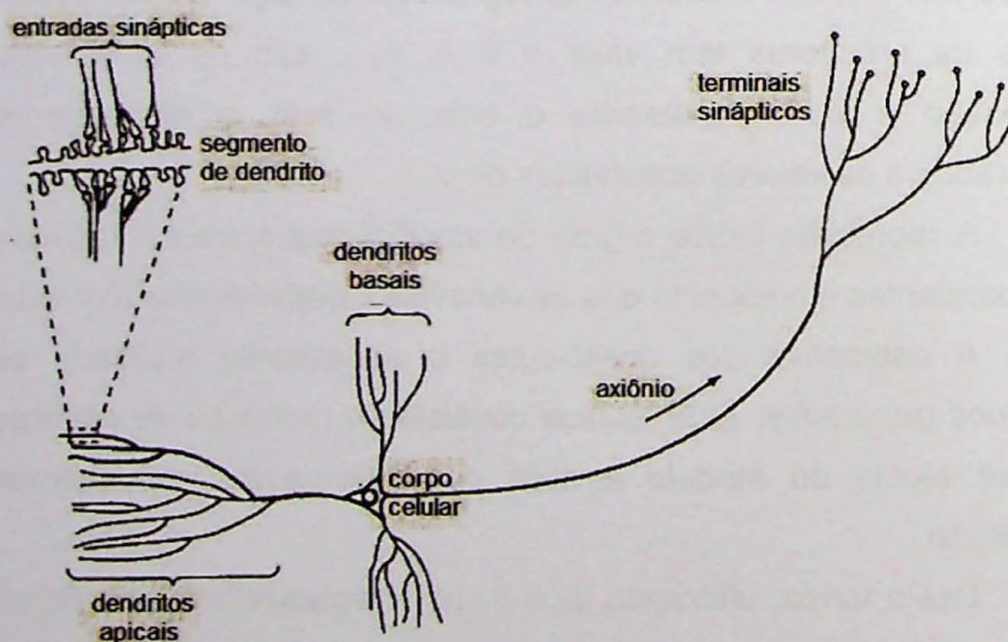


Figura 5: Esquemático de comunicação entre neurônios (modificado de Haykin, 2001).



A RNA é inspirada em redes neurais humanas onde cada neurônio (Figura 5) possui uma rede de conexões e transmitem informações (sinapses) de um neurônio para o outro por meio de axônios. Desta forma, substâncias químicas são lançadas das sinapses e entram pelos dendritos, aumentando ou diminuindo o potencial elétrico do corpo da célula. Caso o potencial seja alto ou maior do que um determinado valor o neurônio é ativado.



Figura 6: Esquemático do funcionamento de uma rede neural artificial.

De forma similar a rede neural artificial é um conjunto de vários neurônios interconectados pelos axônios e que podem ser ativados ou não dependendo do valor do potencial proveniente das sinapses (Figura 6).

McCulloch e Pitts (1949) realizaram os primeiros estudos envolvendo Redes Neurais Artificiais, estabelecendo as bases da neuro-computação por meio do desenvolvimento de algoritmos matemáticos similares ao funcionamento dos neurônios biológicos.

Elementos importantes para o desenvolvimento das redes neurais foram posteriormente descobertos com os trabalhos de Hebb (1949) que propôs que a conectividade do cérebro é continuamente modificada conforme o organismo vai aprendendo tarefas funcionais diferentes e que agrupamentos neurais são criados por tais modificações.

Novos modelos e implementações foram propostas a partir de 1949, sendo a mais relevante sugerida por Rosenblatt (1957), que introduziu o conceito de perceptron, construindo o que seria a base da inteligência artificial.

Atualmente existem diversas variações na arquitetura de uma rede neural como mostra a Figura 7. Na Figura 8 é ilustrado um modelo mais detalhado, com quatro



camadas, sendo duas delas camadas ocultas. Além disso, a imagem mostra os processos conhecidos como *back propagation* e *feed forward*.

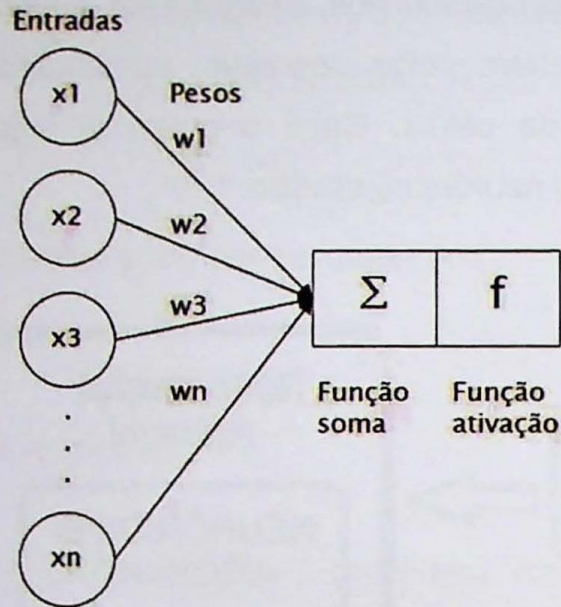


Figura 7: Arquitetura de uma rede neural artificial com apenas uma camada.

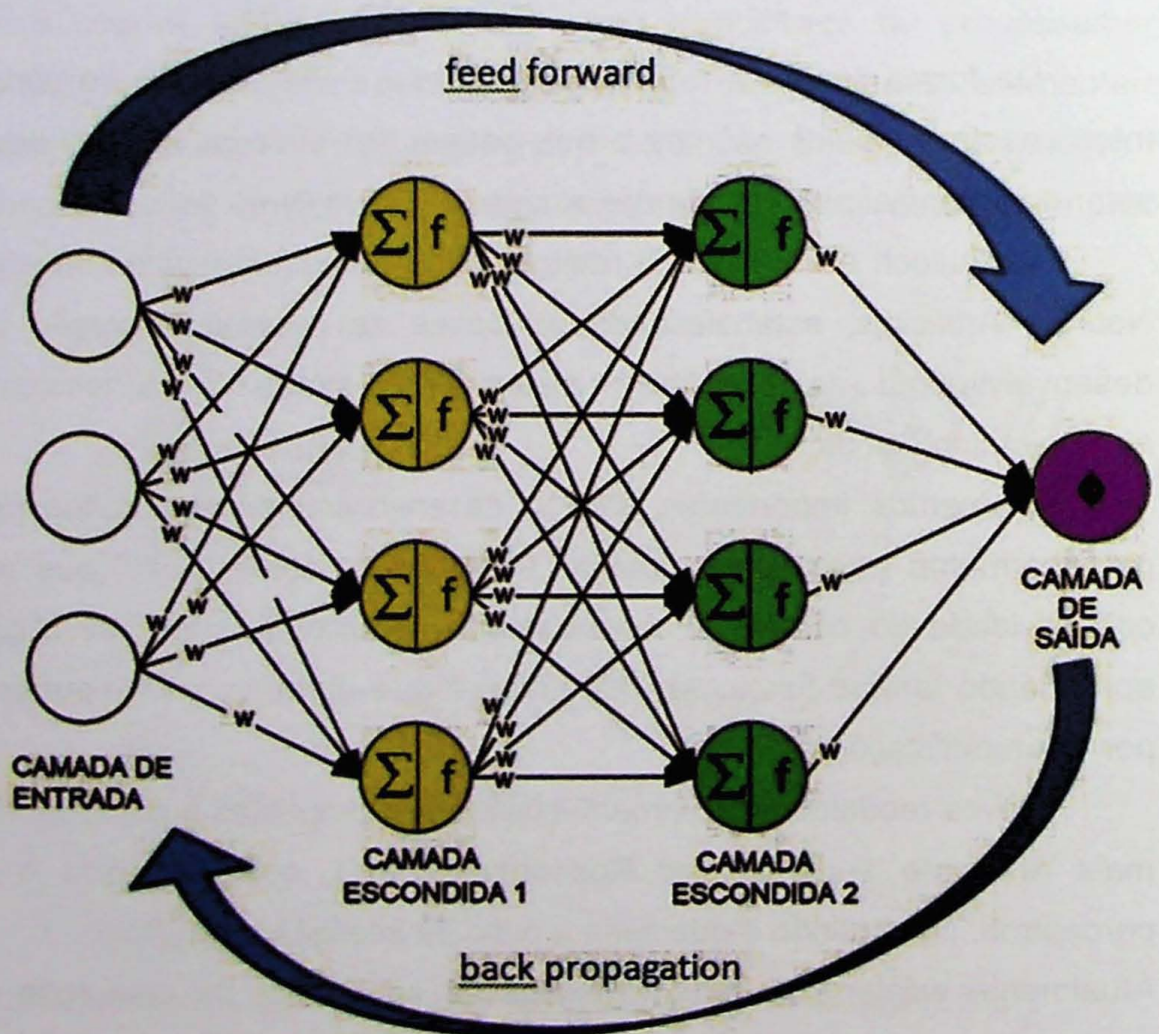


Figura 8: Exemplo esquemático da arquitetura de uma rede neural artificial multicamada.



Para cada entrada existe um valor e um peso que ao ser conectado com um neurônio na camada seguinte passa por uma somatória.

Uma vez calculada a somatória foi então aplicada uma função de ativação em todas as camadas. Após uma série de testes envolvendo uma série de funções, verificou-se que a melhor função para a finalidade deste projeto é a função tangente hiperbólica (Figura 9), conforme mostra a expressão abaixo:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3.7)$$

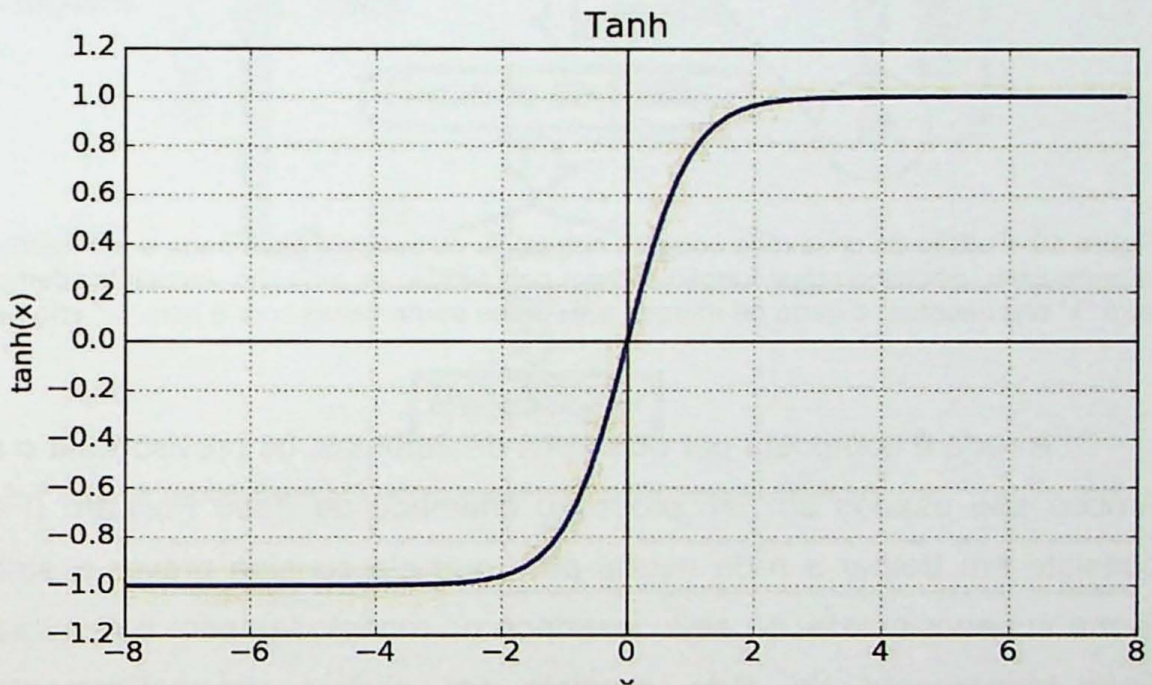


Figura 9: Gráfico da função tangente hiperbólica.

O processo de calcular o valor da somatória e em seguida passar pela função de ativação ocorre de forma sucessiva, cíclica e automatizada, até que a camada de saída aponte um valor final, conforme mostrado na figura 7.

Uma etapa fundamental deste processo é o valor atribuído aos pesos. Um peso positivo é chamado de sinapse excitadora, enquanto um peso negativo é uma sinapse inibidora, desta forma os pesos amplificam ou reduzem o sinal de entrada (Figura 10). O papel da função de ativação é de atuar como um filtro “corta baixa”, ativando um sinal superior a um valor de corte e interrompendo a propagação caso esteja abaixo deste valor.



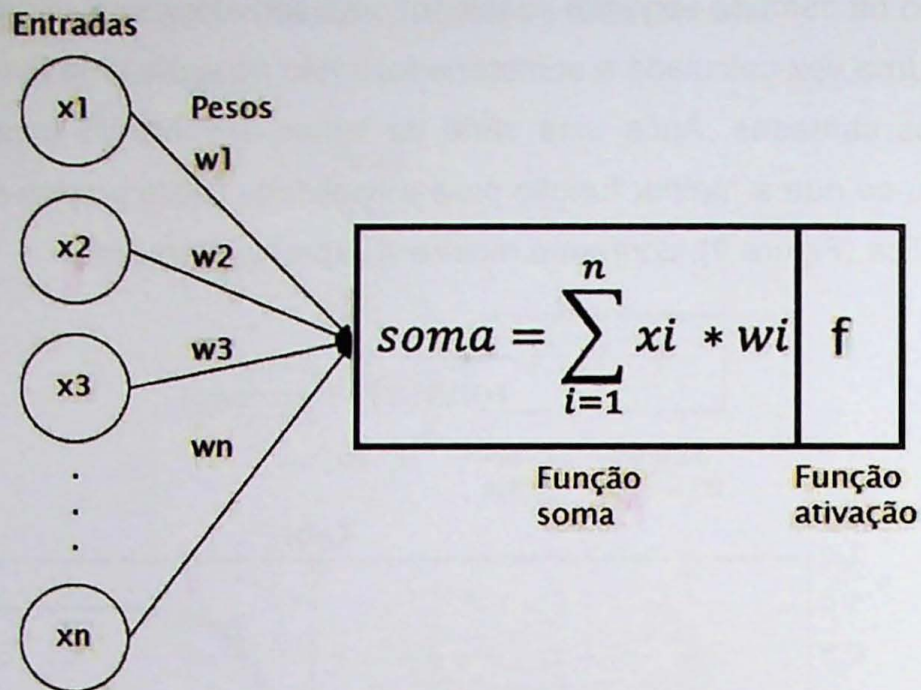


Figura 10: Padrão de conexões entre os neurônios da camada de entrada e um neurônio da camada intermediária, passando pela função soma e pela função de ativação. As células demarcadas com a letra "x" representam o dado de entrada e as setas demarcadas com a letra "w" são os pesos.

A rede é composta por dois tipos de atributos, os previsores e o atributo meta. Ambos são usados em um processo chamado de *Feed Forward* (Figura 11) que consiste em treinar a rede neural para que ela consiga prever o atributo meta de forma supervisionada, ou seja, inserindo no modelo também o resultado conhecido. Este treinamento da rede consiste em ajustar sistematicamente os pesos, comparando com o resultado esperado. Este processo é realizado de maneira iterativa até que os resultados previstos estejam próximos o suficiente o quanto se queira dos resultados esperados. Esse processo de comparação é feito pela diferença entre o valor esperado e o predito.



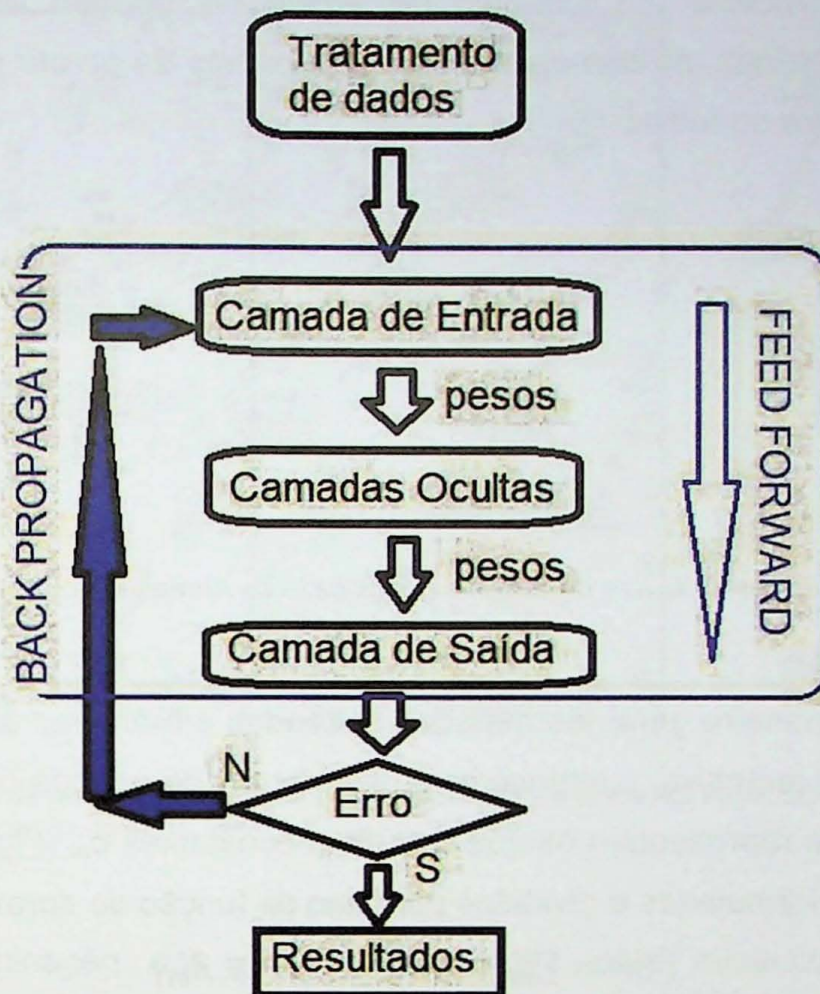


Figura 11: Fluxo de processamento e treinamento da rede neural artificial.

### 3.1.3 Árvores de decisão

Árvores de decisão estão entre os métodos de aprendizagem de máquinas mais promissores e poderosos desenvolvido nas últimas décadas. Trata-se de um método supervisionado capaz de realizar regressões e classificações, de forma simples com alto índice de sucesso (Russell e Norvig, 2003).

As árvores de decisão são estruturas de dados inspiradas em aspectos botânicos de árvores reais e possuem elementos análogos como nós (de decisão), folhas (ou nó terminal) e raízes. Neste contexto, o processo de aprendizagem se assemelha a forma como os seres humanos tomam decisões, ou seja, testes (decisões) são executados em cada nó da árvore, partindo de uma raiz e passando para um nível hierárquico inferior, até atingir um nó folha ou nó terminal, que é a região onde se espera chegar a uma resposta para o problema (Mayrink, 2017). A



Figura 12 mostra um exemplo de árvore de decisão onde a primeira elipse representa a raiz, as demais representam os nós da árvore e os retângulos são os nós terminais ou folhas.

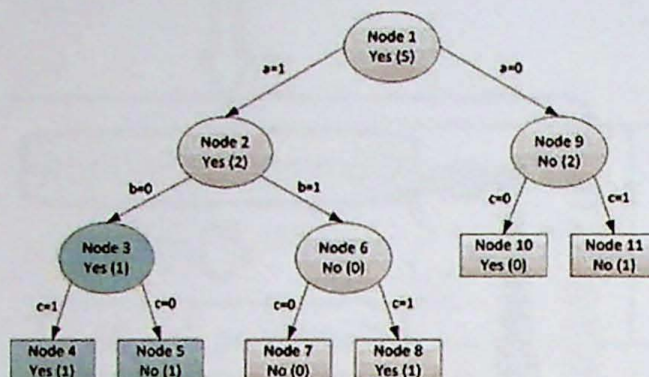


Figura 12: Exemplo de árvore de decisão (modificado de Ahmed et al., 2018).

De maneira geral, os métodos baseados em árvores de decisão particionam, de maneira recursiva, o espaço com os atributos de entrada ( $x_1$  e  $x_2$ ) criando regiões ( $R_1, R_2$ ) que representam os nós e contém constantes  $c_m$  (Figura 13). Tais partições podem ser formuladas e divididas por meio da função de aproximação:

$$\hat{f}(x) = \sum_{m=1}^M c_m I\{(x_1 \text{ e } x_2) \in R_m\} \quad (3.8)$$

onde  $\hat{f}(x)$  representa a predição para a constante  $c_m$ .

O objetivo do algoritmo é automatizar a seleção de atributos e respostas de forma que consiga criar as melhores partições e encontrar a solução ideal para o problema, em outras palavras, busca-se um algoritmo autônomo que estabeleça regras ideais para que a árvore tome as melhores decisões, que a conduzam às melhores predições. Para isso, deve-se minimizar a função custo ou função objetivo:

$$Q(f) = \sum_{i=1}^N [y_i - f(x_i)]^2 \quad (3.9)$$

Existem diversos algoritmos de indução de árvores de decisão, como por exemplo, AID (Morgan e Sonquist, 1963), *Classification and regression tree* (CART), introduzido por Breiman et al.(1984), C4.5 (Quinlan, 1993) e outros igualmente capazes de estabelecer as regras de decisão com base em uma amostra de dados. Neste trabalho foi adotado o algoritmo CART para regressão com uso do algoritmo Gradient Boosting, que será descrito mais adiante.



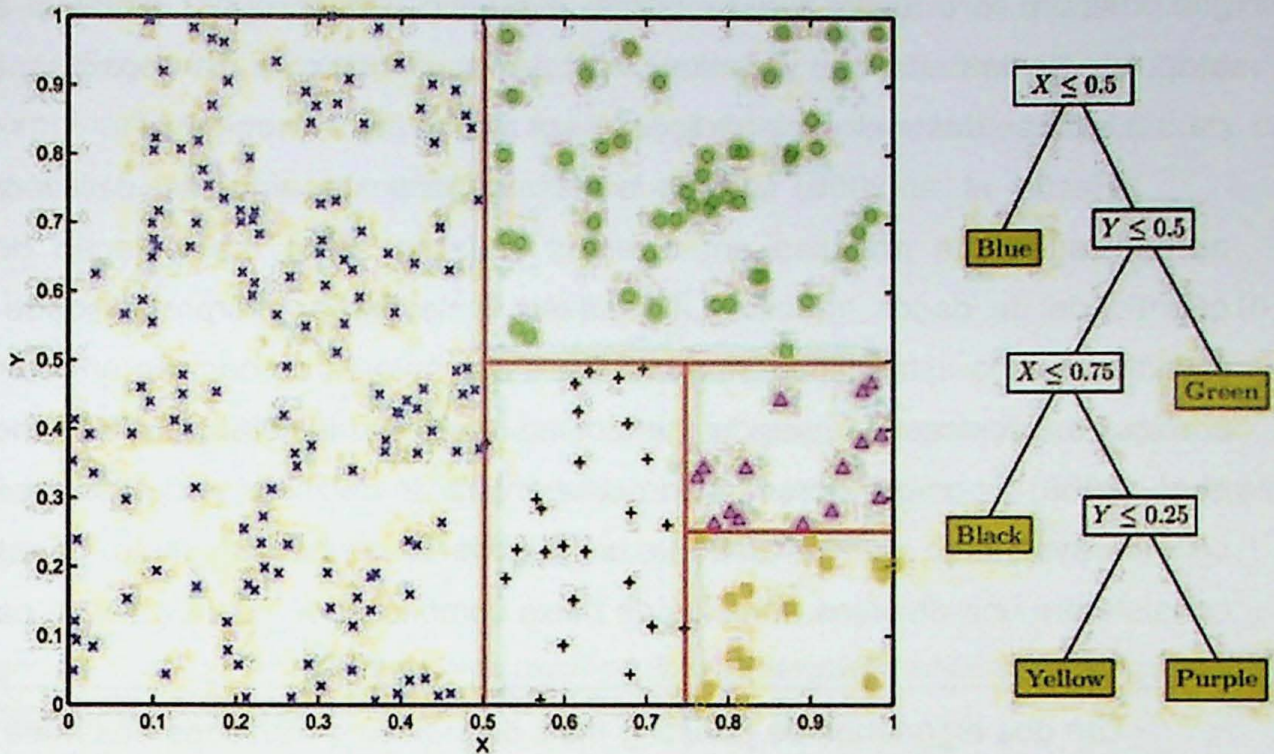


Figura 13: Exemplo de uma árvore de decisão. (a) partição de entradas associada (b) aos nós da árvore. (modificado de Kang et al., 2016).

A determinação das melhores partições pela minimização da função aproximação pelo método dos mínimos quadrados pode ser um problema computacionalmente difícil de ser resolvido (Mayrink, 2016) e por esta razão podem ser adotadas estratégias melhores, como por exemplo, a “Greedy”, sugerida por Hastie (2001; 2009). Ela consiste inicialmente em dividir a amostra em duas partes  $R_1$  e  $R_2$ , sendo as partições determinadas por:

$$R_1(j,s) = \{x | x_j \leq s\} \text{ e } R_2(j,s) = \{x | x_j > s\} \quad (3.10)$$

Desta forma, é necessário determinar os valores das constantes  $s$  e  $j$ . Isso pode ser feito minimizando a equação abaixo:

$$\min_{j,s} \left[ \min_{c_1} \sum_{x_1 \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_1 \in R_2(j,s)} (y_i - c_2)^2 \right] \quad (3.11)$$

Desta forma,  $c_1$  e  $c_2$  dependem da escolha de  $j$  e  $s$  e podem ser determinados diretamente por:

$$\hat{c}_1 = \text{med}(y_i | x_i \in R_1(j,s)) \quad (3.12)$$

e

$$\hat{c}_2 = \text{med}(y_i | x_i \in R_2(j,s)) \quad (3.13)$$



Assim, o trabalho realizado pelo algoritmo é buscar, por tentativa e erro os atributos de entrada que permitam calcular as constantes de aproximação ideais para a criação das melhores partições para a solução do problema.

Hastie et al.(2009) avaliou o desempenho de algumas das técnicas de aprendizagem de máquina em diversos quesitos como manipulação de grande quantidade de dados, robustez à variáveis irrelevantes, interpretabilidade e poder preditivo, sendo esta última um ponto fraco nas árvores de decisão em comparação com outras técnicas de aprendizagem como redes neurais artificiais. Segundo Hastie et al. (2009) o poder de predição de uma árvore de decisões pode ser aperfeiçoado consideravelmente com o uso de um algoritmo conhecido como *boosting*, que consiste em unir diversos modelos de baixa complexidade criando assim os comitês de previsão.

Um dos algoritmos de *boosting* mais difundidos e utilizados nas mais diversas aplicações é o *Gradient Boosting*, primeiramente apresentado por Friedman (2001). Trata-se de um processo iterativo aditivo, em que um novo termo é adicionado ao modelo a cada iteração minimizando os erros de predição (Mayrink, 2016).

Hastie et al. (2009) faz uma apresentação completa da metodologia, aqui serão evidenciadas apenas as passagens fundamentais para a compreensão do método. Conforme mencionado anteriormente o algoritmo realiza uma partição do espaço em regiões  $R_j$ , sendo  $j=1, 2, 3, \dots, J$ , conforme representados pelo nós terminais da árvore. Desta forma,  $Y_j$  é uma constante e a regra de predição é:

$$x \in R_j \rightarrow f(x)=y_j \quad (3. 14)$$

Assim a árvore pode ser representada por:

$$T(x,\theta) = \sum_{j=1}^J y_j I(x \in R_j) \quad (3. 15)$$

onde  $\hat{\theta}=\{R_j, y_j\}$  e pode ser determinado pela minimização da função:

$$\hat{\theta} = \arg \min \quad \sum_{j=1}^J \sum_{x_i \in R_j} L(y_i, \gamma_i) \quad (3. 16)$$

onde  $L(y_i, \gamma_i)$  é a função perda ou *loss function*.

Desta forma, a otimização do problema pode ser dividida em duas partes, a estimativa de  $\gamma_j$  e a de  $R_j$ . Como  $R_j$ ,  $\gamma_j = \bar{y}_j$  uma das aproximações que podem ser usadas para otimizar  $R_j$ , utilizando o critério de otimização é:

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^N \hat{L}(y_i, T(x_i, \theta)) \quad (3. 17)$$



Deste modo, fazendo  $R_j = R_j$ ,  $\gamma_j$  pode ser estimado usando o critério original.

Segundo a premissa de que o boosting é estratégia para “juntar” ou “combinar” uma vasta gama de modelos definidos por Hastie et al.(2001) como “fracos”, a soma destes modelos pode ser definida como:

$$f_M(x) = \sum_{m=1}^M T(x, \theta_m) \quad (3.18)$$

Esta equação é denominada “*forward stagewise manner*” e busca a solução da equação anterior pela adição iterativa de novas funções (ou novas árvores). O algoritmo 1 expressa a implementação computacional para este processo.

Algoritmo 1: Estrutura computacional do algoritmo *Forward stagewise additive modeling*.

**Inicializar  $f_0(x)=0$**

**de  $m=1$  até  $M$  faça:**

Calcule:  $(\beta_m, \gamma_m) = \arg \min_{\beta, \gamma} \sum_{i=1}^N L(y_i, f_{m-1}(x_i) + \beta b(x_i, \gamma))$   
 e  $f_m(x) = f_{m-1}(x) + \beta_m b(x, \gamma_i)$

Em cada passo deste processo, deve ser resolvida a equação:

$$\hat{\theta}_m = \arg \min_{\theta_m} \sum_{i=1}^N L(y_i, f_{m-1}(x_i) + T(x_i, \theta_m)) \quad (3.19)$$

onde as constantes  $\theta_m = \{R_{jm}, \gamma_{jm}\}_1^{j_m}$  da árvore seguinte até o modelo recorrente  $f_{m-1}(x)$ .

A partir das regiões  $R_{jm}$  o valor ótimo das constantes  $\gamma_{jm}$  em cada região é dada pela expressão direta:

$$\hat{\gamma}_{jm} = \arg \min_{\gamma_{jm}} \sum_{i=1}^N L(y_i, f_{m-1}, \gamma_{jm}) \quad (3.20)$$

a equação (9) pode ser numericamente otimizada por qualquer função perda diferenciável. Neste caso, a função  $f(x)$  preditora de  $y$  usada foi:

$$L(f) = \sum_{i=1}^N L(y_i, f(x_i)) \quad (3.21)$$

Nesta etapa, objetiva-se minimizar  $L(f)$  em relação a  $f$ , onde  $f(x)$  é dada pela expressão (3.19). Esta minimização pode ser dada por:

$$\hat{f} = \arg \min_f L(f) \quad (3.22)$$

onde  $f \in \mathbb{R}^N$  são parâmetros da cuncao de aproximacao:

$$f = \{f(x_1), f(x_2), \dots, f(x_n)\} \quad (3.23)$$



A equação 3.37 pode ser otimizada pela soma das componentes veroriais:

$$\widehat{f}_M = \sum_{m=0}^M \widehat{h}_m, \quad \widehat{h}_m \in \mathbb{R}^N \quad (3.24)$$

onde  $\widehat{h}_m$  é a função “passo” ou *step function*, que pode ser resolvida de forma iterativa.

O método “*Steepest Descent*” permite trocar a função passo  $\widehat{h}_m$  por  $-\rho_m g_m$  onde  $\rho_m$  é um escalar denominado de comprimento do passo ou “step length” e  $g_m$  é o gradiente da função perda  $L(f)$  (GUARIDO, 2018). Desta forma, os componentes do gradiente são:

$$g_{im} = \left[ \frac{\partial L(x_i, f(x_i))}{\partial f(x_i)} \right] \quad (3.39)$$

e o comprimento do passo é dado por:

$$\rho_m = \arg \min_{\rho} L(f_{m-1} - \rho g_m) \quad (3.40)$$

onde a solução iterativa é dada pela formula de recorrência:

$$f_m = f_{m-1} - \rho_m g_m \quad (3.41)$$

A ideia central do método “*Steepest Descent*” é fazer com que a função perda diminua rapidamente por meio da iteração acima.

O algoritmo 2 apresenta a estrutura genérica do código para o algoritmo *Gradient Boosting* para regressão. A parte inicial diz respeito a um nó onde é inicializada a constante e os componentes do gradiente negativo são calculados.

Algoritmo 2: Estrutura computacional do método Gradient Boosting(modificado de Hastie et al. 2001).

---

Inicializar:  $f_0(x) = \arg \min_{\gamma} \sum_{i=1}^N L(y_i, \gamma)$   
 de  $m=1$  até  $M$  faça:  
   de  $i=1$  até  $N$  faça:  
      $r_{im} = - \left[ \frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]$   
   ajuste a árvore aos alvos  $r_{im}$ , para regiões  $R_m=1,2,\dots,J_m$   
   de  $j=1$  para  $J_m$  faça:  
     
$$\gamma_{jm} = \arg \min_{\gamma} \sum_{x \in R_{jm}} L(y_i, f_{m-1}(x_1) + \gamma)$$
  
   Atualizar:  
    $f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$   
 Imprima:  $\widehat{f}(x) = f_M(x)$

---



Tabela 1: Funções “perda” e gradientes (modificado de Hastie et al., 2001).

Tipo	Função Perda	$-\partial L(y_i, f(x_i))/\partial f(x_i)$
Regressão	$1/2  y_i - f(x_i) ^2$	$y_i - f(x_i)$
Regressão	$ y_i - f(x_i) $	$ y_i - f(x_i) $
Classificação	Desvio	$I(y_i - G_k) - p_k(x_i)$

Guarido (2018) explicou que ao contrario do “forward stagewise manner”, os parâmetros são calculados a cada iteração pelo ajuste do gradiente na árvore, como mostra a expressão:

$$\hat{\theta}_m = \arg \min_{\theta} \sum_{i=1}^N [-g_{im} - T(x_i; \theta)]^2 \quad (3.42)$$

Desta forma, o gradiente é dado pela derivada parcial da função perda.

### 3.1.4 Máquinas de vetores de suporte (SVM)

As máquinas de vetores de suporte, do inglês *Support Vector Machine* (SVM) consistem em uma técnica de aprendizagem de máquina proposta inicialmente por Cortes e Vapnik (1995), que é capaz de separar e classificar conjuntos de dados. Em geral, ela apresenta melhor desempenho, ou seja, melhor capacidade de predição para modelos de classificação em comparação com outras técnicas. Além disso, a SVM's apresentam boa performance em problemas que envolvem alta complexidade. Outra vantagem da técnica é o fato de não ser muito influenciada por ruídos dos atributos de entrada.

Assim como ocorre com as demais técnicas de aprendizado de máquina, a SVM tem uma gama enorme de aplicações, como por exemplo, processamento sísmico (Contreras e Verdin, 2006), predição de propriedades de rocha (Liu e Sacci, 2003; Liu, 2008), predição de permeabilidades (BAOUCHE; CHAOUCHI, 2015) e geoestatística (LIU, 2009).

A técnica se baseia na criação de hiperplanos de separação e no aprendizado das melhores localizações para estes hiperplanos. A localização do hiperplano é



determinada pela "margem máxima", que é calculada em função da distancia entre o hiperplano de separação e os vetores de suporte. Estes vetores representam a distancia entre cada dado e o hiperplano e desta forma, servem para reconstruir o hiperplano, conforme mostra o esquemático ilustrado na Figura 14. O algoritmo calcula e escolhe o melhor hiperplano de separação, auxiliando na determinação dos suportes. Portanto, o aprendizado do modelo consiste exatamente em encontrar os suportes ideais. Quanto mais espessa for a margem, maior a generalização da previsão que o algoritmo será capaz de fazer.

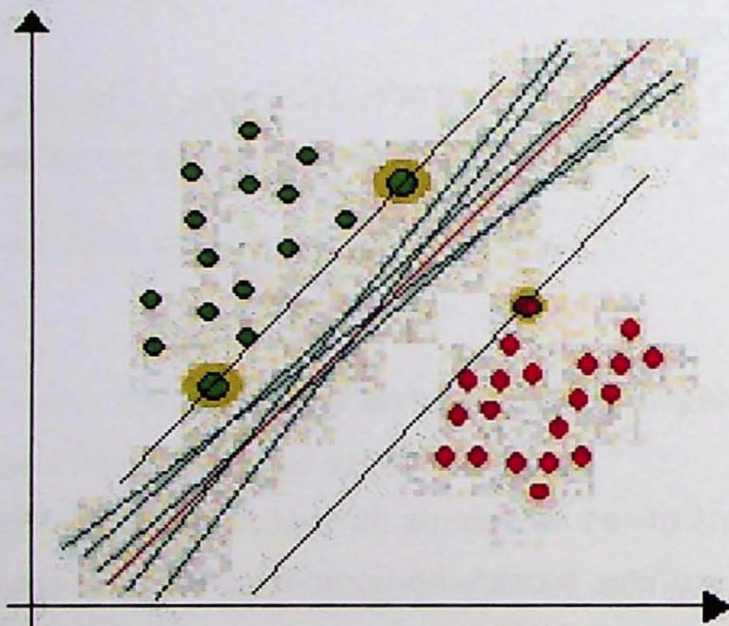


Figura 14: Hiperplanos de separação (em azul) de um conjunto de dados com duas classes. A reta vermelha representa o hiperplano ideal (Julio Ardito, 2019).

Uma das técnicas para a criação do hiperplano é a "envoltória convexa" ou "*Convex hulls*", que consiste no cálculo iterativo da distancia entre polígonos (Figura 15) que contém cada grupo de variáveis categóricas do problema. Outra técnica usada é a chamada "abordagem matemática", que consiste na aplicação de produtos entre matrizes.

A abordagem matemática é mais utilizada no cálculo dos hiperplanos e para isso utilizam a função:

$$\vec{w} * \vec{x} + b = 0 \quad (3.43)$$



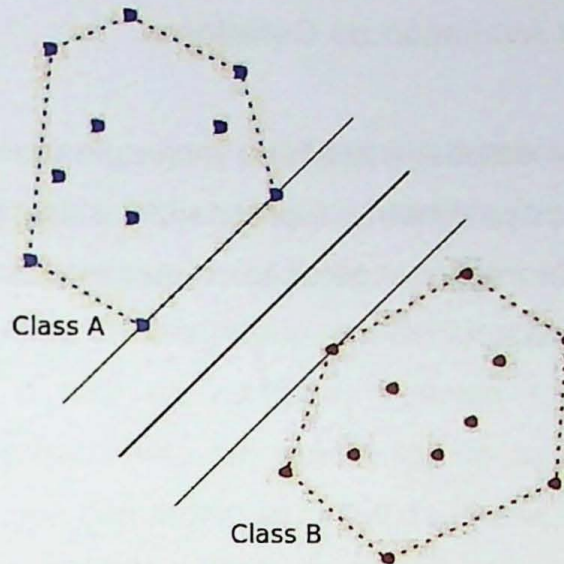


Figura 15: Exemplo de envoltória convexa. (Modificado de Hughes e Correl, 2016).

Um dos conceitos mais relevantes desta técnica é o de *erros e custo*, onde a minimização de erro pode ser determinada pela equação:

$$\min \frac{1}{2} |w|^2 + c \sum_i a_i \quad (3.44)$$

onde  $c$  é a punição por classificação incorreta.

Desta forma, quanto maior o valor de  $c$ , mais eficiente será o hiperplano de separação e menor será o erro.

Outra vantagem das SVM's é a possibilidade de abordar problemas lineares e não lineares. No caso de problemas não lineares a principal técnica utilizada no processo de classificação é o chamado "*kernel trick*".



## 3.2 Técnicas para Avaliação de Geradores

No intuito de estudar e avaliar a rocha geradora, faz-se necessária uma série de análises sobre os parâmetros e propriedades desta rocha que podem impactar na capacidade de maturação e geração. algumas destas análises e técnicas serão descritas nesta seção.

### 3.2.1 Pseudo-Sônico

O método consiste em analisar o impacto que a presença de matéria orgânica na rocha causa nas medidas dos tempos de transito e resistividades. Para isso é feito um re-escalamento no perfil sônico, onde  $50\mu\text{s}/\text{ft}$  equivale a uma década de resistividade e assim os perfis são comparados e as regiões ricas em contudo orgânico realçadas.

Deste modo, são feitos *crossplots* do registro Sônico pela resistividade Desta forma são plotados todos os dados de resistividade em função do perfil sônico e uma reta ajustada, chamada de linha base que seria a reta limítrofe de teor orgânico, ou seja, os pontos acima dela são relevantes para geração.

Esta linha é determinada por:

$$D \log R = \log_{10} (R/R_{\text{base}}) + 0,02 (Dt - Dt_{\text{base}}) \quad (3.45)$$

onde R é a leitura do registro de resistividade. Rbase a zona onde a resistividade é mais baixa, Dt o registro sônico e DTbase são os tempos de trânsito dos níveis mais argilosos.

Henderson (1999) aplicou a técnica e encontrou uma correlação com o metamorfismo orgânico (LOM). Desta forma, o teor de carbono orgânico pode ser calculado por:

$$\text{COT} = (D \log R) 10 (2,297 - (0,1688 \times \text{LOM})) \quad (3.46)$$

Onde COT é o teor de carbono orgânico e LOM (level of organic metamorphism), representando o nível de maturidade.



### 3.2.2 Pirólise

Alguns dos arquivos de poços cedidos pela Agência Nacional de Petróleo e Biocombustíveis (ANP) para este projeto continham resultados de análises geoquímicas de laboratório feitas em amostras de calha, amostras laterais e testemunhos. Estes ensaios de laboratório, conhecidos com Pirólise ou *Rock Eval*, consistiram em medir o teor de carbono orgânico total (COT) expresso em miligramas (mg) de hidrocarboneto por grama (g) de rocha, e também medir os parâmetros S<sub>1</sub>, S<sub>2</sub> e S<sub>3</sub>, que demarcam as fases da janela de geração em função da temperatura (Oliveira et al., 2006) e a temperatura máxima do pico de S<sub>2</sub> (TMAX).

A partir destes resultados, foi calculado o índice de hidrogênio (HI) de acordo com a equação (3.47) expresso em mg de HC/g de COT e o índice de oxigênio (OI) conforme a equação (3.48) expresso em mg de CO<sub>2</sub>/g de COT.

$$HI = S_2 / COT * 100 \quad (3.47)$$

$$OI = S_3 / COT * 100 \quad (3.48)$$

Na Tabela 2 é informada a quantidade de dados por parâmetro cedidas pela ANP em um total de 9 poços.

Tabela 2: Quantidade de registros de análises geoquímicas de Pirólise de cada parâmetro.

Resultados das Análises de Rock Eval em 9 poços				
COT	S1	S2	S3	TMAX
891	437	455	509	440



## 4. Processamento e análise de resultados

Nesta seção serão apresentadas todas as etapas de processamento dos dados, além de discutir em detalhes cada um dos resultados alcançados.

Esta tese consistiu basicamente em estabelecer um fluxo de trabalho consistente e alternativo ao processamento convencional. Para isso, dados de perfilagem, como sônico, densidade, neutrão e resistividade, foram processados de duas as formas distintas, a primeira de maneira convencional e a segunda com o uso da sequência sugerida neste projeto. Além disso, foram também discutidas as vantagens e desvantagens de cada metodologia.

Inicialmente, para servir de parâmetro comparativo, foi realizado o fluxo convencional de processamento e análise petrofísica com o uso dos registros de perfilagem dos principais reservatórios da bacia.

O primeiro lote de dados com 18 poços foi solicitado para a ANP no ano de 2016. A maior parte destes dados apresentou algum tipo de problema, como falta de um ou mais perfis ou mesmo arquivos corrompidos. Desta forma, a maior parte dos dados não pôde ser lida diretamente no software IP (Interactive Petrophysics) e por esta razão tiveram que ser digitalizados no *software AutoCad*.

Após a vetorização individual de cada curva, foi desenvolvido um programa, na linguagem de programação AutoLisp, para a leitura, escalonamento e conversão destas curvas em tabelas de dados com formato .ascii e desta forma puderam ser carregadas e lidas pelo IP. A seguir os dados passaram por um controle de qualidade envolvendo comparações com o perfil *caliper* para verificar eventuais zonas de arrombamento e o impacto nos demais perfis. Após esta verificação os perfis foram normalizados para iniciar o processamento e análise de formação convencional.

Uma vez estabelecido o fluxo de trabalho e realizadas as etapas de tratamento de dados, foi necessário recorrer a técnicas de *machine learning* para desenvolver algoritmos que de forma autônoma são capazes de ler os *logs* de perfilagem, encontrar padrões e prever porosidade, permeabilidade e TOC, além de classificar e agrupar o reservatório.



## 4.1 Análises petrofísicas convencionais

Para mostrar que o fluxo de trabalho alternativo desenvolvido é capaz de alcançar resultados similares aos produzidos segundo os fluxos usuais, serão apresentadas aqui todas as etapas de processamento dos perfis de poço e das análises petrofísicas convencionais. Este fluxo de trabalho para avaliação de formação seguiu as etapas apresentadas na Figura 16.

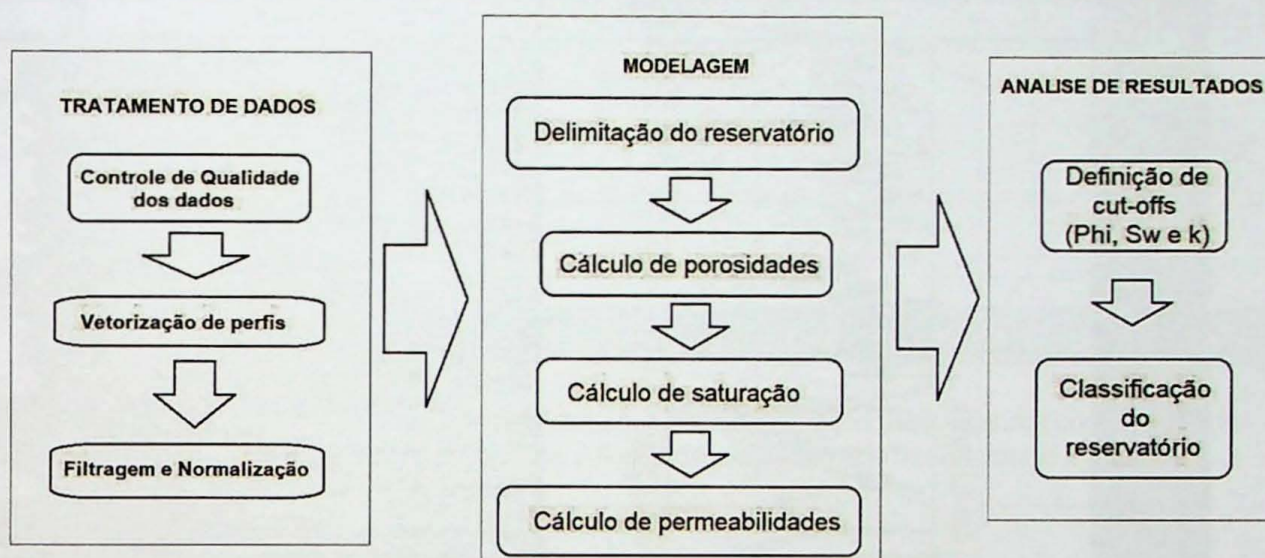


Figura 16: Fluxograma de trabalho convencional utilizado neste trabalho para avaliação das Formações Poti, Cabeças e Jaicós.

A primeira etapa consistiu no tratamento dos dados, onde foi realizada uma triagem preliminar e foram verificados dados com algum sinal de incoerência decorrentes de regiões de arrombamentos. A seguir foram digitalizados no *software* Auto Cad uma série de perfis que haviam sido entregues com algum tipo de conflito ou mesmo corrompidos e desta forma não puderam ser lidos no *software* IP. As curvas *vetorizadas* foram transformadas em arquivos com dados no formato *ascii* por meio de um programa desenvolvido em *AutoLisp*. Por fim, foram realizadas as filtrações, normalizações e carregamento dos dados de todos os poços em um banco de dados. Desta forma, deu-se início à etapa de modelagem e determinação de parâmetros, delimitando o reservatório, considerando as informações e divisões dos perfis compostos e também as assinaturas das eletrofácies de cada formação.

Os dados de cada um dos poços foram submetidos a todos os processos descritos acima. O primeiro destes processos foi a determinação da argilosidade de



cada intervalo. Para isso foi necessário realizar uma delimitação gráfica das zonas de livres de argila (*Gr clean*) e dos intervalos saturados de argila (*Gr clay*), que juntamente com os valores do registro *gamma ray* foram usadas para calcular a curva de argilosidade (Figura 17). Os níveis de argilosidade podem influenciar de maneira relevante a porosidade e principalmente a permeabilidade do reservatório.

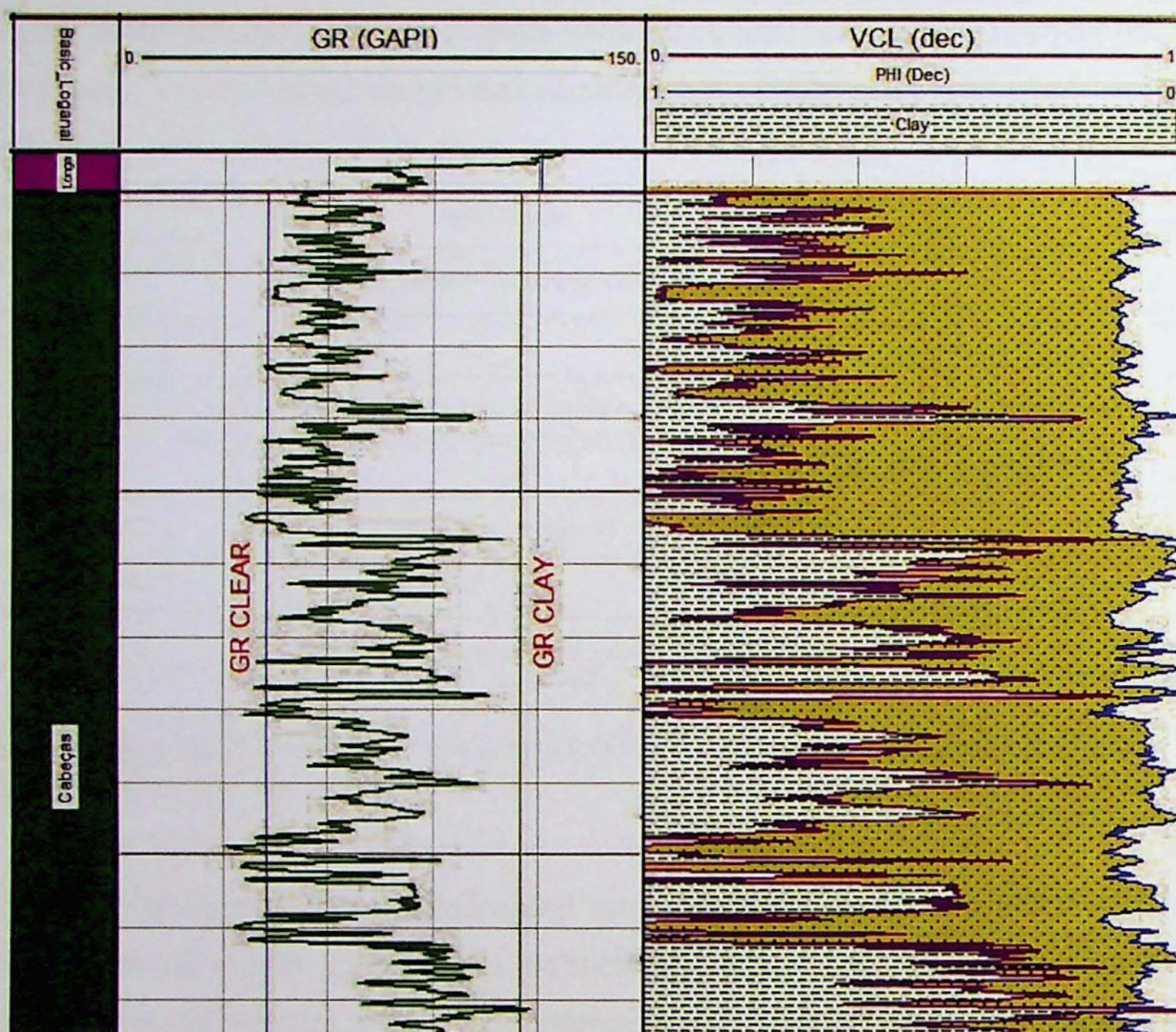


Figura 17: Perfil Gama com valores de limites (clean e clay) para determinação da curva de argilosidade (à esquerda). A área verde hachurada representa o volume de argila na rocha.

Em seguida foi realizado o cálculo da porosidade de cada intervalo. Este cálculo se deu com o uso dos perfis densidade, resistividade, sônico, neutrão, gama e das equações 2.4 a 2.6. Os resultados são apresentados nas Figura 19 e Figura 22. O passo seguinte envolveu a determinação da saturação de água no reservatório e a estimativa de alguns parâmetros como o *fator de formação* ( $a$ ), o *coeficiente de saturação* ( $n$ ) e o *coeficiente de cimentação* ( $m$ ). Para estimar um valor para  $m$  foi usado um método clássico conhecido como *Pickett Plot*, que consiste em plotar um



gráfico da resistividade pela porosidade de um reservatório saturado em água (Figura 18) e ajustar uma reta. O coeficiente angular da reta ajustada representa uma estimativa para  $m$  e a intersecção da reta com o eixo x (resistividade) fornece uma estimativa para resistividade da água da formação ( $R_w$ ).

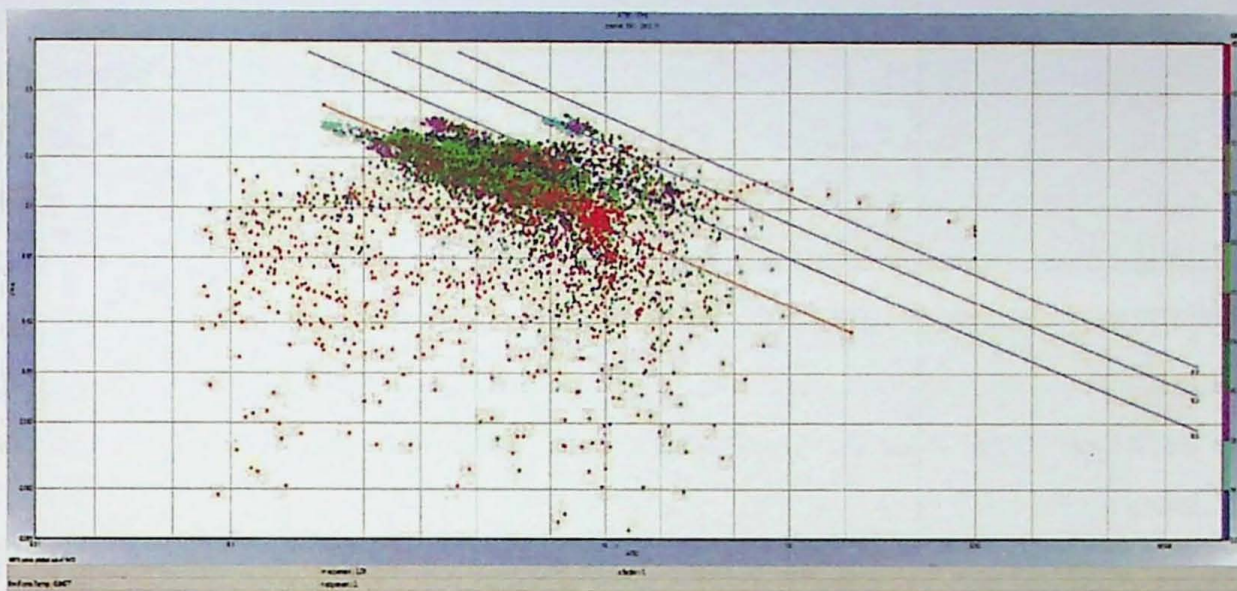


Figura 18: Aplicação do método *Pickett Plot* para determinação do valor de  $m$  (coeficiente de cimentação), por meio da comparação de valores de porosidade e resistividade.

A saturação de água foi determinada por meio do modelo clássico de Archie (Archie, 1942) dada pela equação 2.14 e é apresentado juntamente com os perfis de porosidade e de permeabilidade.

Após determinar os valores de porosidade, saturação da água e estimar outros coeficientes e parâmetros importantes do reservatório, foi então adotado o modelo de Morris e Brigs (1967) para a estimativa de permeabilidade.

Nos dados fornecidos pela ANP foram obtidos os valores de 36 plugs de 3 poços com análises petrofísicas básica em laboratório. Devido a esse pequeno espaço amostral não foi possível a utilização destes dados em estudos quantitativos que pudessem auxiliar na determinação dos valores de permo-porosidade, sendo este um fator importante que motivou a solicitação de um segundo pacote de dados à ANP. O segundo pacote de dados solicitados à ANP, em 2018, apresentou três poços com dados de perfis de permeabilidade obtidos pelo método de Ressonância Magnética Nuclear (RMN) e, assim, contemplando alguns trechos dos reservatórios Poti e Cabeças.



Desta forma, foi realizado um ajuste fino entre o perfil permeabilidade *RMN* e o perfil permeabilidade calculada, de forma que os coeficientes *a*, *b* e *c* eram ajustados manualmente, por tentativa e erro, até que o perfil permeabilidade calculada coincidissem com o perfil permeabilidade *RMN* ao longo do poço, conforme mostra a última coluna da Figura 23. Assim, chegou-se aos valores para os coeficientes  $a=7241$ ,  $b=3.2$  e  $c=2.0$ . Considera-se neste trabalho estes valores mais adequados para as especificidades da Bacia do Parnaíba do que os valores sugeridos por Morris e Briggs (1967) para reservatórios de gás ( $a=6241$ ,  $b=5$  e  $c=2.0$ ). Os valores encontrados estão mais próximos dos sugeridos por Timur (1968). Holmes et al. (2005) realizaram uma série de análises e concluíram que a equação de Timur funciona melhor em reservatórios do tipo “*tight gas*”, especialmente quando o expoente *b* é calibrado com dados de porosidade de laboratório.

As figuras 19 a 23 mostram as permeabilidades calculadas nas formações estudadas. É possível perceber que de um poço para outro os valores variam bastante dada à heterogeneidade dos reservatórios da bacia. A Figura 19 mostra os resultados da avaliação petrofísica realizada na Formação Cabeças nos poços 1, 2, 3, e 4. Com exceção do poço 3, as assinaturas de gama, resistividade, densidade, porosidade e permeabilidade são similares e mostram intervalos portadores de gás significativos.

Na Figura 20 é possível as avaliações petrofísicas da Formação Cabeças em 3 poços (do poço 5 ao 7). Embora as assinaturas dos perfis gama, resistividade e densidade sejam parecidas, os resultados são bem distintos para as estimativas de porosidade, permeabilidade, saturação de água e intervalos portadores de gás. Um dos fatores que podem explicar estas diferenças consiste no fato dos três poços ficarem a uma distância considerável, maior do que 200km entre si, sendo um deles localizado no depocentro da bacia e os outros na margem leste.

A Figura 21 apresenta as avaliações dos poços 3, 4 e 5 da Formação Jaicós, sendo que em 6 e 7 os valores médios dos parâmetros petrofísicos são similares, além disso são poços vizinhos e relativamente distantes do poço 5.

Os resultados das avaliações da Formação Poti são apresentados na Figura 22 e mostram perfis dos poços 22 a 25, sendo os poços 22 e 24, vizinhos.



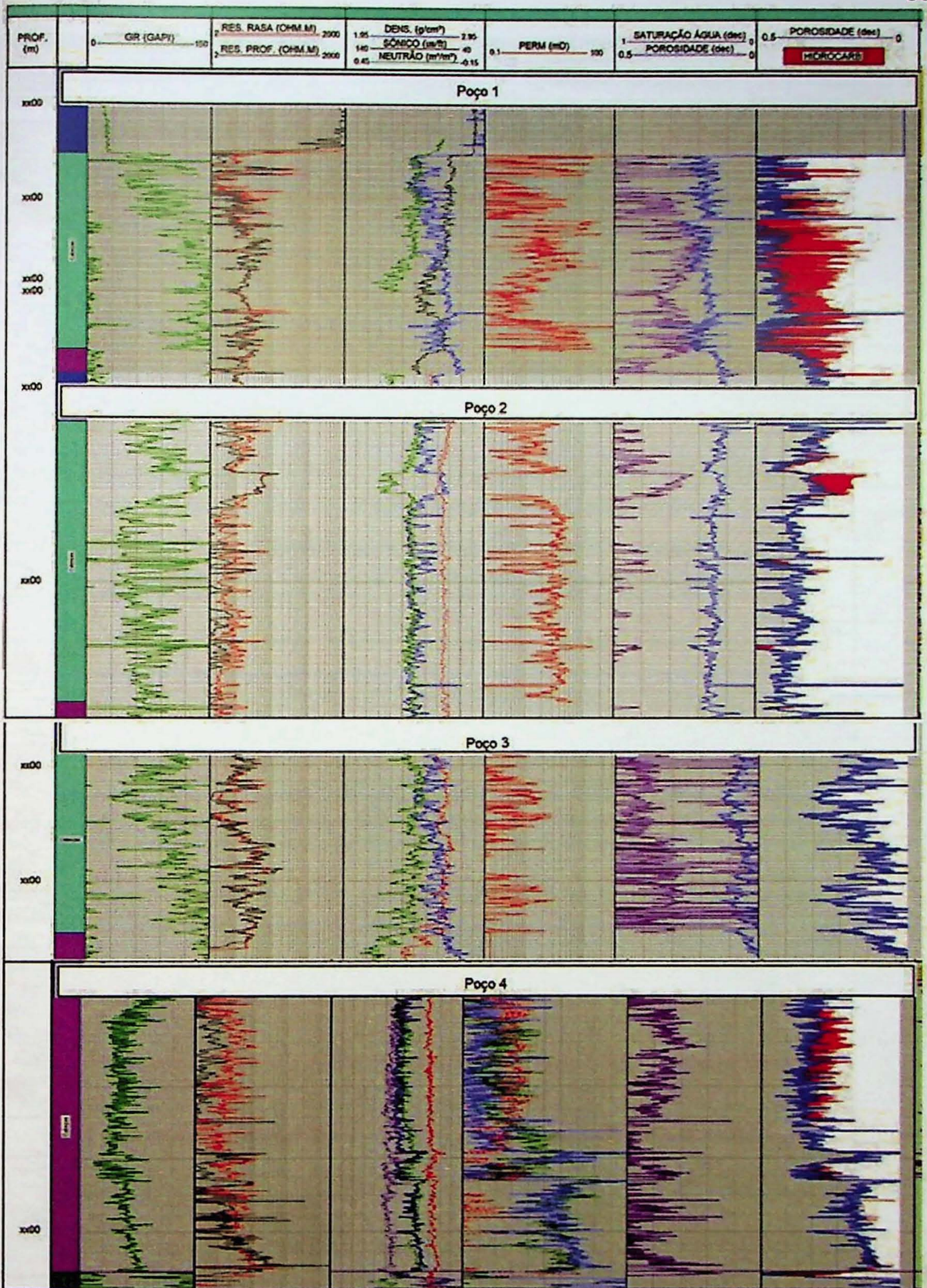


Figura 19: Avaliação petrofísica final da Formação Cabeças nos poços 1, 2, 3 e 4. A coluna 6 contém o perfil de permeabilidade, a coluna 7 mostra a de saturação de água e a 8 mostra porosidade e as regiões portadoras de gás.



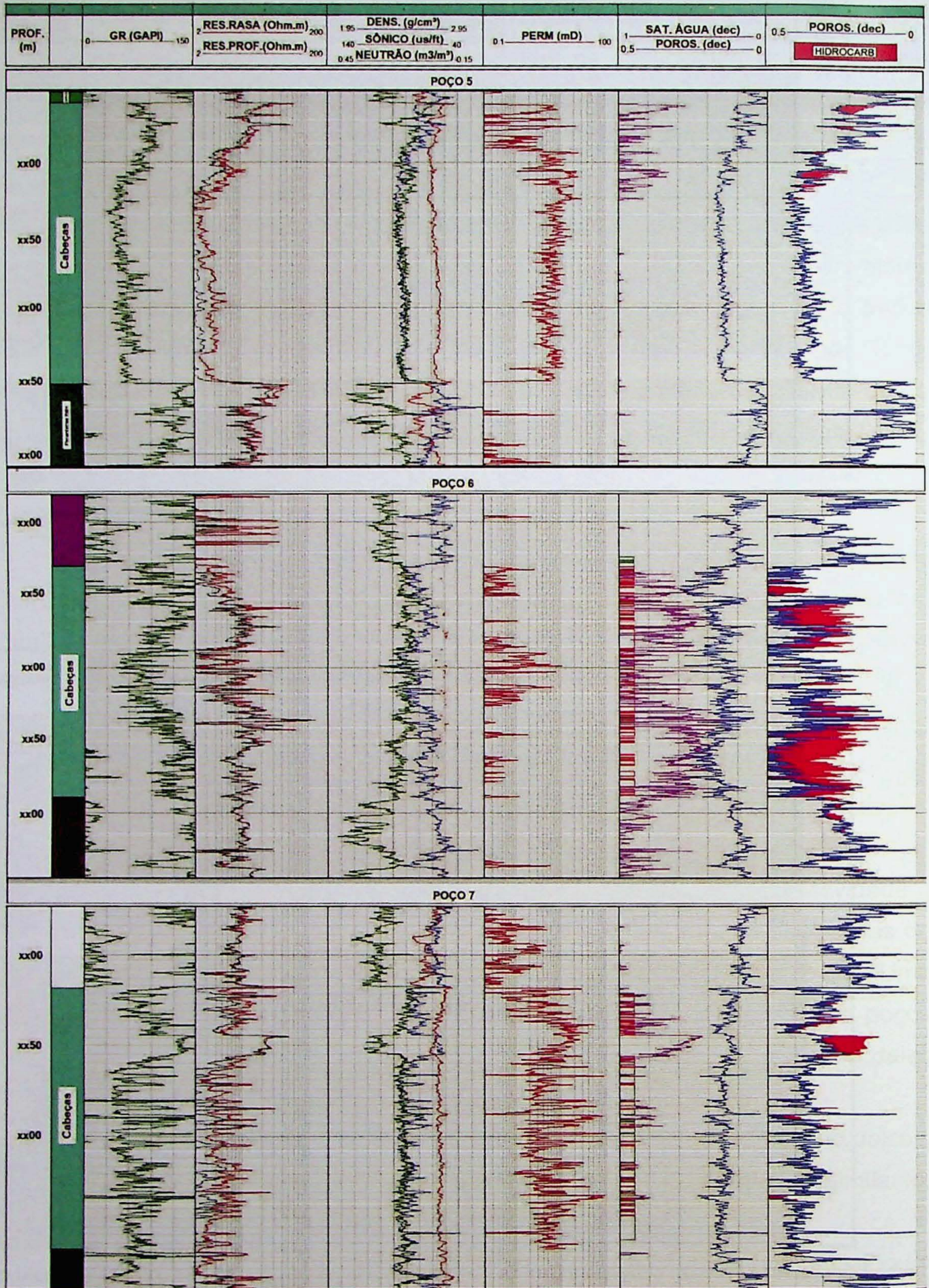


Figura 20: Perfis da avaliação petrofísica final da Formação Cabeças nos poços 5, 6 e 7. A coluna 6 contém o perfil permeabilidade, a coluna 7 mostra a saturação de água, a porosidade e as regiões portadoras de gás (netpay). A coluna 8, mostra em outra escala a porosidade e as zonas com gás.



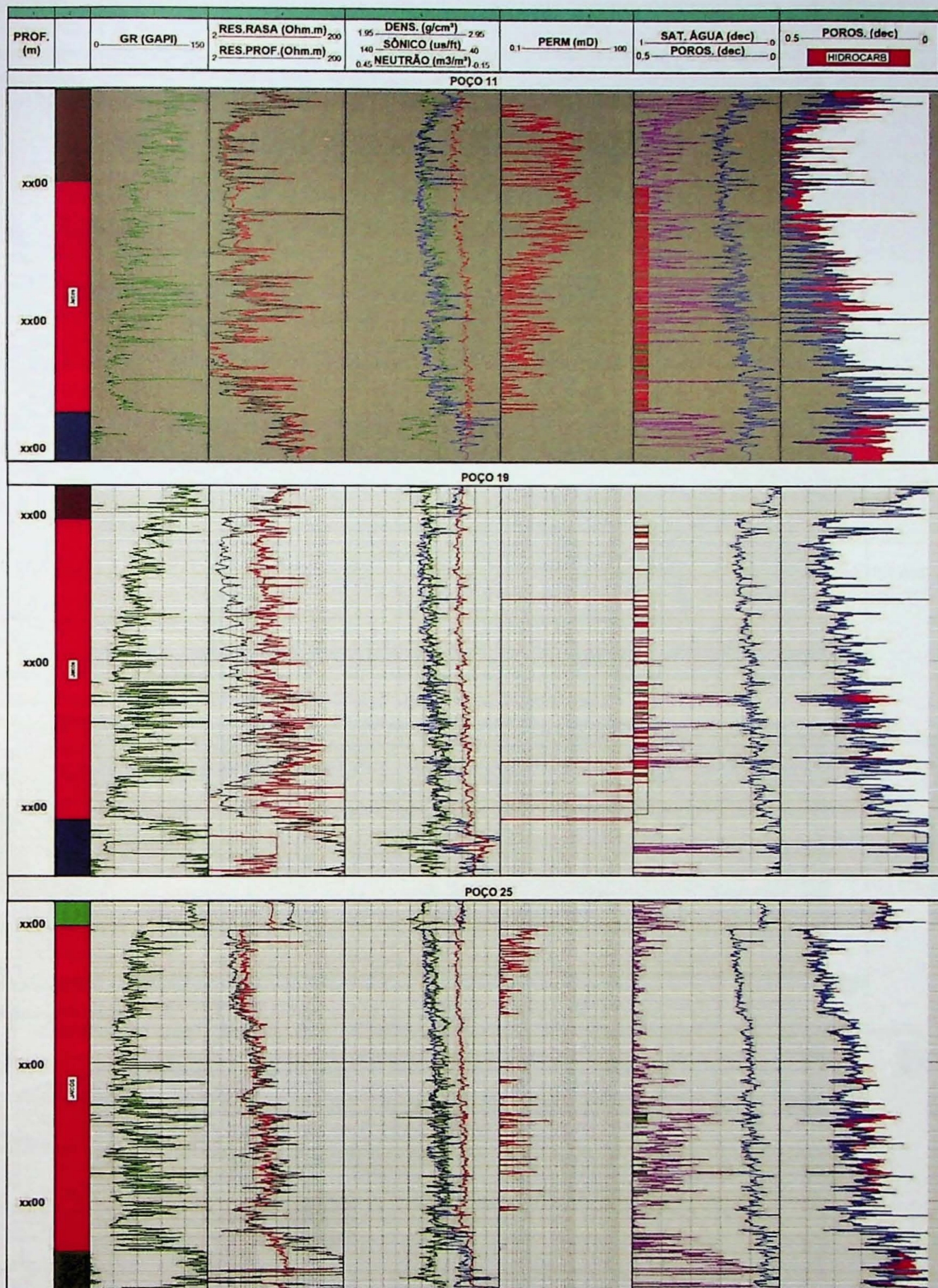


Figura 21: Perfis da avaliação petrofísica final da Formação Jaicós nos poços 11, 19 e 25. A coluna 6 contém o perfil de saturação de água, a coluna 7 mostra a permeabilidade e a 8 mostra porosidade e as regiões portadoras de gás.



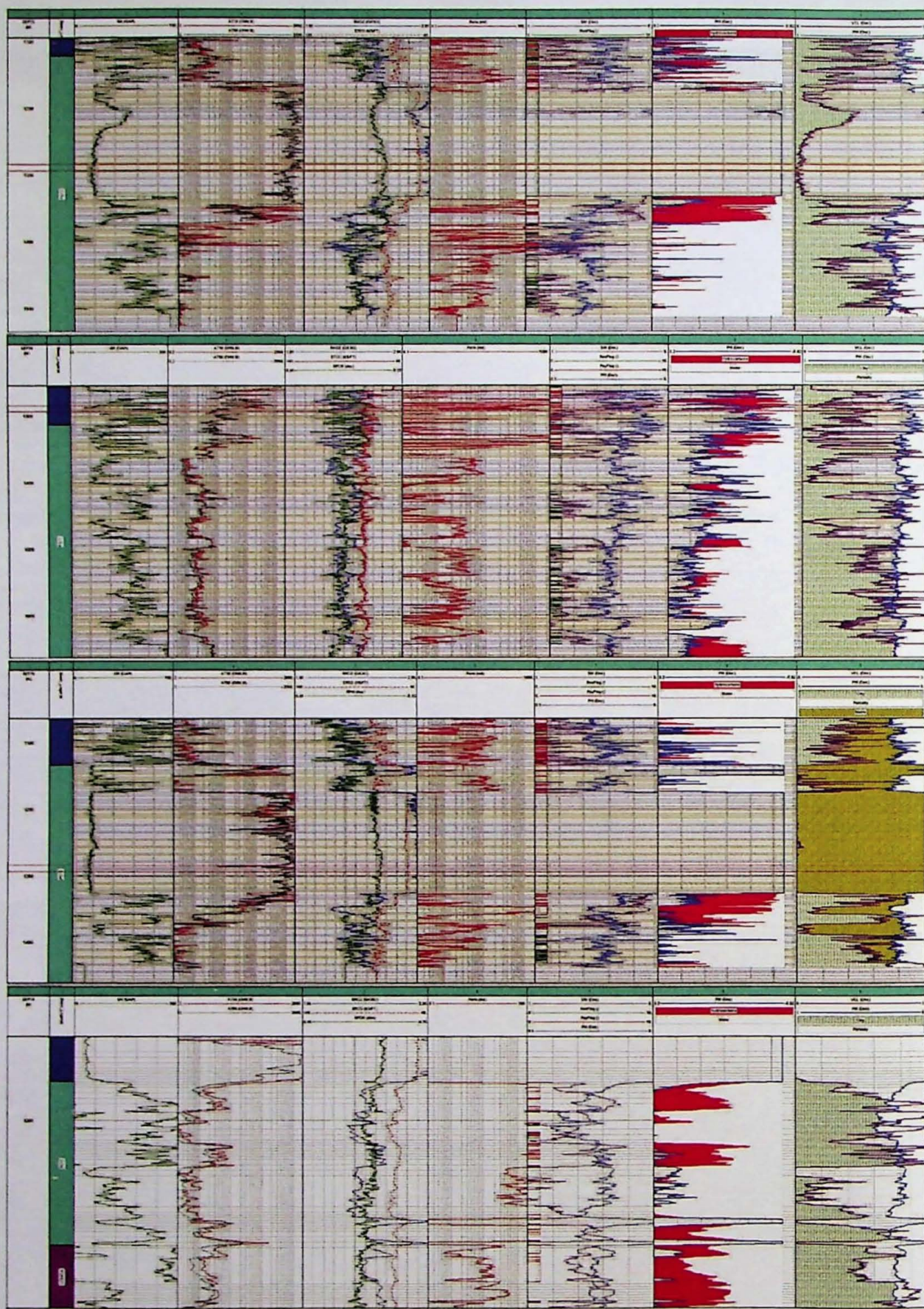


Figura 22: Perfis da avaliação petrofísica final da Formação Poti nos poços 22, 23, 24 e 25. A coluna 6 contém o perfil de saturação de água, a coluna 7 mostra a permeabilidade e a 8 mostra porosidade e as regiões portadoras de gás.



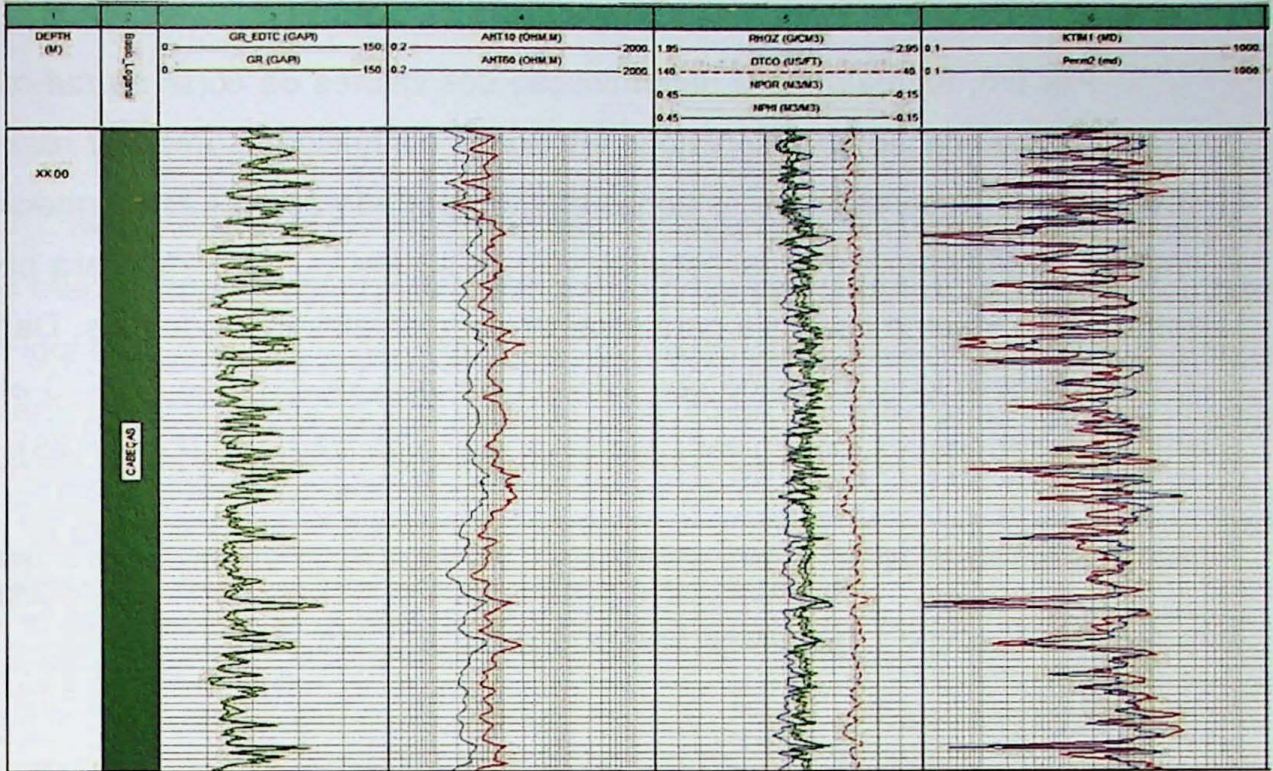


Figura 23: Perfis da avaliação petrofísica com destaque para a última coluna que contém as curvas de permeabilidade calculada (em azul) e medida por ressonância magnética nuclear (em vermelho) da Formação Cabeças no poços 7.

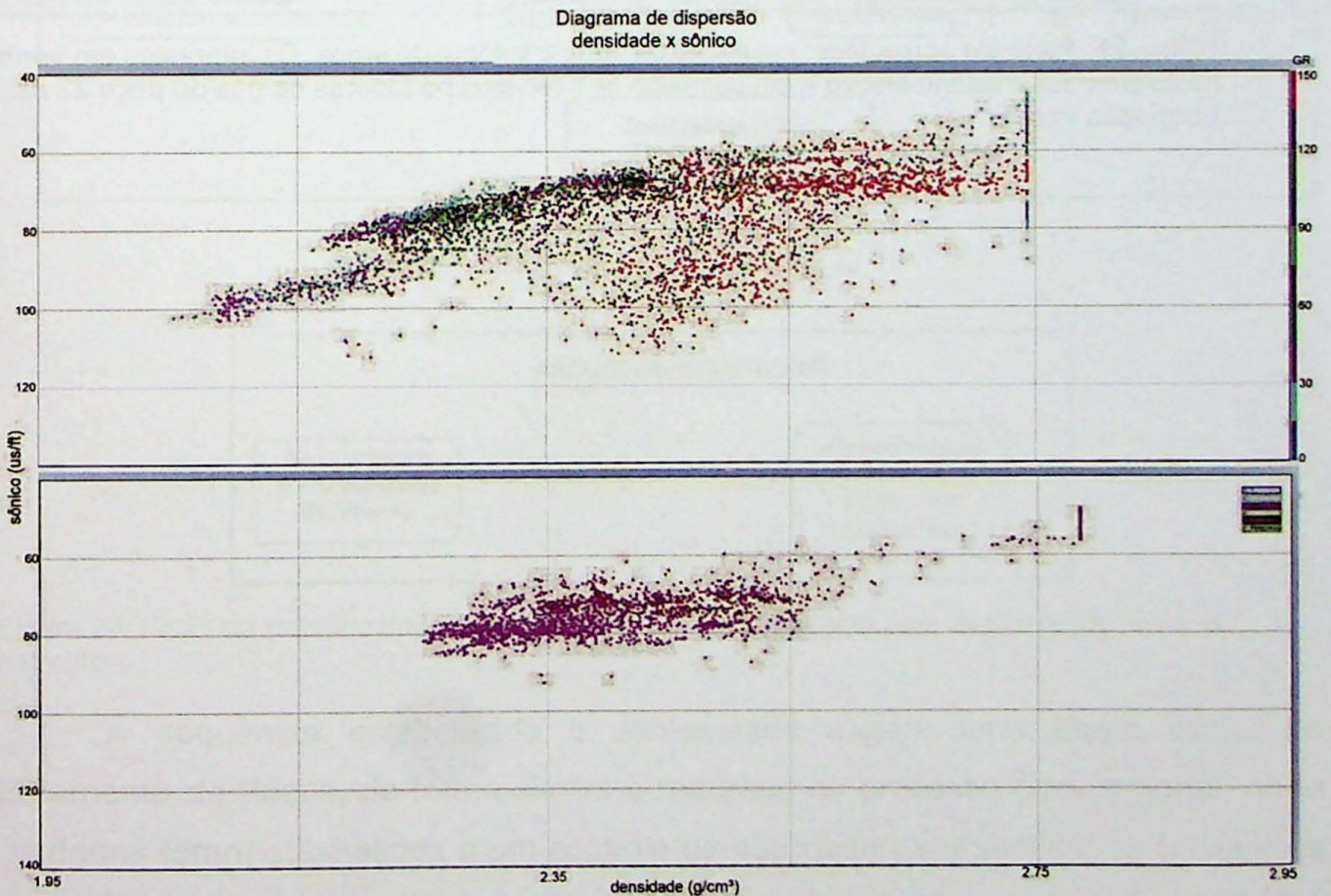


Figura 24: Diagrama de dispersão da densidade pelo perfil sônico de um dos poços da Bacia do Parnaíba. Em A as cores representam a variação Gama. Em B, as cores indicam a formação, sem rosa a Formação Poti e Marrom a Formação Cabeças.



Por fim, foi realizada a determinação dos valores de corte ou *cut-offs* (Figura 25) e a classificação do reservatório em zonas conhecidas como *net reservoir*, que são zonas de reservatório e as zonas efetivas portadoras de gás, conhecidas como *net pay*. Nesta etapa, é usual escolher valores limites, ou de *corte*, para porosidade, saturação e volume de argila para eliminar os intervalos improdutivos. Desta forma, foram estabelecidos *cortes* de porosidade com valores variando entre 3 e 5%, para ambas as formações, porém variando de um poço para o outro (Figura 25).

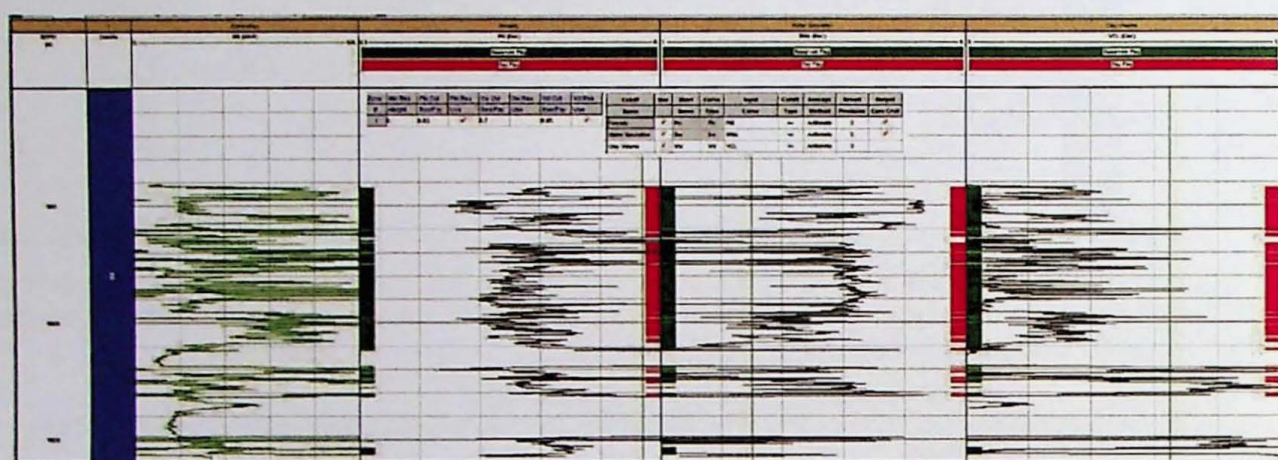


Figura 25: Perfis de porosidade, saturação de água e volume de argila. Os intervalos em verde mostram o reservatório efetivo e em vermelho os intervalos portadores de gás do poço 22 na Formação Poti.



## 4.2 Fluxo de trabalho alternativo

Nesta seção serão discutidos em detalhes os aspectos técnicos de cada uma das etapas do fluxo de trabalho alternativo proposto na presente tese de doutorado. A Figura 26 apresenta as etapas do fluxo alternativo, iniciando com uma etapa de tratamento de dados, e em seguida uma etapa intermediária, de modelagem e treinamento e uma etapa final onde foram feitas as análises de resultados.

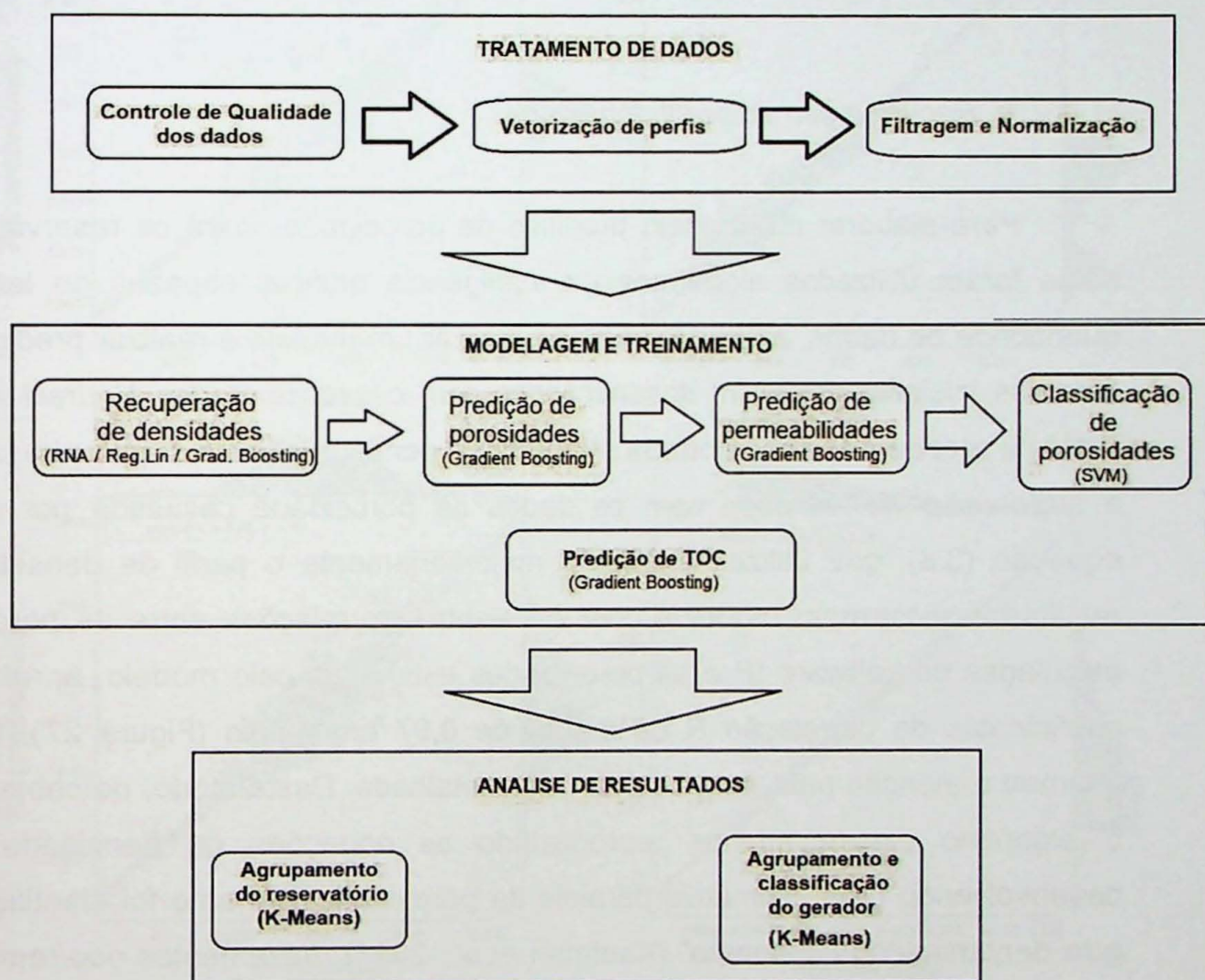


Figura 26: Fluxo de trabalho alternativo desenvolvido neste trabalho para avaliação de sistema petrolífero.

A sequência estabelecida e apresentada sugere uma etapa inicial de tratamento de dados, de forma similar a realizada no processo convencional, onde os dados foram submetidos a um controle de qualidade para verificar se haviam os perfis básicos em cada poço e a integridade dos dados. Ainda nesta etapa, os dados corrompidos foram vetorizados e digitalizados e por fim normalizados. A seguir, na etapa intermediária, a principal desta tese, ocorreu a elaboração e treinamento dos



modelos e foram utilizadas uma compilação de técnicas supervisionadas de *machine learning* para a estimativa de parâmetros petrofísicos como densidade, porosidade, permeabilidade e teor de carbono orgânico.

Por fim, a fase final proposta envolveu a análise de resultados, onde foram usados métodos não supervisionados no intuito de agrupar as informações do reservatório e do gerador por semelhança de características e desta forma estimar as zonas de maior probabilidade de serem portadoras de gás.

#### 4.2.1 Recuperação do perfil densidade

Para elaborar um modelo preditivo de porosidades para os reservatórios da bacia foram utilizados algoritmos de inteligência artificial capazes de ler grande quantidade de dados, aprender padrões, treinar um modelo e realizar previsões. Os modelos inicialmente foram desenvolvidos com o uso de *Redes Neurais Artificiais* (RNA) e receberam como entrada atributos de perfilagem. Em um primeiro momento a supervisão foi realizada com os dados da porosidade calculada por meio da equação (3.6), que utiliza no cálculo majoritariamente o perfil de densidade. Os resultados mostraram modelos com excelentes correlações entre as porosidades calculadas no software IP e as porosidades estimadas pelo modelo, apresentando coeficientes de correlação R da ordem de 0,97 em média (Figura 27), fato que chamou a atenção para uma análise mais detalhada. Deste modo, percebeu-se que o algoritmo estava apenas reproduzindo as equações de densidade e não desenvolvendo uma estimativa paralela da porosidade, portanto foi identificado um erro denominado “vazamento” (Kaufman et al., 2011). Vazamentos ocorrem quando uma resposta é usada como atributo de entrada, ou neste caso, quando se compara resultados de um modelo que usou densidade na entrada com perfis que também usaram densidade como entrada.

Embora cinco perfis tenham sido usados como atributos de entrada da rede neural, os pesos foram ajustados iterativamente de forma a treinar o modelo para geração da porosidade final, compatível com a porosidade de perfil. Analisando os indicadores de importância dos atributos percebemos que a rede neural atribuía um peso muito maior ao perfil densidade, chegando a ser quatro ordens de grandeza



mais importante do que os outros atributos como sônico, raio gama e neutrão. Desta forma, constatou-se que havia um erro de vazamento no uso desta metodologia, pois as duas porosidades, supostamente distintas, que estavam sendo comparadas eram provenientes do mesmo atributo, a densidade.

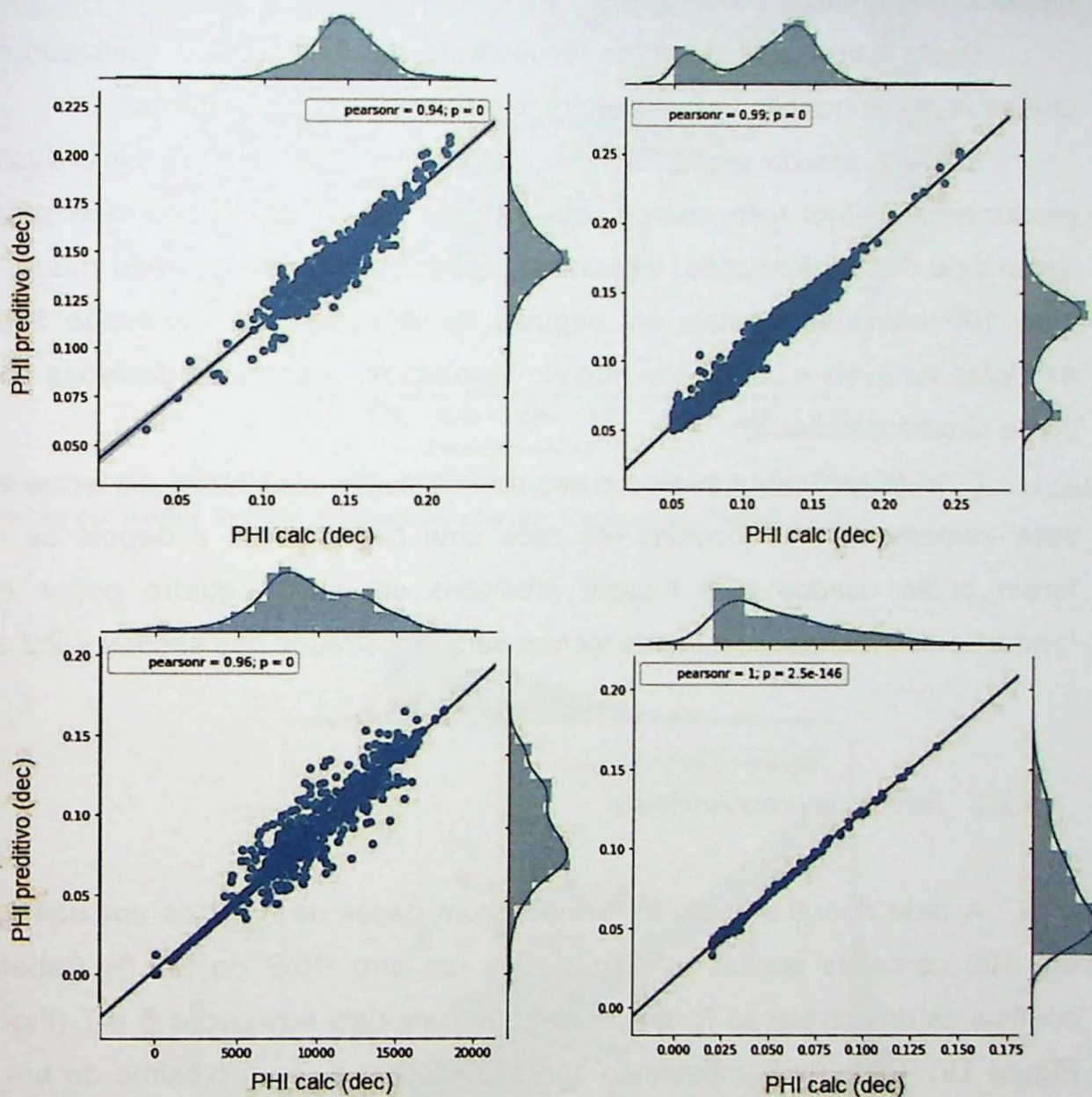


Figura 27: Diagrama de dispersão da porosidade preditiva gerado por Redes Neurais Artificiais. A figura mostra a correlação entre a porosidade calculada e a porosidade predita no intervalo Cabeças dos poços 5, 6, 7 e 9.

Assim, a única maneira de desenvolver uma abordagem alternativa as convencionais seria recuperar o perfil densidade com o uso de técnicas de *machine learning*, usando como entrada somente os demais perfis disponíveis como sônico, resistividade, gama e neutrão. Neste caso, a nova densidade habilitaria a construção



de um modelo preditivo de porosidades. Além disso, diversos dados fornecidos pela ANP eram de poços antigos que não continham o perfil densidade e nestes casos foi possível recuperá-los com êxito. Em outro contexto, e de maneira semelhante, Guarido (2019) aplicou uma técnica de *machine learning* conhecida como *Gradient Boosting* para recuperar com sucesso o perfil *photoelectric* (PE) de alguns poços, baseado em atributos de 8 poços.

Desta forma, o processo de recuperação da densidade foi realizado em duas unidades, na Formação Cabeças e em seguida na Formação Jaicós.

Em um estudo preliminar três métodos foram utilizados para avaliar qual produziria a melhor resposta, ou seja, em que o perfil densidade mais próximo da densidade real da formação. Inicialmente foi desenvolvida uma *rede neural artificial* com 100 camadas ocultas, em seguida foi utilizada uma *regressão linear com múltiplas variáveis* e por fim um método baseado em árvores de decisões conhecido como *Gradient Boosting*.

O teste preliminar avaliou o uso de três poços escolhidos de forma aleatória para treinamento dos modelos em cada uma das técnicas e depois os modelos foram então usados para realizar previsões em outros quatro poços de cada formação. Os resultados de cada técnica serão mostrados nas seções 4.2.2 a 4.2.4.

#### 4.2.2 Redes Neurais Artificiais

A rede neural artificial foi treinada com dados de 6 poços por 654 iterações em 100 camadas ocultas e chegou-se a um erro RMS de 3,9 % (tabela 3). O coeficiente de correlação R obtido ao testar o modelo nos poços 5 e 7 (Figura 28 e Figura 29) foi de 0,65, mostrando que não se chegou nem próximo de um modelo bem treinado.

Testes preliminares revelaram que o desempenho da função de ativação *Logistic* foi superior às funções *ReLU* e *secante hiperbólica* e por isto foi a escolhida para o treinamento do modelo. Além disso, algum parâmetro como taxa de aprendizagem e momento foram obtidos por tentativa e erro e os valores finais usados foram respectivamente 0,0005 e 0,94.



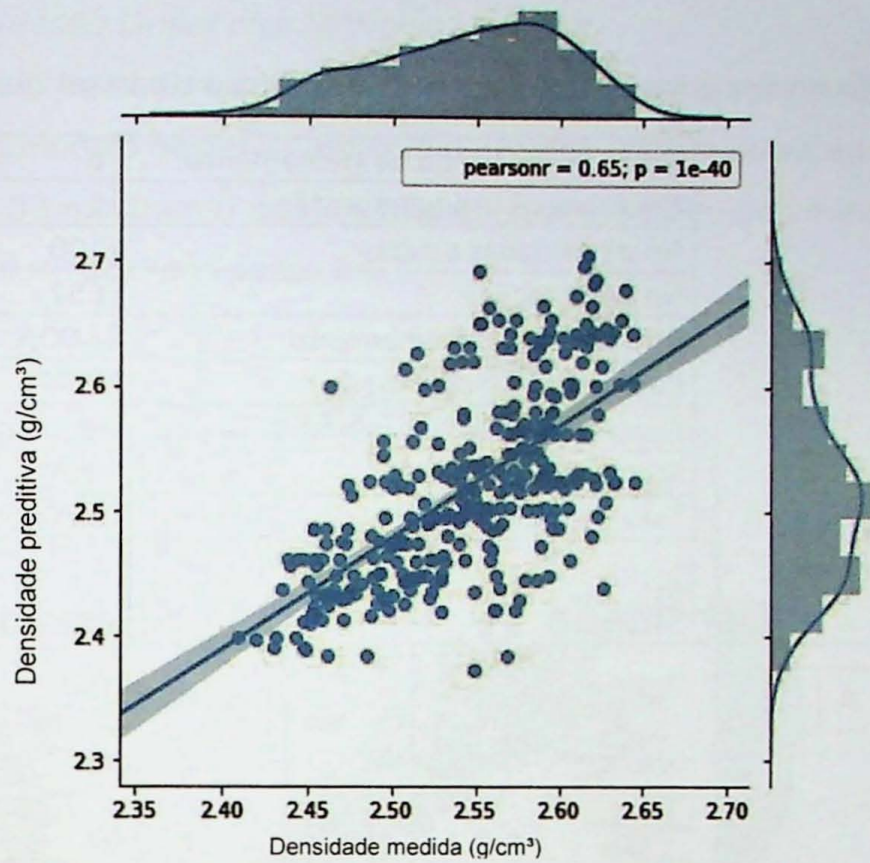


Figura 28: Diagrama de dispersão com a correlação entre a densidade calculada e a porosidade predita por Redes Neurais Artificiais no intervalo Cabeças do poço 5.

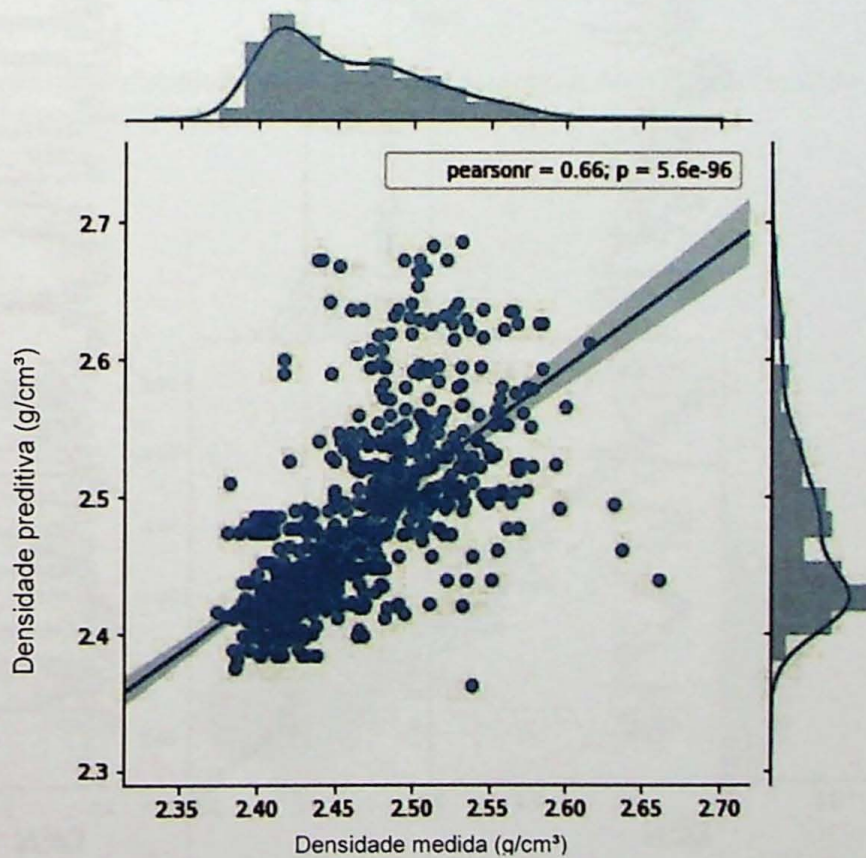


Figura 29: Diagrama de dispersão com a correlação entre a densidade calculada e a porosidade predita por Redes Neurais Artificiais no intervalo Cabeças do poço 7.



Tabela 3: Parâmetros utilizados para treinamento da rede neural artificial.

No de poços usado no treinamento:	6
No de poços usado no teste:	2
No de camadas ocultas:	100
No de iterações:	654
Erro padrão do treinamento:	3,90%
Coeficiente de correlação $r$ :	0,54
Função de ativação:	<i>logistic</i>
Momento:	0,94
Taxa de aprendizagem:	0,0005

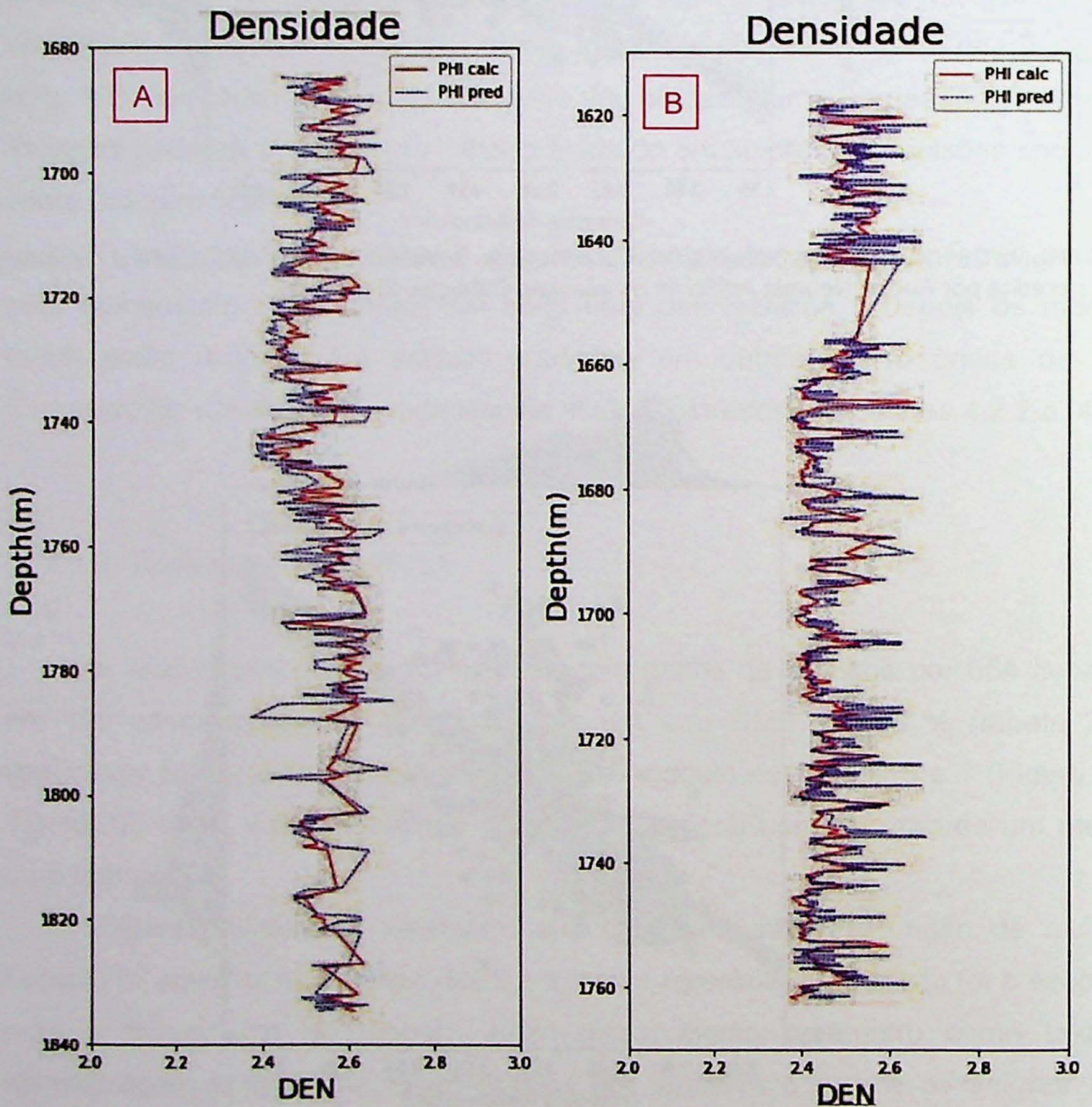


Figura 30: Perfis de densidade. Em azul as densidades previstas pela rede neural artificial e em vermelho as curvas de densidade medidas na perfilagem dos poços 5 (A) e 7 (B).



### 4.2.3 Regressão Linear com Múltiplas Variáveis

O modelo treinado com o uso da regressão linear apresentou suave melhora em relação ao modelo desenvolvido com o uso das redes neurais artificiais. A Figura 31 e a Figura 32 mostram o ajuste realizado com o método dos mínimos quadrados para o teste do modelo nos poços 5 e 7.

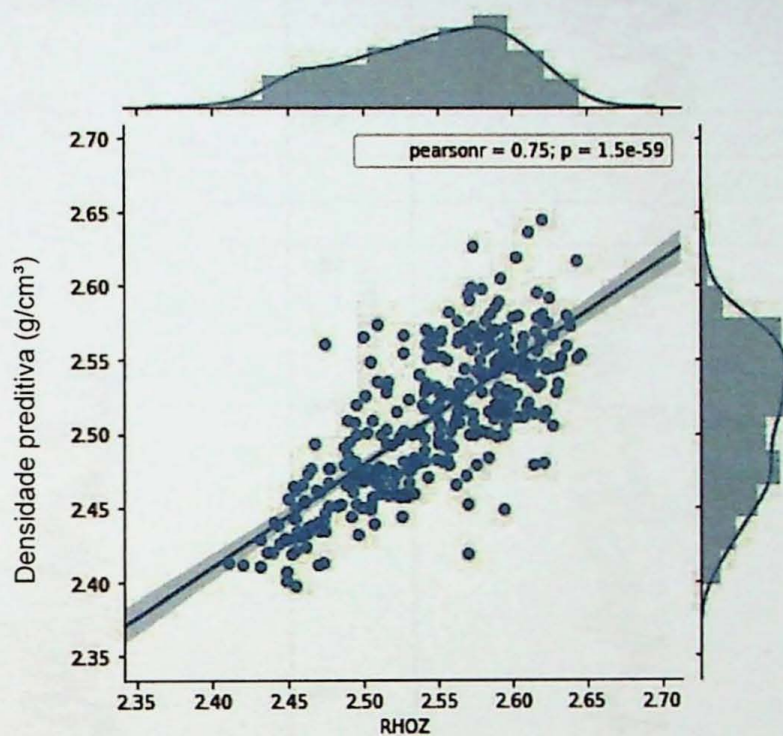


Figura 31: Diagrama de dispersão com a Correlação entre a densidade calculada e a predita por regressão linear com múltiplas variáveis no poço 5.

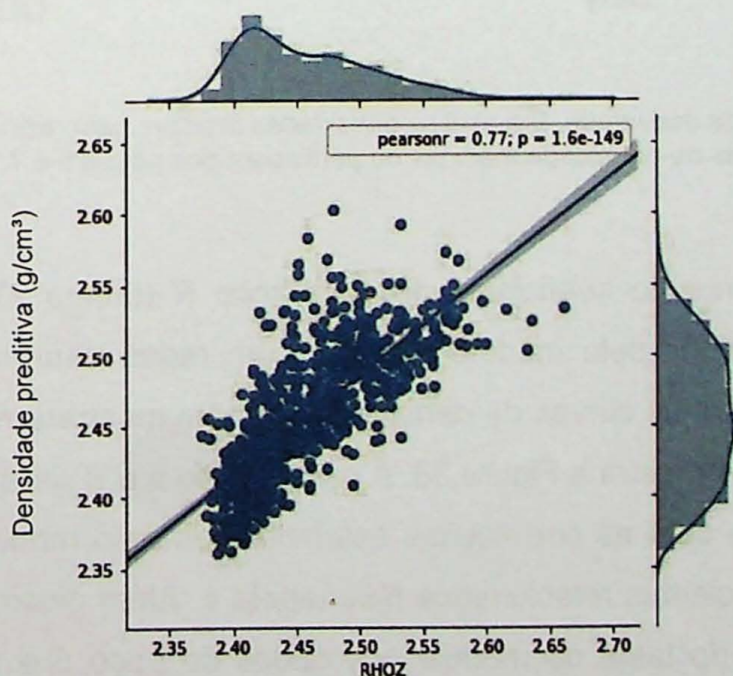


Figura 32: Diagrama de dispersão com a Correlação entre a densidade calculada e a predita por regressão linear com múltiplas variáveis no poço 7.



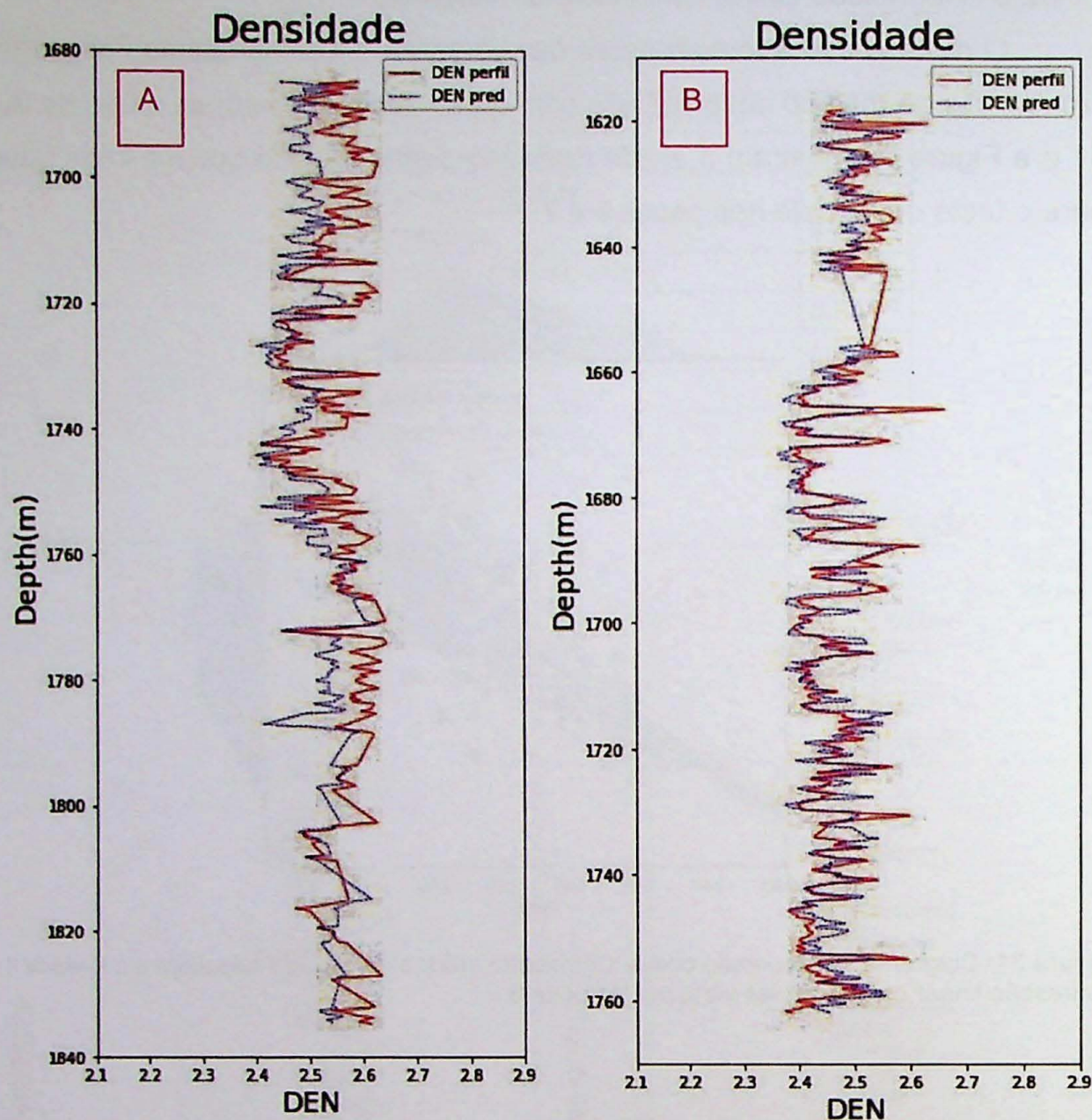


Figura 33: Perfis de densidade. Em azul as densidades previstas pela rede neural artificial e em vermelho as curvas de densidade medidas na perfilagem dos poços 5 e 7.

Os valores do coeficiente de correlação  $R$  (tabela 2) foram superiores aos ajustes produzidos pelo modelo treinado com redes neurais ficando entre 0,75 e 0,77. Além disso, as curvas de densidade também mostraram ajustes sensivelmente melhores, como mostra a Figura 33, A para o poço 5 e B para o poço 7.

As retas com as correlações estabelecidas pelo modelo após o treinamento tem seus coeficientes relacionados pela tabela 4. Além disso, a tabela mostra ainda o erro padrão do teste do modelo nos dados do poço 5 e também nos dados do poço 7, que ficaram respectivamente em 0,011 e 0,0031.



Tabela 4: Métricas e parâmetros utilizados para treinamento da rede neural artificial.

Coeficientes angulares:		
<i>Gama</i>		2,8398e-3
<i>Resistividade</i>		-6,1553e-04
<i>Sônico</i>		1,6314e-04
<i>Neutrão</i>		-8,5925
Coeficiente linear:		2,8398
	Poço 5	Poço 7
MSE erro:	0,0011	0,0031
Variância:	0,56	-0,06



#### 4.2.4 Gradient Boosting

O modelo treinado com uso do método *Gradient Boosting* apresentou os melhores resultados na comparação com os dois métodos usados anteriormente. O treinamento foi realizado com 500 iterações e 4 níveis profundidade. Os parâmetros usados no treinamento foram determinados por tentativa e erro, após o uso de um algoritmo auxiliar, que realizou uma série de iterações com diversos parâmetros até minimizar o desvio padrão. Estes parâmetros são mostrados na tabela 5. A Figura 34 apresenta o desvio padrão dos conjuntos de treino e de teste por iteração.

Tabela 5: Métricas e parâmetros utilizados para treinamento do modelo.

MSE:	0,002	Taxa de aprendizagem:	0,005
Alpha:	0,9	Iterações	500
Prof.:	4		

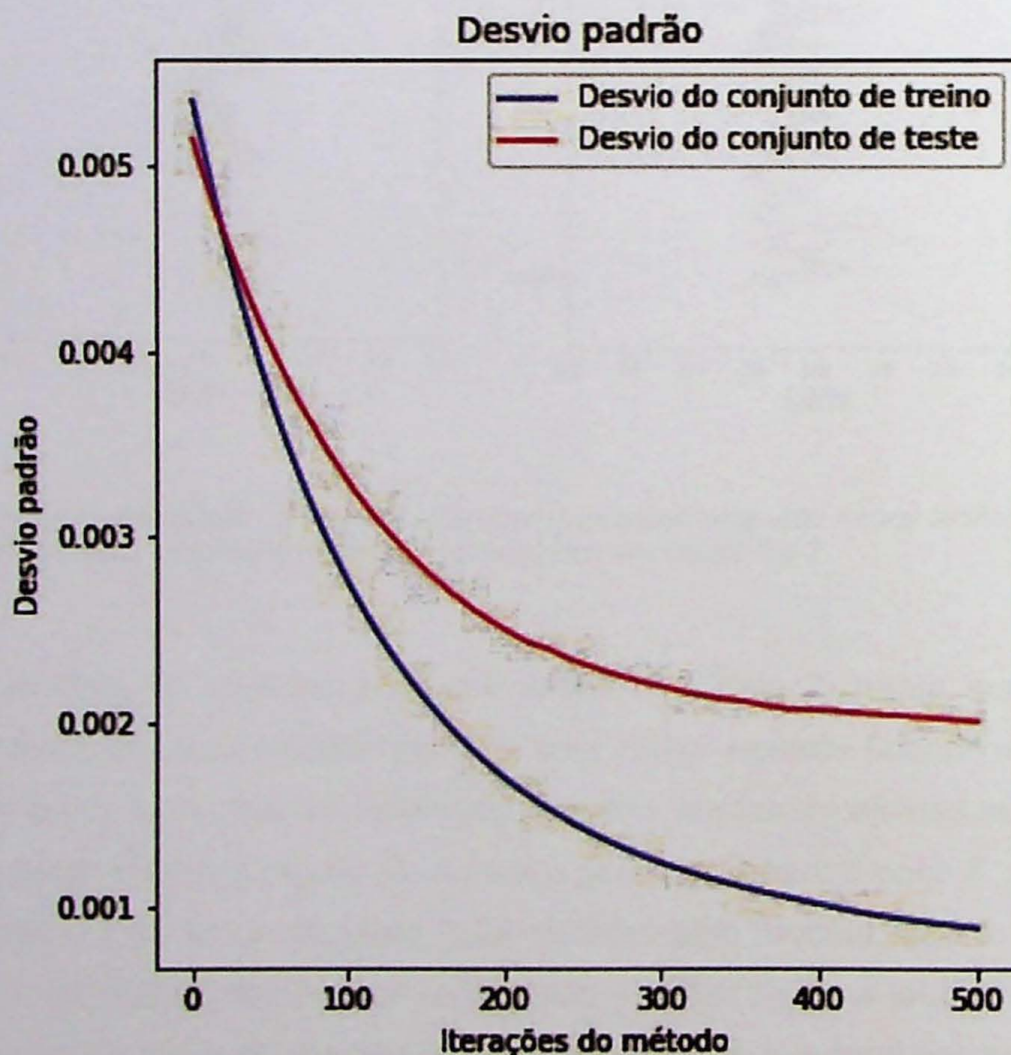


Figura 34: Gráfico do erro padrão em função do número de iterações para o conjunto de dados de testes e para o conjunto de dados de treino.



As Figura 35 e 36 mostram o resultado do modelo treinado com dados de 6 poços e a Figura 37 apresenta as densidades recuperadas por este modelo nos poços 5 e 7. Tanto a correlação, como o ajuste entre a densidade conhecida e a predita são visivelmente melhores, ressaltando o desempenho superior do método em relação aos demais e por isto, o método foi escolhido para recuperar as densidades em os todos os poços, especialmente os mais antigos, onde não havia o perfil densidade.

Uma vez demonstrada à eficácia da técnica para recuperação do perfil densidade, foi realizado um novo treinamento do modelo, desta vez utilizando como entrada atributos de todos os poços que continham o perfil densidade, justamente para se treinar um modelo mais abrangente, com o maior número possível de dados e desta forma, obter um modelo mais refinado, com maior capacidade de predição e menores incertezas.

Desta forma, foi possível testar este modelo, desta vez nos dados de todos os outros poços, incluindo os poços antigos onde a densidade não havia sido perfilada.

Os resultados deste novo modelo treinado são mostrados nas figuras 36 e 37, onde se observa a correlação entre as densidades conhecidas e os dados de densidade preditiva.

De maneira geral, os resultados mostram mais uma vez a eficácia da técnica neste tipo de análise, uma vez que os coeficientes de correlação entre conjuntos de treino e teste ficaram entre 0,85 e 0,91. A Figura 38 mostra respectivamente os testes do novo modelo na Formação Cabeças nos poços 1, 2, 3, 4, demarcado na figura com A, B, C e D. Os resultados mostraram correlações melhores do que as geradas com o uso da Regressão Linear e das Redes Neurais Artificiais, com coeficientes de correlação variando entre 0,85 e 0,91.

A Figura 40 mostra os diagramas de dispersão com as respectivas correlações entre densidades preditas e calculadas da Formação Cabeças nos poços 5 a 8. Os coeficientes de correlação R, em azul, mostraram valores entre 0,88 e 0,90, corroborando o desempenho obtido do modelo nos quatro poços iniciais. Além disso, as retas vermelhas mostram a reta expectativa onde  $X=Y$  e ajuda a visualizar a simetria dos dados, sendo também muito superior neste quesito do que os resultados obtidos com o uso dos dois métodos anteriores.



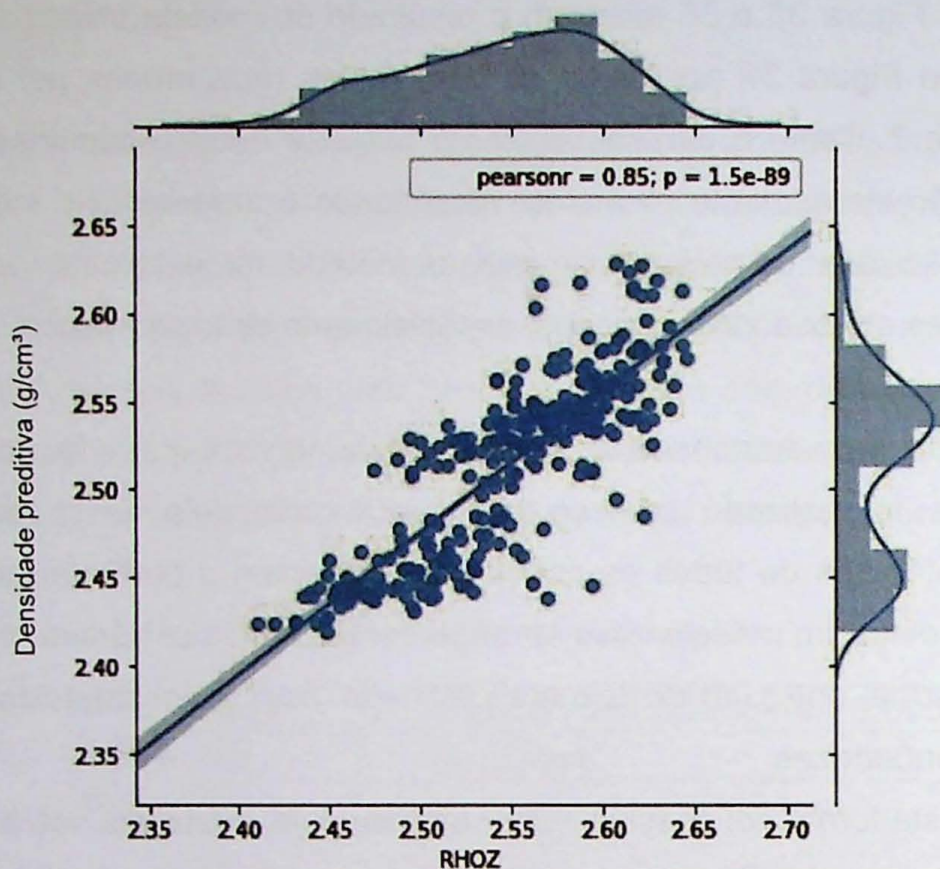


Figura 35: Diagrama de dispersão da densidade calculada pela predita por Gradiente Boosting para os dados do poço 5. A figura mostra o coeficiente de correlação R de 0,85.

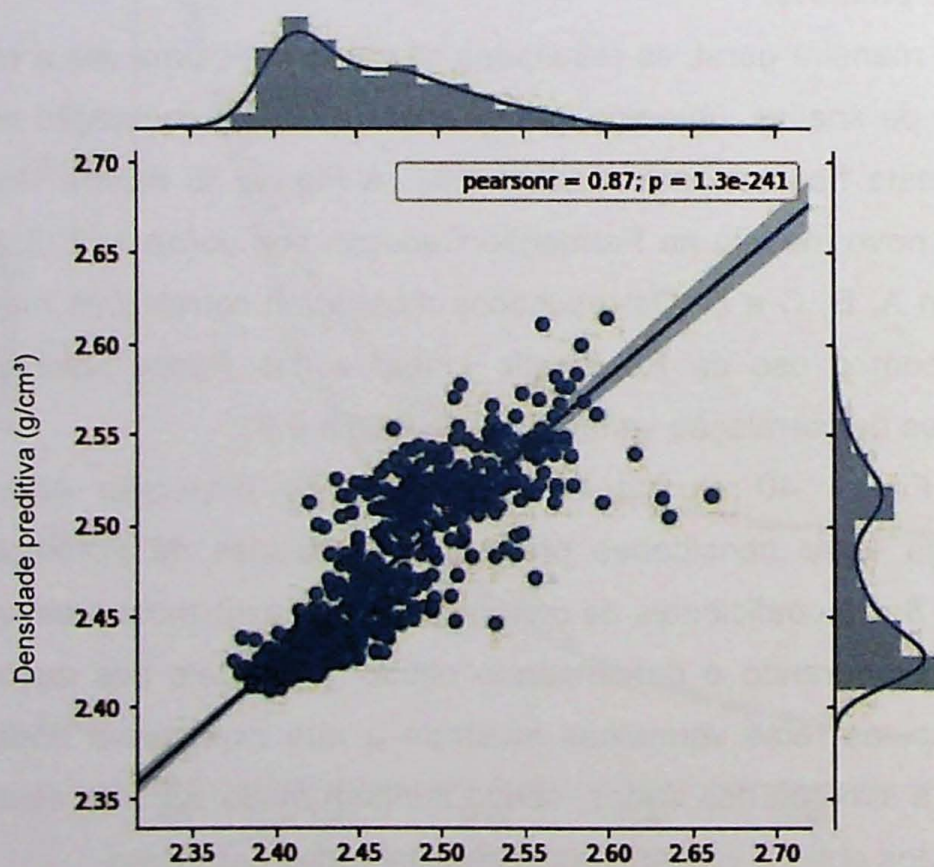


Figura 36: Diagrama de dispersão da densidade calculada pela predita por Gradiente Boosting para os dados do poço 7. A figura mostra o coeficiente de correlação R de 0,87.



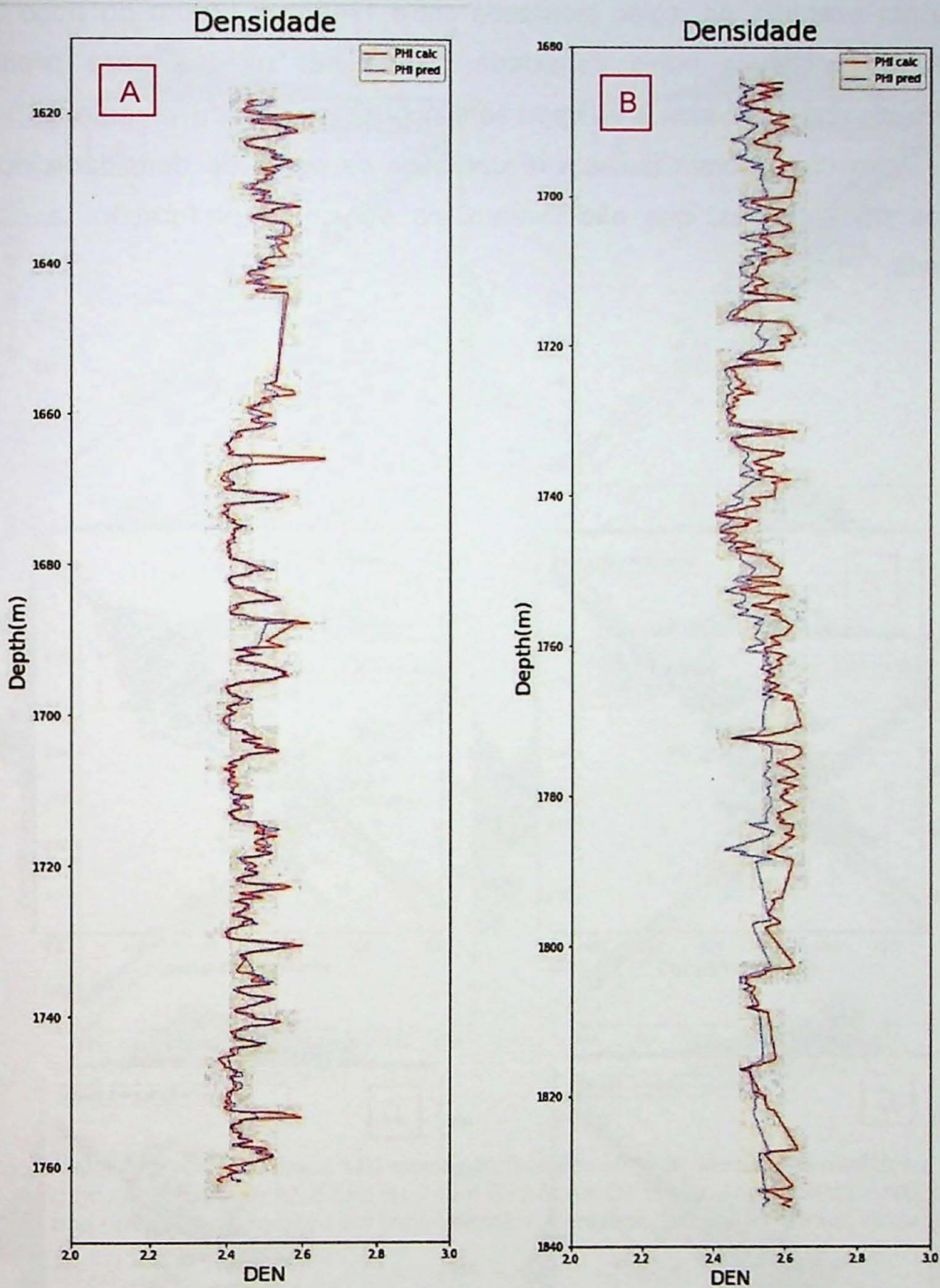


Figura 37: Curvas de densidade pela profundidade dos poços 5 (A) e 7 (B). Em azul as curvas previstas com o uso do método Gradient Boosting e em vermelho as densidades medidas na perfilagem.

Na Figura 40 possível perceber visualmente que os perfis de densidade medida e densidade prevista são muito bem ajustados, ou seja, as curvas permanecem muito próximas ao longo de todo intervalo, mantendo os padrões de sinusidade e resolução. Em raros trechos ocorre uma sensível perda de resolução,



como por exemplo, na região localizada entre 1740m e 1780m do poço 3. Desta forma, esses novos perfis densidade podem ser usados para predição da porosidade nos reservatórios da bacia sem incorrer em erro de vazamento.

Além disso, foram também recuperados os perfis de densidade dos poços antigos da Petrobras, que não tiveram, na época da perfuração, a densidade perfilada.

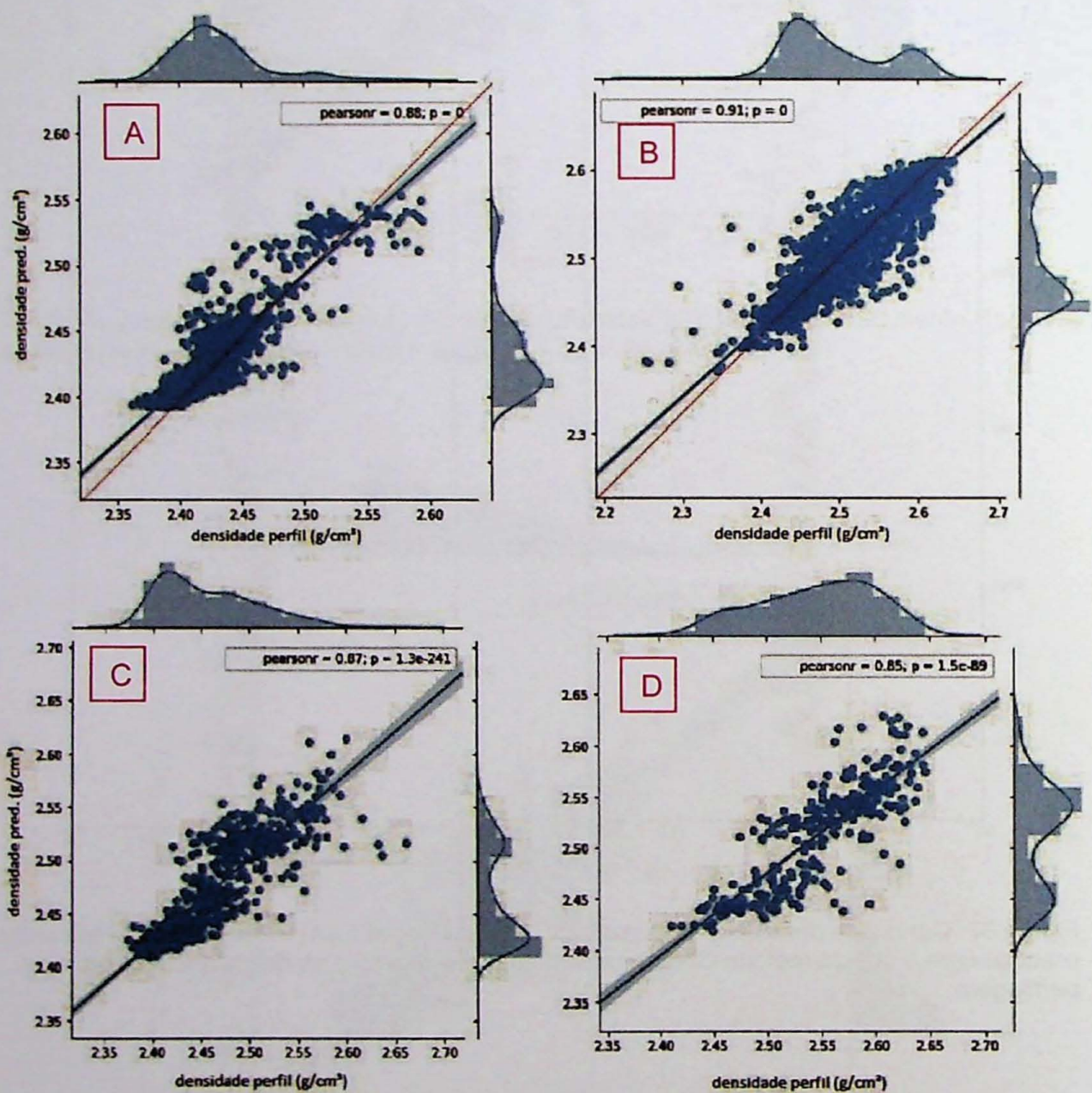


Figura 38: Diagramas de dispersão com as correlações entre as densidades medidas e previstas pelo modelo nos poços 1 (em A), 2 (em B), 3 (em C) e 4 (em D). Em azul é mostrada a reta de regressão linear dos dados de cada poço e a linha vermelha representa a reta expectativa, onde  $x=y$ .



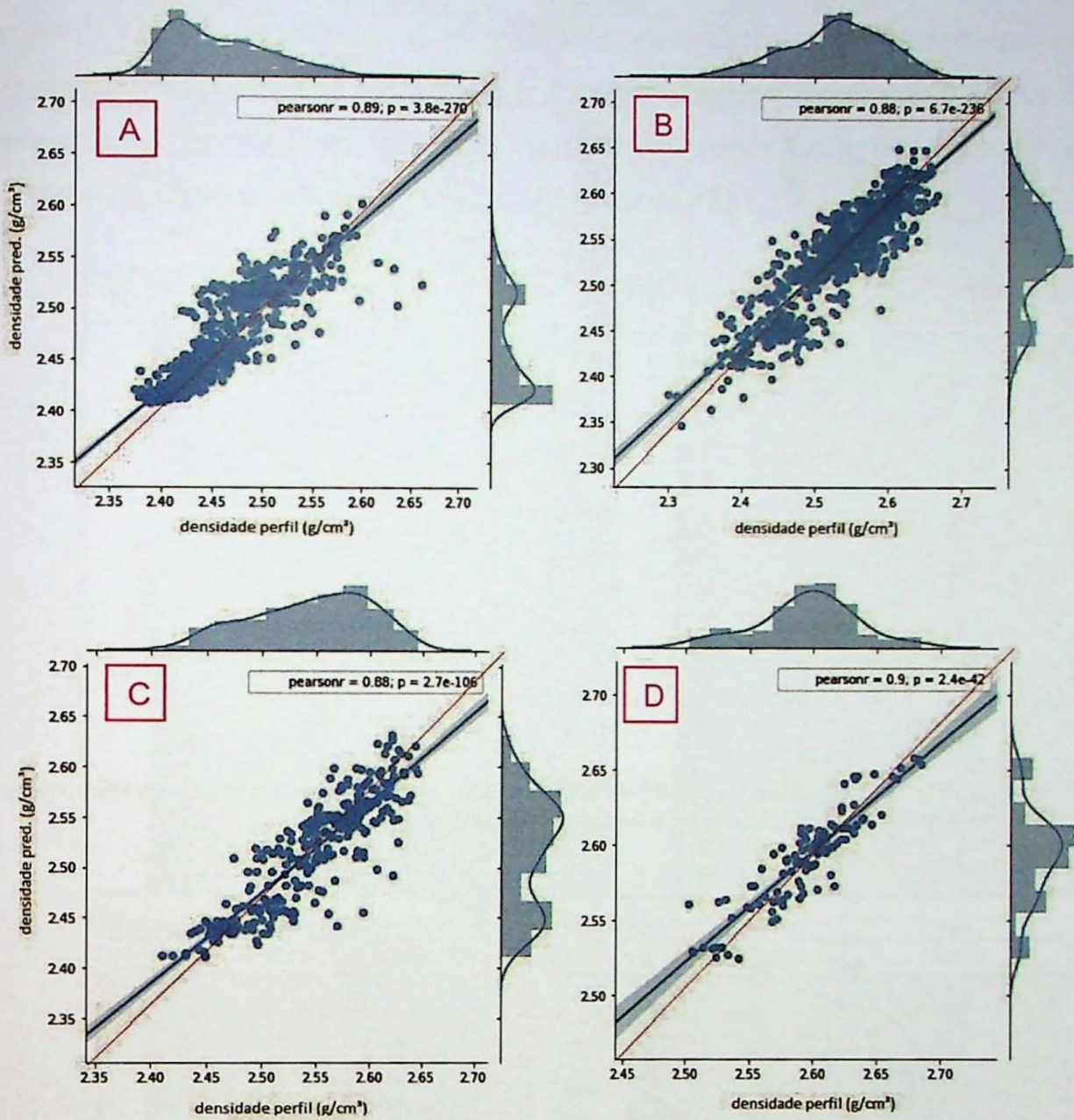


Figura 39: Diagramas de dispersão com as correlações entre as densidades medidas e previstas pelo modelo nos poços 5 (em A), 6 (em B), 7 (em C) e 8 (em D). Em azul é mostrada a reta de regressão linear dos dados de cada poço e a linha vermelha representa a reta expectativa, onde  $x=y$ .



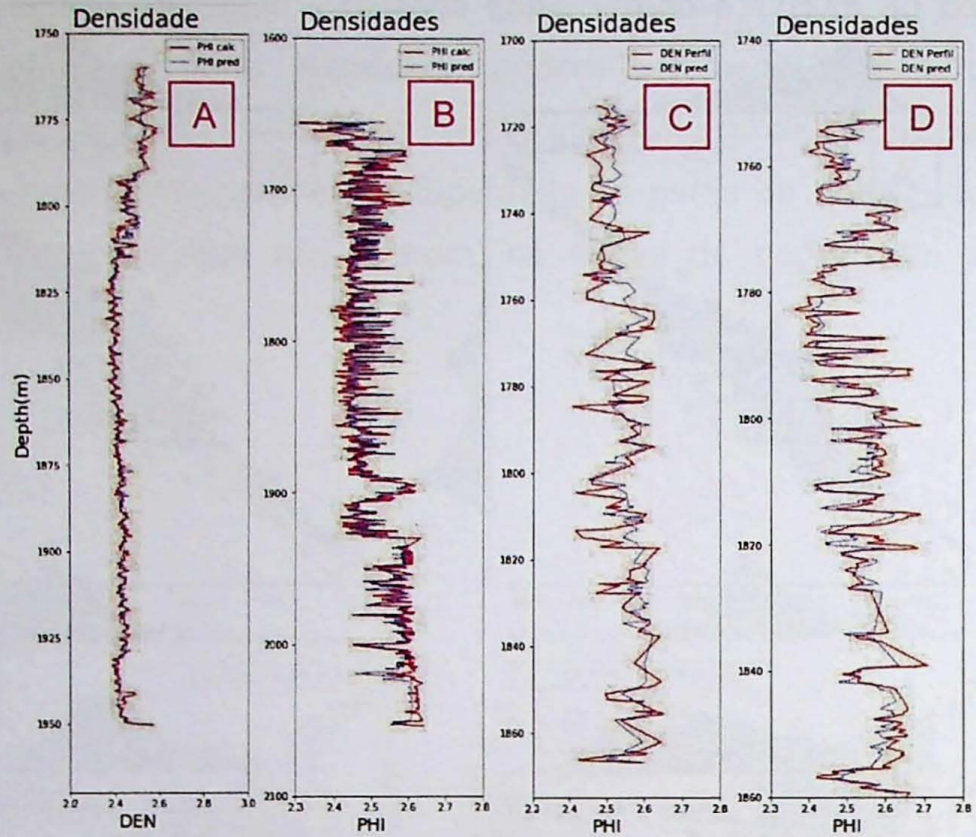


Figura 40: Perfis com as densidades previstas (em azul) e as densidades medidas (em vermelho) dos poços 1 (em A), 2 (em B), 3 (em C) e 4 (em D) da Formação Cabeças.

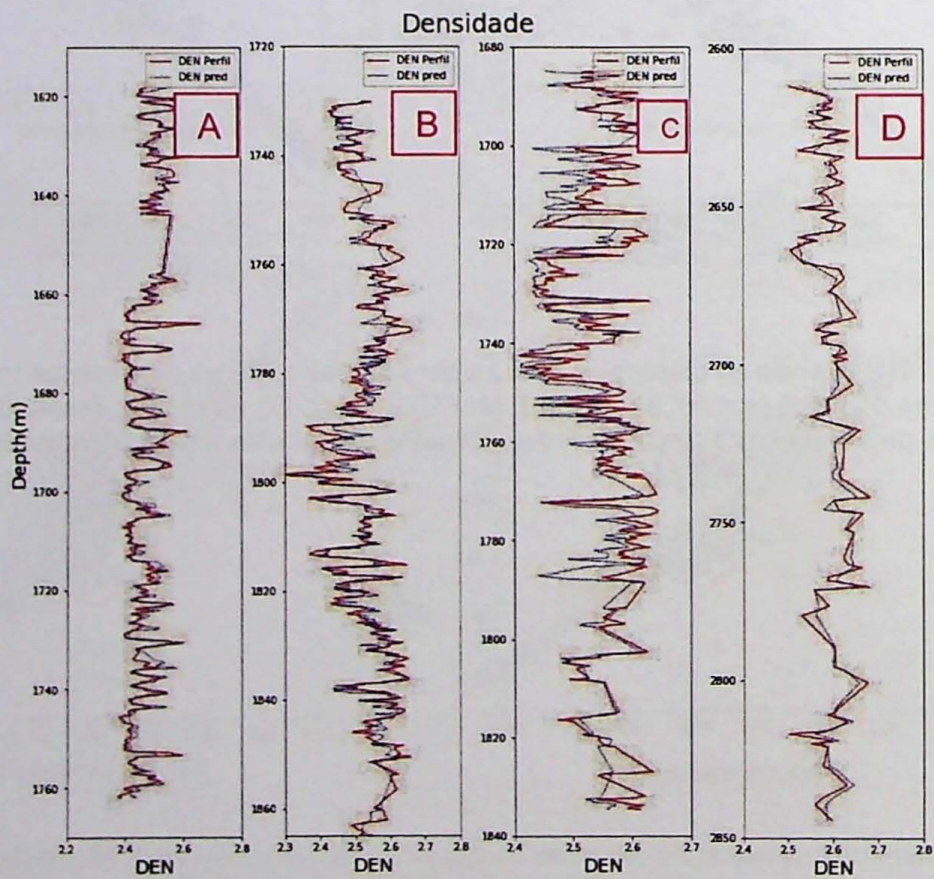


Figura 41: Perfis com as densidades previstas (em azul) e as densidades medidas (em vermelho) dos poços 5 (em A), 6 (em B), 7 (em C) e 8 (em D) da Formação Cabeças.



As curvas de densidade mostradas nas Figura 40 e Figura 41 foram inseridas no IP (Interactive Petrophysics) para calcular novos perfis de porosidade por meio da equação de densidade normalizada (3.6). Os resultados foram comparados com as porosidades conhecidas, ou seja, aquelas calculadas com uso das densidades medidas durante a perfilagem (Figura 42 e Figura 43).

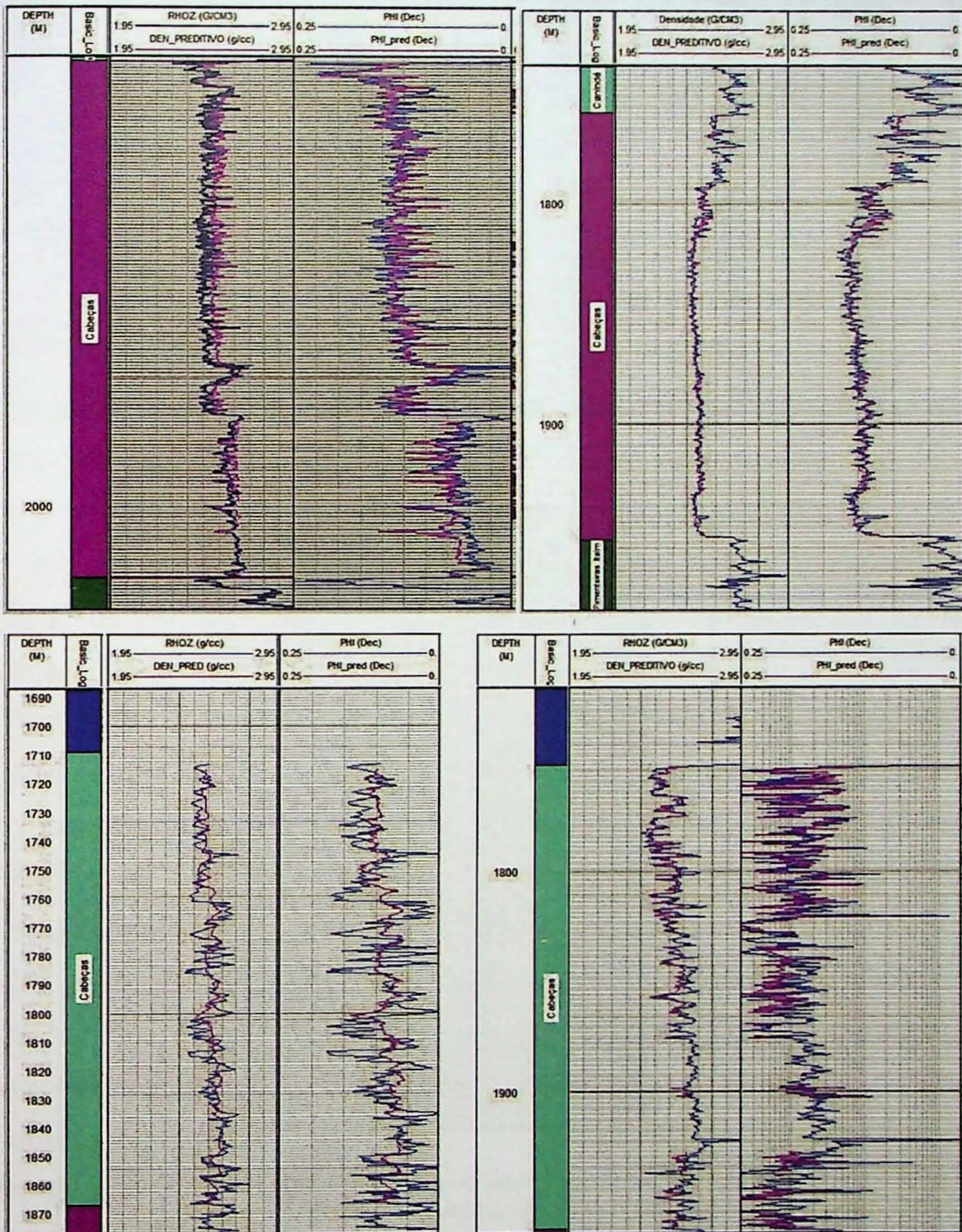


Figura 42: Porosidades calculadas por meio do perfil densidade (em azul) e por meio da densidade predita (em violeta) com o uso do método Gradient Boosting dos poços 1, 2, 3 e 4.



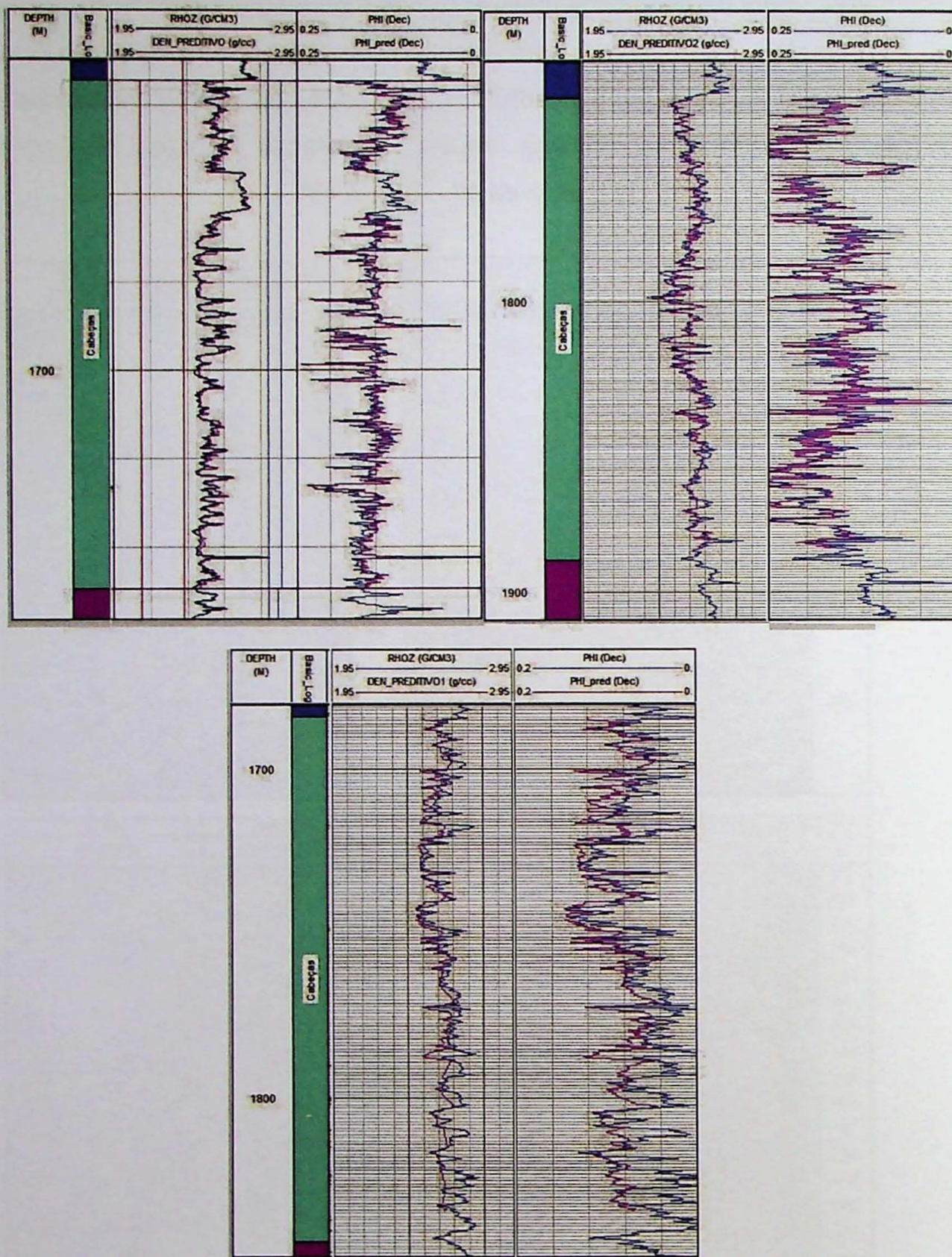


Figura 43: Porosidades calculadas por meio do perfil densidade (em azul) e por meio da densidade predita (em violeta) com o uso do método *Gradient Boosting* dos poços 5, 6 e 7.

Os resultados apresentados nas figuras 42 e 43 mostram alto grau de semelhança e compatibilidade entre as curvas de densidade, o que resultou na semelhança e compatibilidade das porosidades de cada um dos poços 1 a 7 na



unidade Cabeças. As porosidades, que anteriormente já haviam sido calculadas com o uso do perfil densidade medida, desta vez foram calculadas novamente com o uso de perfis da densidade predita por meio do algoritmo *Gradiente Boosting*. De forma geral, em uma análise visual, as curvas de porosidades mostradas nas imagens aparentam ser compatíveis na maior parte dos poços, com exceção do poço 3 (Figura 42), onde a curva de densidade proveniente do modelo não apresentou bons resultados, em comparação com os demais. Isso ocorreu, muito provavelmente, em função de zonas de arrombamento no poço, refletida nos demais perfis e da perda de resolução mostrada nos dados do perfil densidade.

#### 4.2.4 Uso da Densidade Recuperada para Predição de Porosidades

Após conseguir recuperar com êxito os perfis de densidade, fundamentais na determinação da porosidade do reservatório, os trabalhos prosseguiram de modo a atingir o objetivo final, ou seja, a elaboração de um modelo preditivo de porosidades, sem o uso de uma equação pré-estabelecida ou de softwares específicos para petrofísica, utilizando somente algoritmos autônomos baseados em técnicas de machine e deep learning.

A escolha do método *Gradient Boosting* para esta tarefa se deu em função dos melhores resultados para o perfil densidade preditiva terem sido obtidos a partir de um modelo construído com o este método. Além disso, alguns trabalhos como (EROFEEV et al., 2019) relatam performance superior da técnica quando comparada com o desempenho de outros métodos como as Redes Neurais Artificiais.

Os perfis Gama Ray, Resistividade, Sônico, Neutrão e Densidade predita foram escolhidos como *input* para o treinamento do modelo. Como parâmetros do modelo foram escolhidos, por tentativa e erro, em testes preliminares, os valores da apresentados na tabela 6.

Tabela 6: Métricas e parâmetros utilizados para treinamento do modelo.

MSE:	0,005	Taxa de aprendizagem:	0,05
Alpha:	2	Iterações	500
Prof.:	4	Divisão mínima de amostragem	2



Tabela 7: Métricas e parâmetros utilizados para treinamento do modelo.

GR	RES	NPOR	DTCO	RHOZ
17	17	14	16	35

A tabela 7 apresenta o fator de importância que o modelo deu a cada um dos dados de entrada. Embora o perfil densidade ainda seja o mais relevante, cerca de 2 vezes maior do que o segundo perfil mais importante, é também possível observar que houve uma aumento significativo, quase uniforme, na importância dos demais perfis.

No intuito de verificar se seria possível realizar previsões conhecendo apenas dados de um único poço, Inicialmente foi desenvolvido e treinado um modelo com apenas um poço e os demais foram utilizados como *blind test*. A porosidade do modelo preditivo foi comparada com aquela calculada com o uso da equação de densidade calculada no Iterative Petrophysics (IP). Como as porosidades iniciais mostraram com baixa correlação em comparação com as calculadas foi necessário treinar o modelo com mais dados. Foram então usados dados de quatro poços para treino e 4 para testes. Em todas as situações as correlações apresentaram um padrão bipolar de resultados, ou seja, correlações ou boas (Figura 44), cujo o coeficiente de correlação varia de 0.84 a 0.99 ou muito fracas (Figura 45 a Figura 47).



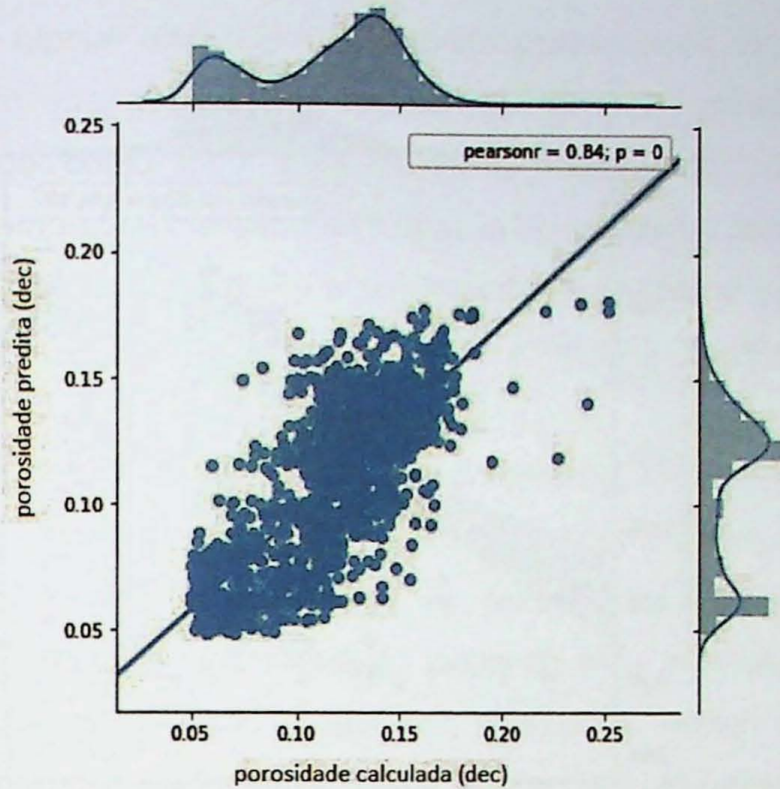


Figura 44: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e preditas pelo modelo no poço 5.

A Figura 44 mostra o diagrama de dispersão entre as porosidades, calculada e predita do poço 5, com R de 0,84, já as Figura 45, 46 e 47 mostram o mesmo tipo de diagrama do teste do modelo em outros poços (6, 7 e 8) porém sem correlação.

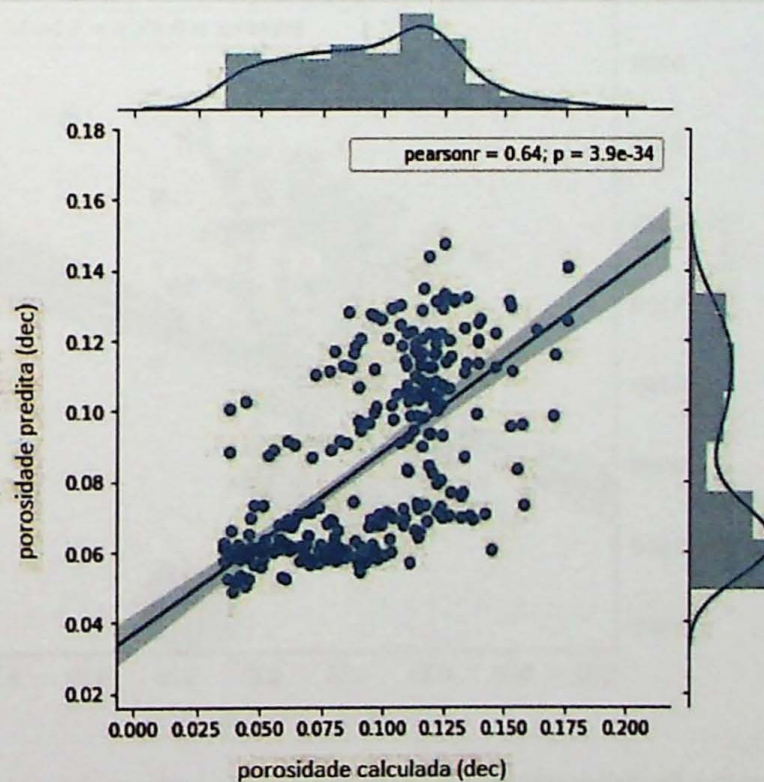


Figura 45: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e preditas pelo modelo no poço 6.



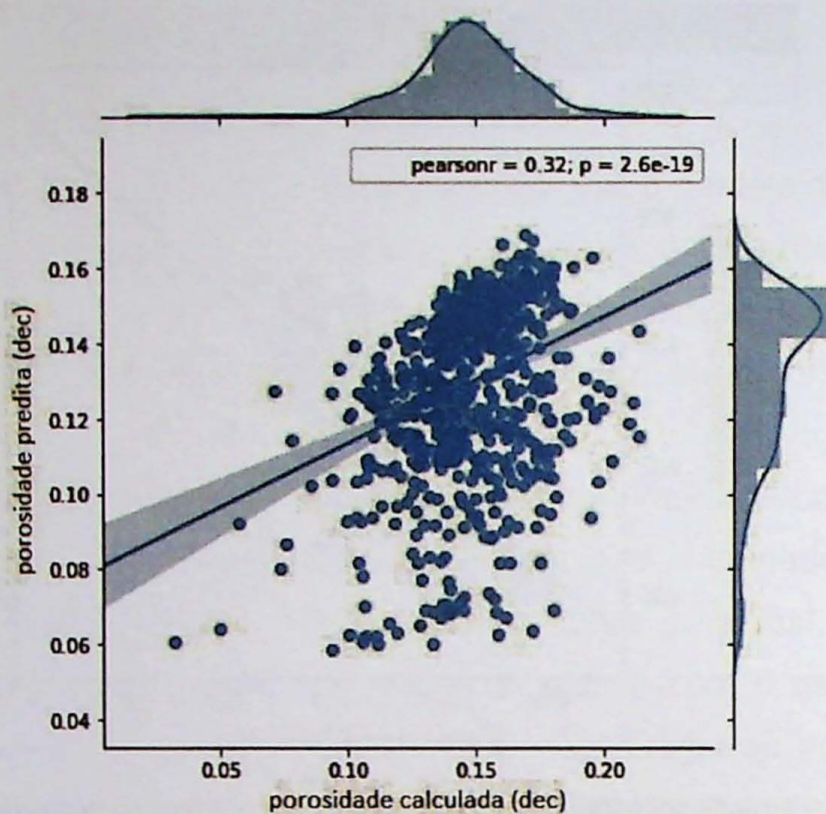


Figura 46: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e previstas pelo modelo no poço 7.

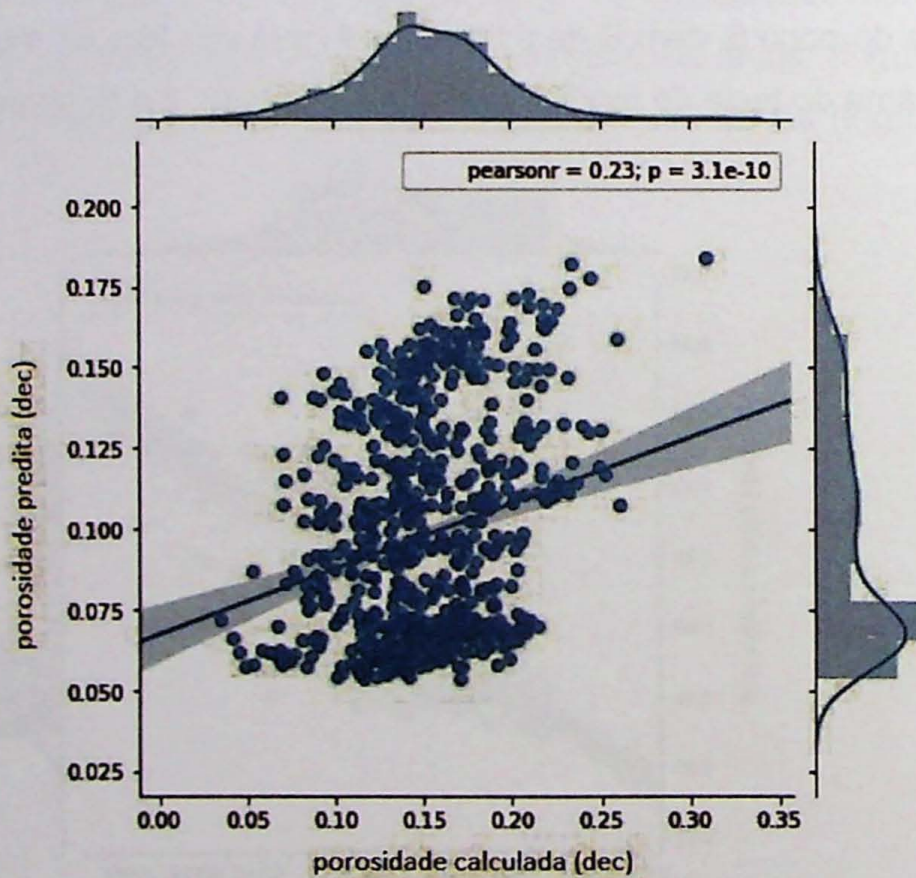


Figura 47: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e previstas pelo modelo no poço 8.



O padrão bipolar discutido no paragrafo anterior pode ter ocorrido devido a existência de um erro associado à recuperação do perfil densidade, que ao ser novamente usado como input para treino do modelo de porosidades parece contribuir ou mesmo potencializar o erro final, culminando em correlações ora muito fracas, ora muito boas. Este efeito é um importante indicador de que se deve ter muito cuidado ao utilizar as técnicas de estatística multivariada de forma indiscriminada.

A Figura 48 mostra os perfis de porosidade resultantes das predições realizadas pelo modelo nos poços 5 e 6 para o intervalo da Formação Cabeças. Na Figura 48(A) é possível observar que as porosidades preditas ficaram muito próximas das porosidades calculadas ao longo de todo intervalo, porém em B, o ajuste entre as porosidades não foi tão bom, alternando trechos bem ajustado com zonas pobremente correlacionáveis, como ocorre, por exemplo, entre 1825m e 1855m.



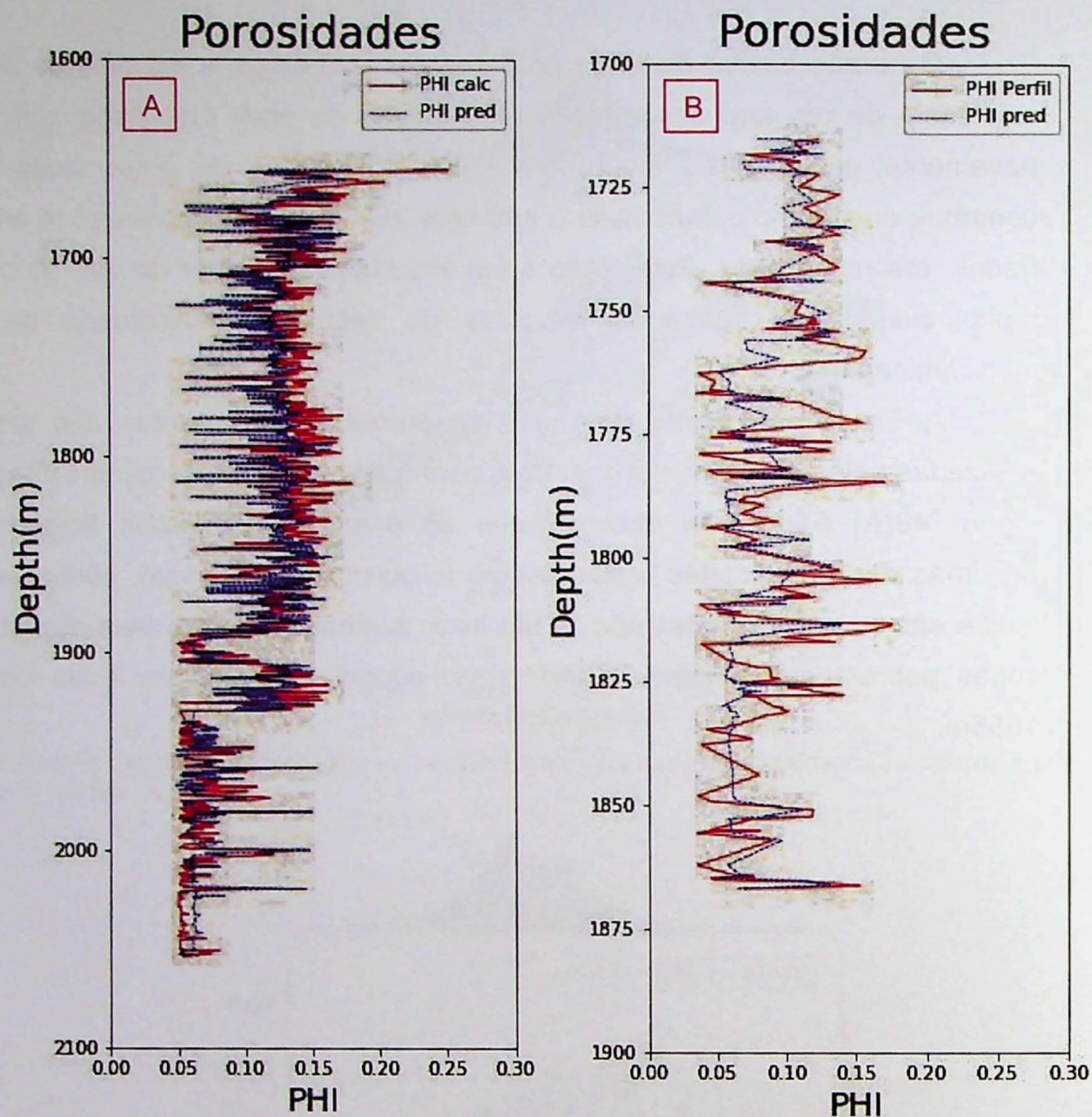


Figura 48: Perfis com as porosidades calculadas (linhas azuis) e preditas (linhas vermelhas) pelo modelo para os poços 5 (em A) e 6 (em B).

É imprescindível neste caso que o atributo de saída do modelo seja o resultado das porosidades determinadas a partir da rocha, em laboratório, ou mesmo de algum outro perfil capaz de indicar de forma direta a porosidade, como por exemplo, o perfil de ressonância magnética. Desta forma, seria possível construir de forma confiável, um modelo preditivo de porosidades para os reservatórios.

Nesta etapa do projeto, devido às conclusões de que não seria possível desenvolver um modelo preditivo adequado sem nenhum outro tipo de dado,



decidiu-se solicitar um novo lote de dados à Agência Nacional do Petróleo e Biocombustíveis (ANP).

Desta forma, foi possível obter junto à ANP um pacote adicional com dados de 40 poços, sendo que 3 deles continham perfis de poço com porosidade e permeabilidade medidas com uso da técnica Ressonância Magnética Nuclear (RMN). Um destes perfis RMN contemplou todo o intervalo da Formação Poti e também da Formação Cabeças, porém os outros dois tinham dados de apenas algumas dezenas de metros em cada uma destas formações. Desta forma, foi desenvolvido um modelo preditivo de porosidades com uso do perfil RMN como rótulo. O processo de treinamento e testes foi realizado de duas formas, sendo que na primeira, os dados dos 3 poços foram divididos aleatoriamente em um conjunto de treino (com 60% dos dados) e outro conjunto de testes (com 40% dos dados), o resultado é apresentado na Figura 49.

A segunda abordagem foi realizada usando a totalidade dos dados dos 3 poços que continham perfis de porosidade por ressonância magnética nuclear (RMN) para treinar o modelo. Uma vez treinado, o modelo foi testado em outros 7 poços e as porosidades preditas em cada um deles foram comparadas com as porosidades calculadas de forma usual, ou seja, por meio do *software Interactive Petrophysic* (IP) com uso da equação 3.6. Os resultados de cada teste são apresentados nas Figura 50 a Figura 55 em forma de diagramas de dispersão com as respectivas correlações lineares.

A Figura 49 mostra os resultados da primeira abordagem, com uma boa correlação entre a porosidade medida na perfilagem e a porosidade predita pelo modelo. Os resultados mostraram um coeficiente de correlação  $R$  de 0,93. Este fato evidencia que os parâmetros usados no treinamento do modelo foram adequados e que o modelo é capaz de realizar estimativas iniciais de porosidade de forma relevante.

A seguir são apresentados os resultados da segunda abordagem. As figuras 50 a 55 apresentam os resultados dos testes do modelo, ou seja, as porosidades preditas em função das porosidades medidas para cada um dos poços 22, 23 a 27. Nos poços 24 e 25 (figuras 52 e 53) a correlação linear entre porosidades predita e medida foi boa, sendo que seus coeficientes de correlação  $R$  variaram entre 0,88 e 0,91, novamente evidenciando a capacidade do modelo em realizar estimativas



preliminares de porosidade. No entanto, nos demais poços, ou seja, nos poços 23, 26 e 27 (Figura 49, Figura 52 e Figura 53) houve uma correlação moderada com coeficientes variando de 0,71 a 0,84. Isso mostra que o modelo, em geral, pode ser viável e é uma alternativa aos métodos convencionais, porém para alcançar uma melhor performance seria necessário um número maior de poços com perfis de porosidade por ressonância magnética nuclear (RMN) para o treinamento do modelo.

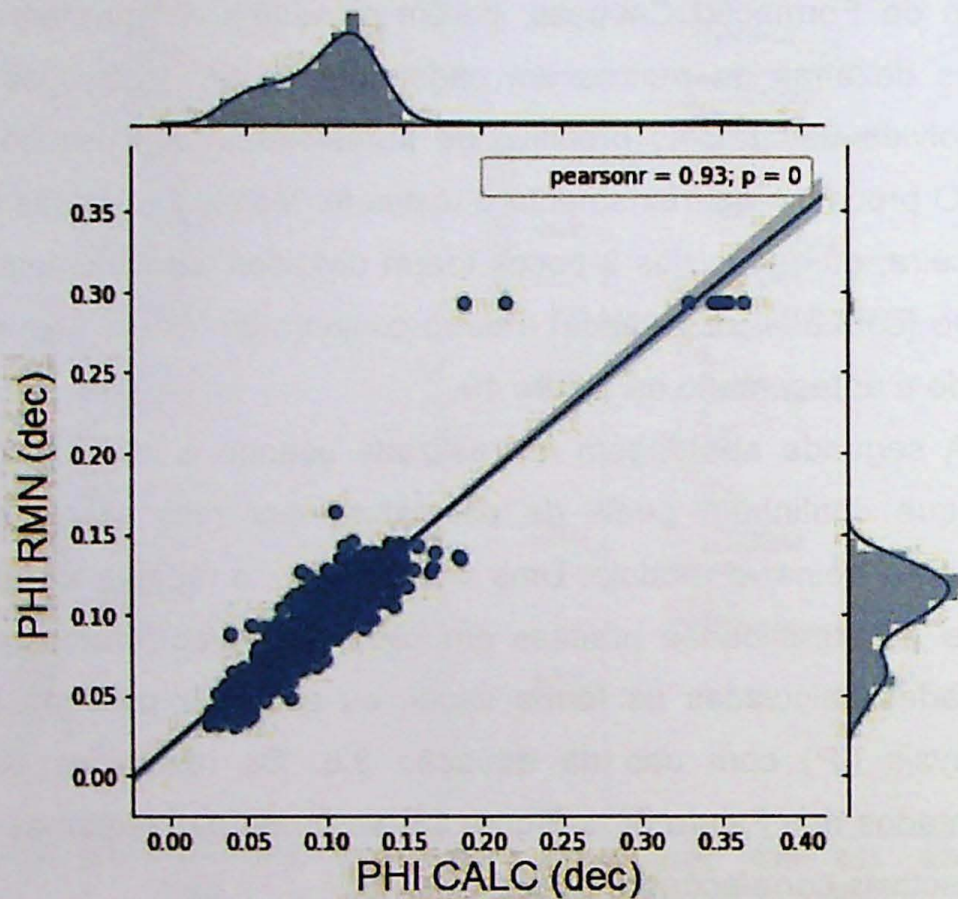


Figura 49: Diagrama de dispersão com a correlação linear entre o conjunto de dados de teste e o conjunto de dados de treino. As porosidades foram previstas com uso de ressonância magnética nuclear.



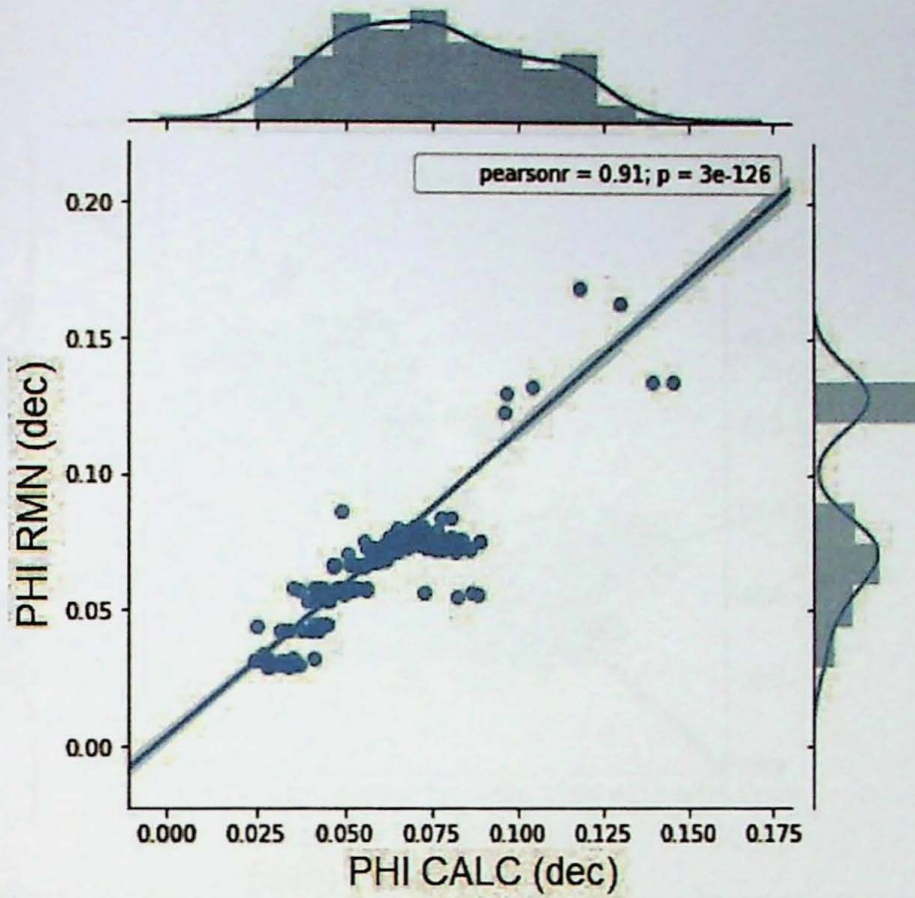


Figura 50: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e as porosidades previstas pelo modelo no poço 22 do intervalo Poti.

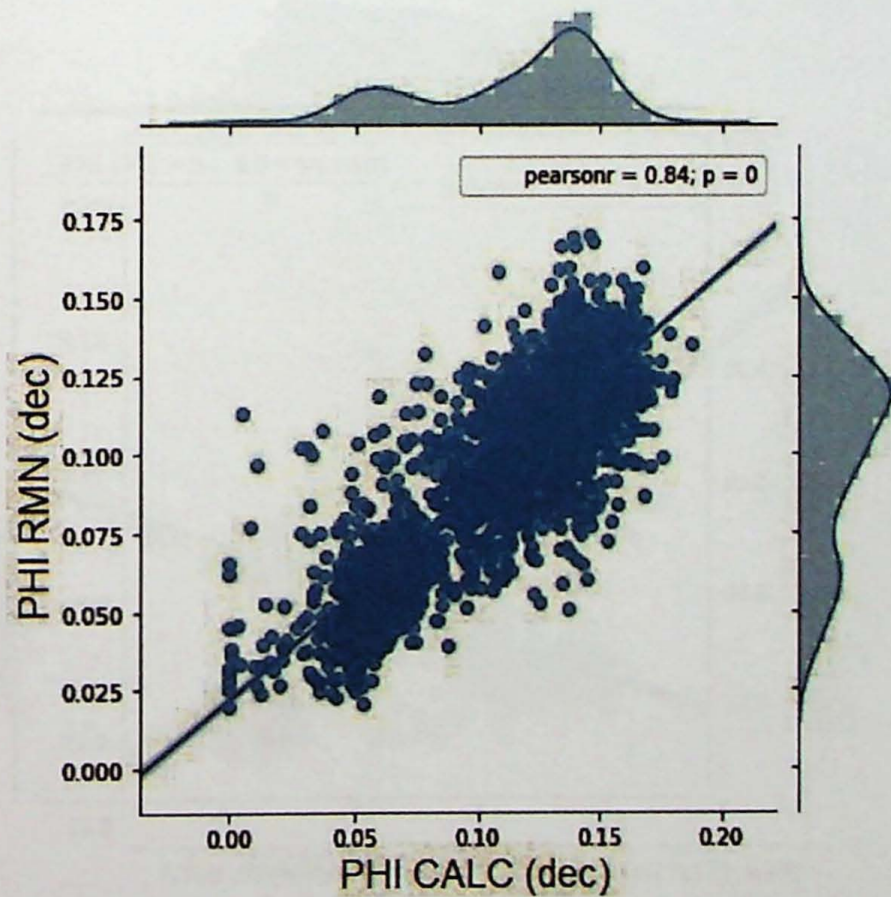


Figura 51: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e as porosidades previstas pelo modelo no poço 23 do intervalo Poti.



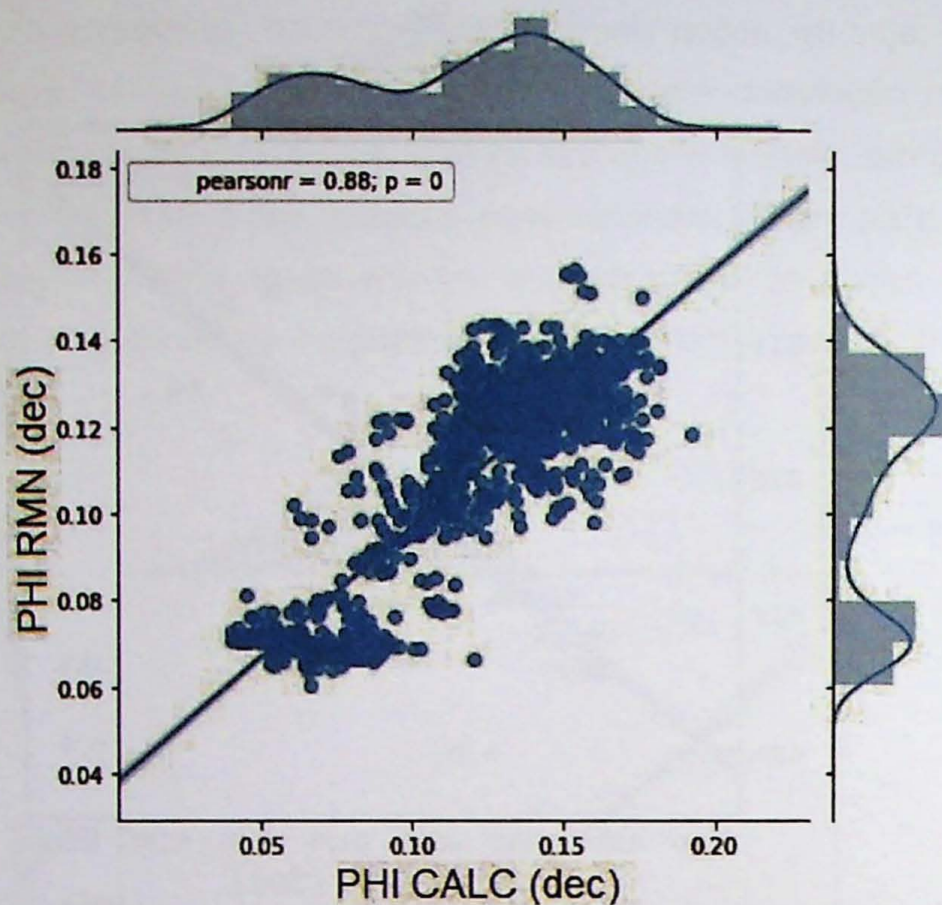


Figura 52: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e as porosidades previstas pelo modelo no poço 24 do intervalo Poti.

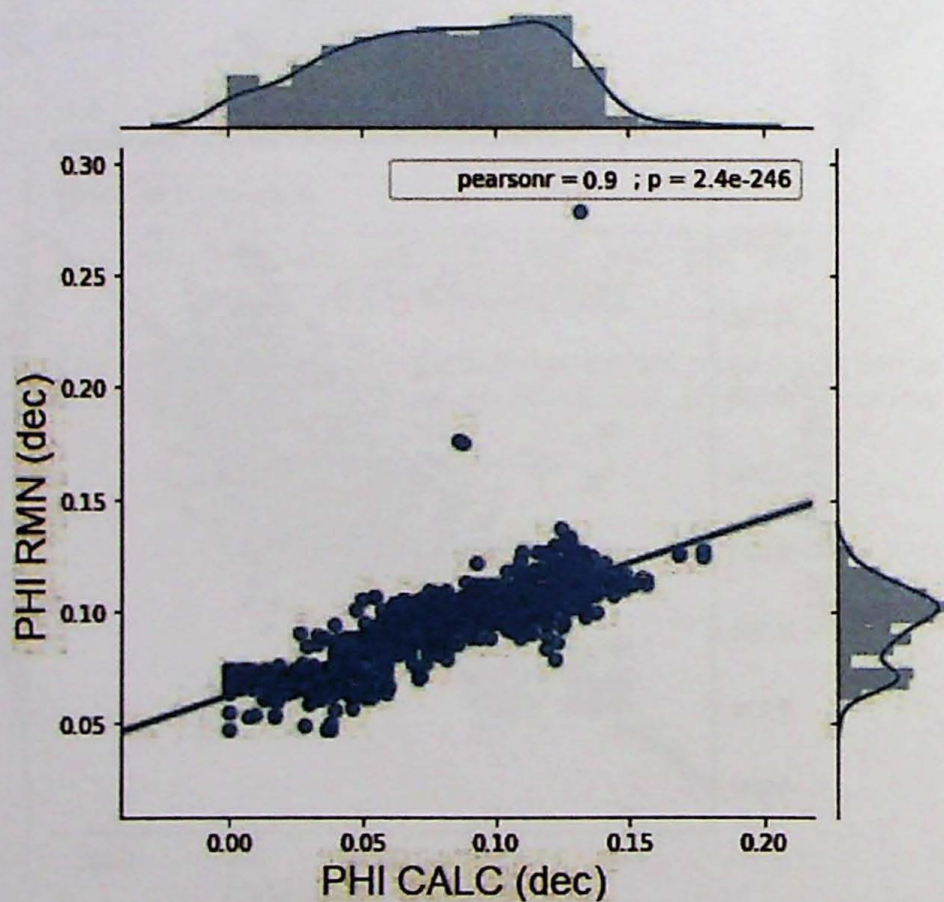


Figura 53: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e as porosidades previstas pelo modelo no poço 25 do intervalo Poti.



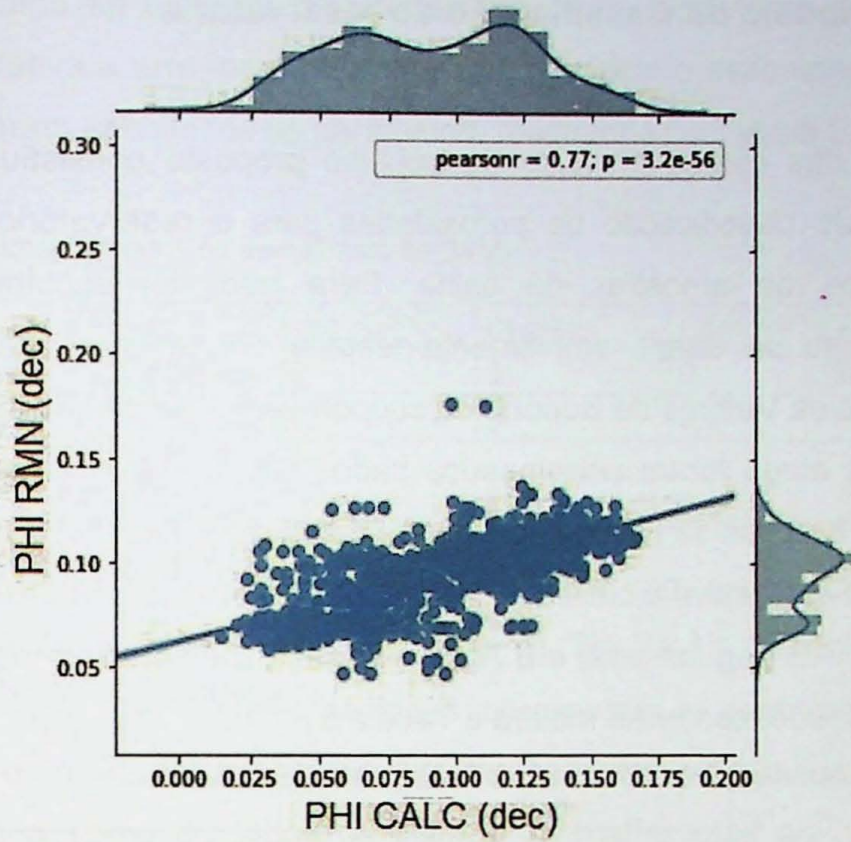


Figura 54: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e as porosidades previstas pelo modelo no poço 26 do intervalo Poti.

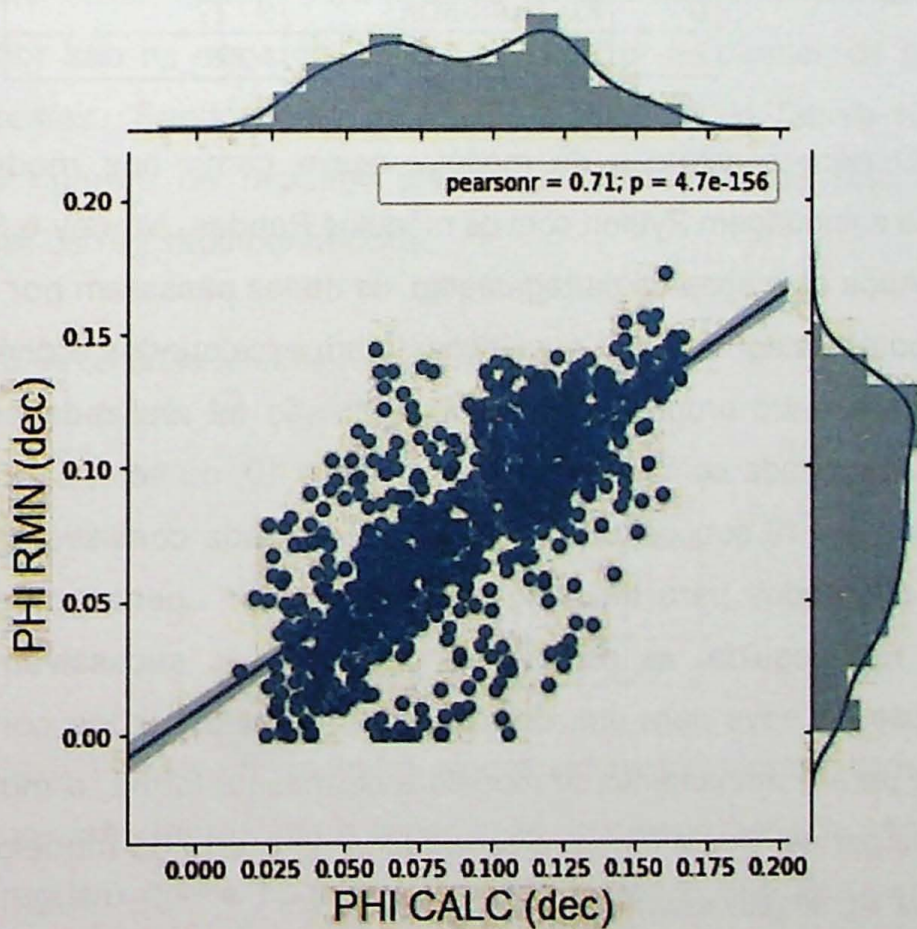


Figura 55: Diagrama de dispersão com a correlação linear entre as porosidades calculadas e as porosidades previstas pelo modelo no poço 27 do intervalo Poti.



### 4.3 Modelo de classificação do reservatório

Uma das etapas do fluxo de trabalho proposto consistiu em desenvolver um modelo de classificação de porosidades para o reservatório Poti, com base em descrições de amostras de calha. Para isso, foi escolhida uma técnica de treinamento de dados amplamente testada em diversas áreas conhecida como Máquinas de Vetores de Suporte ou *support vector machines* (SVM).

Nesta etapa foram selecionados dados de perfis e descrições de amostras de calha de 8 poços da Formação Poti. As descrições das amostras inclusas no pacote de dados continham a classificação categórica: F (fechada), FR (fechada-regular), R (regular), RB (regular-boa) e B (boa) e neste trabalho foram associadas a intervalos de porosidade, conforme mostra a Tabela 8.

Tabela 8: Classificação prévia de porosidade de amostras de calha.

Classe	Classificação	Porosidade (%)
A	FECHADA	0 - 3
B	FECHADA-REGULAR	3 - 6
C	REGULAR	6 - 9
D	REGULAR-BOA	9 - 12
E	BOA	> 12

O para construção do modelo, assim como nos modelos anteriores, foi utilizada a linguagem Python com os módulos Pandas, Numpy e Sklearn.

Após as etapas de carregamento, os dados passaram por uma transformação de atributos categóricos em numéricos, foram escalonados e divididos em um grupo de treino e outro grupo de testes. Esta divisão foi realizada por meio da técnica validação cruzada com o parâmetro  $k \text{ folder} = 10$ , ou seja, subdividindo o conjunto de dados em 10 subconjuntos. A validação cruzada consistiu em usar nove partes (90%) dos dados para treinar o modelo e deixar apenas uma parte (10%) para testes. Em seguida, as partes foram permutadas sucessivamente, mantendo a proporção de nove para um, de modo que todos os dados contribuíssem de forma indireta para o treinamento do modelo e da mesma forma, o modelo foi testado em todas as partes. Desta forma, a capacidade de acerto do modelo é potencializada e se reduz significativamente as incertezas.



Em seguida foi utilizada uma ferramenta conhecida como *grid search*, que busca, por tentativa e erro, os parâmetros que otimizem o desempenho do modelo. Desta forma, foram escolhidos os parâmetros mostrados na Tabela 9:

Tabela 9: Parâmetros utilizados no treinamento da SVM.

	C	Cache-size	Degree	Gamma	Kernel
Default	0	200	3	0	"RBF"
Gridsearch	50	200	3	0,1	"RBF"

Uma das principais ferramentas de avaliação do modelo de classificação é a matriz de confusão. Após o treinamento do modelo com o uso do *GridSearch* foi calculada a matriz de confusão apresentada na Tabela 10. As colunas da matriz mostram as classes conhecidas e as linhas representam as preditas. Desta forma, a matriz mostra quantas predições corretas em cada classe o modelo foi capaz de realizar, sendo a soma de todos os valores da matriz igual ao número total de registros.

A diagonal principal da matriz de confusão contém o número de previsões corretas. Deste modo, quanto mais próximo de zero os elementos fora da diagonal principal menor são os erros do modelo ao predizer as classes de porosidade do conjunto de testes. Sendo assim, na linha **A** e coluna **A** da Tabela 10, o valor 121 representa o número de registros preditos corretamente, ou seja, aqueles que coincidem com os registros conhecidos.

Tabela 10: Matriz de confusão do modelo treinado com SVM.

	A	B	C	D	E
A	121	2	9	0	0
B	2	152	29	2	0
C	2	20	249	0	3
D	0	22	47	57	2
E	0	2	15	2	17

Na Tabela 11 são apresentados alguns estimadores estatísticos da matriz de confusão. A precisão representa a capacidade do classificador de não rotular como positiva um registro que é negativo, ou seja, um dado que não pertence a uma determinada classe. A coluna acertos mostra razão entre o número de predições



corretas e o total. A terceira coluna mostra a média ponderada entre as duas primeiras. A última coluna mostra o total de ocorrência de cada classe.

Os resultados da diagonal principal da matriz de confusão mostram que o modelo obteve uma boa performance na predição e classificação das classes *A*, *B* e *C*. No entanto, as classes *D* e *E* apresentaram as menores porcentagens de acerto e eram também as classes que continham o menor número de registros, especialmente a classe *E*. Isso pode sugerir que uma nova classificação deveria ser feita para a esta base de dados, provavelmente com apenas 4 classes.

Em geral, o modelo obteve um boa performance, com um índice de acerto de 82%, porém para melhores resultados deve-se utilizar registros de perfilagem de um numero mais expressivo de poços no treinamento do modelo.

Tabela 11: Estimadores estatísticos da matriz de confusão.

	Precisão(dec)	Acertos(dec)	Média ponderada (dec)	Ocorrências
A	0,97	0,91	0,94	132
B	0,77	0,82	0,79	185
C	0,7	0,91	0,79	272
D	0,93	0,43	0,59	128
E	0,77	0,47	0,59	36
Total	0,82	0,80	0,77	753

#### 4.4 Implementação e treinamento de um modelo preditivo de permeabilidades com uso do algoritmo *Gradient Boosting*

A permeabilidade é um dos parâmetros mais importantes de um reservatório uma vez que permite conhecer as condições para produção do óleo, gás ou água “*in place*”, porém é também um dos parâmetros mais complexos de se determinar, dado o número de variáveis direta e indiretamente envolvidas. A forma mais eficaz de se determinar a permeabilidade é por meio de medidas diretas em laboratório, porém



estas medidas são caras e pontuais. Outra maneira muito utilizada para estimar os valores de permeabilidade em um reservatório é o cálculo indireto a partir de dados de perfis de poço (Morris e Briggs, 1967; Wyllie e Rose, 1950). Recentemente, alguns autores tem mostrado que a permeabilidade pode ser estimada com certo grau de sucesso por meio do uso de técnicas de inteligência artificial. (Ahmadi e Chen, 2018; Erofeev et al., 2019; Arigbe et al., 2018). Esta forma de estimar permeabilidade é particularmente nova e promissora, porém ainda muito pouco explorada. Nesta etapa do trabalho uma metodologia alternativa para estimativa de permeabilidade é proposta. Para isso, será utilizado o algoritmo *Gradient Boosting*.

Neste estudo foram realizadas duas abordagens. Na primeira foram usados dados de perfilagem de 5 poços da Bacia do Parnaíba contendo apenas os logs de perfilagem convencionais (Gama, Resistividade, Sônico, Densidade e Neutrão) juntamente com perfis de ressonância (RMN) (Tabela 12). Na segunda abordagem, além dos perfis convencionais, foram incluídos dados de cromatografia gasosa e logs de perfuração, como ROP, RPM e peso sobre broca e fluidos (Tabela 13). Em ambos os casos foram desenvolvidos modelos capazes de determinar a permeabilidade nos principais reservatórios da Bacia do Parnaíba, ou seja, Poti e Cabeças e a permeabilidade RMN foi usada como rótulo (resultado final) para treinar do modelo.

Tabela 12: Dado de perfis de poço utilizados no treinamento do modelo A.

	DEPTH	GR	RES	RHOZ	DTCO	NPHI	PHI	PERM	KTIM	TCMR
0	1673.9612	124.9989	4.4825	2.4960	63.7054	0.1160	0.1547	0.3586	0.4205	0.0782
1	1674.1136	110.6623	4.6443	2.5405	63.6209	0.1300	0.1255	0.0697	0.0249	0.0529
2	1674.2660	105.5945	4.6355	2.5951	63.5798	0.1471	0.0880	0.0041	0.0008	0.0389
4	1674.5708	129.3937	5.7209	2.6446	64.8015	0.1657	0.0724	0.0011	0.0008	0.0401

Tabela 13: Dado de perfis de poço utilizados no treinamento do modelo B.

DEPTH	GR	RES	RHOZ	NPOR	DTCO	CMFF	KTIM	TQI	PSB	ROP	HYDC	METH	ETH	PRP	BUT	PEN
1536.0455	60.2486	1.1120	2.4254	0.0773	65.5373	0.1263	62.3320	389.150	9.380	20.560	8274.0000	3953.3799	170.060	29.580	104.210	246.480
1536.2360	35.0503	2.2421	2.4299	0.0741	64.9425	0.1282	137.6862	379.494	8.964	20.916	8269.2002	4002.2700	174.654	37.816	107.714	265.182
1536.4265	35.9322	5.4680	2.4679	0.0754	65.9955	0.1269	128.4053	369.838	8.548	21.272	8264.4004	4051.1599	179.248	46.052	111.218	283.904



Na fase de elaboração do modelo foi escolhida a técnica *Gradient Boosting*, baseada em árvores de decisão. Uma vez treinado, o modelo foi testado nos dados de cada poço selecionado e realizou, em cada conjunto, uma regressão linear.

Na primeira abordagem, foram elaborados modelos com dados de 3 poços para os intervalos Cabeças e dois poços para o intervalo Poti. Os modelos foram testados de duas maneiras distintas. Na primeira os dados foram divididos de forma aleatória em dois conjuntos, um deles denominado conjunto de treino com 60% dos dados e o outro conjunto de testes com 40% dos dados. Na segunda, todo o conjunto contendo dados dos 3 poços foi utilizado para treinar o modelo, que foi testado em dados de outros poços da bacia, escolhidos aleatoriamente.

A Figura 56 mostra os resultados da primeira forma de avaliação (usando o conjunto de dados dividido em treino e teste), em que os resultados da permeabilidade preditiva têm correlação linear moderada, apresentando coeficiente de correlação  $R$  de 0,85, entre as permeabilidades medidas e previstas, no entanto, a segunda forma de avaliação (testar o modelo em outros poços) mostrou correlações fracas entre as permeabilidades medidas com as modeladas em outros poços da bacia (Figura 56 a 59). Isso pode indicar que outros fatores que influenciam direta e indiretamente a permeabilidade variam muito de poço para poço e muitos deles podem não ser contemplados quando o modelo é construído apenas com logs convencionais de perfilagem. Um destes fatores pode ter sido a distância entre os poços, uma vez que se trata de uma bacia com propriedades muito heterogêneas, ao utilizar um poço muito distante, este muito provavelmente terá propriedades muito distintas, como no caso do poço usado na Figura 57.



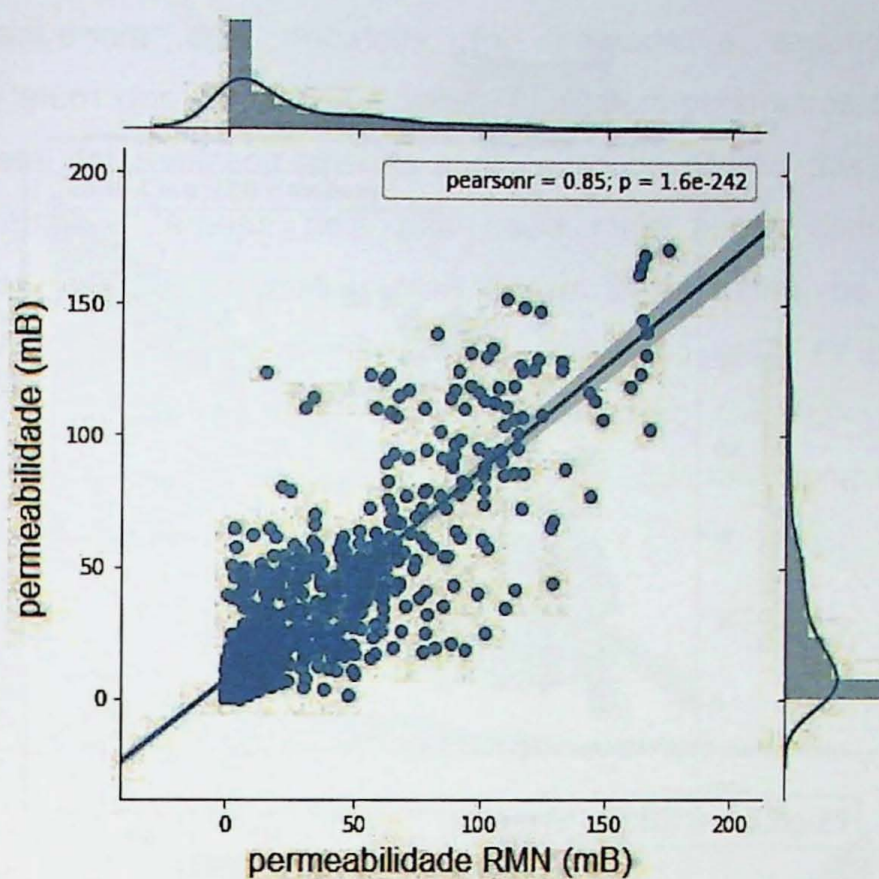


Figura 56: Diagrama de dispersão com a correlação linear entre a permeabilidade medida por ressonância magnética nuclear e permeabilidade predita no intervalo da formação Cabeças usando a primeira abordagem.

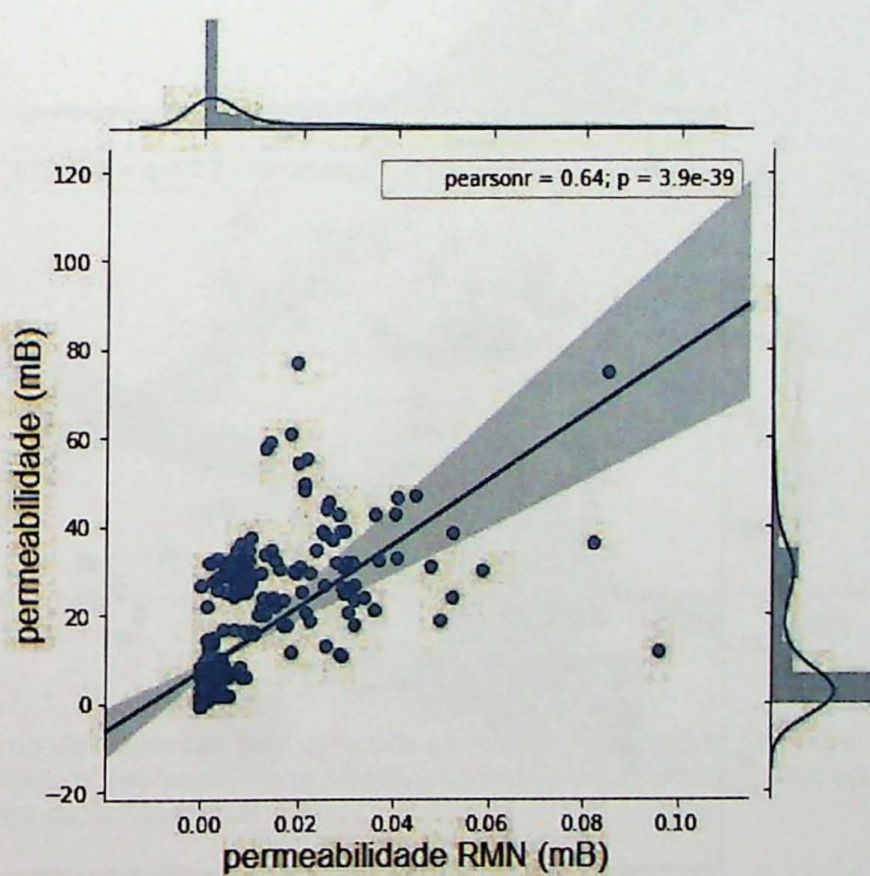


Figura 57: Gráfico da permeabilidade medida com o perfil ressonância magnética pela permeabilidade determinada pelo modelo no poço 2 da formação Cabeças.



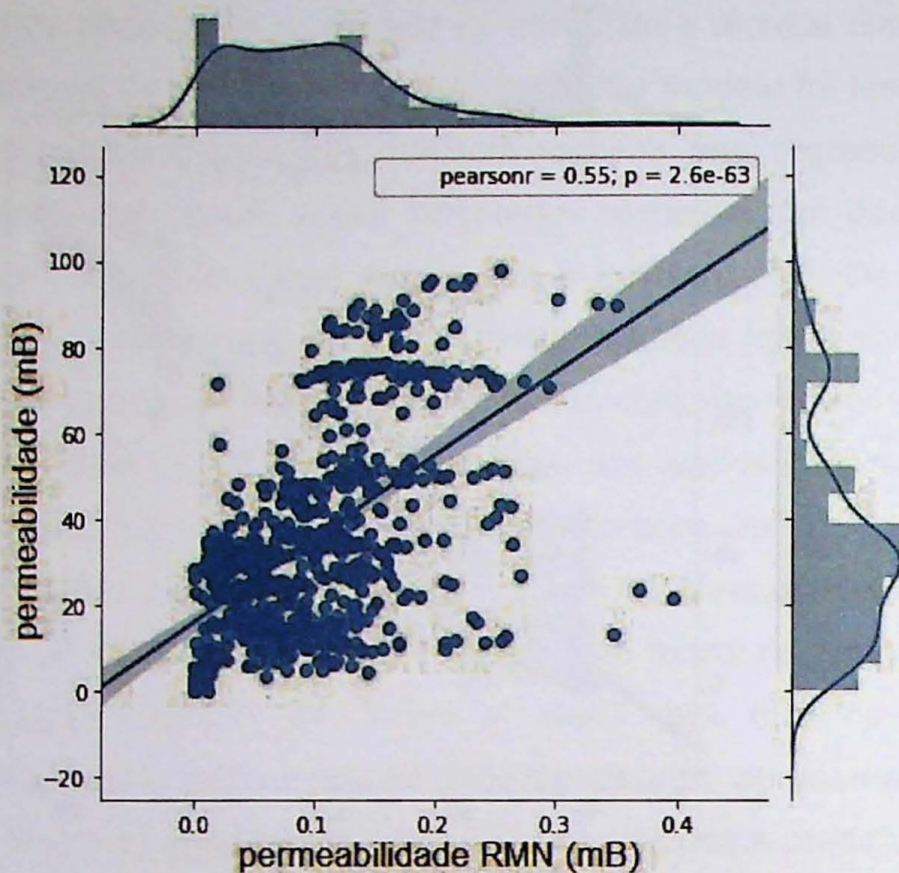


Figura 58: Gráfico da permeabilidade medida com o perfil ressonância magnética pela permeabilidade determinada pelo modelo no poço 3 da formação Cabeças.

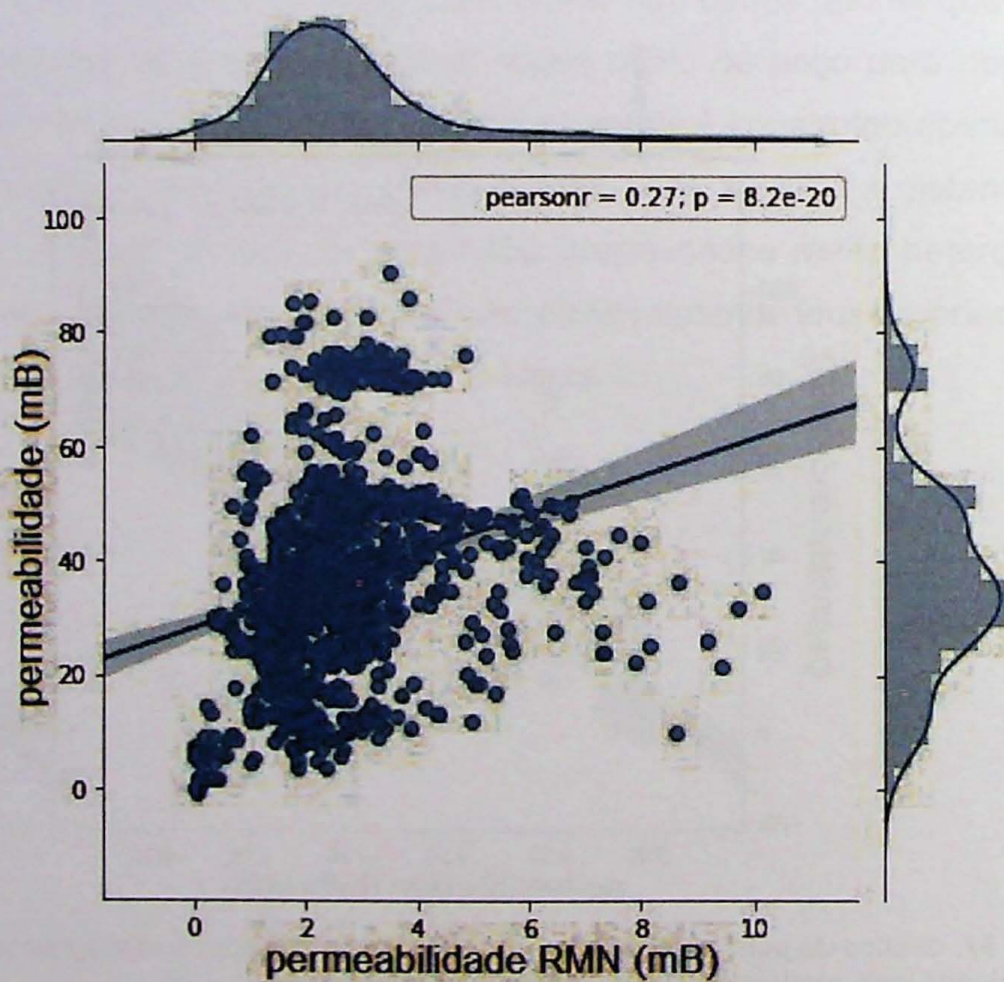


Figura 59: Gráfico da permeabilidade medida com o perfil ressonância magnética pela permeabilidade determinada pelo modelo no poço 4 da formação Cabeças.



Na sequência dos trabalhos, foi realizada a segunda abordagem, considerando além dos perfis convencionais, também parâmetros de perfuração e gás. Nesta fase foi realizado apenas o processo de divisão dos dados em dois conjuntos (treino e testes), pois não havia mais poços com porosidade e permeabilidade RMN para realizar mais testes. Desta forma, os resultados são apresentados na Figura 60, que mostra a correlação entre as permeabilidades medidas e modeladas no intervalo do reservatório Cabeças. Os resultados mostraram, em geral, boa correlação  $R=0,89$  quando comparados com a permeabilidade de perfil RMN.

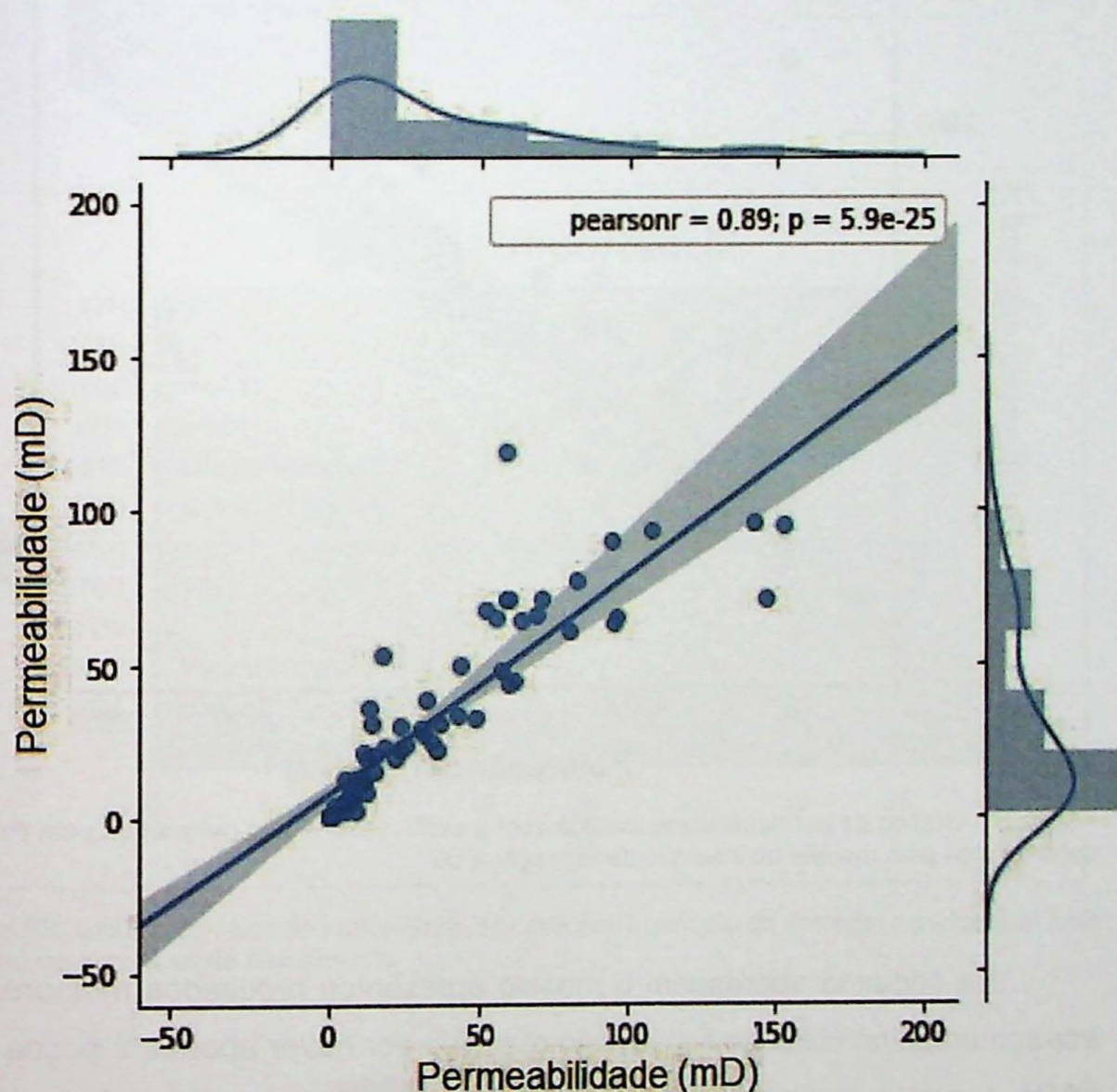


Figura 60: Diagrama de dispersão com dados de permeabilidade medida com o perfil ressonância magnética em função da permeabilidade preditiva no intervalo da Formação Cabeças. A linha azul mostra o coeficiente de correlação R.

O mesmo processo de treinamento do modelo foi realizado também no intervalo Poti, no entanto foram usados dados de apenas dois poços, pois eram os



únicos dados disponíveis para este intervalo. O modelo mostrou um desempenho levemente inferior como mostra a Figura 61, com coeficiente de correlação de 0,83 calculado entre a permeabilidade determinada pelo modelo e a permeabilidade do perfil RMN.

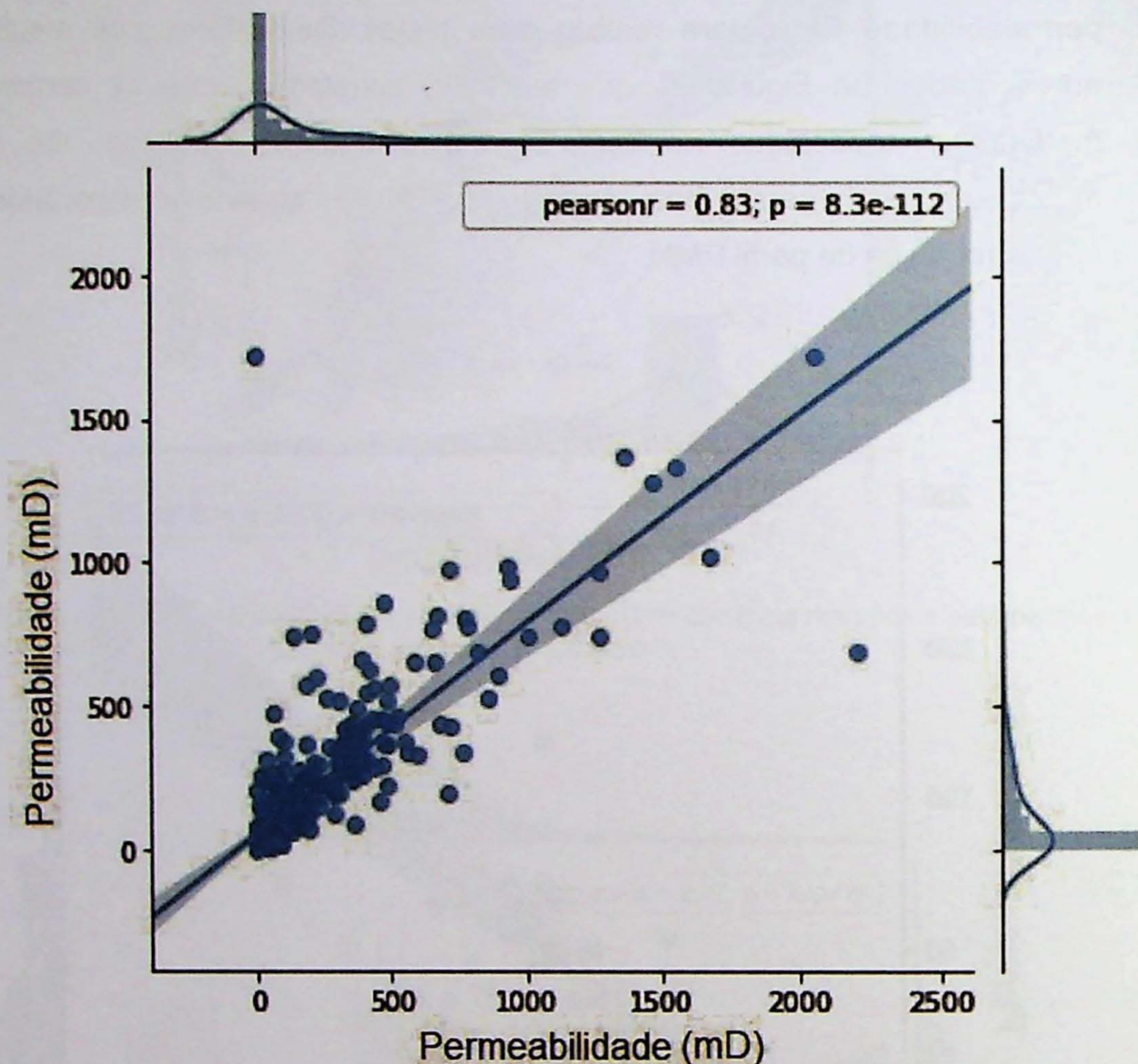


Figura 61: Gráfico da permeabilidade medida com o perfil ressonância magnética pela permeabilidade determinada pelo modelo no intervalo da formação Poti.

Na segunda abordagem o modelo apresentou resultados melhores, porém o espaço amostral utilizado ficou muito reduzido por haver apenas 2 poços com perfis RMN, sendo que em ambos só existem poucas dezenas de metros na formação Cabeças. Para resultados mais conclusivos a respeito desta técnica seriam necessários mais dados de RMN.

O fator de importância de cada perfil na construção do modelo é apresentado na Tabela 14, onde a maior contribuição foi dada ao perfil porosidade RMN e em



seguida empatam os perfis gama e densidade com 10% cada, conforme mostra a figura 62.

Tabela 14: Fator de importância dado pelo modelo a cada perfil utilizado.

<i>GR</i>	<i>RES</i>	<i>RHOZ</i>	<i>NPOR</i>	<i>DTCO</i>	<i>PHI (RMN)</i>
10%	5%	10%	1%	3%	37%
<i>TQI</i>	<i>PSB</i>	<i>ROP</i>	<i>HYDC</i>	<i>METH</i>	<i>ETH</i>
11%	11%	4%	4%	2%	2%

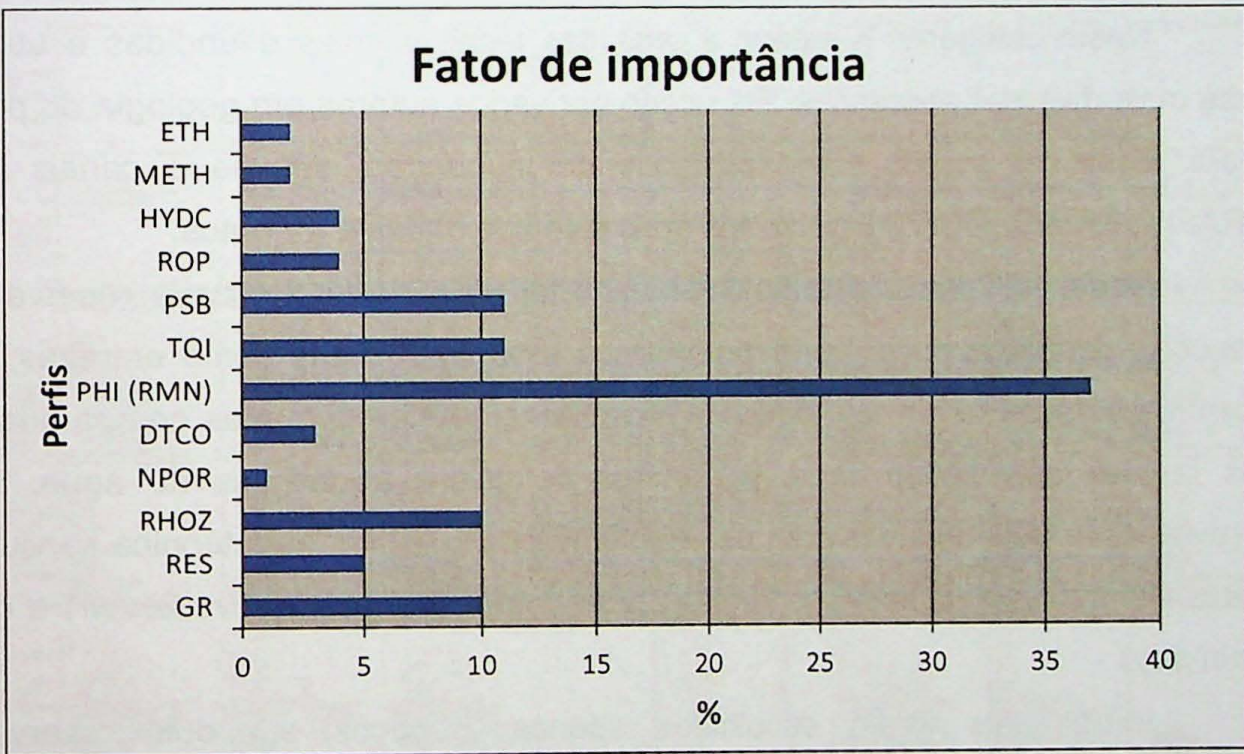


Figura 62: Gráfico do fator de importância de cada perfil (atributo de entrada) considerado pelo modelo no processo de treinamento.



#### 4.5 Delimitação do reservatório e de zonas portadoras de gás (*netpay*) por agrupamento com uso do método K-means

No intuito de identificar e agrupar atributos com características similares foi utilizado uma técnica de *machine learning* conhecida como *K-Means*. Os métodos de “*clusterização*” ou agrupamento diferem das tradicionais técnicas de classificação, onde se tem um conjunto de dados de entrada e se conhece um rótulo de saída, ou seja, o parâmetro a ser determinado. Deste modo, são eficazes quando se deseja identificar grupos ou agrupar dados por semelhança de características. Neste tipo de técnica não é necessário conhecer a resposta final para treinar o modelo e por esta razão são classificados como métodos não supervisionados.

Nesta categoria, K-means é uma das técnicas mais difundidas e utilizadas nas mais diversas aplicações. Foi usado por vários autores em geologia do petróleo para identificar grupos como planícies de inundação, *sabkha* e canais fluviais (RAFIK; KAMEL, 2017), a partir dos mais diversos dados e atributos.

Neste trabalho o algoritmo k-means foi utilizado no intervalo reservatório e recebeu dados de poço, perfis petrofísicos e curvas de gás como entradas. Desta forma, os dados foram agrupados em camadas e estas camadas comparadas com as regiões delimitadas como portadoras de gás e portadoras de água, ambas previamente determinadas com uso do *software* IP. Assim, esta técnica serviu como uma alternativa ao cálculo de regiões portadoras de água (*net reservoir*) e de gás (*net pay*).

Inicialmente foram escolhidos apenas 3 poços em dois reservatórios (Cabeças e Poti) que continham os tradicionais perfis de *wireline*, densidade, neutrão, sônico, gama e resistividade, além dos dados de gás total, metano e perfis de porosidade e permeabilidade provenientes de ressonância magnética RMN. A principal razão de se escolher apenas 3 poços foi que eram os únicos poços com dados de RMN que contemplavam totalmente os intervalos escolhidos.

As Figura 63 e Figura 64 mostram os gráficos de todos os atributos, dois a dois, já agrupados pelo algoritmo, para uma análise quantitativa dos parâmetros dos reservatórios Poti e Cabeças. Na figura 63 é possível perceber correlações lineares fracas a moderadas entre os atributos Gama e Densidade e também entre Gama e



porosidade. Já no reservatório Cabeças, representado pelos atributos da figura 64, nota-se correlações lineares mais fortes entre os dados do Sônico e porosidade, sônico e resistividade e sônico e densidade. Além disso, a porosidade mostra correlação moderada com resistividade, densidade e com gama.

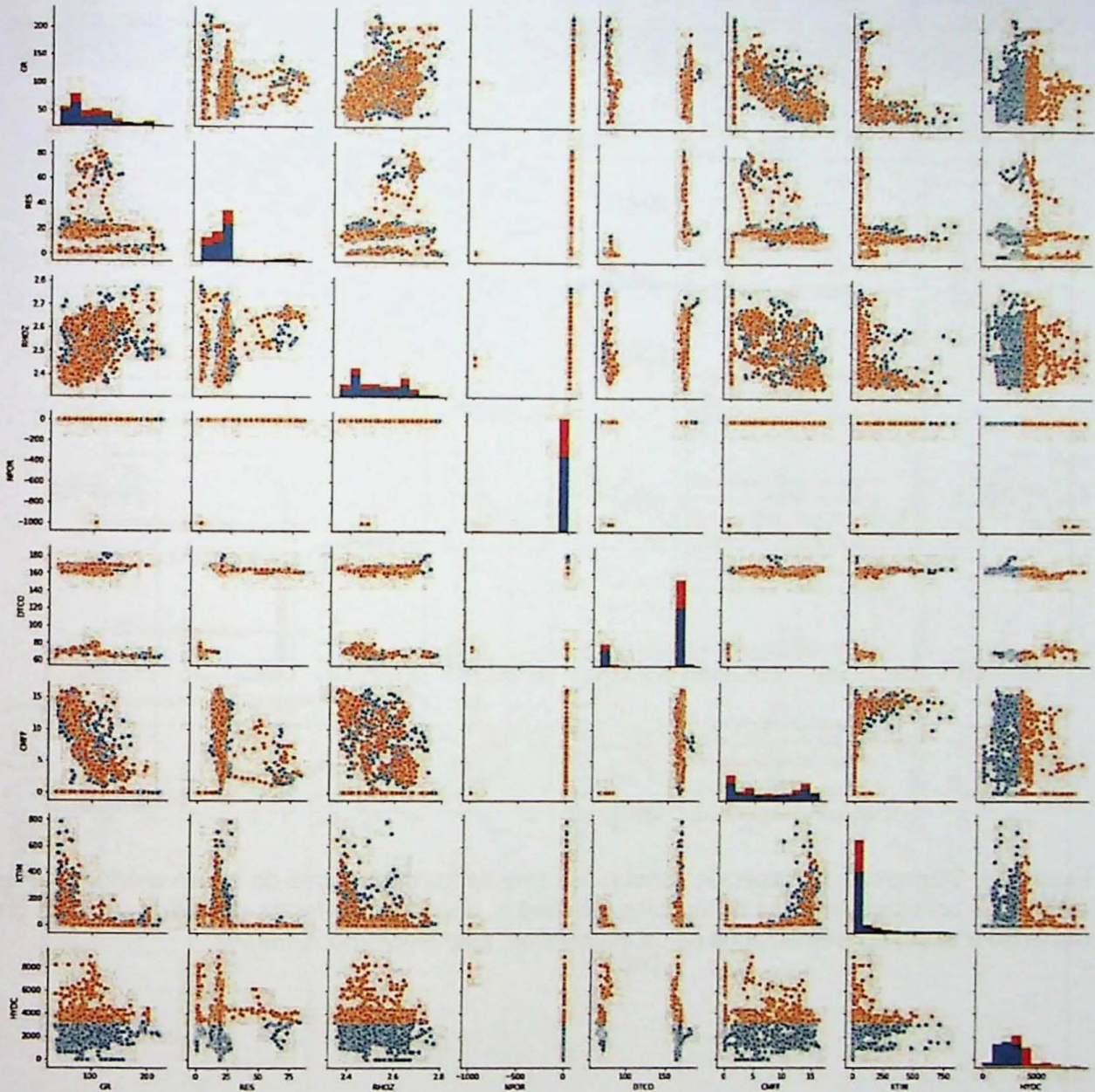


Figura 63: Diagramas de dispersão construídos com dados de 3 poços da Formação Cabeças. Os diagramas contemplam todos os atributos envolvidos, dois a dois e foram divididos em dois grupos (amarelo e azul), por semelhança de características, pelo algoritmo K-means.

Duas abordagens foram realizadas para definir quais os parâmetros mais relevantes na determinação das zonas de net pay por agrupamento. Na primeira foram utilizados os dados que incluem porosidade e permeabilidade RMN (Tabela 15). Na segunda foram excluídos os dados de porosidade e permeabilidade RMN (Tabela 16). A segunda abordagem foi realizada para testar sua viabilidade e



compatibilidade, uma vez que é feita com dados de perfis usuais. Já a primeira dependeria de uma perfilagem adicional que inclui ressonância magnética, porém este não é um ensaio comum de ser realizado devido aos custos elevados.

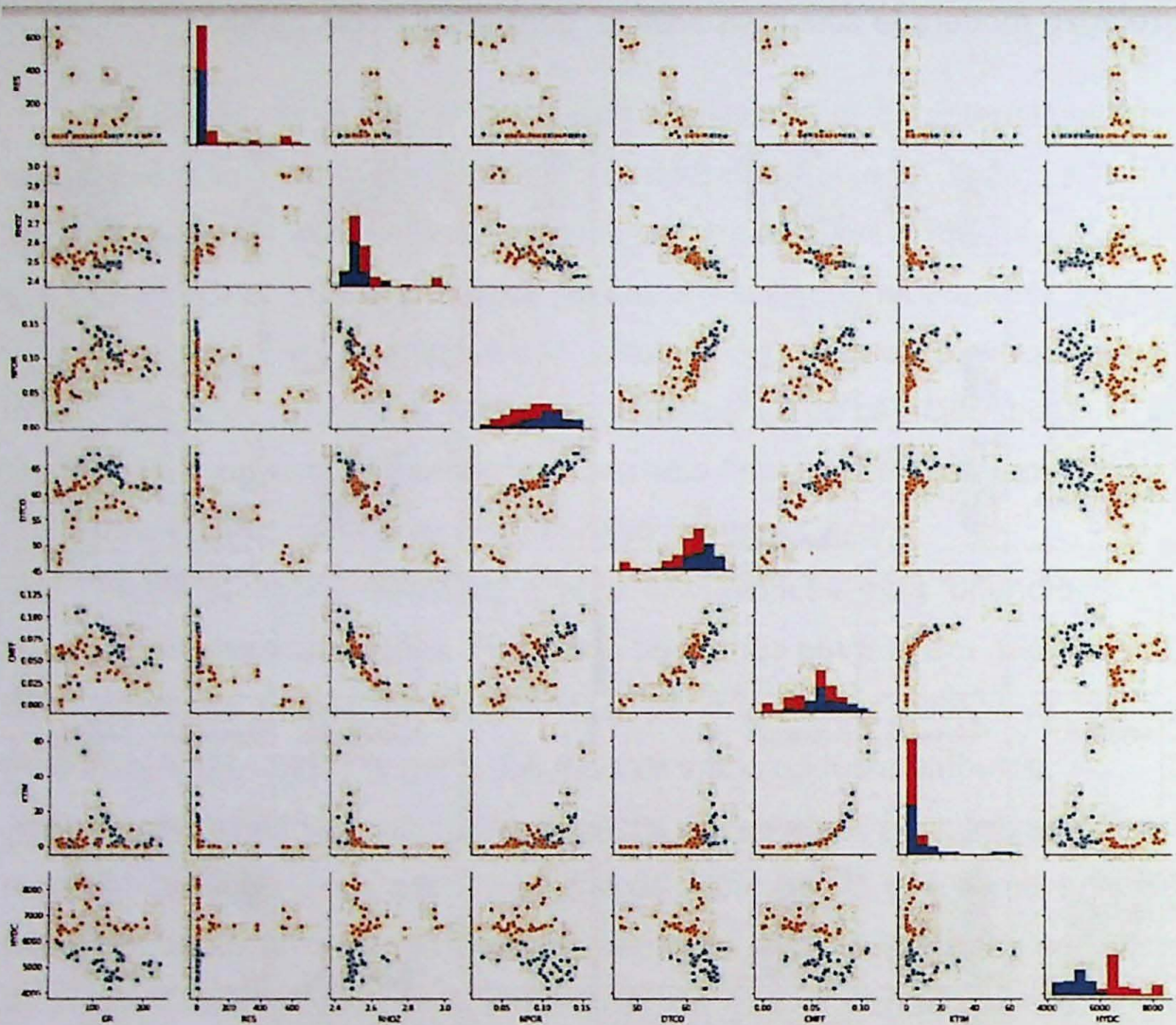


Figura 64: Diagramas de dispersão construídos com dados de 3 poços do reservatório Poti. Os diagramas contemplam todos os atributos envolvidos, dois a dois e foram divididos em dois grupos (amarelo e azul), por semelhança de características, pelo algoritmo K-means.

Tabela 15: Atributos de entrada utilizados no treinamento do modelo A.

DEPTH	GR	RES	RHOZ	NPOR	DTCC	HYDC	METH
1536.0455	60.2486	1.1120	2.4254	0.0773	65.5373	8274.0000	3953.3799
1536.2360	35.0503	2.2421	2.4299	0.0741	64.9425	8269.2002	4002.2700
1536.4265	35.9322	5.4680	2.4679	0.0754	65.9955	8264.4004	4051.1599



Tabela 16: Atributos de entrada utilizados no treinamento do modelo B.

GR	RES	RHOZ	NPOR	DTCO	CMFF	KTIM	HYDC
60.2486	1.1120	2.4254	0.0773	65.5373	0.1263	62.3320	8274.0000
35.0503	2.2421	2.4299	0.0741	64.9425	0.1282	137.6862	8269.2002
35.9322	5.4680	2.4679	0.0754	65.9955	0.1269	128.4053	8264.4004

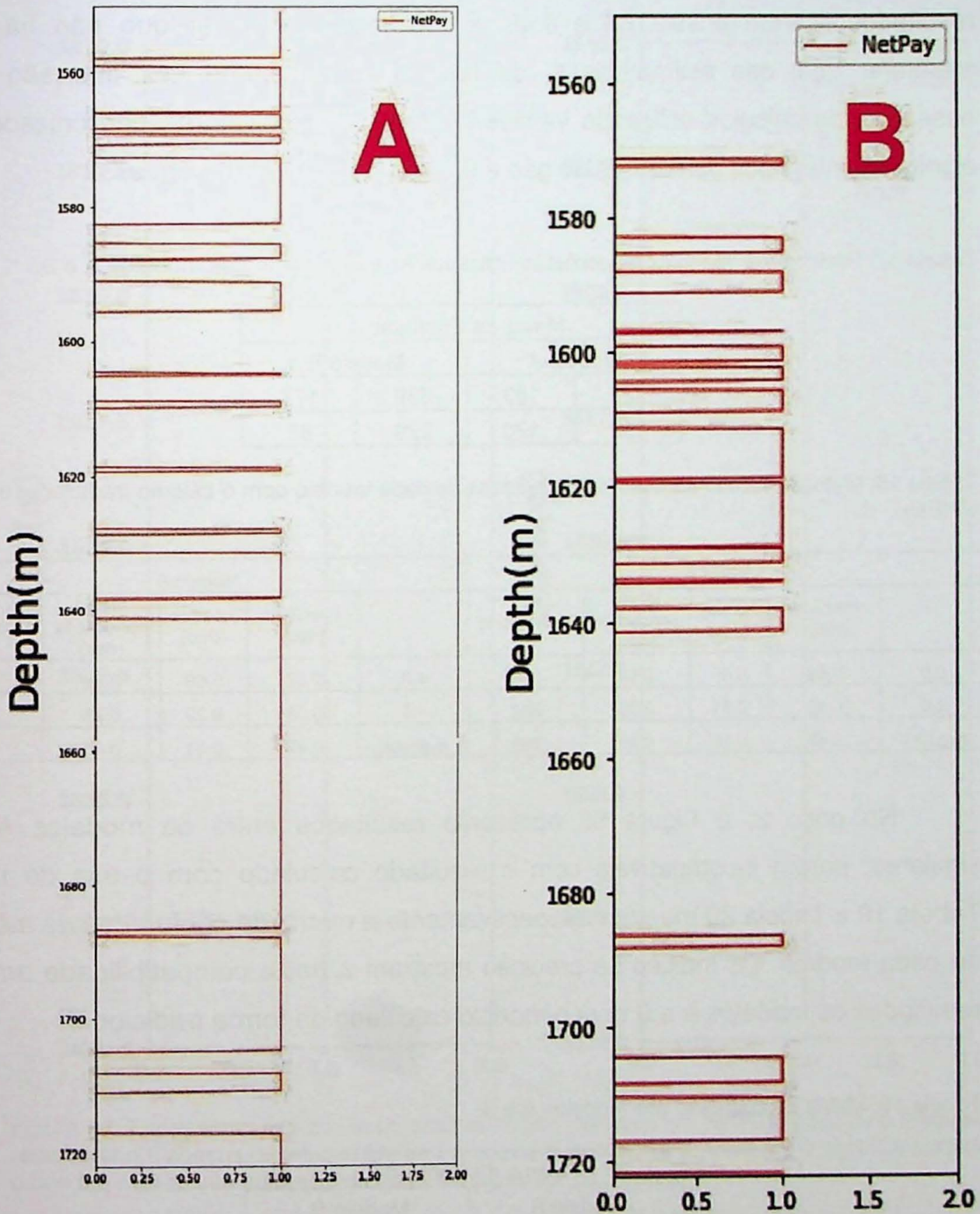


Figura 65: Zonas de Net pay em perfis do poço 1 no intervalo do reservatório Cabeças do modelo A e do modelo B. A ausência de net pay aqui é interpretada como net gross.



Depois do treinamento, o modelo foi testado em outros três poços como “*blind test*”. As Figura 65, Figura 66 e Figura 67 mostram perfis com o zoneamento da Formação Cabeças realizado pelos modelos A e B em cada um dos poços escolhidos para teste. Os resultados dos dois modelos foram comparados com a espessura de *netpay* calculada pela forma convencional, ou seja, com o uso do IP. A Tabela 17 mostra a matriz de confusão de cada modelo testado no poço 1 e a Tabela 18 apresenta as respectivas métricas. Os índices de compatibilidade entre os resultados ficaram entre 0,44 e 0,49. É importante observar que não há como mensurar qual das estimativas é, de fato, a correta, uma vez que são todas baseadas de cálculos utilizando variáveis diferentes. Os valores padronizados 1,0 significam intervalos portadores de gás e 0,0 são portadores de água.

Tabela 17: Parâmetros da matriz de confusão utilizados no treinamento dos modelos A e B.

Matriz de Confusão			
Modelo A		Modelo B	
333	183	339	177
262	120	229	83

Tabela 18: Métricas da comparação dos resultados de cada modelo com o cálculo tradicional de NetPay.

Modelo A					Modelo B				
	Precisão (dec)	Acertos (dec)	Média ponderada (dec)	Ocorrências		Precisão (dec)	Acertos (dec)	Média ponderada (dec)	Ocorrências
0,0	0,56	0,65	0,60	516	0,0	0,57	0,68	0,59	516
1,0	0,40	0,31	0,35	382	1,0	0,38	0,29	0,26	382
avg/total	0,49	0,50	0,49	898	avg/total	0,49	0,51	0,45	898

No poço 2, a Figura 66 apresenta resultados entre os modelos A e B similares, porém incompatíveis com o resultado calculado com o uso do IP. As Tabela 19 e Tabela 20 mostram respectivamente a matriz de confusão e as métricas de cada modelo. Os índices de precisão mostram a baixa compatibilidade entre os resultados os modelos A e B com o modelo calculado da forma tradicional.

Tabela 19: Matriz de confusão dos modelos A e B.

Matriz de Confusão			
Modelo A		Modelo B	
333	173	339	167
252	150	209	113



Tabela 20: Métricas para avaliação dos modelos A e B.

Modelo A					Modelo B				
	Precisão (dec)	Acertos (dec)	Média ponderada (dec)	Ocorrências		Precisão (dec)	Acertos (dec)	Média ponderada (dec)	Ocorrências
0,0	0,56	0,65	0,62	516	0,0	0,55	0,66	0,59	516
1,0	0,44	0,36	0,41	382	1,0	0,39	0,32	0,36	382
avg/total	0,53	0,55	0,55	898	avg/total	0,51	0,53	0,50	898

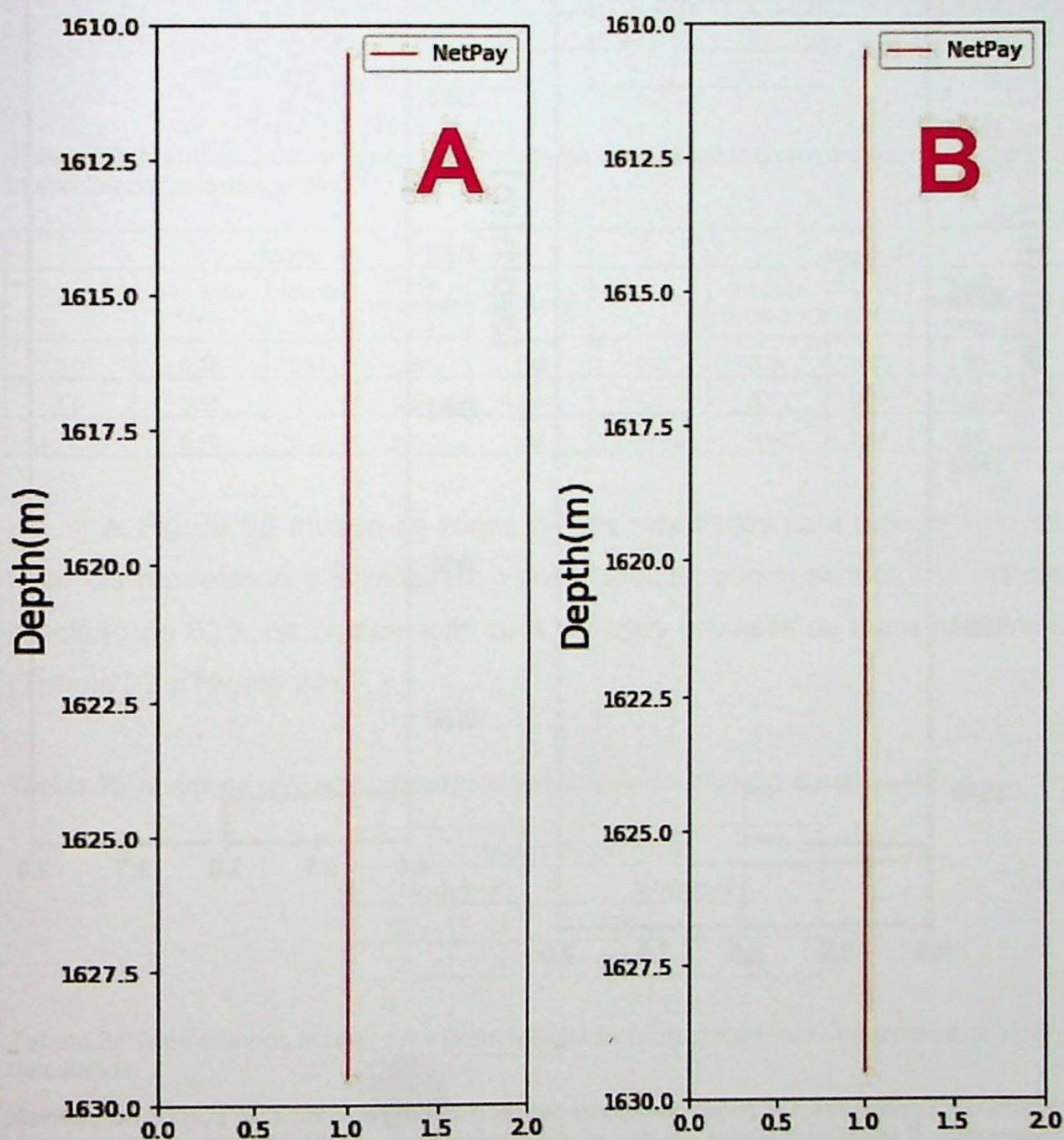


Figura 66: Espessuras das zonas portadoras de gás (Net pay) em perfis do poço 2 no intervalo do reservatório Cabeças estimadas por dois modelos distintos, A e B. O valor 1,0 significa portador de gás e 0,0 representa ausência de gás.



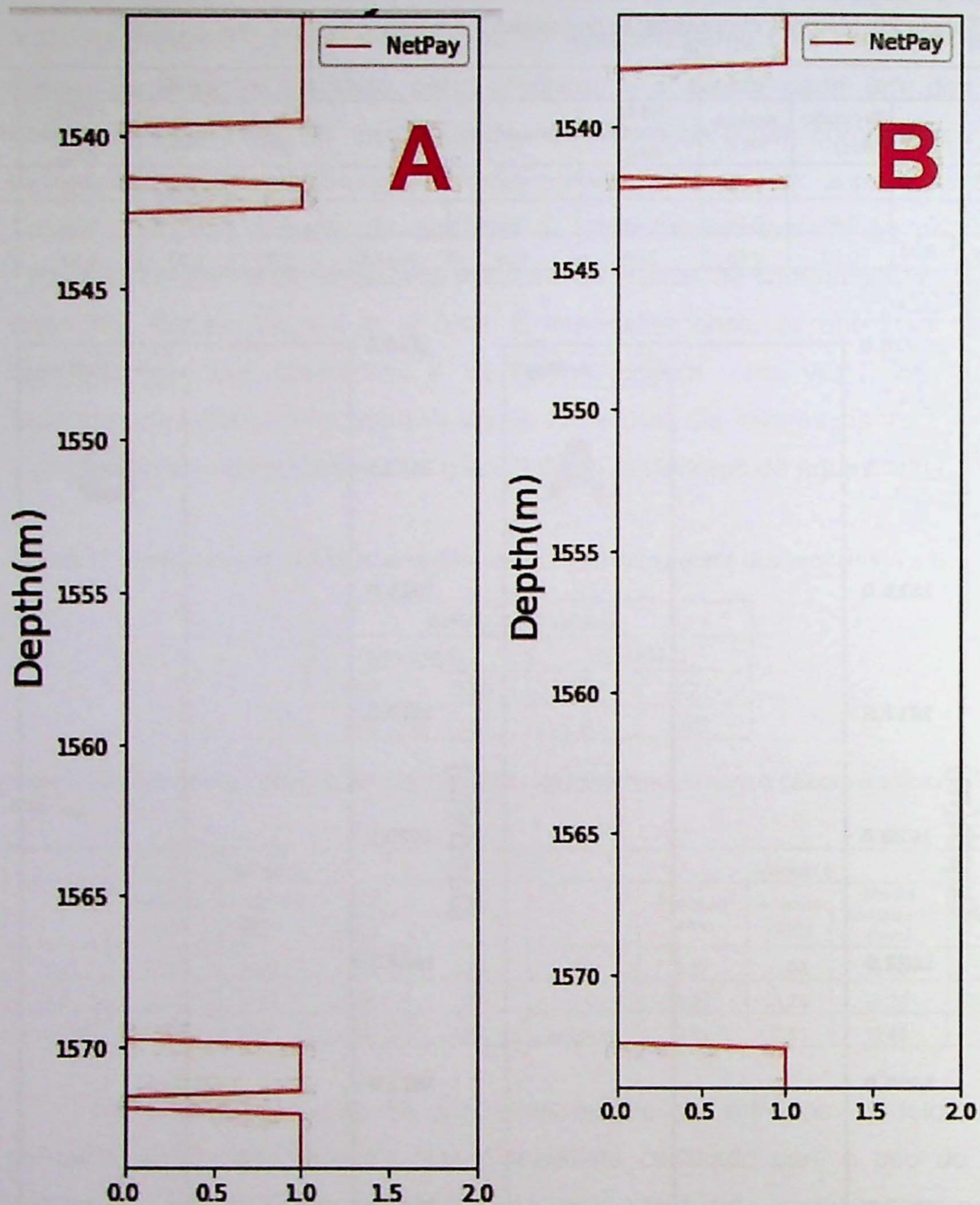


Figura 67: Espessuras das zonas portadoras de gás (Net pay) em perfis do poço 3 no intervalo do reservatório Cabeças estimadas por dois modelos distintos, A e B. O valor 1,0 significa portador de gás e 0,0 representa ausência de gás.

A Figura 67 apresenta os resultados do zoneamento do reservatório modelado duplamente (A e B) para o poço 3. Novamente é possível observar uma



alta compatibilidade entre os modelos, porém as métricas apresentadas na Tabela 21 mostra incompatibilidade de ambos os modelos com o resultado da *netpay* calculada de maneira tradicional. Na Tabela 22 são apresentados os *scores* dos modelos A e B.

Tabela 21: Parâmetros da matriz de confusão utilizados no treinamento dos modelos A e B.

Matriz de Confusão			
Modelo A		Modelo B	
89	18	63	14
0	8	0	0

Tabela 22: Métricas da comparação dos resultados de cada modelo com o cálculo tradicional dos intervalos portadores de gás.

Modelo A					Modelo B				
	Precisão (dec)	Acertos (dec)	Média ponderada (dec)	Ocorrências		Precisão (dec)	Acertos (dec)	Média ponderada (dec)	Ocorrências
0,0	0,88	0,85	0,84	108	0,0	0,94	0,82	0,90	77
1,0	0,52	0,72	0,64	44	1,0	0,00	0,00	0,00	0
avg/total	0,70	0,75	0,74	152	avg/total	0,94	0,82	0,90	77

A Figura 68 mostra as zonas de gás modeladas para o poço 1 no intervalo Poti. Os modelos A e B mostram compatibilidade, porém as métricas indicam uma precisão de 53% na comparação com a *netpay* estimada de maneira convencional (Tabela 23 e Tabela 24).

Tabela 23: Matriz de confusão utilizados no treinamento dos modelos A e B.

Matriz de Confusão			
Modelo A		Modelo B	
52	14	339	177
2	8	229	83

Tabela 24: Métricas dos modelos A e B em relação na comparação com os intervalos de NetPay calculados.

Modelo A					Modelo B				
	Precisão (dec)	Acertos (dec)	Média ponderada (dec)	Ocorrências		Precisão (dec)	Acertos (dec)	Média ponderada (dec)	Ocorrências
0,0	0,96	0,74	0,83	71	0,0	0,57	0,69	0,59	516
1,0	0,21	0,69	0,28	6	1,0	0,39	0,32	0,32	382
avg/total	0,51	0,53	0,73	77	avg/total	0,53	0,55	0,45	898



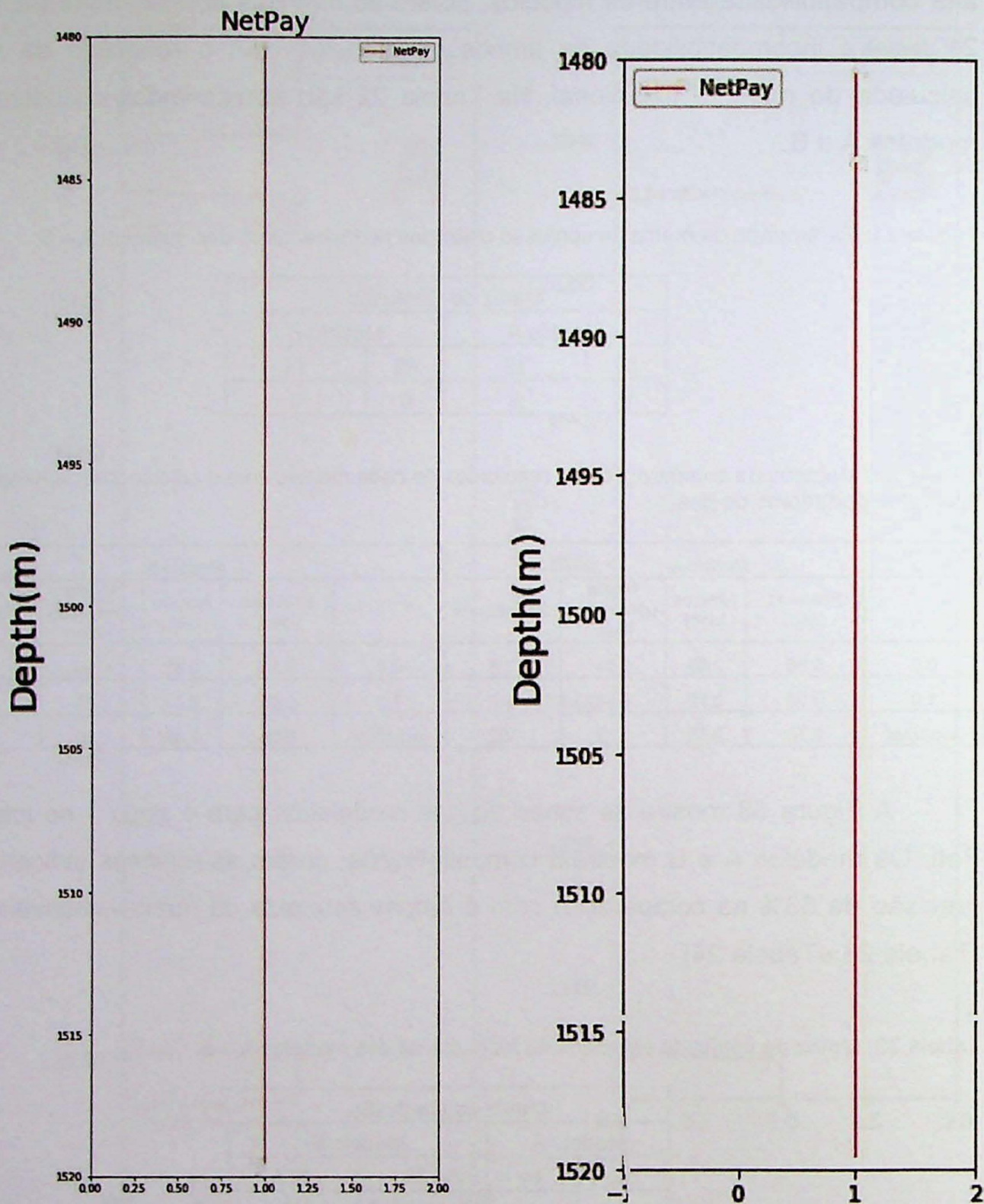


Figura 68: Espessuras das zonas portadoras de gás (Net pay) em perfis do poço 1 no intervalo do reservatório Poti estimadas por dois modelos distintos, A e B. O valor 1,0 significa portador de gás e 0,0 representa ausência de gás.

Por fim, os modelos foram testados com os dados do poço no intervalo do reservatório Poti e se mostraram incompatíveis (Figura 69) para o poço 2.



O modelo A apresentou 90% de precisão quando comparado com o modelo calculado, porém o modelo B mostrou-se incompatível (Tabela 25 e Tabela 26).

Tabela 25: Parâmetros da matriz de confusão utilizados no treinamento dos modelos A e B.

Matriz de Confusão			
Modelo A		Modelo B	
94	1	63	14
12	13	0	0

Tabela 26: Métricas da comparação dos resultados de cada modelo com o cálculo tradicional.

Modelo A					Modelo B				
	Precisão (dec)	Acertos (dec)	Média ponderada (dec)	Ocorrências		Precisão (dec)	Acertos (dec)	Média ponderada (dec)	Ocorrências
0,0	0,89	0,99	0,94	95	0,0	1,00	0,82	0,90	95
1,0	0,93	0,52	0,67	25	1,0	0,00	0,00	0,00	25
avg/total	0,90	0,89	0,88	120	avg/total	0,00	0,21	0,07	120

Para o poço 3 da Formação Poti os resultados das métricas são apresentados nas tabelas 27 e Tabela 28.

Em geral o modelo A testado com os dados do poço 3 para o reservatório Poti foi o que apresentou melhor performance, na comparação com as espessuras portadoras de gás estimadas pelo software.

Com exceção dos resultados do poço 3, os modelos A e B mostraram alto grau de compatibilidade entre si em todos os testes, o que parece indicar que o uso da porosidade e da permeabilidade do perfil ressonância não é tão necessário.

Além disso, em geral, os modelos mostraram baixo grau de compatibilidade quando correlacionados com o modelo de *NetPay* calculado de forma tradicional.

Neste caso, é impossível apontar qual deles é mais “correto”, ou seja, qual deles representa melhor a realidade da geologia da rocha reservatório estudada, uma vez que não há informações sobre a resposta real de cada intervalo reservatório.

Tabela 27: Métricas da comparação dos resultados de cada modelo com o cálculo tradicional de *NetPay*.

Matriz de Confusão			
Modelo A		Modelo B	
52	19	339	177
2	4	229	83



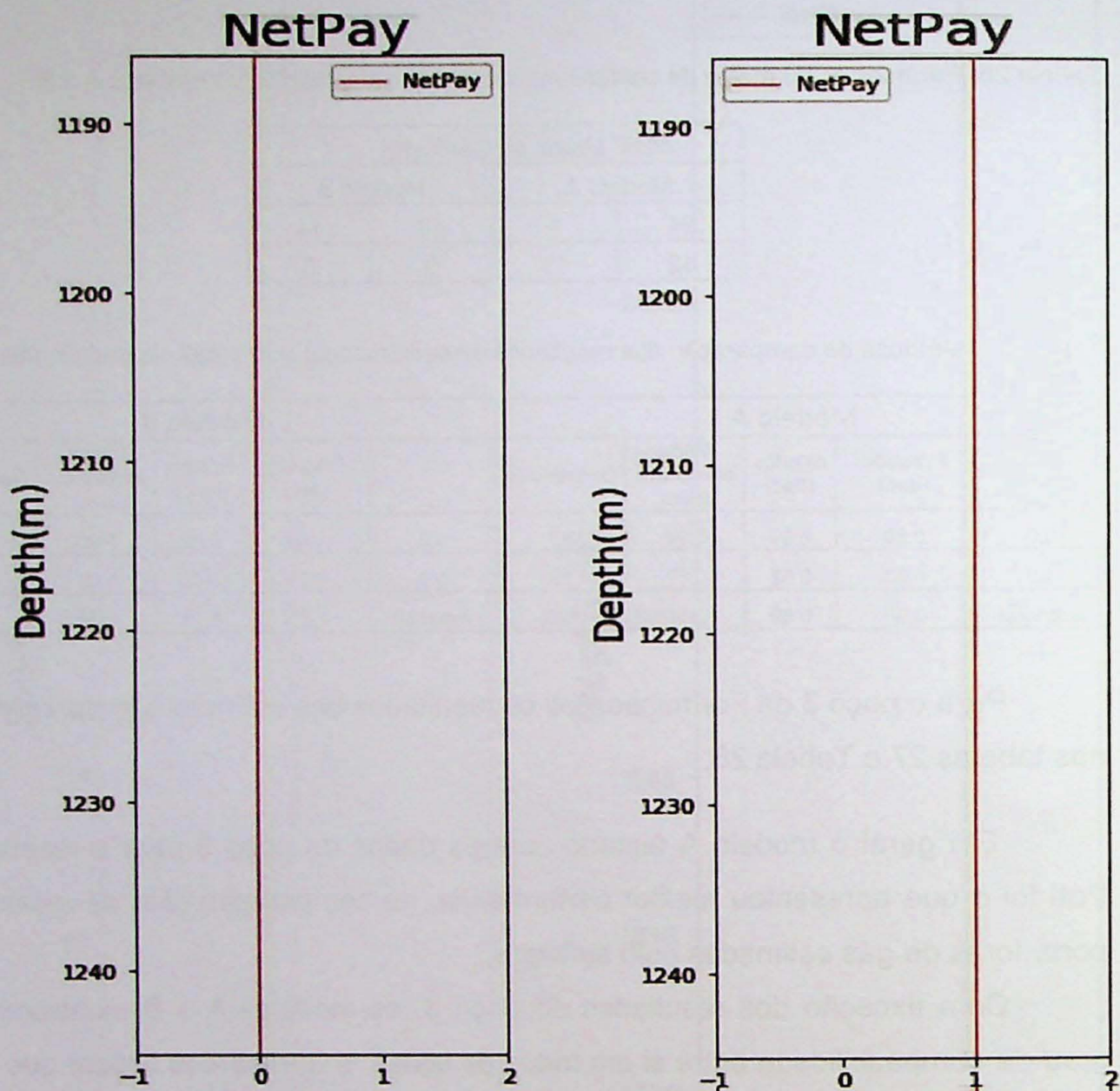


Figura 69: Espessuras das zonas portadoras de gás (Net pay) em perfis do poço 3 no intervalo do reservatório Poti estimadas por dois modelos distintos, A e B. O valor 1,0 significa portador de gás e 0,0 representa ausência de gás.

Tabela 28: Métricas da comparação dos resultados de cada modelo com o cálculo tradicional de NetPay.

	Modelo A				Modelo B				
	Precisão (dec)	Acertos (dec)	Média ponderada (dec)	Ocorrências		Precisão (dec)	Acertos (dec)	Média ponderada (dec)	Ocorrências
0,0	0,00	0,00	0,00	247	0,0	0,21	1,00	0,35	247
1,0	0,79	1,00	0,88	902	1,0	0,00	0,00	0,00	902
avg/total	0,62	0,79	0,69	1149	avg/total	0,05	0,21	0,08	1149



Embora os modelos agrupados tenham apresentado, na maior parte dos testes, precisões entre 51 e 60% na comparação com os intervalos portadores de gás calculados, é importante notar que são grandezas subjetivas, calculadas de formas diferentes. Portanto, estas duas estimativas (*netpay* calculada e o resultado do agrupamento) somente foram comparados nesta etapa do trabalho, pois havia uma hipótese de que poderiam representar a mesma informação, ou seja, os intervalos portadores de gás. Assim, após esta primeira avaliação mais testes devem ser realizados para afirmar se o agrupamento é uma técnica capaz, de fato para substituir com precisão a estimativa de *netpay*.



#### 4.6 Análises petrofísicas dos reservatórios

A partir dos resultados dos modelos de porosidades, permeabilidades e da estimativa da espessura média do intervalo saturado com hidrocarbonetos foram elaborados e analisados mapas da distribuição destes parâmetros na bacia do Parnaíba. Os valores médios de cada parâmetro foram obtidos por média ponderada, sendo considerado como peso a espessura de cada valor.

A Figura 70 mostra a distribuição das porosidades da Formação Poti. Na porção central, as porosidades variam entre 15 e 25%, chegando a valores de cerca de 35% nas proximidades do depocentro da bacia. Para esta formação foram usados dados de 56 poços interpolados com uso de máxima quadratura.

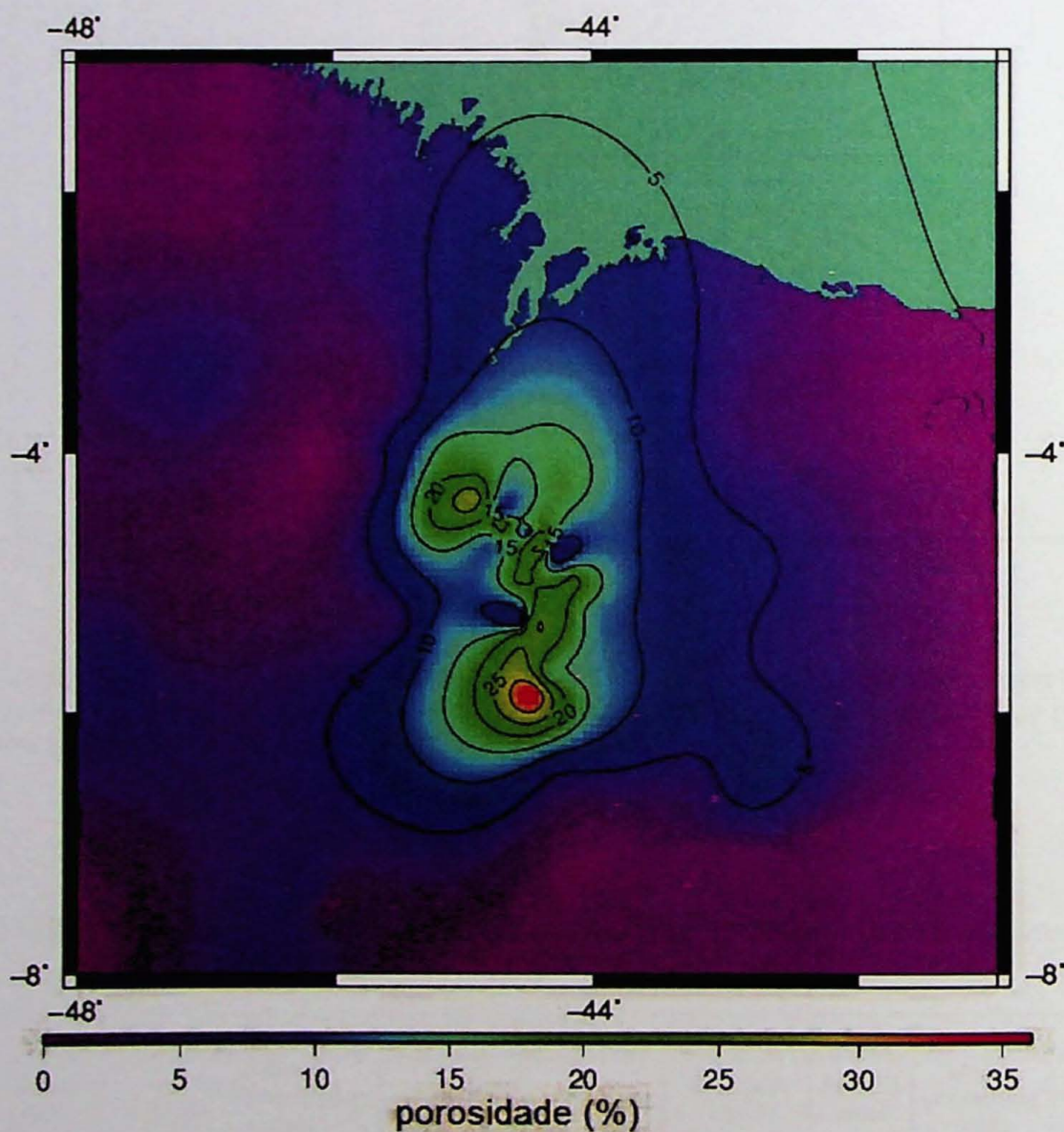


Figura 70: Mapa com a distribuição de porosidades médias da Formação Poti na Bacia do Parnaíba.



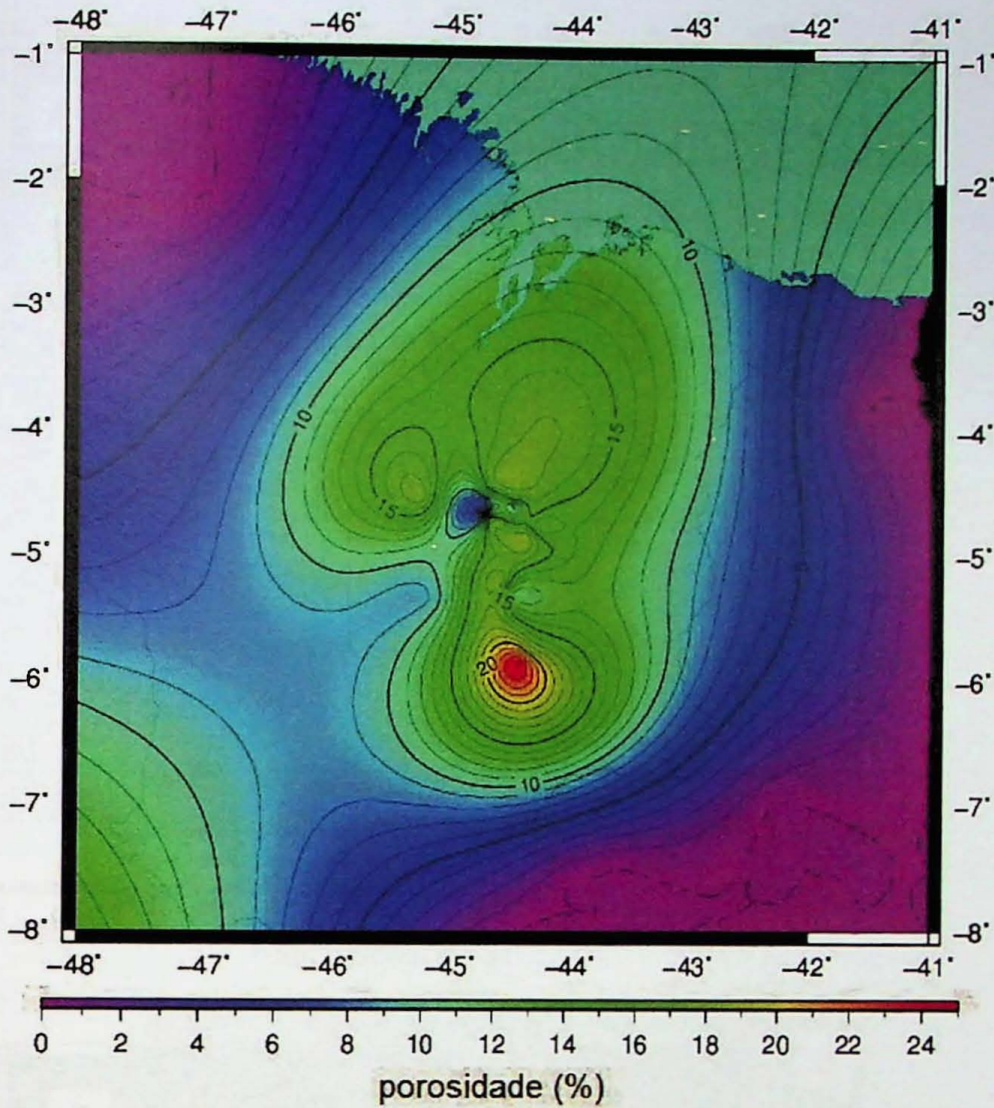


Figura 71: Mapa com a distribuição de porosidades médias da Formação Cabeças na Bacia do Parnaíba.

Na Formação Cabeças a porosidade também pode ser dividida em dois grupos, sendo um mais periférico, com valores entre zero e 4% e um grupo central com valores entre 8 e 20% (Figura 71). Assim como ocorre com a formação Poti, as maiores porosidades encontram-se localizadas nas vizinhanças do depocentro da bacia.

As Figura 72 A e B mostram respectivamente as distribuições de permeabilidades da formação Poti e Cabeças pela bacia dos Parnaíba. Embora ocorram eventuais “picos” de permeabilidade alcançando até 3000mD, nestes mapas foram apenas amostradas as permeabilidades médias de cada formação. Na formação Poti as permeabilidades máximas são da ordem de 100 mD, próximo ao depocentro. Já na formação Cabeças os valores chegam a 90 mD, sendo as áreas de maiores permeabilidades encontradas mais ao norte. Na Figura 73 vemos as distribuições da espessura de portadores de gás das Formações Poti e Cabeças.



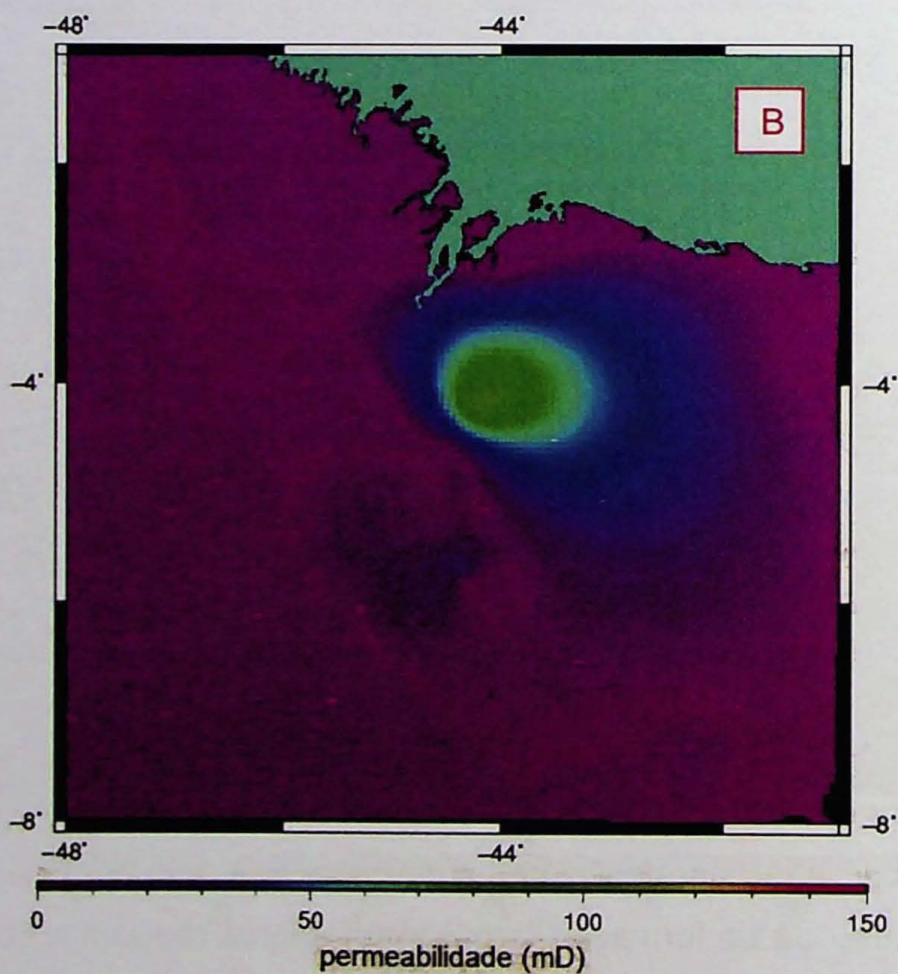
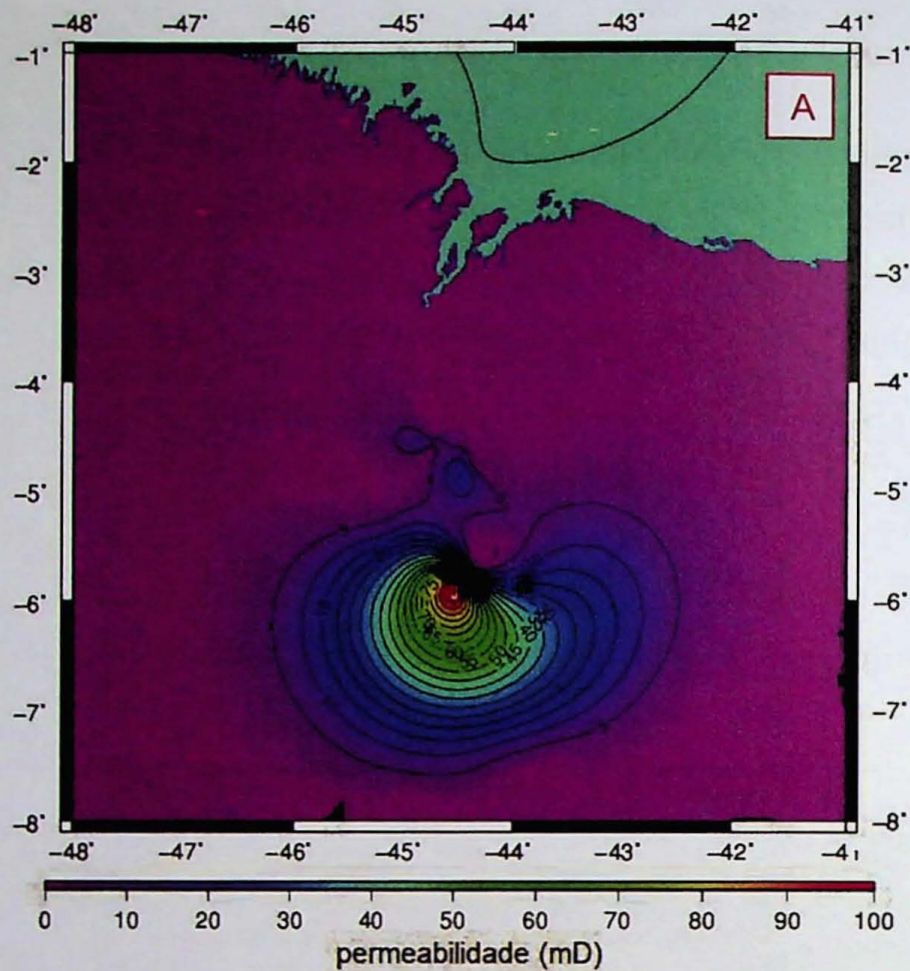


Figura 72: Mapa com a distribuição de permeabilidades médias da Formação Poti (em A) e da Formação Cabeças (em B) da bacia do Parnaíba.



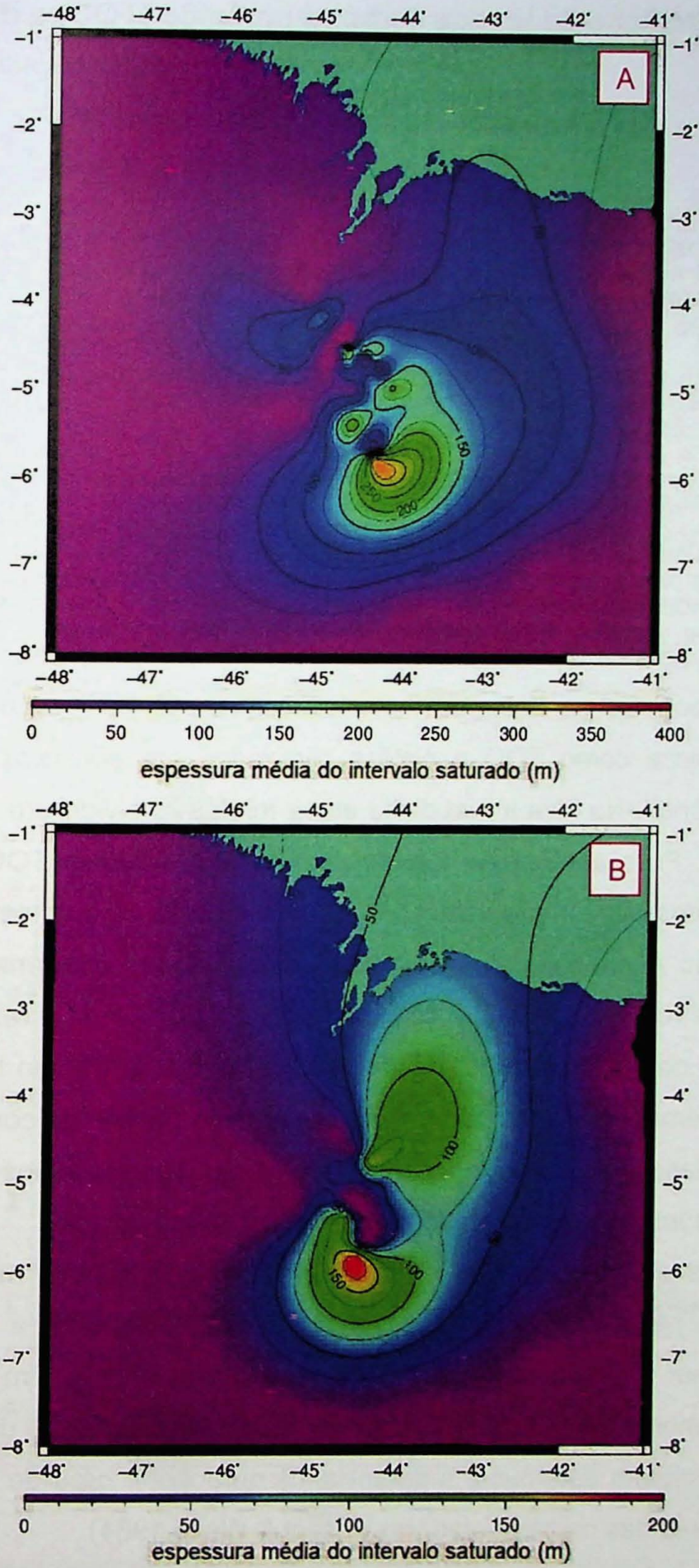


Figura 73: Mapa com a distribuição das espessuras médias portadoras de gás pela Bacia do PArnaíba. Em A são representadas as espessuras da da Formação Poti e em B da Formação Cabeças (em B).



#### 4.6 Predição do teor de carbono orgânico (TOC) e do parâmetro Tmax

De maneira geral, define-se rocha geradora como qualquer rocha de granulometria fina com conteúdo de carbono orgânico suficiente para gerar hidrocarbonetos em determinadas condições de temperatura e pressão, permitir sua migração. A caracterização das rochas geradoras de petróleo constitui uma informação essencial para a avaliação dos riscos que envolvem a exploração petrolífera, principalmente em bacias maduras (Coelho et al, 2014).

Considerando que todo o gás proveniente de campos produtores dependem essencialmente de uma rocha geradora em estágio de geração (Rojas et al, 2011) faz-se necessário um estudo detalhado da desta rocha geradora.

O presente trabalho pretende avaliar o potencial de geração de hidrocarbonetos na bacia do Parnaíba com o uso de *logs* de perfilagem e dados geoquímicos como TOC e pirólise realizados em amostras laterais de 7 poços exploratórios. Na fase inicial desta etapa foi desenvolvido um modelo para predição de TOC. Para analisar a melhor forma de se estimar o TOC foram testadas três abordagens, sendo que duas envolveram o uso de algoritmos supervisionados para regressão e predição. Ambas utilizaram resultados das análises laboratoriais de pirólise e *rock eval*, ou seja, com dados de S1, S2, Tmax e resíduo, porém somente em uma delas foi utilizado também as curvas de perfilagem tradicionais. A terceira abordagem foi realizada por meio de um método conhecido como "Deltalog de R", ou DlogR desenvolvido por Passey et al.(1990). Os resultados das predições foram comparados com os resultados das análises de laboratório.

As análises de *rock eval* foram feitas pela Petrobras e alguns resultados cedidos para este estudo. A primeira análise realizada nas amostras visou determinar a porcentagem de carbono orgânico (TOC). Em seguida as amostras com melhores índices de TOC foram enviadas para análises de pirólise e *Rock-Eval*, que permitiram determinar o potencial de geração, o grau de maturação e o tipo de querogênio das rochas geradoras (Tissot & Welte 1984).

Baseados nos perfis dos poços disponíveis foram analisados os registros de gama (GR) e selecionado os intervalos com maior radioatividade. Nestes intervalos foram analisados os perfis Sônico e resistividade e aplicados os métodos de



Schmoker (1979) e DlogR (Passey et al., 1990) para estimar o conteúdo orgânico e avaliar o potencial gerador.

#### 4.6.1 Abordagem 1

O método escolhido para as abordagens 1 e 2 foi o *Gradient Boosting*, sendo que na primeira abordagem apenas os resultados dos ensaios de pirolise dos 7 poços foram usados. Além disso, os dados foram randomicamente divididos em dois conjuntos, sendo um deles de treino e outro de teste. Para uma razão de 60%/40% o modelo treinado foi capaz de alcançar um coeficiente de correlação de 0,97 no intervalo da Formação Pimenteiras (Figura 75). Alterando a razão para 80%/20% o coeficiente de correlação foi 0,96 (Figura 74). Embora, o número de dados utilizados na elaboração do modelo tenha sido muito limitado, o modelo apresentou uma performance relevante em ambos os testes.

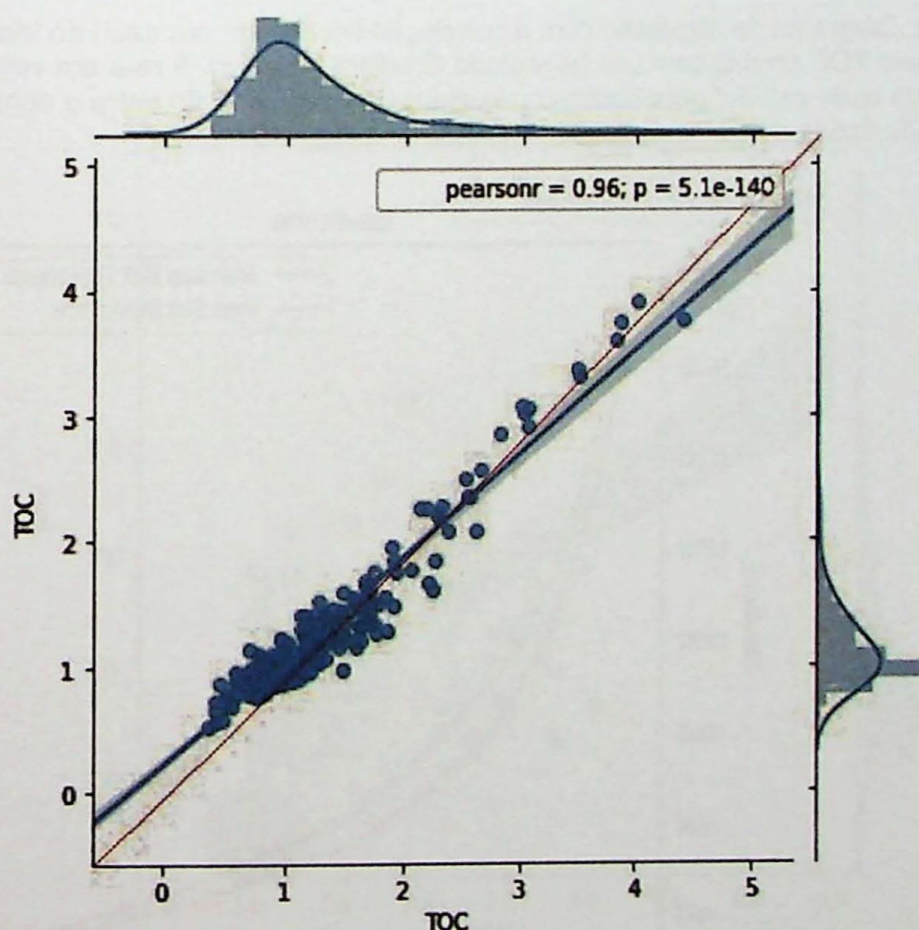


Figura 74: Diagrama de dispersão com a correlação linear (reta em azul) do teor de carbono orgânico medido pelo TOC predito com uso do método *Gradient Boosting*. A reta em vermelho mostra a função de expectativa onde  $x=y$ . A figura ilustra os resultados da comparação entre o conjunto de treino com o conjunto de dados de testes na razão 80 / 20.



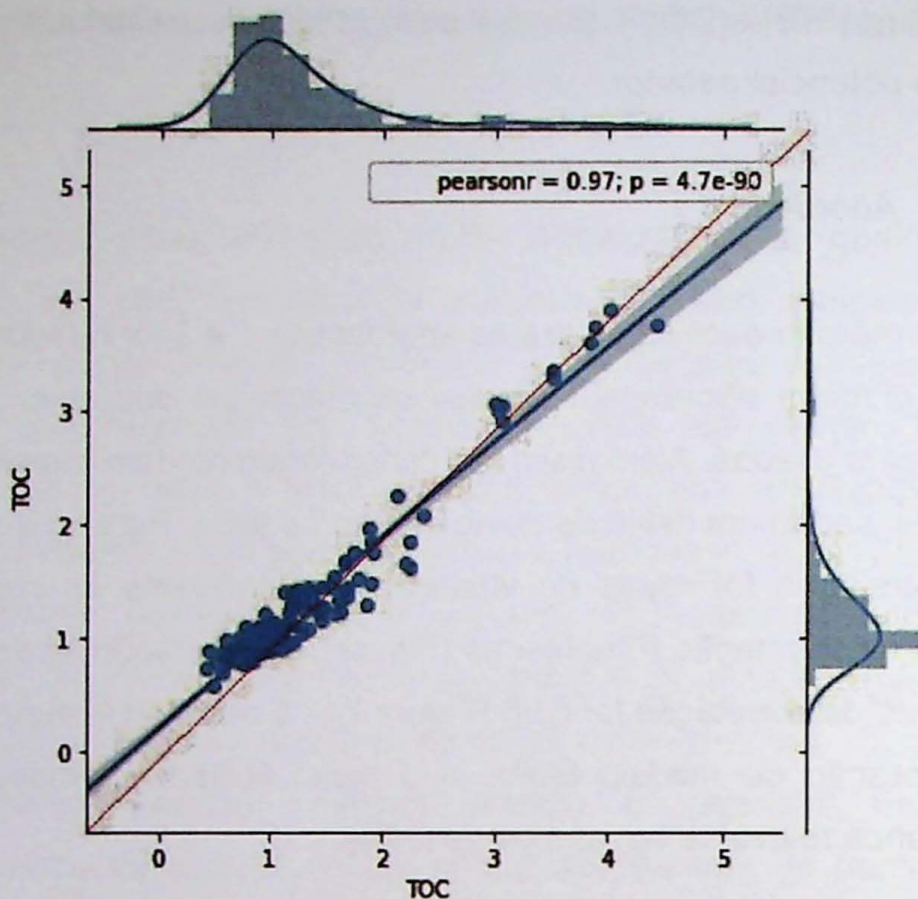


Figura 75: Diagrama de dispersão com a correlação linear (reta em azul) do teor de carbono orgânico medido pelo TOC predito com uso do método *Gradient Boosting*. A reta em vermelho mostra a função expectativa onde  $x=y$ . A figura ilustra os resultados da comparação entre o conjunto de treino com o conjunto de dados de testes na razão 60 / 40.

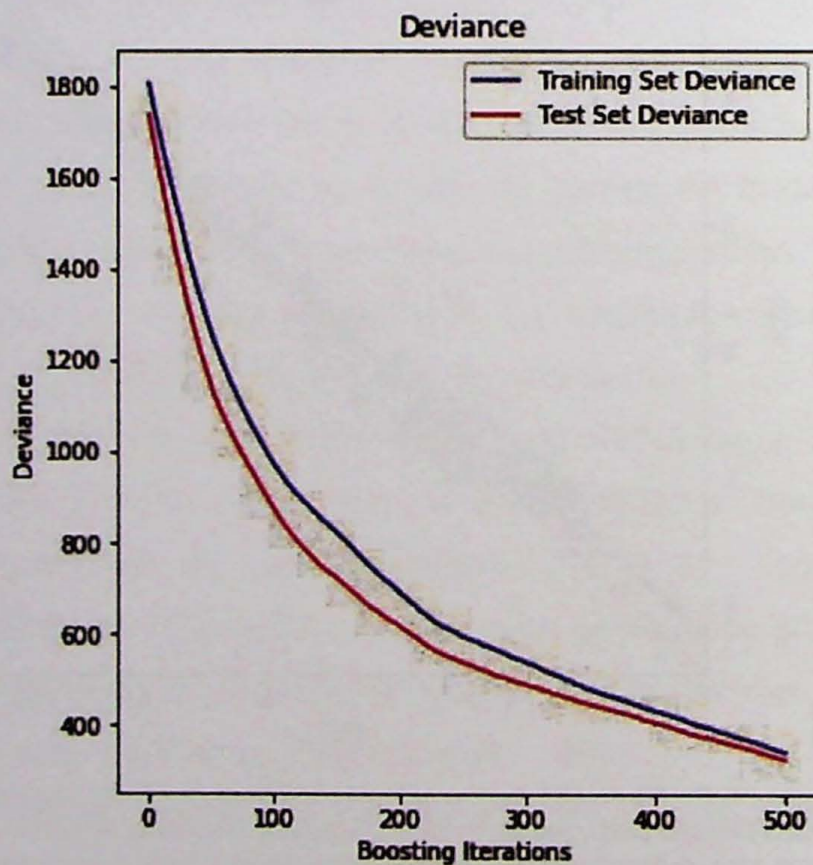


Figura 76: Variância do erro das predições em função do número de iterações realizadas pelo modelo durante o processo de treinamento.



Na elaboração deste modelo foram adotadas 500 iterações, pois o desvio entre o conjunto de dados de treino e o conjunto de dados de teste cai expressivamente conforme aumenta o número de iterações (Figura 76). Desta forma, adotando a partir deste número de iterações a relevância entre as razões utilizadas como treino e teste decai significativamente.

#### 4.6.2 Abordagem 2

A segunda abordagem consistiu no uso dos logs de perfuração usuais como gama, resistividade, sônico, densidade e neutrão, além dos resultados das análises laboratoriais. Desta vez, os resultados apresentados na Figura 77 mostram que quando o modelo, treinado com perfis convencionais, é testado em alguns poços, como este, por exemplo, o poder de predição do método cai significativamente.

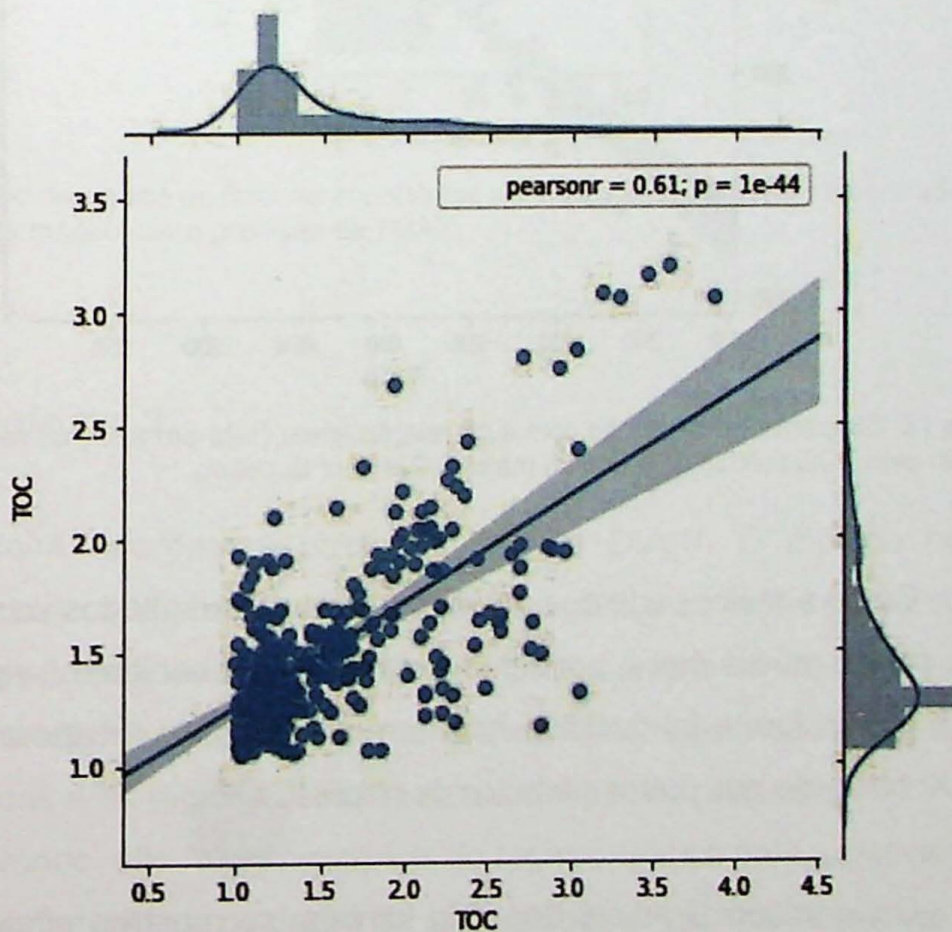


Figura 77: Diagrama de dispersão com a correlação linear (reta em azul) do teor de carbono orgânico medido pelo TOC predito com uso do método *Gradient Boosting*.



Em outros casos, especialmente quando o poço está localizado geograficamente em uma região mais próxima dos poços usados para treinar o modelo, como é o caso do teste apresentado na Figura 78, os resultados são melhores, compatíveis com a abordagem 1.

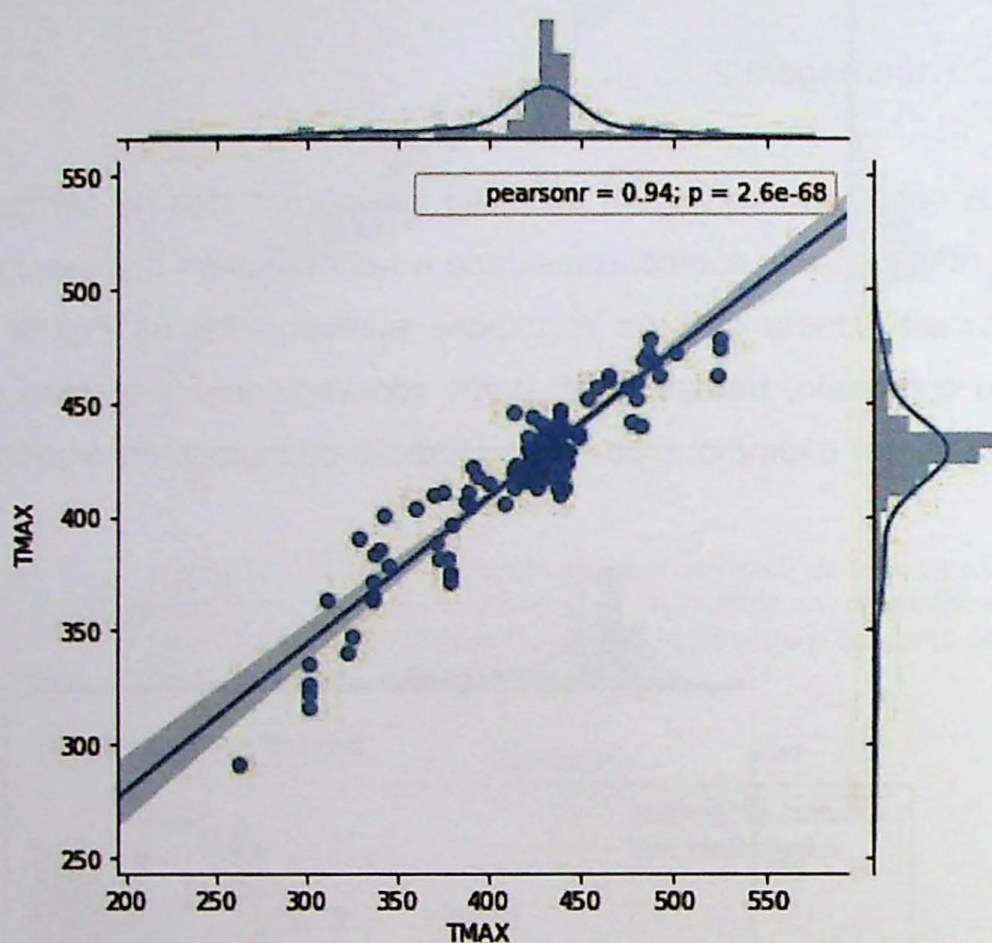


Figura 78: Diagrama de dispersão com a correlação linear (reta em azul) do teor de carbono orgânico medido pelo TOC predito com uso do método *Gradient Boosting*.

Como a maior parte dos dados apresentam resultados para TOC, resíduo, S1 e S3, porém muitos deles, por alguma razão desconhecida, não possuem valores de TMAX, desta forma foi decidido criar um modelo para recuperação dos valores de TMAX, baseado nos outros atributos de pirolise.

O algoritmo *Gradient Boosting* também se mostrou eficiente para recuperar dados de Tmax. No total havia 52 amostras que tiveram TOC, S2 e Hi e Ho medidos, porém por alguma razão estavam sem valores de TMAX. Desta forma, para estes casos foi desenvolvido um modelo com uso do *Gradient Boosting* para recuperação



dos 52 dados faltantes de TMAX na formação Pimenteiras. Os resultados mostraram correlação de 0,94 entre os dados preditos e os dados reais e desta forma foi realizada a recuperação, conforme mostrado na Figura 78.

A Figura 79 apresenta o fator de importância atribuído pelo modelo no processo de treinamento. O elemento de maior relevância na predição de TMAX segundo o modelo é a S3, seguido do S2, resíduo e TOC, sendo este último o menos relevante dos quatro atributos de entrada utilizados na predição de TMAX.

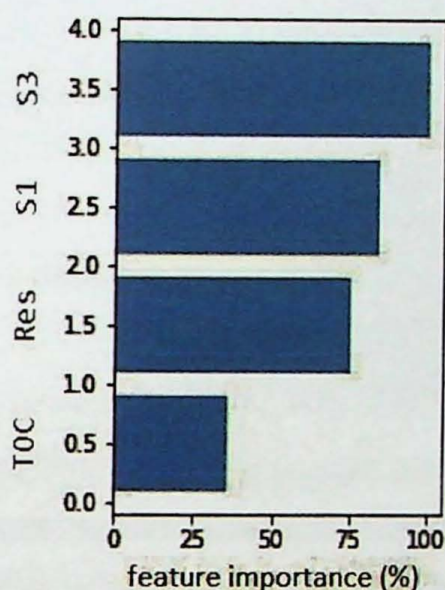


Figura 79: Gráfico de barras do fator de importância normalizado dos atributos de entrada encontrados pelo modelo para predição de TMAX.

#### 4.6.3 Abordagem 3

A terceira abordagem utilizou o método DlogR. O método consiste em analisar o impacto que a presença de matéria orgânica na rocha causa nas medidas dos tempos de transito e resistividades. Para isso é feito um re-escalamento no perfil sônico, onde  $50\mu\text{s}/\text{ft}$  equivale a uma década de resistividade e assim os perfis são comparados e as regiões ricas em conteúdo orgânico realçadas.

Deste modo, são feitos *crossplots* do registro Sônico pela resistividade (Figura 80). A figura mostra todos os dados de resistividade em função do sônico e uma reta ajustada, chamada de linha base que seria a reta limítrofe de teor orgânico, ou seja, os pontos acima dela são relevantes para geração. Além disso, podem ser interpretados como um *shale play*.



A Figura 81 apresenta, na ultima coluna, a sobreposição do perfil pseudo-sônico sobre o perfil sônico. As áreas pintadas em azul claro, na coluna 6, representam as regiões do reservatório onde o teor de matéria orgânica é mais expressivo. É importante observar que os picos do perfil DtR coincidem com os pontos de maior TOC das amostras laterais (Coluna 4).

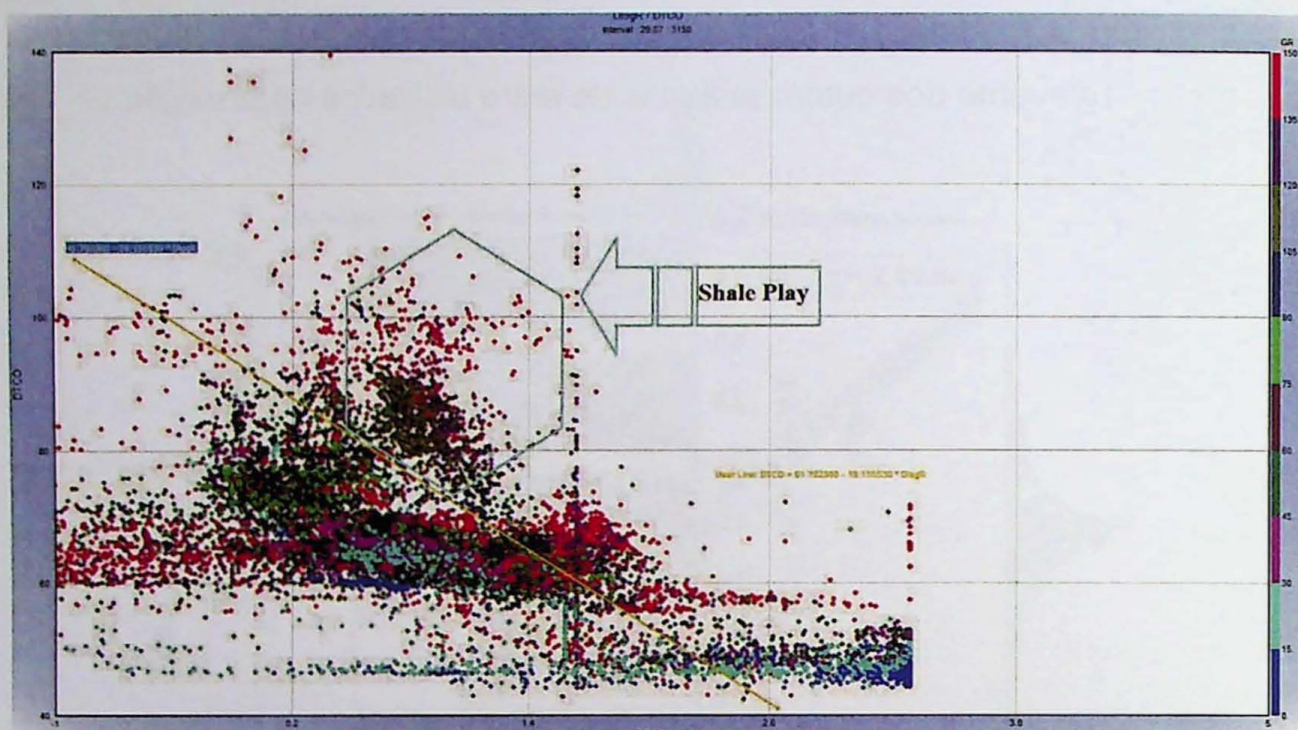


Figura 80: Diagrama de dispersão do perfil "pseudo-sônico" (DlogR) em função da resistividade, com destaque para a linha base (em amarelo) que representa a gênese do log pseudo-sônico.

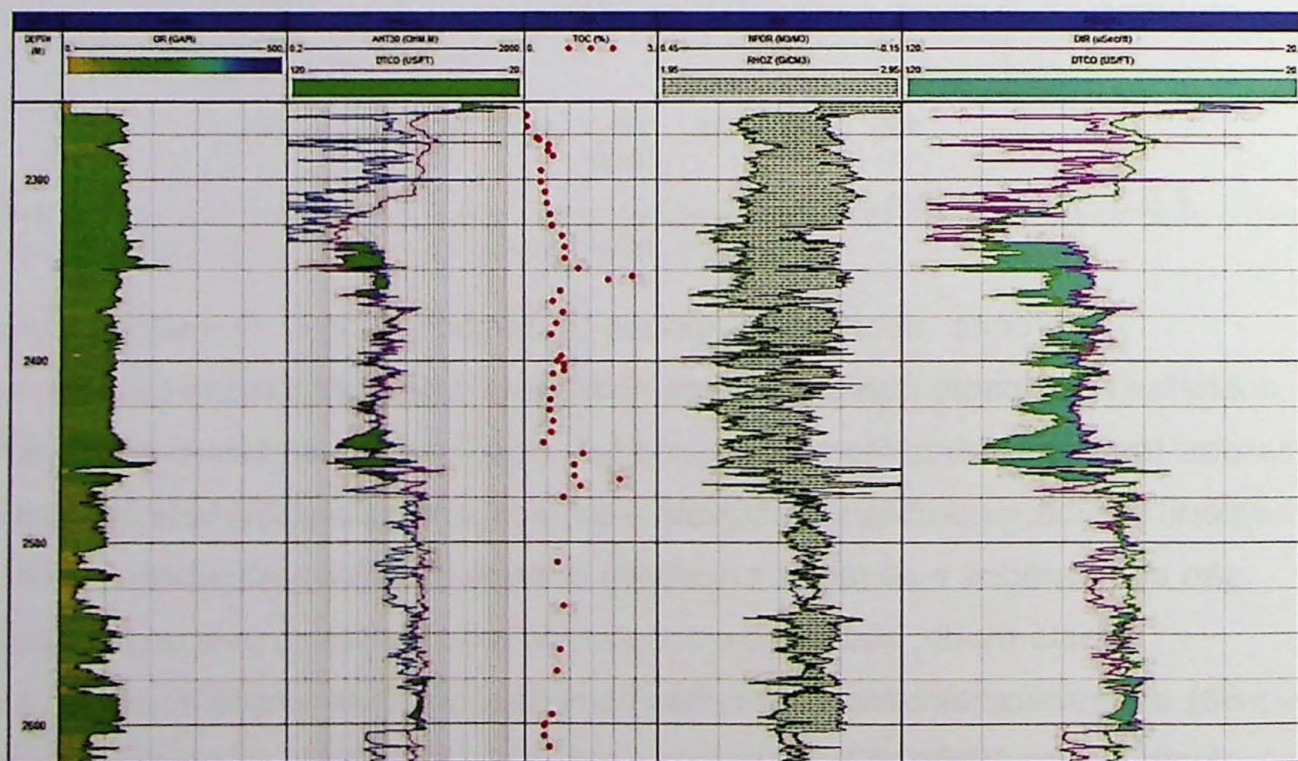


Figura 81: Análises do teor de carbono orgânico (TOC) pelos perfis gama, sônico e "pseudo-sônico". A coluna 4 mostra os valores de TOC calculados em laboratório e a coluna 6 mostra a comparação do perfil sônico com o DlogR, onde as áreas hachuradas em azul mostram as regiões com maiores teores de carbono orgânico.



Na primeira abordagem, embora o método tenha mostrado excelente desempenho, existe uma limitação, que é a necessidade de dados de laboratório, algo bastante restrito na bacia do Parnaíba. Mais testes devem ser realizados, especialmente em intervalos geradores de outras bacias que disponham de um maior volume de dados.

Embora as três abordagens tenham funcionado e mostrado que são boas alternativas para a predição do teor de carbono orgânica, a segunda destoou das demais sem apresentar resultados expressivos. Isso pode ter ocorrido em função do limitado espaço amostral e principalmente por não ter outros perfis como gama espectral que poderia contribuir significativamente se usado como input do método.

O baixo desempenho desta abordagem não invalidaria a técnica, pelo contrário, testes com mais dados devem ser feitos para se obter interpretações mais conclusivas.

A terceira, embora seja a mais trabalhosa, utiliza apenas perfis e em situações em que existem somente os dados de perfilagem disponíveis é a técnica mais funcional para predição do teor de carbono orgânico. Além disso, é uma técnica amplamente testada e utilizada no estudo de geradores.

#### *4.6.4 Análises do potencial de geração da bacia*

Após realizar a primeira etapa, para avaliar técnicas de predição de TOC, fez-se necessário realizar outras análises para entender fatores como os níveis de maturação do querogênio, a assinatura isotópica do gás e o potencial de geração de hidrocarbonetos das formações geradoras da bacia do Parnaíba.

Para esta finalidade, foi necessário mapear e compreender a distribuição das espessuras das formações geradoras da bacia e foi realizada uma interpolação por mínima quadratura. As isópacas resultantes são apresentadas nas Figura 82, sendo a Formação Longa, mostrada em 82a, as isópacas da Formação Pimenteiras em 82c e as espessuras da Formação Tianguá em 82e. Nos itens **b** e **d** são mostradas as isópacas das soleiras.



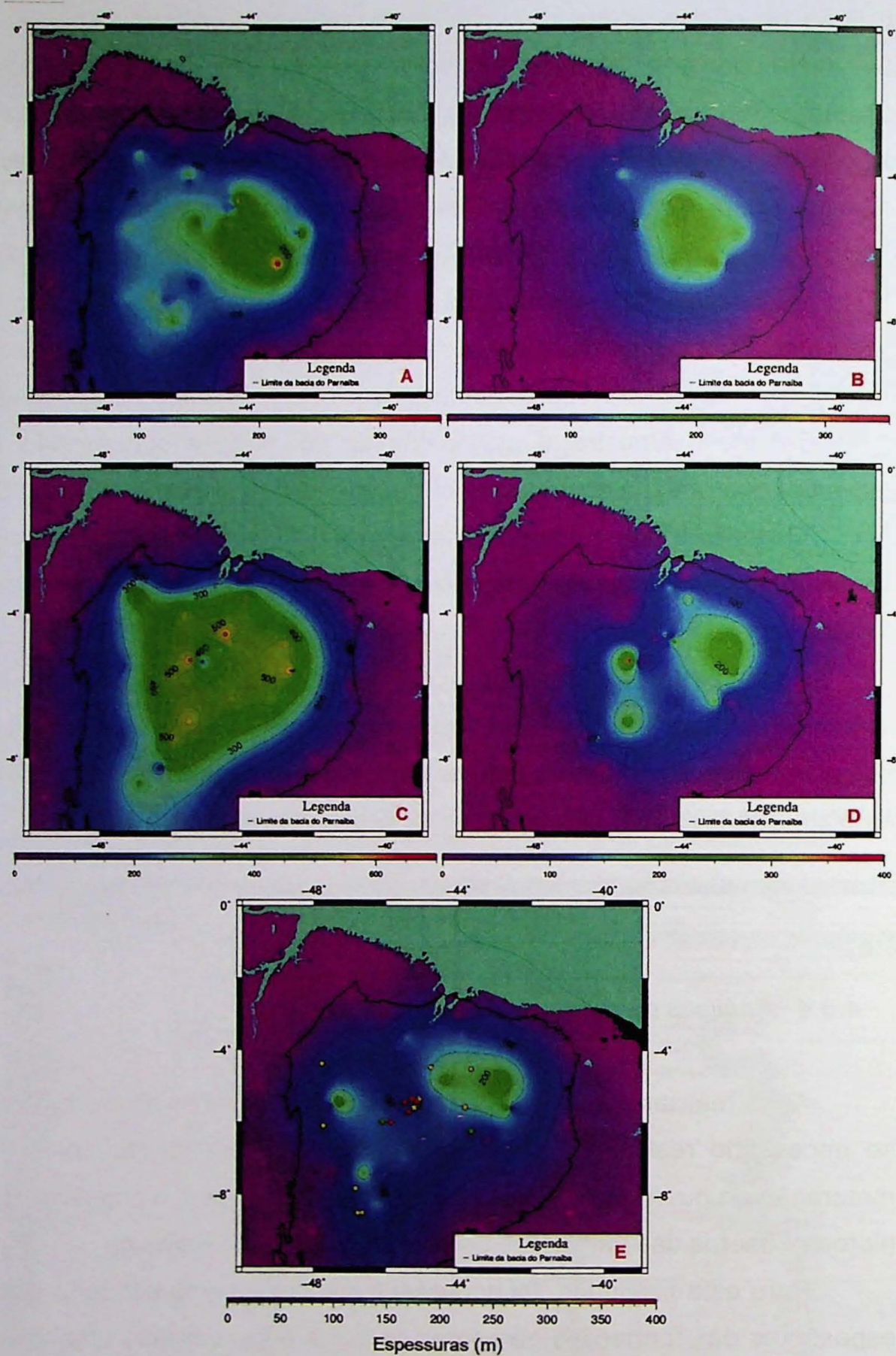


Figura 82: Mapas com isópacas das formações geradoras da Bacia do Parnaíba. A) Formação Longá, B) Soleira da formação Longá, C) Formação Pimenteiras, D) Soleira da formação Pimenteiras, E) Formação Tianguá.



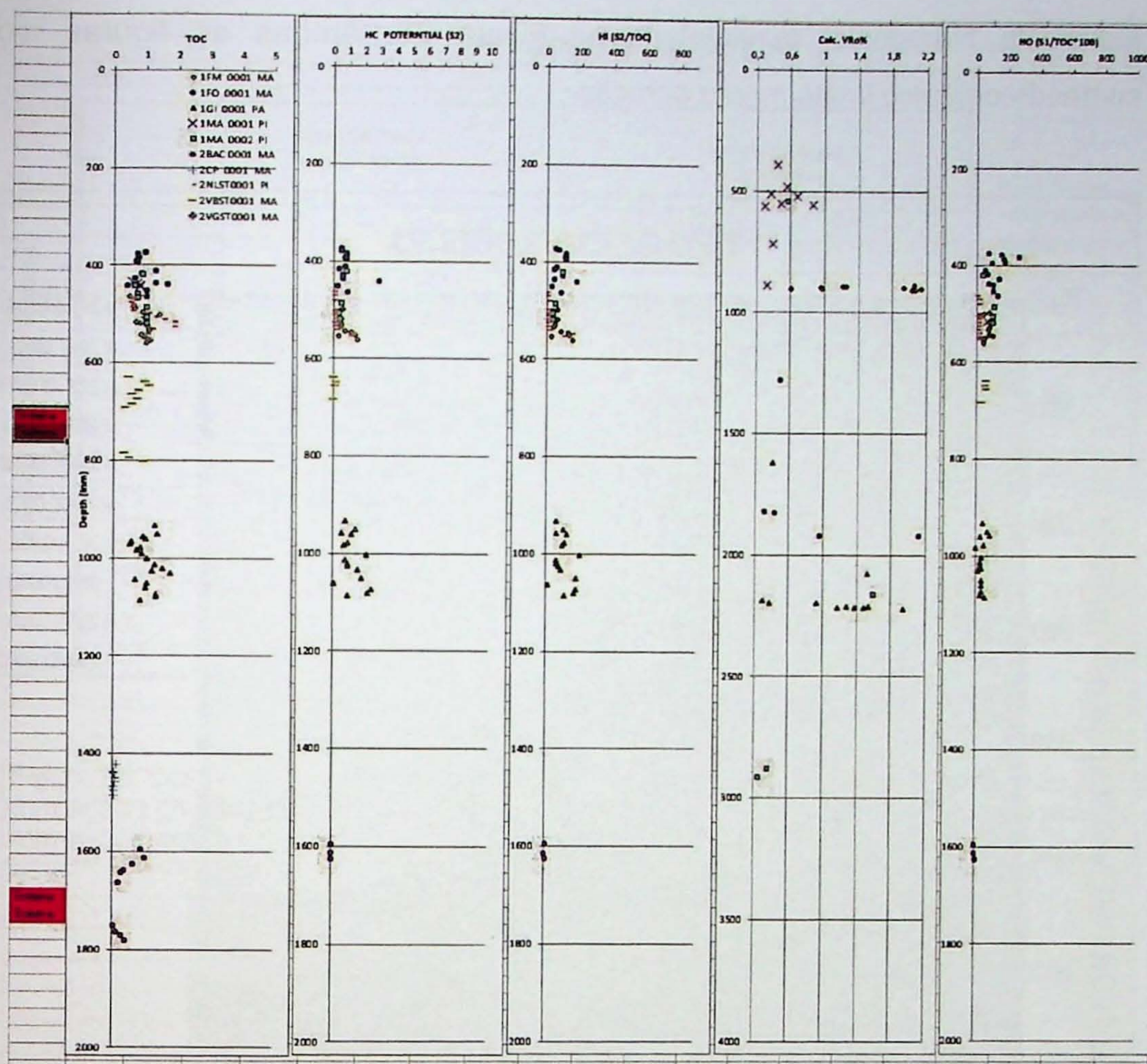


Figura 83: Gráficos dos valores de carbono orgânico, S<sub>2</sub>, índice de hidrogênio, refletância de vitrinita calculada e do índice de oxigênio da formação Longá. Os valores mostrados foram retirados de resultados dos ensaios de pirólise em amostras.

Os resultados das análises de TOC e pirólise das amostras e testemunho da formação Longá são apresentados na Figura 83 e mostram valores entre 0,2 e 2% para o TOC, menores próximos às soleiras e gradativamente maiores conforme a distancia para a soleira aumenta, conforme mostra a coluna 1 da figura. Estes resultados estão de acordo com a afirmação de Rodrigues (1995), de que nos intervalos com ausência de intrusões não houve evolução térmica significativa para promover geração de hidrocarbonetos nestas áreas.

O diagrama de Van Krevelen para os dados da formação Longá (Figura 84) indicam querogênio do tipo 3 e 4, o que é corroborado tanto pelo diagrama de carbono orgânico total apresentado na Figura 85 como pelo diagrama de T<sub>max</sub> pelo



índice de hidrogênio, apresentado na Figura 86. Ambas as figuras mostram conteúdo orgânico imaturo para geração.

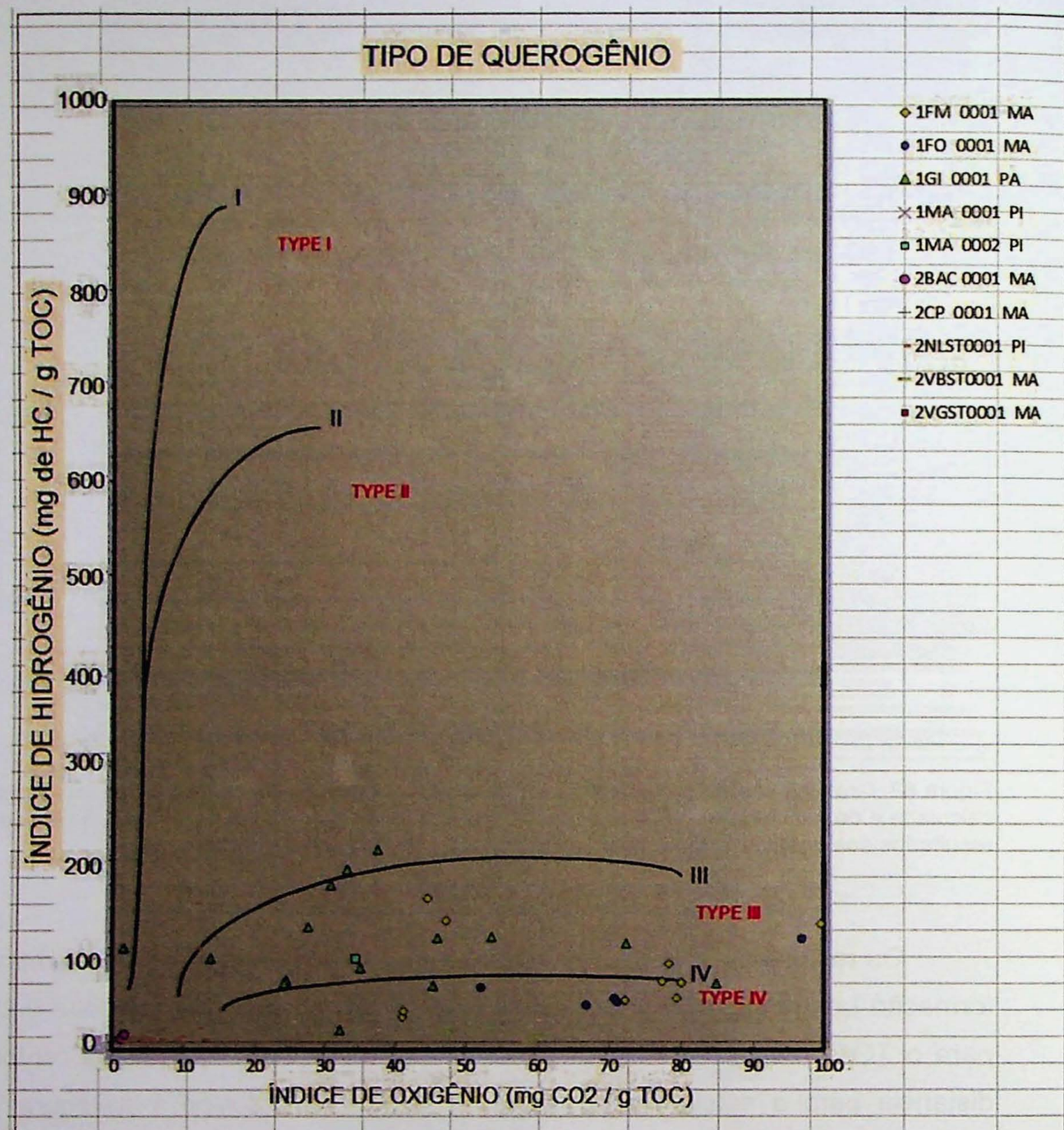


Figura 84: Diagrama de Van Krevelen para análise do tipo de matéria orgânica. O diagrama mostra os valores do índice de hidrogênio em função do índice de oxigênio da formação Longá.

A Figura 87 apresenta as distribuições das razões hidrogênio-oxigênio,  $T_{max}$ , teor de carbono orgânico e do índice de hidrogênio da formação Longá pela bacia.



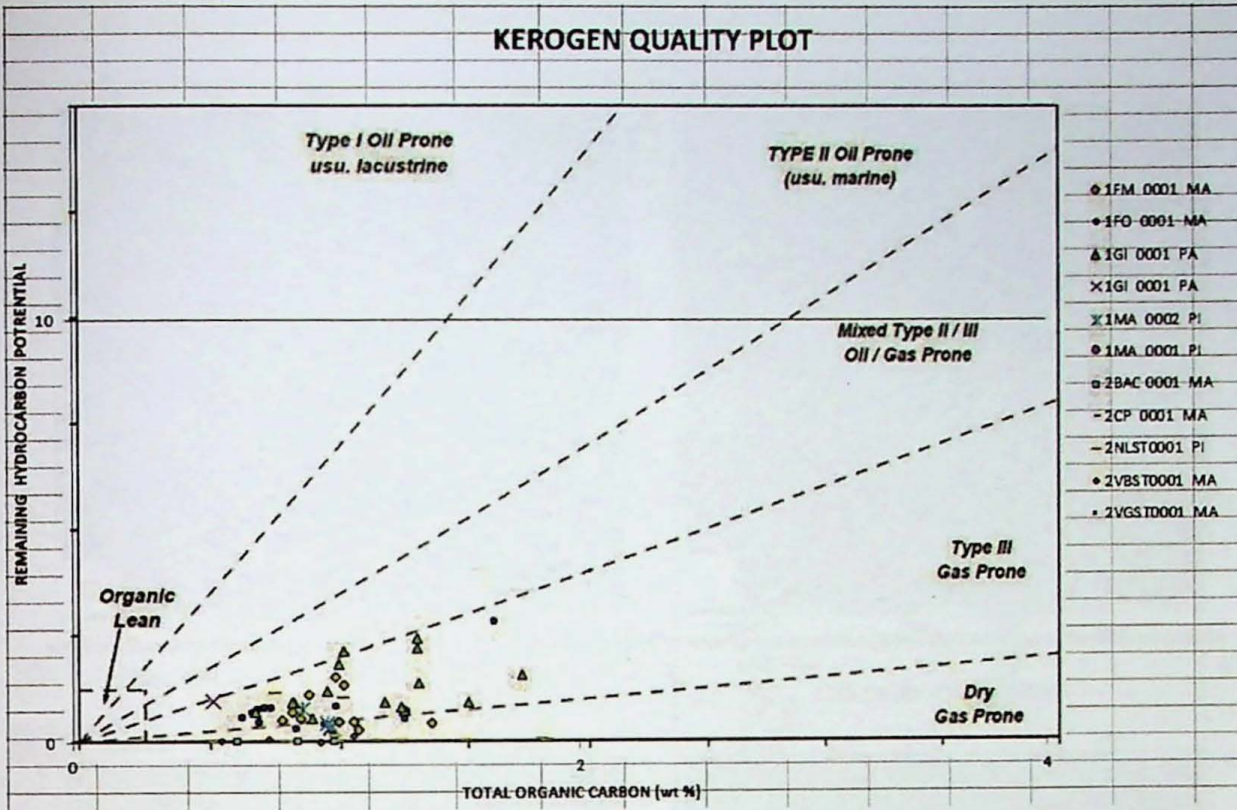


Figura 85: Diagrama de carbono orgânico total pelo potencial remanescente de geração (S2) da formação Longá. O diagrama é usado para classificar e avaliar a qualidade do Querogênio da formação geradora.

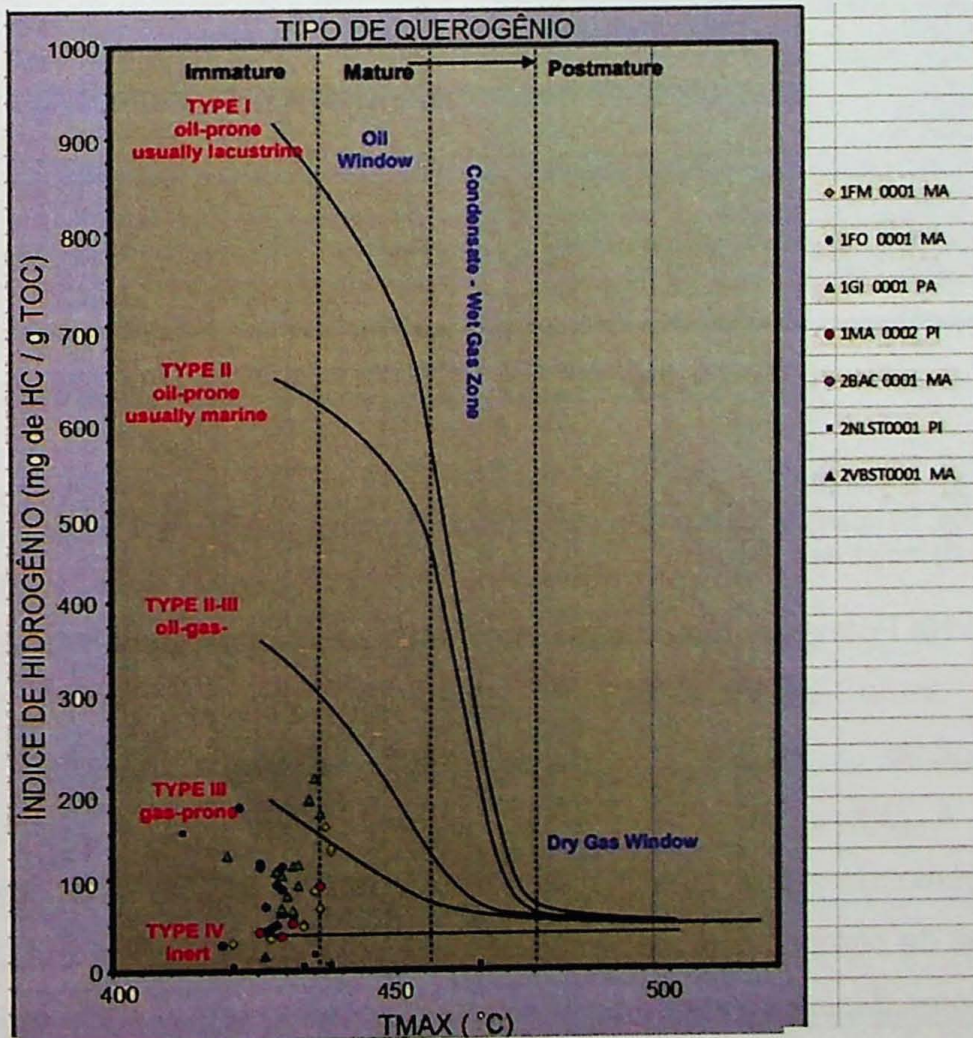


Figura 86: Diagrama de Índice de hidrogênio pelo TMAX para análise e classificação do tipo de querogênio do intervalo gerador Longá.



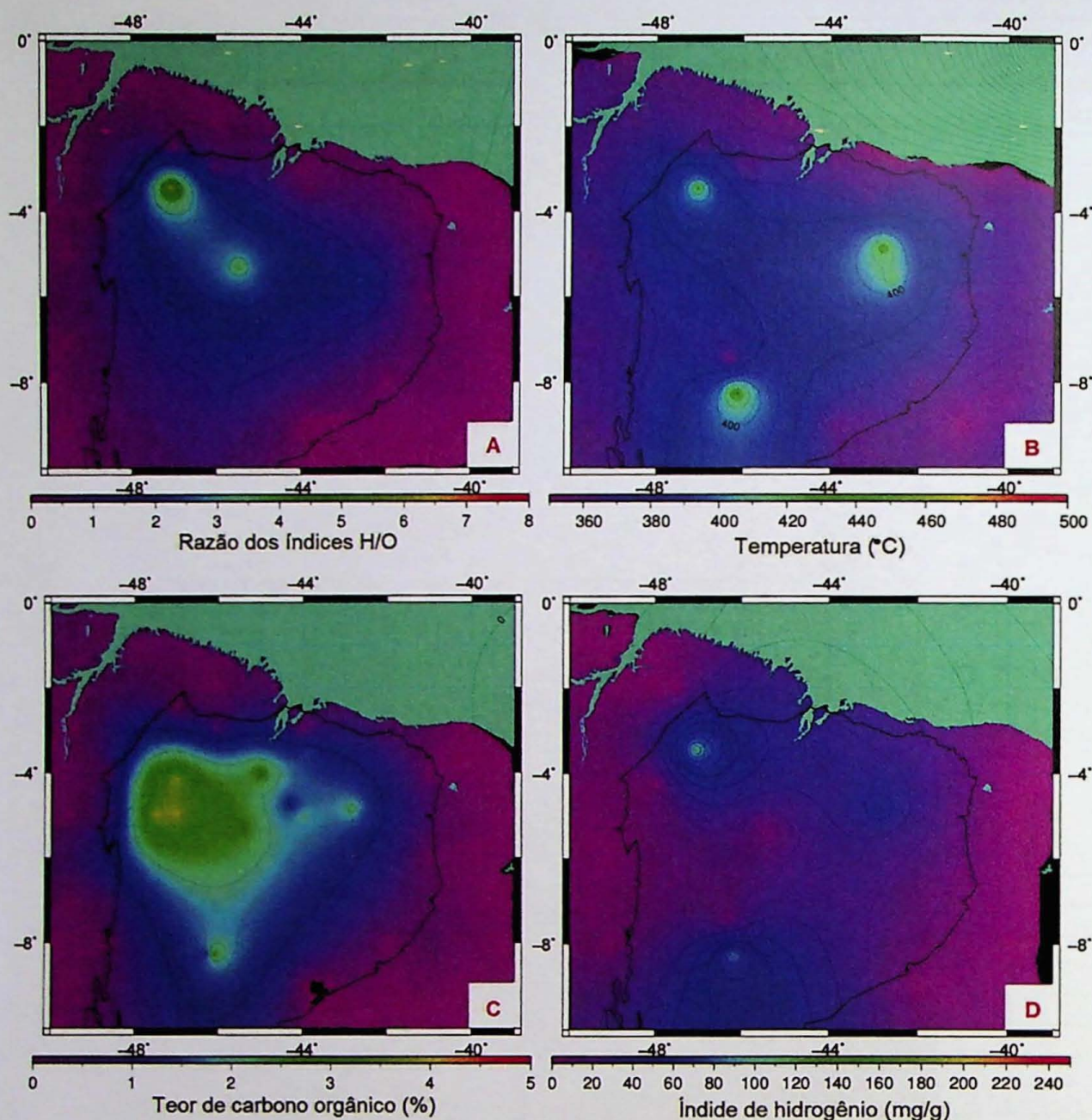


Figura 87: Isólicas da formação Longá, onde: A) Razão dos índices de hidrogênio pelos índices de oxigênios, B) interpolação dos valores de TMAX nos poços estudados, C) Distribuição dos valores de TOC e D) Índice de Hidrogênio.

Na Formação Pimenteiras encontram-se os principais intervalos geradores da bacia, tanto em condições de maturidade térmica, como na qualidade do querogênio.

A Figura 88 apresenta os valores de carbono orgânico, S<sub>2</sub>, índice de hidrogênio, do índice de oxigênio da formação Pimenteiras. A primeira coluna mostra valores de carbono orgânico entre 1 e 4%, com médias entre 2 e 2,5%. Além disso, os valores diminuem com a proximidade da soleira, o que pode indicar que a soleira acelerou, ou mesmo, atuou como catalizador na geração de hidrocarbonetos.



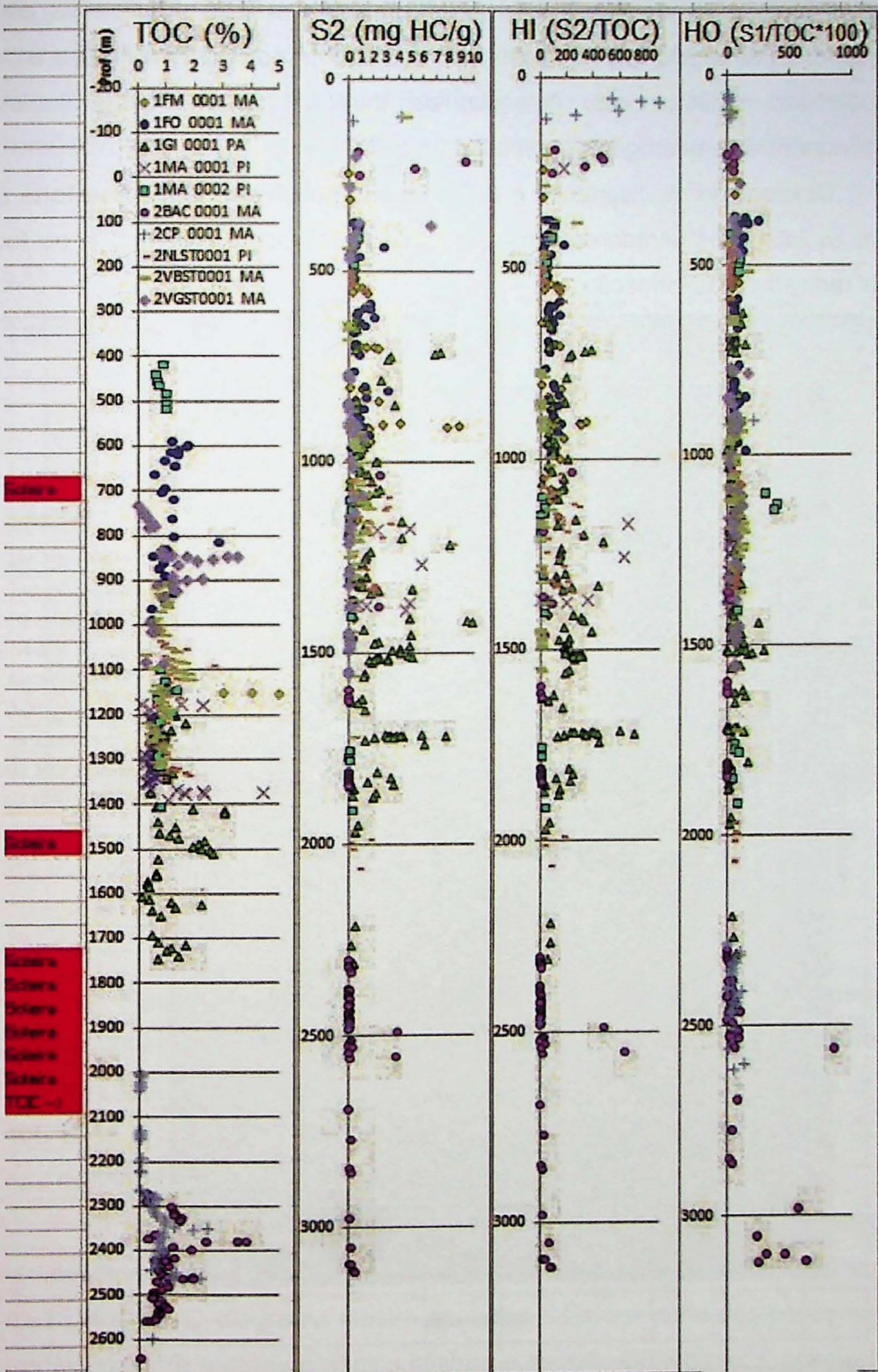


Figura 88: Gráficos dos valores de carbono orgânico, S2, índice de hidrogênio, do índice de oxigênio da formação Pimenteira. Os valores mostrados foram retirados de resultados dos ensaios de pirólise em amostras.



O diagrama de *Van Krevelen* apresentado na Figura 89 mostra querogênio predominantemente do tipo 3 e 4, porém uma quantidade significativa de dados localizam-se na região correspondente ao querogênio do tipo 2. A Figura 91 do teor de carbono orgânico pelo potencial de geração remanescente indica uma predominância de querogênio do tipo 2 e 3.

Observando os diagramas e os perfis são possíveis notar uma forte relação entre os intervalos geradores com maior teor de carbono orgânico e os folhelhos mais radioativos da formação.

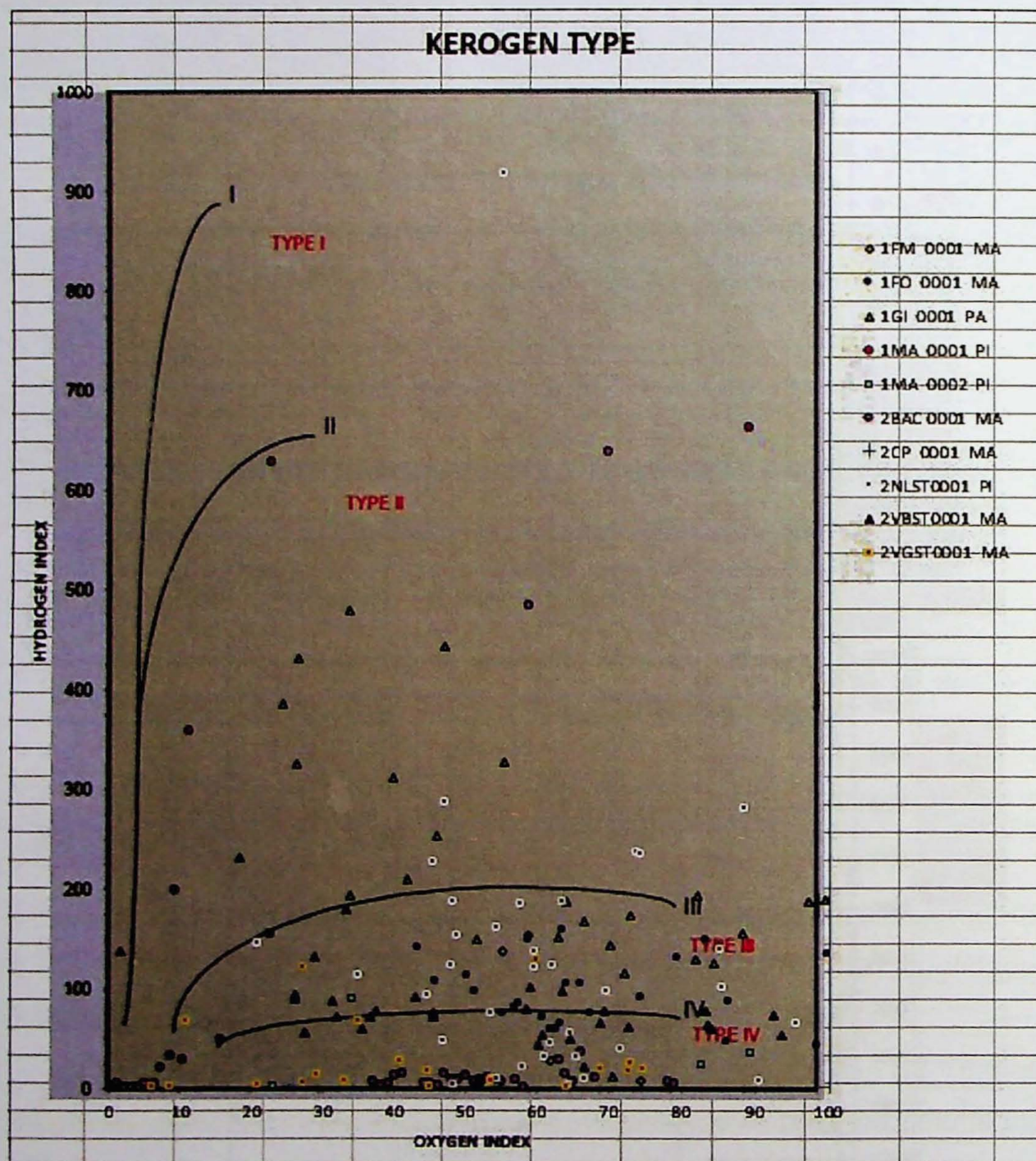


Figura 89: Diagrama de Van Krevelen para análise do tipo de matéria orgânica. O diagrama mostra os valores do índice de hidrogênio em função do índice de oxigênio da Formação Pimenteiras.



Rodrigues (1995) dividiu a formação em três zonas de acordo com o potencial de geração de cada uma (Figura 90). Os folhelhos do tipo A tem TOC entre 1 e 3% e querogênio do tipo 3. Os do tipo B apresentam TOC entre 1 e 3,5% e querogênio do tipo 2 e 3. Já o C, segundo Rodrigues (95), correspondem aos folhelhos mais radioativos da bacia, com TOC entre 1 e 5% e querogênio tipo 2 (Figura 89).

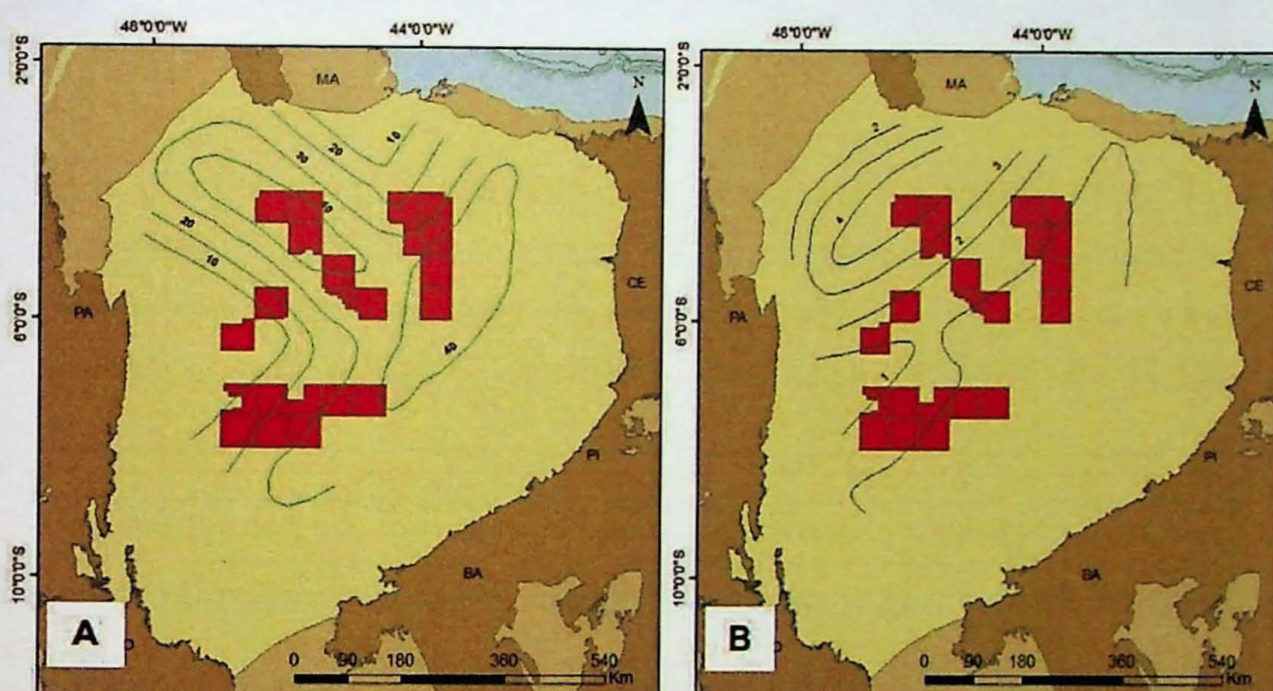


Figura 90: Isólitais dos folhelhos radioativos (em A) e do teor de carbono orgânico (TOC) (em B) do tipo C. (modificado de Rodrigues, 95).

A Figura 91 mostra o diagrama de Tmax pelo índice de hidrogênio para os dados das amostras da Formação Pimenteiras. Neste diagrama é possível observar a predominância de amostras na janela que compreende o intervalo entre 430 e 450°C de Tmax e índice de hidrogênio entre 50 e 300 mg de HC por TOC, ou seja, nas regiões compatíveis com gás/querogênio do tipo 3.

O tipo do querogênio também foi avaliado segundo o diagrama da temperatura TMAX em função do índice de hidrogênio (HI), conforme apresentado na Figura 92. O diagrama mostra os dados da maior parte dos poços na janela de transição entre matéria orgânica imatura e madura. Além disso, é possível observar, assim como apresentado nos diagramas anteriores, uma maior incidência de amostras localizadas nas regiões correspondentes ao querogênio do tipo 3, porém



com incidência significativa de amostras ocupando a região de matéria orgânica do tipo 2.

Os poços situados a NE do depocentro mostraram Tmax mais altos e matéria orgânica do tipo 2-3, porém poços mais próximos do depocentro da bacia, mostram valores entre 10 a 20 graus menores e matéria orgânica do tipo 3, conforme aponta as distribuições destes parâmetros apresentadas na Figura 93. Além disso, os resultados das análises de laboratório indicam que os índices de hidrogênio, bem como os valores de S2 são maiores nas regiões das bordas NE e SW da bacia.

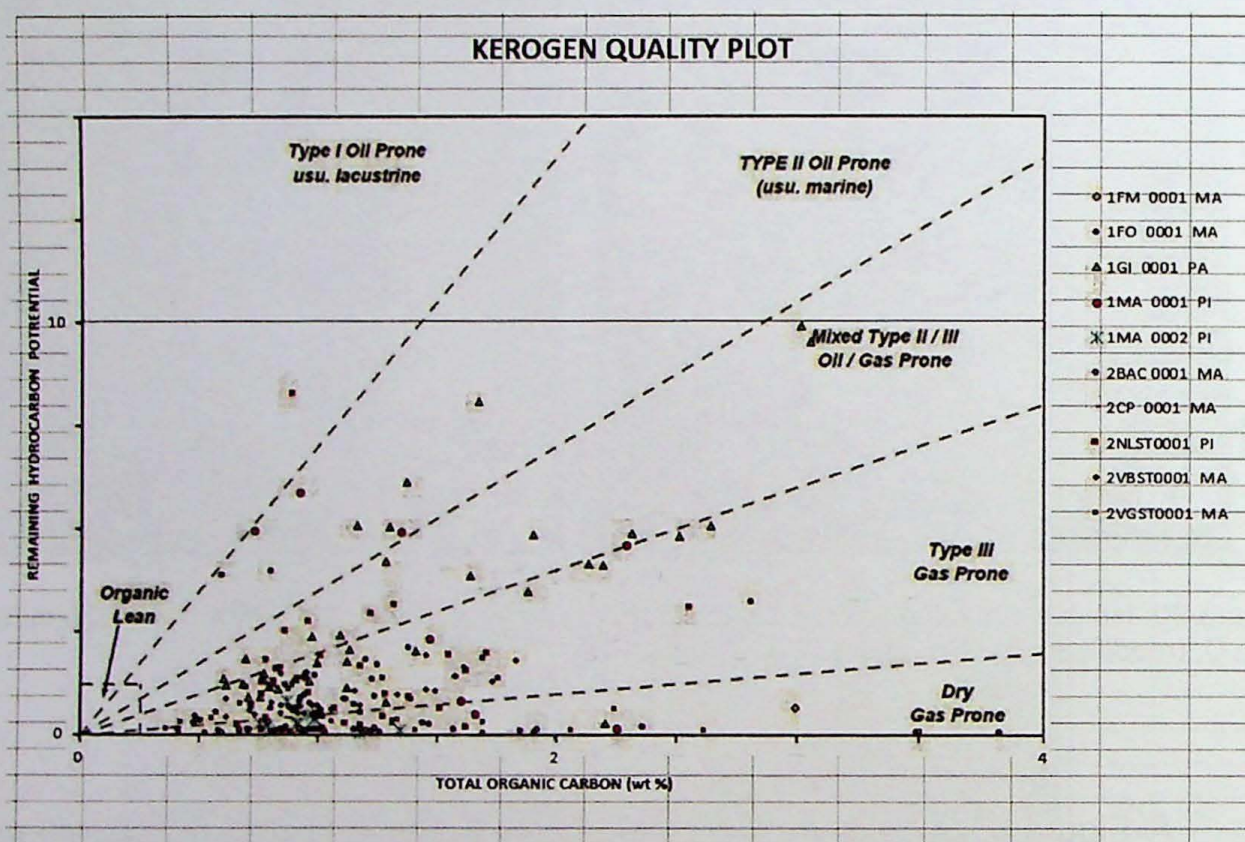


Figura 91: Diagrama do carbono orgânico total em função do potencial remanescente de geração (S2) da Formação Pimenteiras. O diagrama é usado para classificar e avaliar a qualidade do Querogênio da formação geradora.

Dados de refletância de vitrinite entre 0,5 e 0,7% dos intervalos próximos a soleiras, parecem indicar, conforme sugerido por Miranda (2014), um modelo de geração controlado pela distancia entre rocha geradora e as intrusões de diabásio.



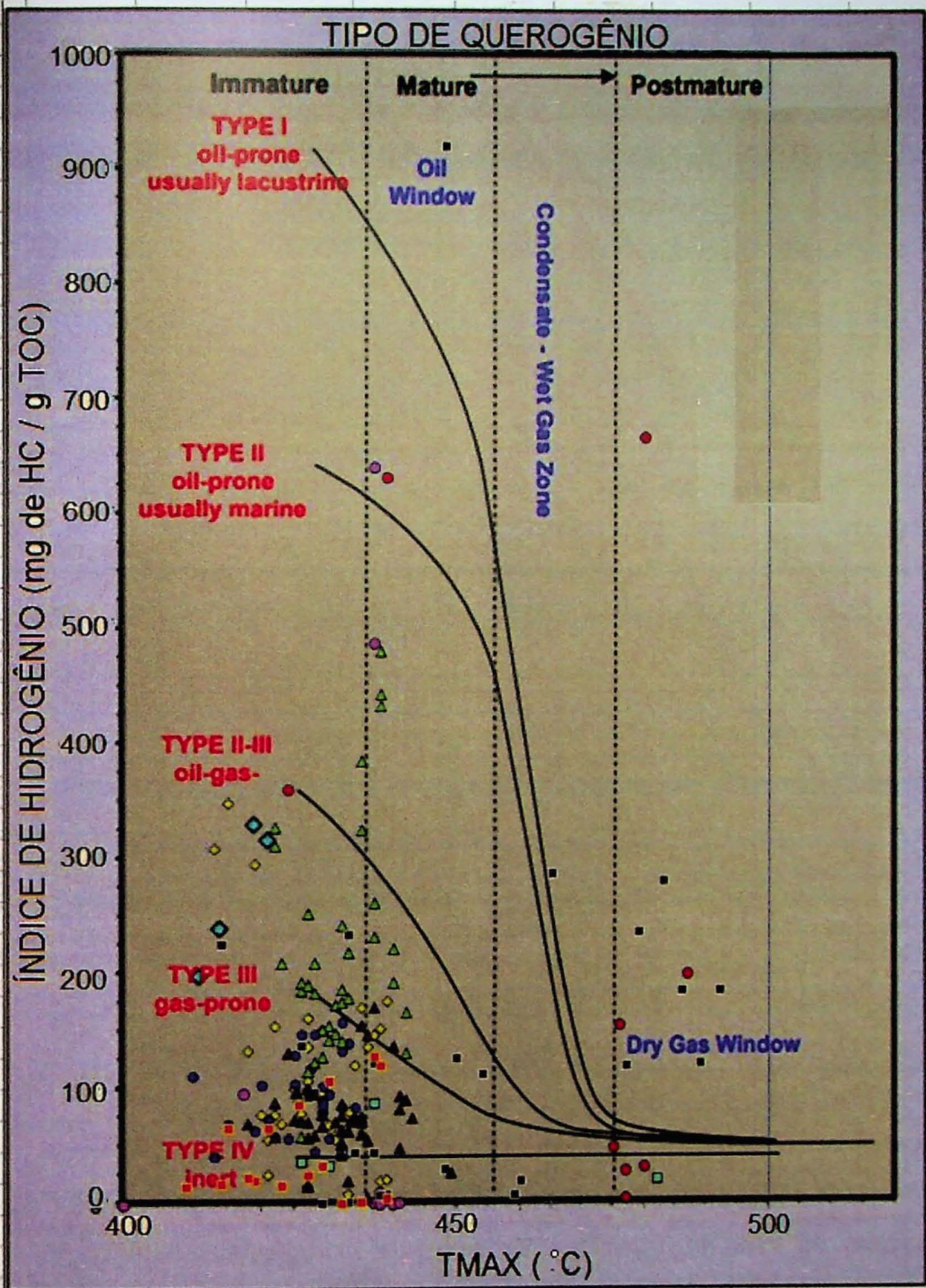


Figura 92: Diagrama do Índice de hidrogênio pelo TMAX para análise e classificação do tipo de querogênio do intervalo gerador Pimenteiras. As formas e cores dos pontos correspondem a dados de diversos poços espalhados pela bacia.



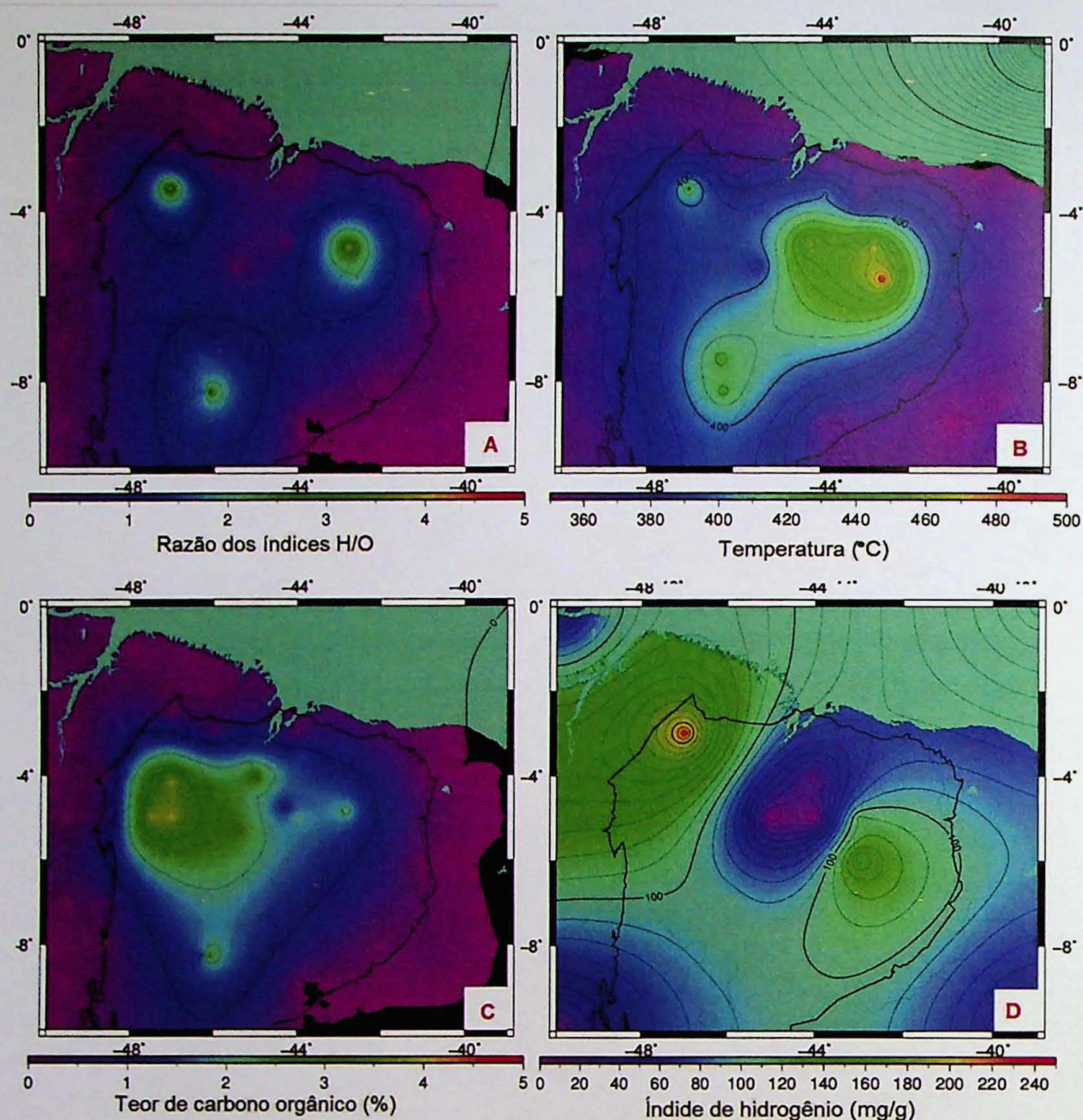


Figura 93: Isólicas da formação Pimenteiras onde: A) Razão dos índices de hidrogênio pelos índices de oxigênios, B) interpolação dos valores de TMAX nos poços estudados, C) Distribuição dos valores de TOC e D) Índice de Hidrogênio.

A mesma análise foi realizada na Formação Tianguá e a Figura 94 apresenta valores de carbono orgânico, S<sub>2</sub>, índice de hidrogênio, refletância de vitrinite calculada e dos índices de oxigênio da formação Tianguá. Os valores de TOC não superam 2% e os resultados de S<sub>2</sub> indicam valores inferiores a 1%.

O diagrama de Van Krevelen dos dados da Formação Tianguá apresentado na Figura 95 aponta o querogênio desta unidade como sendo predominantemente do tipo 4, na maior parte da bacia está fora da janela de maturação, o que é corroborado pelos dados de T<sub>max</sub>.



A Figura 96 apresenta as distribuições realizadas por mínima quadratura das razões hidrogênio-oxigênio, Tmax, teor de carbono orgânico e do índice de hidrogênio da formação Tianguá pela bacia. A distribuição do teor de carbono orgânico indicado na figura 97-C mostra que os maiores valores de até 2% estão localizados da porção central da bacia até a borda nordeste.



Figura 94: Gráficos dos valores de carbono orgânico, S2, índice de hidrogênio, refletância de vitrinita calculada e do índice de oxigênio da formação Tianguá. Os valores mostrados foram retirados de resultados dos ensaios de pirólise em amostras.



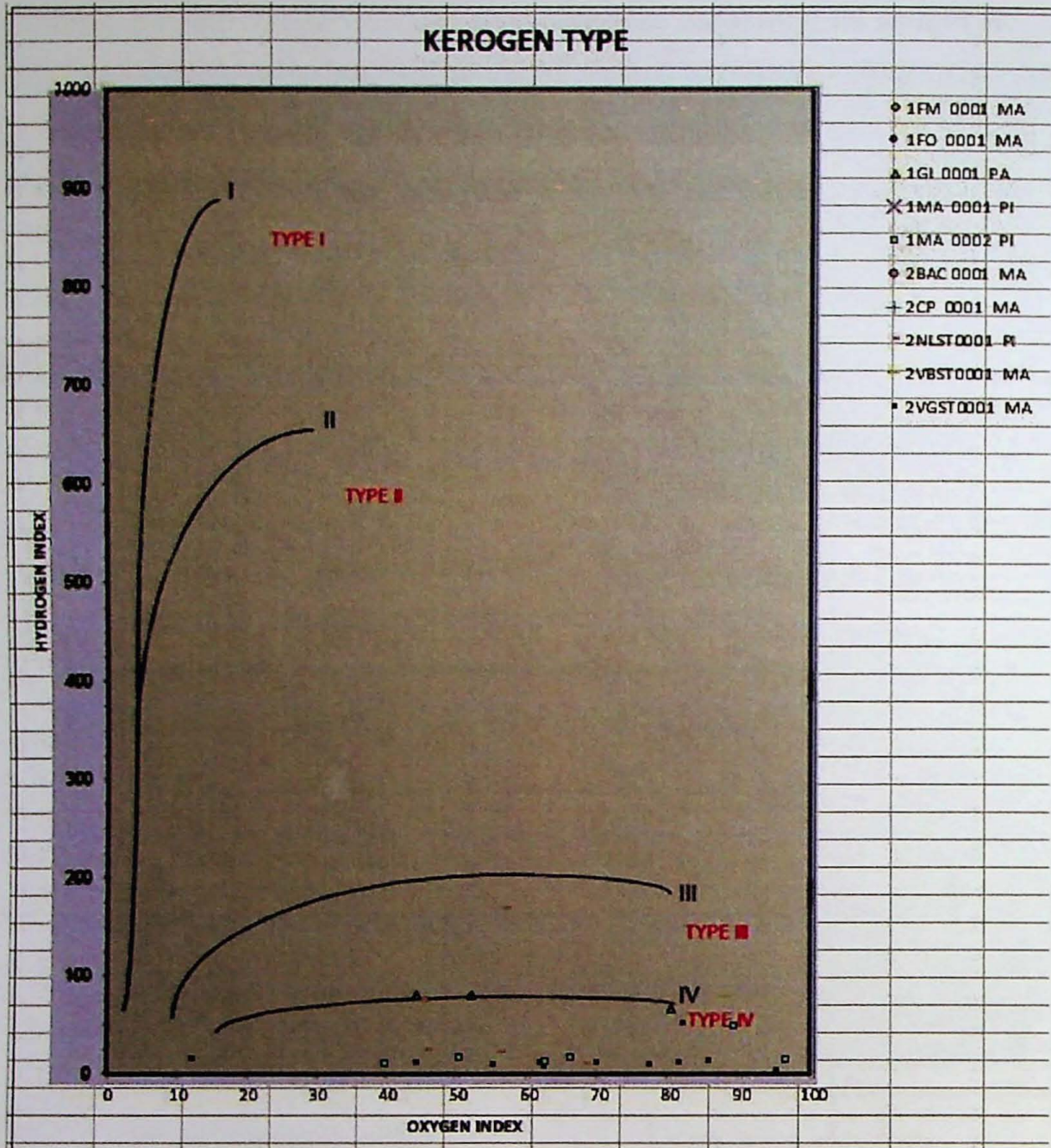


Figura 95: Diagrama de Van Krevelen para análise do tipo de matéria orgânica. O diagrama mostra os valores do índice de hidrogênio em função do índice de oxigênio da Formação Tianguá.



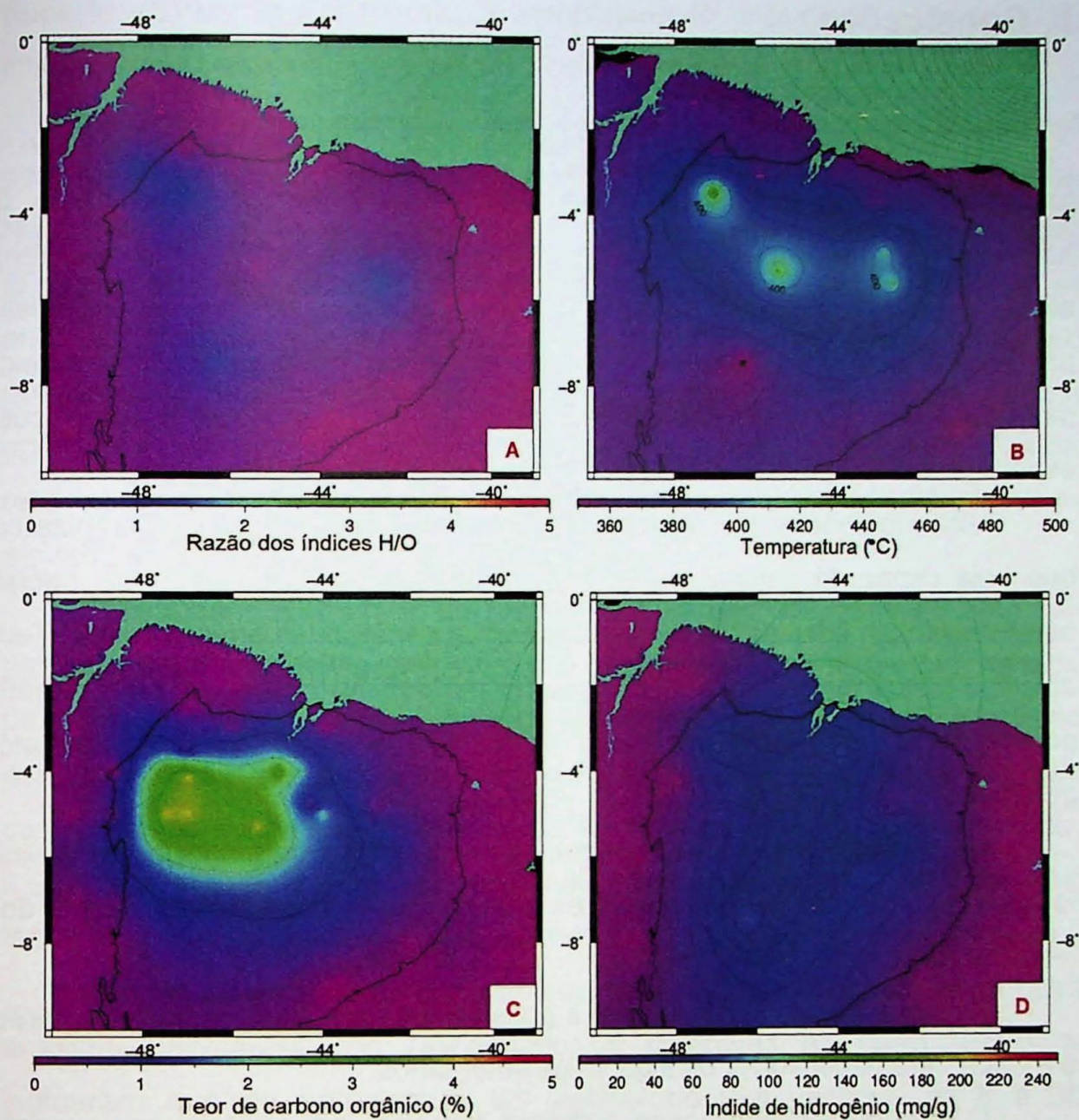


Figura 96: Isólicas da formação Tianguá onde: A) Razão dos índices de hidrogênio pelos índices de oxigênios, B) interpolação dos valores de TMAX nos poços estudados, C) Distribuição dos valores de TOC e D) Índice de Hidrogênio.



## 5. Conclusões

O processamento e a análise petrofísica de reservatórios são etapas fundamentais para a compreensão do sistema petrolífero de uma bacia. Embora funcione, o processo pelo qual estas etapas são usualmente realizadas envolvem altos custos de software, a expertise de um profissional altamente especializado e por vezes demanda de muito tempo, o que pode implicar em custos extras de mobilização de equipes para testes adicionais em campo ou mesmo custos de espera (“stand by”) de sonda até que decisões sejam tomadas.

Visando desenvolver uma sequência de processamento que otimize custos e que seja capaz de estimar parâmetros para subsidiar a tomada de decisões foi estabelecido um fluxo de trabalho alternativo aos métodos convencionais. Testes em diversos poços revelaram a viabilidade do método na obtenção de estimativas de porosidade e permeabilidade, bem como na classificação de reservatórios e geradores.

Além disso, o processo mostrou-se modular, ou seja, pode ser aplicado de forma integral ou parcial, muitas vezes em um intervalo de tempo menor do que é feito com os métodos convencionais.

Outra vantagem significativa é a possibilidade de utilizar softwares livres como o *Python* para desenvolver os algoritmos autônomos.

O método supervisionado *Gradient Boosting* se mostrou mais assertivo e adequado do que as *Redes Neurais Artificiais* e também melhor do que a *Regressão linear com múltiplas variáveis* na recuperação do perfil densidade. Assim, a densidade foi predita pelo modelo com índices de correlação variando entre 85% a 92%, na comparação com a densidade medida. Desta forma, o modelo foi usado para produzir perfis em poços antigos que não tinham o perfil densidade.

O mesmo método garantiu previsões de porosidade com coeficientes de correlação variando entre 77 e 93% com supervisão de perfis de ressonância magnética nuclear (RMN). Estes modelos poderiam ser aprimorados se houvessem mais poços com este tipo de perfil ou ainda se houvessem mais dados de porosidade de laboratório.

A classificação da porosidade do reservatório foi feita com base em amostras de calha e o treinamento foi realizado com uso do método Máquinas de Vetores de



Suporte (SVM). Desta forma, o reservatório foi classificado de acordo com 5 intervalos de porosidade e a precisão do modelo foi de 82%.

Além disso, modelos também foram treinados para predição da permeabilidade, com supervisão do perfil RMN e duas abordagens foram realizadas. Na primeira, o conjunto de dados foi dividido em treino e teste e o modelo treinado foi capaz realizar predições com correlações variando entre 82 e 89%. Na segunda abordagem, um poço foi aleatoriamente escolhido para treinar o modelo e os resultados mostraram correlações fracas, o que reforça a tese de que para aperfeiçoar os modelos preditivos de permeabilidade são necessários mais dados de ressonância ou uma grande quantidade de amostras com permeabilidades medidas em laboratório.

Além da classificação do reservatório, também foi realizada uma estimativa das zonas portadoras de gás com o uso do método K-means. Desta forma, os dados de perfilagem e cromatografia gasosa foram agrupados, por semelhança de características e os resultados foram comparados com as zonas de gás calculadas como uso do software IP. O índice de semelhança entre as duas variou de 0,46 a 0,90, sendo que os valores reais não são conhecidos e, portanto, ambas as técnicas, a convencional e a proposta neste trabalho são apenas estimativas.

Os resultados de cada parâmetro inferido pelos modelos foram usados em análises pervasivas dos reservatórios e geradores da bacia. Estas análises indicaram que as porosidades do reservatório Poti variaram de 0 a 35%, as permeabilidades ficaram entre 0 e 100mD e a espessura do intervalo saturado em gás atingiu cerca de 300m na porção central da bacia, sendo a média entre 50m e 120m. A formação Cabeças apresentou porosidades e permeabilidades médias de 14% e 50mD respectivamente, enquanto que a espessura média dos intervalos saturados com gás é de 50m. Já o reservatório siluriano Jaicós apresentou máximos de porosidades de 14% e permeabilidades variando entre 0,050 e 5 mD. Os intervalos com gás tiveram suas espessuras variando, em geral, entre 0 e 15m, atingindo pontualmente valores de 50m.

A bacia do Parnaíba, talvez pela complexidade geológica, ou mesmo por questões relacionadas à viabilidade da exploração comercial, possui uma enorme carência de informações petrofísicas e geoquímicas de laboratório consideradas essenciais para o desenvolvimento de modelos preditivos mais precisos e assertivos. Desta forma, mais dados, especialmente de laboratório são necessários



para aprimorar as análises e modelos desenvolvidos neste trabalho. Eventualmente, os processos aqui descritos poderiam ser testados em uma bacia offshore, com maior volume e diversidade de dados.



## Referências<sup>1</sup>

- ANIFOWOSE, F. A.; LABADIN, J.; ABDULRAHEEM, A. Prediction of petroleum reservoir properties using different versions of adaptive neuro-fuzzy inference system hybrid models. *Int. J. Comput. Inf. Syst. Ind. Manage. Appl.*, v. 5, p. 413–426, 2013. Disponível em: <[http://www.mirlabs.org/ijcisim/regular\\_papers\\_2013/Paper117.pdf](http://www.mirlabs.org/ijcisim/regular_papers_2013/Paper117.pdf)>.
- ASQUITH, G. B.; JACKA, A. D. *Petrophysics of bimodal porosity: Lower Cretaceous Rodessa limestone, Running Duke field, Houston County, Texas.* 1992.
- Archie, G.E.. "The electrical resistivity log as an aid in determining some reservoir characteristics". *Petroleum Transactions of AIME.* 146: 54–62. doi:10.2118/942054-g. 1942.
- Archie, G.E.. "Electrical resistivity an aid in core-analysis interpretation". *American Association of Petroleum Geologists Bulletin.* 31 (2): 350–366. 1947.
- BAOUCHE, R.; CHAOUCHI, R. Permeability Prediction from CMR using the SVM classifier in shaly sand formations through the Well Takw-1 in Devonian and Ordovician . Illizi basin . p. 1–12, 2015.
- BIZZI, Luiz Augusto et al. *Geologia, Tectônica e Recursos Minerais do Brasil: texto, mapas e SIG.* CPRM, 2003.
- BREIMAN, L., Friedman, J., Stone, C.J., Olshen, R.A. *Classification and Regression Trees.* 1984.
- DE CASTRO, David Lopes et al. Geophysical evidence of pre-sag rifting and post-rifting fault reactivation in the Parnaíba basin, Brazil. *Solid Earth*, v. 7, n. 2, p. 529-548, 2016.
- DE ALMEIDA, Fernando Flávio Marques; CARNEIRO, Celso Dal Ré. Origem e evolução da Serra do Mar. *Revista Brasileira de Geociências*, v. 28, n. 2, p. 135-150, 2017.
- CAMPBELL, D. F.; ALMEIDA, LA de; SILVA, S. de O. Relatório preliminar sôbre a geologia da Bacia do Maranhão: Cons. Nac. Petr., Bol, v. 1, 1949.
- CAROZZI, Albert V. Análise ambiental e evolução tectônica sinsedimentar da seção siluro-eocarbonífera da Bacia do Maranhão. Centro de Pesquisas e Desenvolvimento Leopoldo A. Miguez de Mello, Divisão de Informação Técnica e Patentes, 1975.

---

<sup>1</sup> De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.



CAPUTO, MÁRIO VICENTE; LIMA, EGLEMAR Conde. Estratigrafia, idade e correlação do Grupo Serra Grande-bacia do Parnaíba. In: Anais 33° Congresso Brasileiro de Geologia, Rio de Janeiro, 1984. p. 740-753.

CAPUTO, M.V.; IANNUZZI R. & FONSECA V.M.M. Bacias sedimentares brasileiras: Bacia do Parnaíba. *Phoenix*, 81:1-6. 2005.

CHOQUETTE, Philip W.; PRAY, Lloyd C. Geologic nomenclature and classification of porosity in sedimentary carbonates. *AAPG bulletin*, v. 54, n. 2, p. 207-250, 1970.

CRAIN, E. R. *Petrophysical Handbook*. Alberta, Canada, v. 55, 2000.

DELLA FÁVERA, J. C. Tempestitos da Bacia do Parnaíba. Porto Alegre. 242p. 1990. Tese de Doutorado. Tese (Doutorado em Geociências), Programa de Pós-Graduação em Geociências, Instituto de Geociências/Universidade Federal do Rio Grande do Sul.

DELLA FAVERA, Jorge C. Devonian storm-and tide-dominated shelf deposits, Parnaiba Basin, Brazil. *AAPG Bulletin*, v. 66, n. 5, p. 562-562, 1982.

ELKATATNY, S. et al. New insights into porosity determination using artificial intelligence techniques for carbonate reservoirs. *Petroleum*, v. 4, n. 4, p. 408–418, 2018. Disponível em: <<https://doi.org/10.1016/j.petlm.2018.04.002>>.

EROFEEV, A. et al. Prediction of Porosity and Permeability Alteration Based on Machine Learning Algorithms. *Transport in Porous Media*, v. 128, n. 2, p. 677–700, 2019.

GALUSHKIN, Y.I. 1997. Thermal effects of igneous intrusions on maturity of organic matter: a possible mechanism of intrusion. *Organic Geochemistry*, 26(11-12): 645-658. [http://dx.doi.org/10.1016/S0146-6380\(97\)00030-2](http://dx.doi.org/10.1016/S0146-6380(97)00030-2).

GÓES, A. M. (USP). A Formação Poti (Carbonífero Inferior) da Bacia do Parnaíba. Tese, p. 171, 1995.

GÓES, A. M. O.; FEIJÓ, F. J. Bacia do Parnaíba *Boletim de Geociências da Petrobras*, 1994. .

GRAHN, Y.; CAPUTO, M. V. Early Silurian glaciations in Brazil. *Palaeogeography, Palaeoclimatology, Palaeoecology*, v. 99, n. 1–2, p. 9–15, 1992.

GUARIDO, M. Machine learning in geoscience: facies classification with features engineering, clustering, and gradient boosting trees. v. 30, n. 2017, p. 1–24, 2018.

DEQUECH, VICTOR. Contribuição para a geologia do meio norte, especialmente Piauí e Maranhão,



- Brasil. In: Congresso Pan-Americano de Engenharia de Minas e Geologia. 1946. p. 69-108.
- ELLIS, Darwin V.; SINGER, Julian M. Well logging for earth scientists. Dordrecht: Springer, 2007.
- FRIEDMAN, Jerome; HASTIE, Trevor; TIBSHIRANI, Robert. The elements of statistical learning. New York: Springer series in statistics, 2001.
- HAYKIN, Simon S. et al. Neural networks and learning machines/Simon Haykin. New York: Prentice Hall. 2009.
- HASTIE, Trevor et al. Multi-class adaboost. **Statistics and its Interface**, v. 2, n. 3, p. 349-360, 2009.
- CHIU, Min-Chie et al. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. **Journal of Applied Sciences**, v. 8, n. 16, p. 2554-2558, 1949.
- IP HELP MANUAL, Interactive Petrophysics 4.0 User Manual, 2001, Senergy Software Ltd.
- JANIKIAN, L., ALMEIDA, R., GALEAZZI, C., TAMURA, L., ARDITO, J., CHAMANI, M. Variability of fluvial architecture in a Silurian sag basin margin: sheet-braided and meandering ancestor rivers in the Serra Grande Group, northeastern Brazil. 2019.
- KEGEL, Wilhelm. Lamelibrânquios da formação poti (Carbonífero inferior) do piauí. Divisão de geologia e mineralogia, 1954.
- KUMAR, Abhishek; SINHA, Satish K. Total organic carbon prediction for shale gas exploration using statistical clustering, multiple regression analysis. In: 10th Biennial International Conference and Exposition. 2013. p. 293.
- KURODA, M. C. Redes Neurais Aplicadas ao Estudo de Rochas Reservatório. p. 134, 2012.
- LIMA, Enjôlras de A. Medeiros; LEITE, Jairo Fonseca. Projeto estudo global dos recursos Minerais da Bacia sedimentar do Parnaíba. Integração geológico-metalogenética. 1978.
- LIU, Y. Support Vector Machine ( SVM ) -based Geostatistical Inversion : An Integrated Solution to Honor Well-log Data and Seismic Attributes. p. 2-5, 2009.
- Mayrink, Victor. Avaliação do Algoritmo Gradient Boosting em aplicações de previsão de carga elétrica a curto prazo. Dissertação de Mestrado. 2017.
- MCCULLOCH, W. S.; PITTS, Walter. How nervous structures have ideas. Transactions of the American Neurological Association, p. 10-16, 1949.
- MESNER, J. C., WOOLDRIDGE LCP. Maranhão Paleozoic basin and Cretaceous coastal basins,



north Brazil. *Bull Am Assoc Petrol Geol* 48: 1475-1512. 1964.

MIRANDA, FSDE. Pimenteiras Shale: Characterization of an atypical unconventional petroleum system, Parnaíba Basin, Brazil. *Search and Discovery Article*, v. 10369, 2014.

MILANI, Edison J.; ZALAN, Pedro V. An outline of the geology and petroleum systems of the Paleozoic interior basins of South America. *Episodes-News magazine of the International Union of Geological Sciences*, v. 22, n. 3, p. 199-205, 1999.

MILANI, Edson José. Geodinâmica fanerozóica do Gondwana sul-ocidental e a evolução geológica da Bacia do Paraná. *Geologia do Rio Grande do Sul, CIGO/UFRGS*. Porto Alegre, p. 275-302, 2000.

MOURA, R., KLEIN, M., ZYLSTRA, J., COELHO, C. A., & SINHA, B. Inference for multivariate regression model based on synthetic data generated using plug-in sampling. Washington USA: US Census Bureau. 2018.

QUADROS, Luiz Padilha. Distribuição bioestratigráfica dos Chitinozoa e Acritarchae na bacia do Parnaíba. *Petrobrás, Centro de Pesquisas e Desenvolvimento Leopoldo A. Miguez de Mello, Setor de Informação Técnica e Propriedade Industrial*, 1982.

PAIVA, G. Estratigrafia de sondagem N° 125: anexo N° 6. Serviço de fomento mineral, bol 18. Rio de Janeiro., p. 107. 1937.

PASSEY, Q.R., CREANEY, S., KULLA, J.B., MORETTI, F.J. and STROUD, J.D. 1990. A practical model for organic richness from porosity and resistivity logs. *American Association of Petroleum Geologists Bulletin*, v. 74, no. 12, p. 1777-1794.

PETRI, S., FULFARO, V.J. *Geologia do Brasil (Fanerozóico)* São Paulo: TA Queiroz. Universidade de São Paulo. 1983.

PLUMMER, F. B.; PRICE, L. I.; GOMES, F. A. *Estados do Maranhão e Piauí*. Brasil, Conselho Nacional do Petróleo. Relatório de, p. 87-134, 1946.

QUINLAN, J. R., C4.5: *Programs for Machine Learning*, Morgan Kaufmann, Los Altos, 1993.

LI, Lili; LIU, Xiyu; ZZHUANG, Bo. Fuzzy C-means algorithm based on simulated annealing Particle Swarm Optimization [C]. *Computer Engineering and Applications*, v. 44, n. 30, p. 170-172, 2008.

LIU, Yongshe; JOURNAL, André G. A package for geostatistical integration of coarse and fine scale data. *Computers & Geosciences*, v. 35, n. 3, p. 527-547, 2009.

RAFIK, B.; KAMEL, B. Prediction of permeability and porosity from well log data using the nonparametric regression with multivariate analysis and neural network, Hassi R'Mel Field, Algeria. *Egyptian Journal of Petroleum*, v. 26, n. 3, p. 763-778, 2017.



RODRIGUES, R. Estudo sedimentológico e estratigráfico dos depósitos silurianos e devonianos da Bacia do Parnaíba. Rel. Int, v. 273, 1967.

RODRIGUES, R. A geoquímica orgânica na Bacia do Parnaíba. Porto Alegre-RS, 225p. Tese de doutorado—Universidade Federal do Rio Grande do Sul, 1995.

ROSENBLATT, Frank. The perceptron, a perceiving and recognizing automaton Project Para. Cornell Aeronautical Laboratory, 1957.

ROY, A.; MATOS, M.; MARFURT, K. Automatic Seismic Facies Classification with Kohonen Self Organizing Maps - A Tutorial. Spgindia.Org, n. December, p. 6–14, 2010. Disponível em: <<http://spgindia.org/geohorizon/dec2010/automatic.pdf>>.

RUSSELL, Stuart; NORVIG, Peter. Inteligência Artificial: uma abordagem moderna. Ed. Campus, 2ª Edição. São Paulo, Brazil, 2003.

SACCHI, Mauricio D.; LIU, Bin. Minimum weighted norm wavefield reconstruction for AVA imaging. Geophysical Prospecting, v. 53, n. 6, p. 787-801, 2005.

SAFTIĆ, B. et al. Interdependence of petrophysical properties and depth: Some implications of multivariate solution on distinction between the lower pontian hydrocarbon-bearing sandstone units in the western part of the sava depression. Geologia Croatica, v. 54, n. 2, p. 259–277, 2001.

SANTOS, V. H. (UFPE); FILHO, M. F. L. (UFPE); NEUMANN, V. H. (UFPE). Estudo faciológico em afloramentos e analogia de depósitos fluviais na bacia do parnaíba, ne do brasil. Estudos Geológicos, v. 14, n. 1996, p. 62–79, 2004.

SMALL, H.L. Geologia e suprimento d'água subterrânea no Piauí e parte do Ceará. Série Geologia, Publicação 32, Rio de Janeiro, Inspetoria de Obras Contra as Secas, 146p. 1914.

Tiny Perm II: Portable Air. Permeameter, User's Manual, 2010. NER. New England Research Inc. 2010.

Tissot, B.P.; Welter, D.H., 1984. Petroleum Formation and Occurrence. SpringerVerlag, Berlin; Heidelberg; New York, Tokyo. 679pp.

ZALÁN, Pedro Victor et al. Bacia do Paraná. Origem e evolução das bacias sedimentares, p. 135-168, 1990.

VAZ, P. T. et al. Bacia do Parnaíba: Boletim de Geociências da Petrobras, v. 15. 2007.

VERMA, A. K. et al. Porosity and Permeability estimation using neural network approach from well log data. GeoConvention Vision Conference, Canada, v. 41276, 2012.







## **Anexo A – Trabalhos de Campo e Estudo Faciológico**

Nesta seção será apresentado um estudo de porosidade e permeabilidade de amostras de afloramento do Grupo Serra Grande desenvolvido durante o período deste doutorado.

Para o início do estudo do sistema permo-poroso do grupo Serra Grande foram efetuadas diversas medidas de permeabilidade nos afloramentos rochosos da porção leste da bacia do Parnaíba, situada no estado do Ceará e Piauí. As medidas de permeabilidade do primeiro ponto de estudo (denominado afloramento AF1) foram realizadas in-situ, pois as rochas estavam secas, porém nos demais locais havia uma saturação de água significativa de modo que os blocos tiveram de ser retirados e medidos em laboratório, sempre utilizando o mesmo equipamento e metodologia.

Após medida a permeabilidade, foi necessário convertê-la da unidade de leitura do equipamento para mili Darcy, por meio de uma equação fornecida pelo fabricante. Todos estes valores são apresentados nas tabelas 1, 2, 3 e 4.

Em cada ponto de amostragem foram determinadas as coordenadas geográficas por meio de um GPS diferencial da marca Sokkia, mantendo um ponto base fixo com coleta de dados superior a 4 horas e um ponto móvel em cada local de coleta de amostras. Os pontos foram posteriormente processados com o uso do software Spectrum Survey versão 3.2 da Sokkia. O sistema geodésico de referência utilizado foi o SAD69.

Os resultados apresentados serão comparados com os resultados de permeabilidade e porosidade futuramente realizados em ensaios laboratoriais das mesmas amostras. Além disso, serão realizados ensaios de petrofísica avançada que permitiram inferir a anisotropia do reservatório por meio dos parâmetros como fator de formação, coeficiente de cimentação e expoente de saturação e desta forma, subsidiar a caracterização do reservatório.



Tabela 1: Permeabilidade (mD) das amostras do ponto de estudos 1 (AF1) da bacia do Parnaíba.

<b>Dados de permeabilidade das amostras do grupo Serra Grande</b>								
<b>Amostra</b>	<b>Permeab. horizontal (mD)</b>	<b>Permeab. vertical (mD)</b>	<b>H/V</b>	<b>Estrutura</b>	<b>Facies</b>	<b>N Coordi.</b>	<b>E Coord.</b>	<b>Z Coord (m)</b>
SG-I- 1	3,282	3,282	1,000	Cruzada acanalada	Arenito médio	3°55'2.24"S	41°14'25.13"O	326,31
SG-I- 2		3,282	0,000	Cruzada acanalada	Arenito fino	3°55'2.24"S	41°14'25.13"O	326,31
SG-I-3	16,414	1,249	13,140	Cruzada acanalada	Arenito fino	3°55'2.24"S	41°14'25.13"O	326,31
SG-I-4	8,621	7,827	1,101	Cruzada baixo ang.	Arenito médio	3°55'2.24"S	41°14'25.13"O	326,31
SG-I-5	1,249	-	-	Cruzada acanalada	Arenito médio	3°55'2.24"S	41°14'25.13"O	326,31
SG-I-6	22,648	16,414	1,380	Cruzada acanalada	Arenito fino	3°55'2.55"S	41°14'27.42"O	321,37
SG-I-7	6,248	4,528	1,380	Cruzada baixo ang.	Arenito médio	3°55'2.55"S	41°14'27.42"O	321,37
SG-I-8	2,378	4,528	0,525	Cruzada acanalada	Arenito médio	3°55'2.55"S	41°14'27.42"O	321,37
SG-I-9	4,528	4,528	1,000	Cruzada acanalada	Arenito médio	3°55'2.55"S	41°14'27.42"O	321,37
SG-I-10	8,621	6,248	1,380	Cruzada acanalada	Arenito médio	3°55'2.55"S	41°14'27.42"O	321,37
SG-I-11	3,282	8,621	0,381	maciça	Arenito médio	3°55'2.54"S	41°14'27.29"O	321,58
SG-I-12	8,621	11,896	0,725	maciça	Arenito médio	3°55'2.54"S	41°14'27.29"O	321,58
SG-I-13	16,414	-	-	Cruzada acanalada	Arenito médio	3°55'2.54"S	41°14'27.29"O	321,58
SG-I-14	4,528	28,374	0,160	Cruzada baixo ang.	Arenito médio	3°55'2.52"S	41°14'27.17"O	321,75
SG-I-15	11,896	-	-	Cruzada acanalada	Arenito médio	3°55'2.52"S	41°14'27.17"O	321,75
SG-I-16	11,896	11,896	1,000	Cruzada acanalada	Arenito médio	3°55'2.52"S	41°14'27.17"O	321,75
SG-I-17	4,528	-	-	Cruzada baixo ang.	Arenito médio	3°55'2.52"S	41°14'27.17"O	321,75
SG-I-18	0,069	2,794	0,025	Cruzada baixo ang.	Arenito médio	3°55'2.52"S	41°14'27.17"O	321,75
SG-I-19	16,414	18,078	0,908	Cruzada acanalada	Arenito médio	3°55'2.52"S	41°14'27.17"O	321,75
SG-I-20	3,282	8,084	0,406	Cruzada acanalada	Arenito médio	3°55'2.49"S	41°14'26.88"O	322,16
SG-I-21	4,528	-	-	Cruzada baixo ang.	Arenito médio	3°55'2.49"S	41°14'26.88"O	322,16
SG-I-22	11,896	-	-	Cruzada acanalada	Arenito médio	3°55'2.49"S	41°14'26.88"O	322,16
SG-I-26	16,414	11,896	1,380	Cruz. Acan.	Arn méd	3°55'2.39"S	41°14'26.51"O	324,01
SG-I-27	3,282	14,903	0,220	Cruz. Acan.	Arn méd	3°55'2.39"S	41°14'26.51"O	324,01
SG-I-28	2,378	-	-	Cruz. Acan.	Arn méd	3°55'2.39"S	41°14'26.51"O	324,01
SG-I-29	1,249	-	-	Cruz. Acan.	Arn méd	3°55'2.39"S	41°14'26.51"O	324,01
SG-I-30	4,528	43,121	0,105	Cruz. Acan.	Arn méd	3°55'2.39"S	41°14'26.51"O	324,01



Tabela 2: Permeabilidade (mD) das amostras do ponto de estudos 2 (AF2) da bacia do Parnaíba

<b>Dados de permeabilidade das amostras do grupo Serra Grande</b>								
Amostra	Permeab. horizontal (mD)	Permeab. vertical (mD)	H/V	Estrutura	Facies	N Coordi.	E Coord.	Z Coord (m)
SG-II-01	2,885	2,705	1,067	Cruzada a canalada	Arenito médio	4°3'29.36"S	40°50'21.08"O	738,20
SG-II-02	6,664	4,111	1,621	Cruzada a canalada	Arenito médio	4°3'29.55"S	40°50'21.13"O	740,89
SG-II-03	0,095	0,559	0,170	Cruzada a canalada	Arenito médio	4°3'29.69"S	40°50'21.15"O	739,45
SG-II-03A	0,577	2,378	0,243	Cruzada a canalada	Arenito médio	4°3'29.69"S	40°50'21.15"O	739,45
SG-II-03B	3,614	6,050	0,597	Cruzada a canalada	Arenito médio	4°3'29.69"S	40°50'21.15"O	739,45
SG-II-04	9,195	32,273	0,285	Cruzada a canalada	Arenito médio	4°3'30.64"S	40°50'21.32"O	744,52
SG-II-05	0,997	4,676	0,213	Cruzada a canalada	Arenito médio	4°3'30.70"S	40°50'21.33"O	746,62
SG-II-06	0,796	4,987	0,160	Cruzada Composta	Arenito médio	4°3'30.70"S	40°50'21.33"O	746,62
SG-II-07	0,475	-	-	Cruzada a canalada	Arenito médio	4°3'31.12"S	40°50'21.52"O	749,18
SG-II-08	0,405	0,935	0,433	Cruzada a canalada	Arenito médio	4°3'31.12"S	40°50'21.52"O	749,18
SG-II-09	4,829	2,230	2,166	Cruzada a canalada	Arenito médio	4°3'31.12"S	40°50'21.52"O	749,18
SG-II-10	1,249	3,733	0,335	Cruzada a canalada	Arenito médio	4°3'31.40"S	40°50'21.90"O	751,40
SG-II-11	1,376	0,997	1,380	Cruzada a canalada	Arenito médio	4°3'31.40"S	40°50'21.90"O	751,40
SG-II-12	0,392	2,159	0,182	Cruzada a canalada	Arenito médio	4°3'31.40"S	40°50'21.90"O	751,40
SG-II-13	1,780	3,178	0,560	Cruzada a canalada	Arenito médio	4°3'31.40"S	40°50'21.90"O	751,40
SG-II-14	1,376	2,619	0,525	Cruzada a canalada	Arenito médio	4°3'31.12"S	40°50'22.60"O	745,59
SG-II-15	5,493	24,155	0,227	Cruzada a canalada	Arenito médio	4°3'31.59"S	40°50'22.94"O	757,36
SG-II-16	1,616	3,077	0,525	Cruzada a canalada	Arenito médio	4°3'31.55"S	40°50'22.78"O	753,59
SG-II-17	2,885	10,458	0,276	Cruzada a canalada	Arenito médio	4°3'31.55"S	40°50'22.78"O	753,59
SG-II-18	0,541	1,565	0,346	Cruzada a canalada	Arenito médio	4°3'31.79"S	40°50'23.13"O	758,04
SG-II-19	2,303	1,134	2,031	Cruzada a canalada	Arenito médio	4°3'31.79"S	40°50'23.13"O	758,04

Além das coletas de amostras e coordenadas, foram realizadas descrições de facies das seções levantadas em campo e medidas de paleocorrente. As litofacies fluviais estudadas neste trabalho foram previamente interpretadas e classificadas por mim segundo os critérios de Miall (1996) e seus códigos são apresentados na tabela 5.



Tabela 3: Permeabilidade (mD) das amostras do ponto de estudos 1 (AF1) da bacia do Parnaíba

<b>Dados de permeabilidade das amostras do grupo Serra Grande</b>								
Amostra	Permeab. horizontal (mD)	Permeab. vertical (mD)	H/V	Estrutura	Facies	N Coordi.	E Coord.	Z Coord (m)
SG-I-31	11,896	-	-	Cruzada acanalada	Arenito médio	3°55'2.39"S	41°14'26.51"O	324,006
SG-I-32	-	-	-	Cruzada cavalgante	Arenito médio	3°55'2.39"S	41°14'26.51"O	324,006
SG-I-33	-	10,458	-	Cruzada baixo ang.	Arenito médio	3°55'2.35"S	41°14'26.20"O	324,4832
SG-I-34	1,669	21,931	0,076	Cruzada baixo ang.	Arenito médio	3°55'2.35"S	41°14'26.20"O	324,4832
SG-I-35	6,248	31,251	0,200	Cruzada acanalada	Arenito médio	3°55'2.35"S	41°14'26.20"O	324,4832
SG-I-36	1,249	-	-	Cruzada acanalada	Arenito médio	3°55'2.35"S	41°14'26.20"O	324,4832
SG-I-37	-	-	-	Cruzada cavalgante	Arenito médio	3°55'2.35"S	41°14'26.20"O	324,4832
SG-I-38	-	-	-	Cruzada cavalgante	Arenito médio	3°55'2.35"S	41°14'26.20"O	324,4832
SG-I-39	-	-	-	Cruzada cavalgante	Arenito médio	3°55'2.35"S	41°14'26.20"O	324,4832
SG-I-40	7,827	28,374	0,276	Cruzada baixo ang.	Arenito médio	3°55'2.37"S	41°14'25.98"O	324,4464
SG-I-41	2,378	-	-	Climbing ripples	Arenito médio	3°55'2.37"S	41°14'25.98"O	324,4464

Tabela 4: Permeabilidade (mD) das amostras do ponto de estudos 5 (SG-V) da bacia do Parnaíba

<b>Dados de permeabilidade das amostras do grupo Serra Grande</b>								
Amostra	Permeab. horizontal (mD)	Permeab. vertical (mD)	H/V	Estrutura	Facies	N Coordi.	E Coord.	Z Coord (m)
SG-V-01	34,420	2,705	-	Cruzada acanalada	ARN fino a médio	3°55'6.16"S	41°13'56.20"O	373,07
SG-V-03	0,334	0,334	1,000	Cruzada acanalada	ARN fino a médio	3°55'6.19"S	41°13'56.05"O	374,43
SG-V-04	0,635	0,723	0,879	Cruzada acanalada	ARN fino a médio	3°55'6.16"S	41°13'56.11"O	372,46
SG-V-05	0,418	0,935	0,447	maciça	ARN fino a médio	3°55'6.12"S	41°13'56.22"O	374,59
SG-V-07	3,077	24,945	0,123	maciça	ARN fino a médio	3°55'6.25"S	41°13'56.00"O	372,91
SG-V-08	0,771	0,577	1,336	maciça	ARN fino a médio	3°55'6.26"S	41°13'56.01"O	373,44
SG-V-09	0,199	7,827	0,025	maciça	ARN fino a médio	3°55'6.26"S	41°13'56.01"O	374,74



Tabela 5: Modelo de descrição de fácies segundo critérios de Miall (1996).

Nomeclatura Miall, 1996	Litofácies	Descrição	Estrutura
St	Arenito médio	Arenito médio, cinza, mal selecionado, muito cimentado	Estratificação cruzada acanalada
Sp	Arenito fino a médio	Arenito fino a médio, cinza, mal selecionado, muito cimentado	Estratificação cruzada tabular de baixo ângulo
Sr	Arenito fino a médio	Arenito fino a médio, castanho, texturalmente maduro, muito cimentado	Climbing Ripples

As interpretações aqui apresentadas não representam um modelo faciológico, pois ainda serão submetidas a um estágio de revisão criteriosa por um geólogo/estratigrafo e, portanto, serão modificadas e aprimoradas para a versão final da tese de doutorado. O motivo de constarem neste texto é apenas informar a banca sobre o andamento dos dados adquiridos em campo.

#### Descrição e Dados das Atividades Realizadas em Campo

As atividades de Campo envolveram os estudos de 6 afloramentos, aqui informalmente numerados como afloramentos 1 a 6, sendo os afloramentos 1, 2 e 5 pertencentes à Formação Ipu e os numerados como 3, 4 e 6 pertencentes à Formação Jaicós.

O afloramento 1 (AF1) mostrado na figura 4.2 juntamente com seu modelo interpretado de facies foi localizado próximo à divisa do Piauí com o Ceará na BR 220 e foi descrito como arenitos médios com eventuais clastos grosseiros, bastante cimentado, mal selecionado com estruturas tabulares na parte inferior atribuídas a forma de leito longitudinais. Na parte superior observou-se estruturas acanaladas atribuídas a barras e canais.

No afloramento 2 localizado na mesma rodovia BR 220, cerca de 1km de distancia do AF1, foram descritos arenitos médios da formação Jaicós, com boa seleção e estratificações cruzadas tabulares e acanaladas.

No AF3, localizado na ladeira da Lapa, a cerca de 5km do município de São Benedito - CE foram descritos arenitos médios com estratificação cruzada acanalada associadas a barras e canais.

No AF4, a cerca de 400m de distância do AF3 foram descritos arenitos finos a médios, arcózeos da formação Jaicós. Já o AF5, a cerca de 15km do município de



Ipueiras – CE, forma descritos arenitos médios a grossos, com clastos subangulares e pouco conteúdo de argila, logo abaixo de uma espessa camada (15m) de diamectitos.

No afloramento 6, localizado no sangrador da barragem do rio Guaíba, na região da cidade de Picos, foi possível amostras e descrever arenitos finos a médios com intercalações esparsas de siltito com boa seleção e estratificações cruzadas, por vezes, compostas.



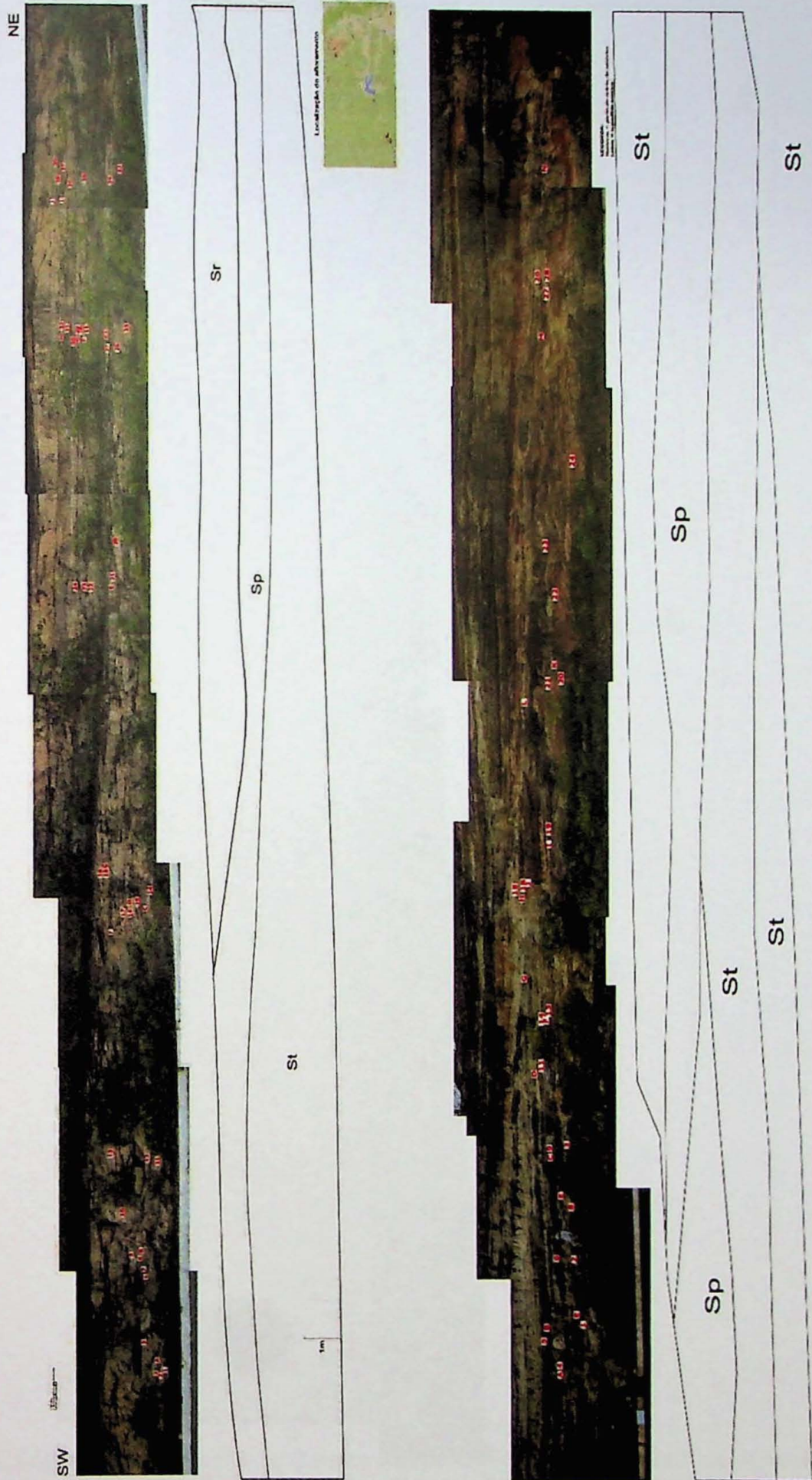


Figura 4.2: Fotomosaicos e modelos de sistemas deposicionais dos afloramentos AF1 e AF2 da formação Ipú.



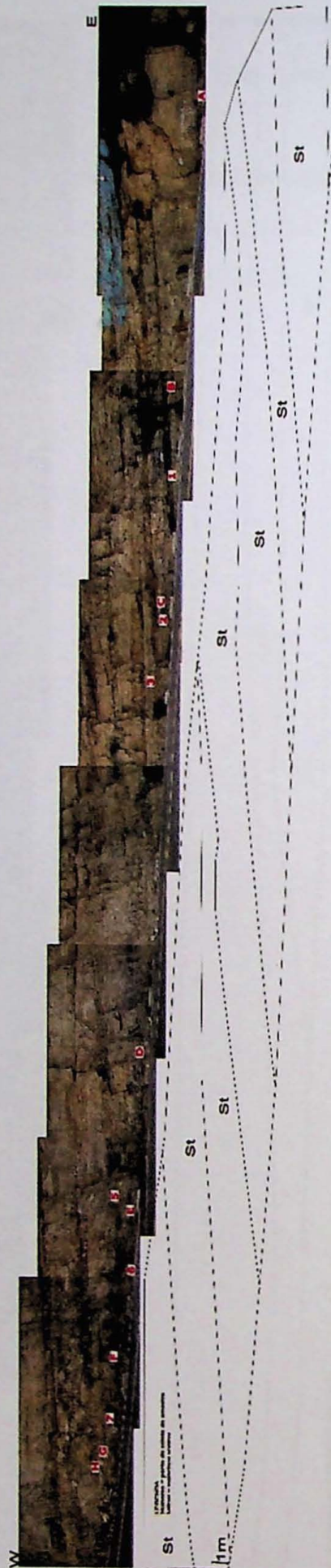


Figura 4.3: Fotomosaicos e modelos de sistemas deposicionais dos afloramentos AF3 e AF4 da formação Ipú.



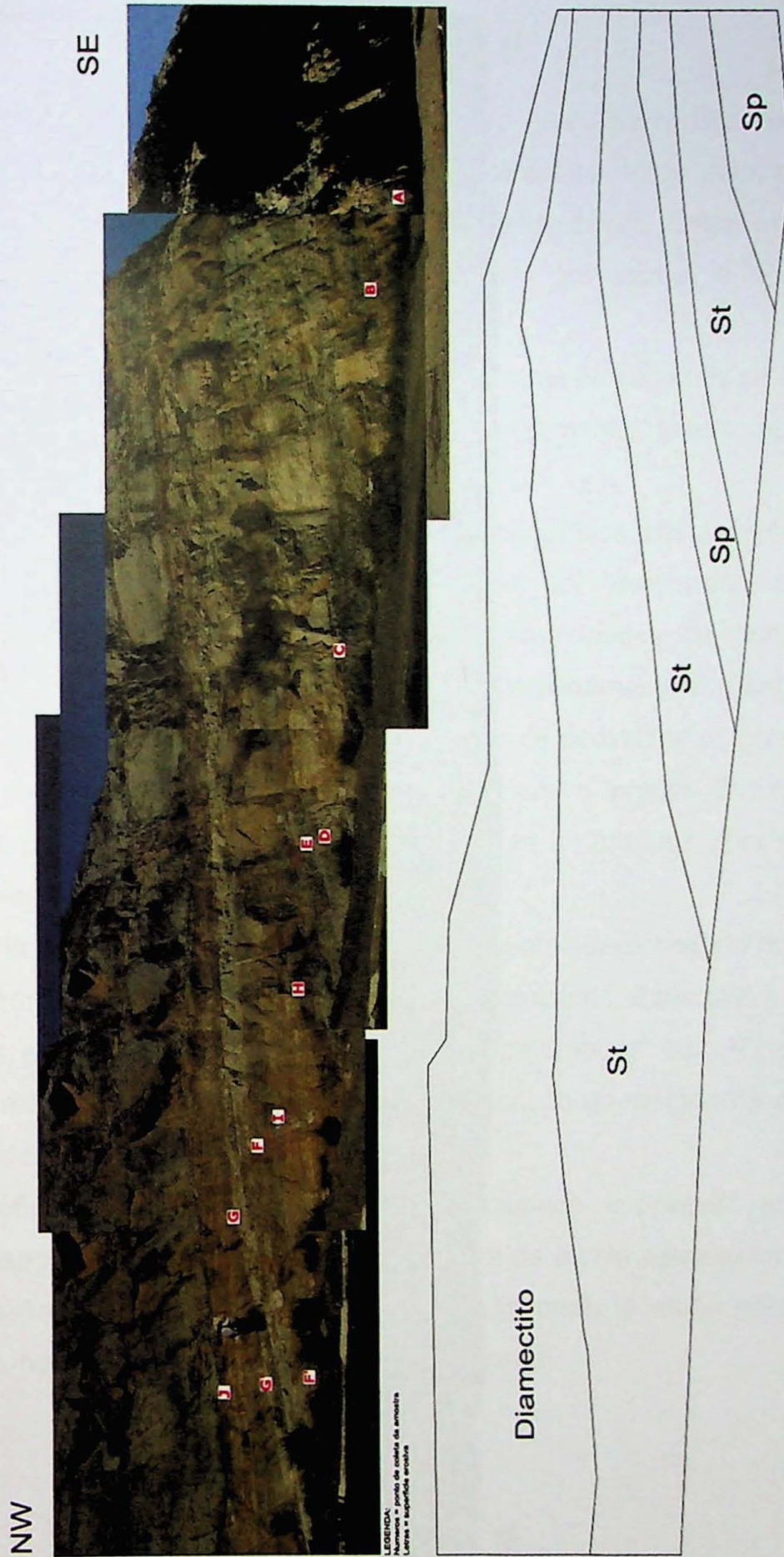


Figura 4.4: Fotomosaico e modelo de sistema deposicional do afloramento AF5 da formação Ipú.





Figura 4.5: Fotomosaico e modelo de sistema deposicional do afloramento AF6 da formação Jaicós.



## *Petrofísica Básica*

Um primeiro lote de amostras coletadas no grupo Serra Grande foi enviada para laboratório com objetivo de fazer a petrofísica básica, ou seja, determinar a porosidade, permeabilidade, densidade e volume de poros. Estes ensaios foram realizados com o uso de um permo-porosímetro profissional e operado pelos petrofísicos da empresa Weatherford.

Os demais lotes das amostras coletadas durante os trabalhos de campo estão aguardando verba para a plugagem e posteriormente serem enviados aos laboratórios parceiros para realização dos ensaios.

A tabela 6 apresenta os resultados de petrofísica básica realizados pelo laboratório da Weatherford. Nesta tabela podem ser observados os dados de porosidade e de permeabilidade, que foram determinados de duas formas, a primeira com a amostra submetida a pressão de confinamento e a segunda com a pressão ambiente. A tabela mostra ainda os dados de densidade de grão, volume de poros e as medidas de cada plug, como peso, altura e largura. Em cada amostra foram feitas plugs verticais e horizontais de modo a contribuir com a análise da anisotropia regional.

A partir da tabela 5 foi plotado o gráfico da porosidade medido com a pressão de confinamento pela porosidade determinada à ambiente. Esse plot é apresentado no gráfico 1 e mostra a perfeita correlação dos parâmetros, com  $R^2$  da ordem de 95%. Esse resultado comprova eficiência da aplicação do método na determinação da porosidade.

O gráfico 2 apresenta um plot da porosidade à pressão ambiente pela permeabilidade ar, na qual foi feita uma tentativa de ajuste exponencial dos dados, porém o parâmetro  $R^2$  de 0,02 ilustra que a dispersão é muito grande entre os dados e não há correlação com esse espaço amostral.



Tabela 6 Resultado da análise de petrofísica básica

Amostra	Pres conf	Permeabilidade (mD)		Porosidade (%)		Dens. Grão	Vol. Poro	Altura,	Diâm.	Peso,
	psi	ao Ar	Klink	Amb	NCS	gm/cc	cc	cm	cm	g
PSG-09-A1 V	500	3,66	2,9	10,1	9,6	2,64	2,45	5,363	2,464	60,640
PSG-09-A1 H	500	13,8	11,3	12,4	12,2	2,63	2,13	3,734	2,442	40,298
PSG-09-A3 H1	500	5,97	4,56	9,8	9,4	2,63	0,88	1,970	2,463	22,282
PSG-09-A3 H2	500	4,95	3,99	9,9	9,3	2,63	2,76	6,174	2,483	70,807
PSG-09-A3 V	500	5,55	4,32	9,8	9,2	2,64	1,52	3,531	2,450	39,609
PSG-09-A5 H1	500	11,80	9,69	9,9	9,3	2,62	2,17	4,952	2,458	55,420
PSG-09-A5 H2	500	35,5	30,5	9,9	9,4	2,62	2,52	5,658	2,459	63,473
PSG-09-A5 V	500	7,59	6,10	9,3	8,6	2,62	2,06	5,036	2,470	57,302
PSG-09-A10 V	500	25,4	21,5	12,5	11,8	2,62	2,77	4,915	2,471	54,123
PSG-09-A11 V	500	0,214	0,155	15,5	15,1	2,65	3,02	4,258	2,455	45,166
PSG-09-A11 H1	500	4,86	3,78	15,9	15,0	2,64	2,53	3,665	2,438	37,952
PSG-09-A12 V	500	0,018	0,009	9,0	8,4	2,69	1,91	4,792	2,461	55,779
PSG-09-A12 H1	500	17,2	14,3	12,8	12,2	2,64	2,79	4,808	2,474	53,138
PSG-09-A13 V	500	0,174	0,124	7,1	6,6	2,63	1,36	4,281	2,478	50,458
PSG-09-A13 H1	500	0,202	0,146	7,2	6,7	2,63	1,97	6,147	2,478	72,426
PSG-09-A14 V	500	6,20	4,82	16,8	16,2	2,65	2,24	2,934	2,451	30,502
PSG-20A V	500	0,035	0,020	6,3	6,0	2,64	0,82	2,927	2,452	34,159
PSG-20A H1	500	0,294	0,219	7,3	7,0	2,64	1,09	3,260	2,469	38,190
PSG-20B H1	500	0,0120	0,0055	4,3	3,7	2,64	0,44	2,523	2,458	30,236
PSG-23A V	500	213,	195,	13,9	13,2	2,63	3,92	6,325	2,456	67,937
PSG-23A H1	500	174,	158,	13,2	12,8	2,64	2,05	3,446	2,439	36,922
PSG-23B V	500	1830,	1760,	21,3	20,4	2,64	2,03	2,235	2,392	20,855

Dada a anisotropia do sistema é necessária uma quantidade significativamente maior de dados para estabelecer alguma correlação. Uma tentativa semelhante foi realizada no plot da porosidade à pressão ambiente em função da distribuição da densidade de grãos e é apresentada no gráfico 3. Novamente observamos a necessidade de se obter mais dados para melhorar estatisticamente nosso modelo de correlação.



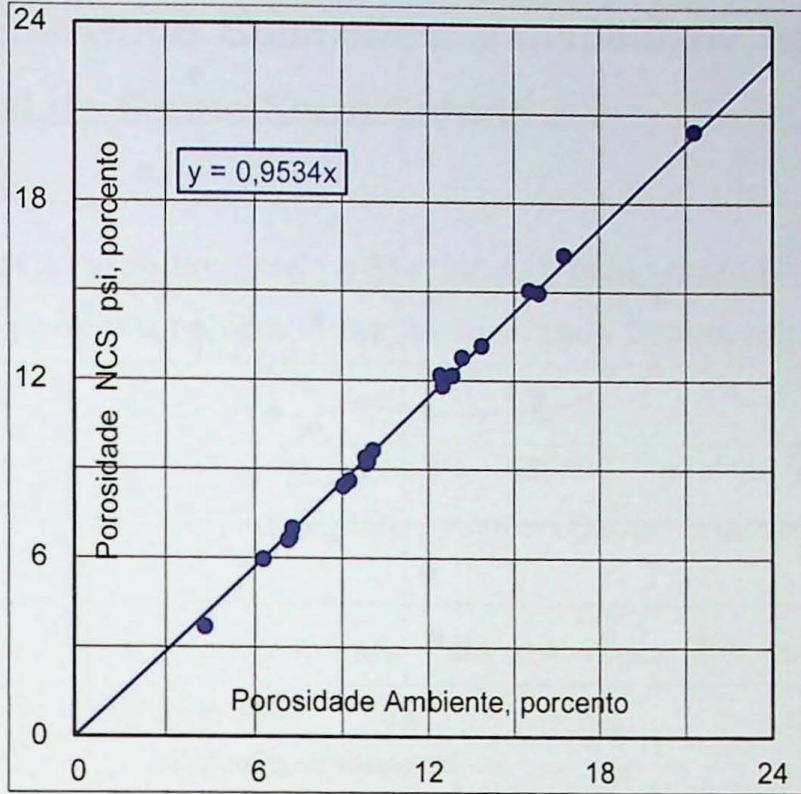


Gráfico 1: Porosidade da pressão de confinamento pela porosidade à pressão ambiente.

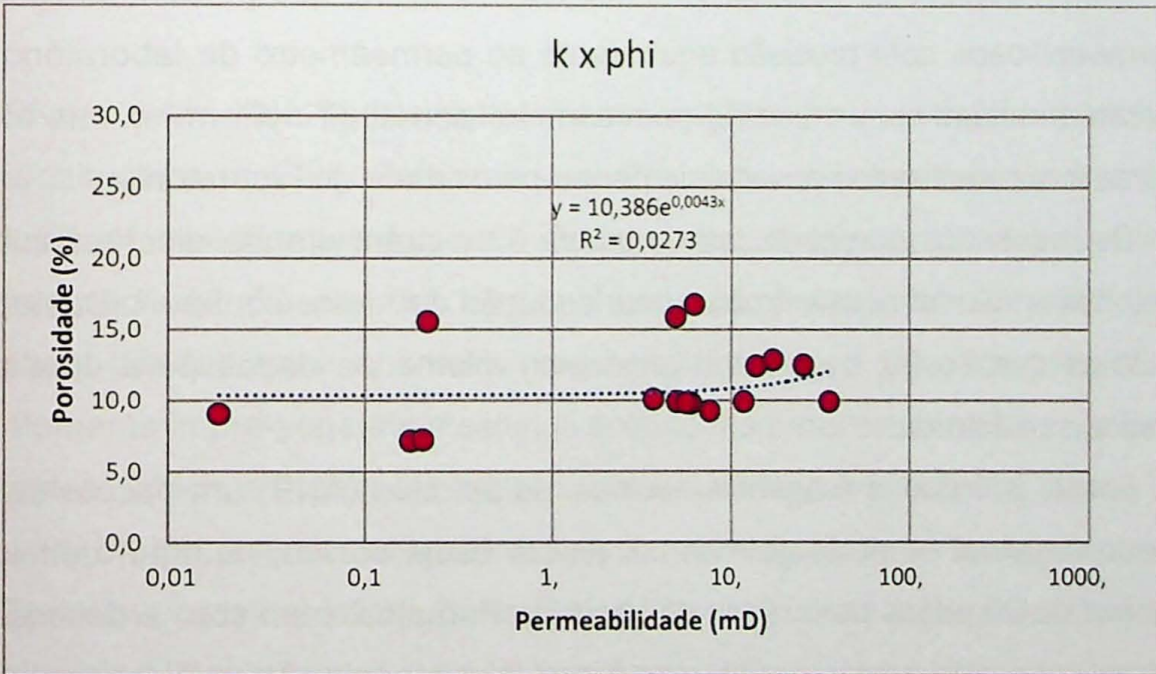


Gráfico 2: Porosidade à pressão ambiente pela permeabilidade ar.



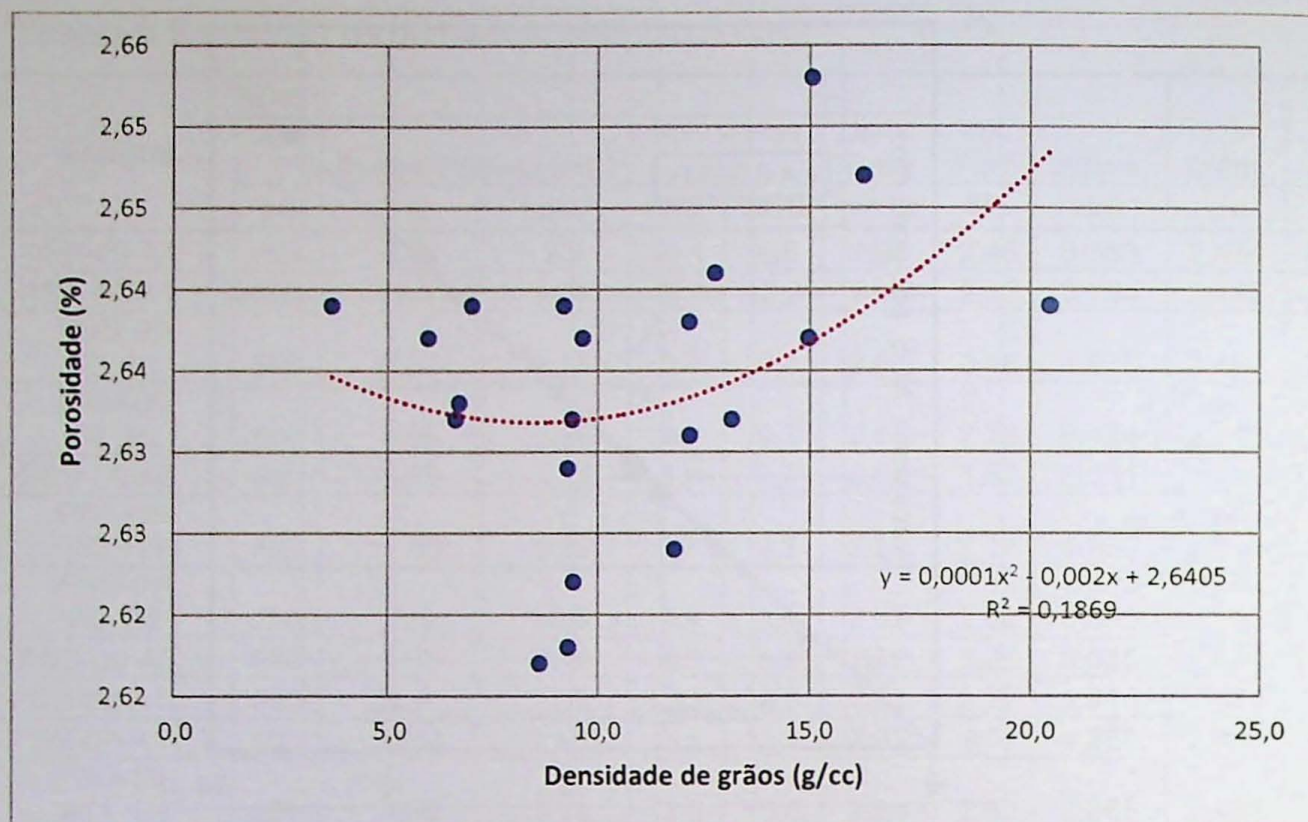


Gráfico 3: Porosidade à pressão ambiente em função da distribuição da densidade de grãos.

Com o objetivo de se obter uma equação de ajuste que permita inferir valores de permeabilidade com precisão equivalente ao permeômetro de laboratório foram realizadas medidas com o uso do permeômetro portátil (TinyPerm) e nas mesmas amostras foram realizados ensaios de permo-porosidade de laboratório.

Os resultados apresentados na tabela 6 mostram grande discrepância entre os métodos e não foi possível obter uma equação de ajuste confiável. Desta forma, concluímos que se faz necessário um maior volume de dados para uma análise estatística satisfatória.

Foram solicitadas à agência nacional de petróleo (ANP) um pacote de dados contendo registros de perfilagem como, resistividade, sônico, neutrão, densidade e raio gama de 20 poços perfurados na bacia do Parnaíba, bem com a descrição das amostras coletadas nestes poços e análises laboratoriais. Além disso, solicitamos cerca de 68 linhas sísmicas distribuídas pela bacia.



## **Anexo B – Artigo Submetido: Variabilidade da Arquitetura Fluvial do Grupo Serra Grande**

**Variability of fluvial architecture in a Silurian sag basin margin: sheet-braided and meandering ancestor rivers in the Serra Grande Group, northeastern Brazil**

Liliane Janikian, Renato Paes de Almeida, Cristiano Padalino Galeazzi, Larissa Natsumi Tamura, Julio Cesar Ardito, Marley Antonio Carrari Chamani

**Keywords:** Meandering ancestors; Early Paleozoic rivers; Bar surface orientation; Cliniform modeling; Paleocurrents.

**SHORT TITLE:** Fluvial architecture in a Silurian sag basin

### **Abstract**

Rivers with little to no influence of vegetation and their implications for the architecture of Early Paleozoic and Precambrian fluvial deposits are an important topic of investigation for both clastic sedimentology and Earth System evolution perspectives. Despite the long-held concept of a dominant style of wide, shallow sheet-braided sandy channels, growing evidence for a great variability of architectural elements in pre-vegetation settings point to the predictable occurrence of other fluvial styles, which are possibly related to areas of low-slope alluvial plains. The present work brings sedimentological descriptions and fluvial style interpretations of Silurian fluvial successions of the Serra Grande Group in northeastern Brazil, developed in a glaciated basin margin (therefore with little to no effect of vegetation). The comparison of the resulting channel body architectures integrated to paleocurrent data and a diffusion-based numerical model enabled the evaluation of the controls on fluvial architecture variability in unvegetated alluvial plains.



## Introduction

The evolution of land plants through the Phanerozoic had major impacts on the dynamics and architecture of fluvial systems, recorded as changes in the proportion of preserved fluvial styles through time (Schumm, 1969; Cotter, 1978; Davies *et al.*, 2011; Gibling *et al.*, 2014; Ielpi *et al.*, 2018). In that sense, the first appearance of vascular plant fossils in the Silurian (e.g. Kenrick and Crane, 1997; Steemans *et al.*, 2014) correlates with a marked decrease of the sheet-braided style fluvial architecture, characterized by relatively thin and wide sediment bodies interpreted as channels much shallower and wider than the latter dominant meandering and braided streams (Cotter, 1978; Long, 1978; Fuller, 1985; Davies and Gibling, 2010; Davies *et al.*, 2011; Birdland *et al.*, 2014). Accordingly, the frequency of preserved meandering river deposits seems to increase from the Silurian to the present, in parallel to the adaptive evolution of land plants to new environments (Fuller, 1985; Davies *et al.*, 2011; Gibling and Davies, 2012; Gibling *et al.*, 2014; McMahon and Davies, 2018).

In fact, there is great variability in preserved fluvial architecture even prior to the land colonization by plants, with the local preservation of all common modern architectural elements in Precambrian successions (Long, 2011), including the presence of point-bar deposits in pre-vegetation strata (Casshyap, 1968; Button and Tyler, 1981; Sweet, 1988; Santos *et al.*, 2014; Ielpi and Rainbird, 2015). Variations in valley slope may be the origin of such variability (Almeida *et al.*, 2016a, Ganti *et al.*, 2019). Distal, low-slope alluvial plains, developed in conditions corresponding to those of modern meandering channels, are predicted to contain braided rivers with deeper channels, finer-grained bedload and greater preservation of overbank deposits. These distal systems coeval to the proximal sheet-braided pre-vegetation rivers are termed “meandering ancestors” (Almeida *et al.*, 2016a).

The present work documents the architecture of channelized bodies developed at different valley slopes along longitudinal profiles of coeval early-Silurian fluvial successions of the Serra Grande Group in northeastern Brazil. At that point of land-plant evolution (Vegetation Stage 3 of Davies and Gibling, 2010), fluvial systems were not yet sufficiently affected by root systems to display any change in architecture relatively to Precambrian rivers (Davis and Gibling, 2010). In the



Parnaíba Basin eastern margin, fluvial successions punctuated by a marine transgression marker bed enable the reconstruction of proximal-to-distal fluvial paleogeographies and the description of ancient low-slope alluvial deposits developed before vegetation-induced bank stabilization made meandering rivers dominant in these settings.

## Study sites and methods

The Parnaíba Basin (Figure 1) is a cratonic sag basin with sedimentary and minor volcanic successions spanning from the Silurian to the Cretaceous (Vaz *et al.*, 2007). The Serra Grande Group is the lowermost unit of the basin fill succession, overlying precambrian igneous and metamorphic basement rocks (Pedreira *et al.*, 2003). At the basin eastern margin, the group is composed of the Ipu, Tianguá and Jaicós formations (Caputo and Lima, 1984). The Ipu Formation consists of alluvial conglomerate and sandstone, locally bearing diamictite lenses interpreted as a result of glacial advance in a high-latitude setting when Gondwana was placed close to the South Pole (Caputo and Lima, 1984). These alluvial deposits grade into marine shales and, locally, to coastal sandstone of the Tianguá Formation, which marks a major transgressive event and sets a stratigraphic datum for regional correlation. Chitinozoans and miospores reveal a Llandovery age (late Aeronian to Early Thelychian) for the uppermost Ipu, Tianguá and lowermost Jaicós formations (Grahm *et al.*, 2005), thus enabling a precise age determination for the here described successions. The upper Jaicós Formation, above the investigated strata, is attributable to early Devonian (latest Pragian-earliest Emsian, Grahm *et al.*, 2005) biozones. The contact of the Tianguá Formation with the overlying fluvio-eolian Jaicós Formation sandstone and conglomerate is concordant but abrupt. The Serra Grande Group is capped by a major unconformity, which floors the overlying a Devonian deposits.

The Serra Grande Group is extensively exposed along the escarpment that marks the Parnaíba Basin eastern margin. The present study focuses on three main areas (Figure 1): the northeasternmost exposures of the Serra Grande Group (Area I), in an approximately east-west, 100 km long section linking the municipalities of São João da Fronteira (Area Ia), Tianguá (Area Ib); a second area further south, near



Ipueiras (Area II); and a third area near the municipality of Picos, in the southeastern border of the Parnaíba Basin (Area III). Facies and architectural elements analyses (Miall, 1996) were performed to interpret the spatial and temporal variability in fluvial styles. Systematic paleocurrent data collection of decimeter and centimeter-scale cross-strata sets indicating flow direction and sense were performed through the direct measurement of exposed planes and the reconstruction of planes from two non-parallel lines exposed at outcrop faces. Paleocurrent data were integrated with measurements of cross-strata set bounding surfaces and used to reconstruct bar surface orientation (see Almeida *et al.*, 2016b). Stratigraphic sections were correlated using the Tianguá Formation maximum flooding surface as the regional datum.

Interpretation of the controls on the spatial and temporal variability of fluvial styles were tested with the help of diffusion-based models of river profiles and coastal-marine clinoforms (Paola, 2000). The model calculates sediment surface elevation in a long profile through time, considering local deposition and erosion as a function of the downstream variation in slope. In the model, the volume of sediment imputed at each time step is kept constant. The diffusion coefficient is a function of local discharge, which increases downstream proportionally to catchment area increase, thus leading to a concave-up profile. Clinoforms are the result of attributing a different, much lower, diffusion coefficient to the marine area. Sea-level vertical position varies according to a simplified hypothetical curve given by a constant slow subsidence (0.07 m/ka) and one short-term pulse of eustatic rise at 1.2 m/ka. Different scenarios were built to represent deposition in Areas I and II. Temporal evolution of local slope was recorded and compared to grain-size and architecture observed in the actual stratigraphic sections.

## Fluvial styles and their spatial distribution

Descriptions of sedimentary facies, their spatial organization and the geometry of the sediment bodies in the studied successions lead to the description of eight architectural elements, and the recognition of three fluvial facies associations: *Proximal sheet-braided rivers*, *Proximal high-slope rivers*, *Distal low-slope rivers*, and a coastal-marine facies association (Table 1). Regarding the fluvial deposits, the Ipu



Formation presents *Proximal sheet-braided rivers* in Area II, *Proximal high-slope rivers*, in Area Ia, and *Distal low-slope rivers*, in Area Ib (Figure 2a); whereas the Jaicós Formation presents *Distal low-slope rivers* association in Areas I, II and III (Figure 2b). The combination of paleocurrent measurements and bar surface reconstructions in the fluvial deposits allowed the reconstruction of the fluvial systems in the Ipu and Jaicós formations. Reconstructions of bar surfaces considering pairs of cross strata and their bounding surfaces (see Almeida *et al.*, 2016b) reveal azimuth differences between mean reconstructed bar surfaces and current direction varying from 3° to 56°, thus pointing to downstream accretion direction for all measured bar deposits (Figure 2c).

The Ipu Formation records two coeval, laterally equivalent fluvial systems in Areas I and II, with paleocurrents towards west-northwest, perpendicular to the regional isopachs and to the basin margin. The deposits in these two areas differ in grain size and depositional architecture, suggesting different catchment areas, sediment fluxes and fluvial styles. In Area II, the fluvial deposits are dominated by pebbly coarse sandstone and a typical sheet-braided architecture (Figure 3a), with local intercalation of glacial deposits (Figure 3b), suggesting deposition close to the basin margin and/or high alluvial plain. These glacial deposits are overlain by glacial-marine fine-grained deposits of the Tianguá Formation (Figure 3c), contrasting with the coastal sandstones with wave-generated structures and both *Cruziana* and *Skolithos* ichnofacies (Figure 3d, 3e) directly overlying fluvial deposits in Area I. In Area I, a few tens of kilometers to the north, fluvial deposits reveal the passage from proximal deposits in Area Ia, dominated by *Proximal high-slope rivers* (Figure 4), to more distal deposits in Area Ib, dominated by *Distal low-slope rivers*. The former is characterized by several meters of fining-upward cycles dominated by the deposits of downstream-migrating unit and compound bars, including discrete channelized bodies and abundant lenses of cross-bedded sandstones (Figure 5). Such architecture indicates deeper channels (less than 1:20 depth-to-width ratio in preserved sediment bodies), with larger discharges than the typical sheet-braided architecture of Area I, most probably related to larger catchment areas.

The coastal-marine deposits of the Tianguá Formation separate the Ipu from the Jaicós formations. Paleocurrent data from both the overlying and underlying fluvial successions indicate a marine basin to the west, and therefore a marine transgression from west to east. Nevertheless, abundance of deeper water deposits



to the south suggests an irregular coast, in agreement with the evidence of a larger fluvial system to the north, capable of enough sediment to build aggradational coastal deposits correlated to the marine strata further south.

The deposits of the Jaicós Formation are marked by decimeter to meter-scale fine sediments flanking several meters deep channels containing downstream-migrating unit and compound bars (*Distal low-slope rivers* association) in all areas of exposure (Figure 7 and Figure 8), characterizing the re-establishment of fluvial depositional environments in the eastern border of the basin after the post-glacial transgression. As in the case of the Ipu Formation in Area Ib, these features characterize the “meandering ancestor” style (Almeida *et al.*, 2016a).

## Discussion and conclusions

Considering the hypothesis of alluvial plain slope as the main control on the development and preservation of the wide and shallow channels of the sheet-braided pre-vegetation style (Almeida *et al.*, 2016a, Ganti *et al.*, 2019), the observed spatial variation of fluvial architecture might be explained as the result of different positions on a river long profile and different long profiles in coeval river systems (Figure 8). The shallower channels and coarser grain size of the fluvial deposits in Area II indicate a smaller catchment area compared to Area I, implying near sources that enabled the advance of glaciers from high basement areas into the basin and consequently higher slopes in the alluvial plain and smaller sediment fluxes. The latter is expected in smaller catchment basins (Syvitski and Milliman, 2007) and caused a major transgression in the event of post-glacial eustatic rise, flooding the coastal alluvial plain with fully marine deposits. In contrast, grain size, depositional architecture and paleocurrent directions in Area I suggest river systems with a much larger catchment area, a more distal fluvial system displaying larger and deeper channels, causing the post-glacial transgressive event to be recorded as aggradational coastal deposits due to the higher sediment flux.

Diffusion-based river profile and coastal-marine clinoform models for two scenarios were built to simulate the larger river system in Area I (Figure 9a) and the coeval smaller system in Area II (Figure 9b). The results reveal that the observed stratigraphic sections can be simulated with constant subsidence and one episode of



increased accommodation due to eustatic rise. Differences in sediment input between the systems can be the cause of fully marine sediments in the Tianguá Formation in Area II, represented at position P2 in Figure 9b, coeval to coastal deposits in Area I, represented at position P1 in Figure 9a. Additionally, the temporal evolution of local slopes of fluvial deposits at P1 and P2 reveals that high alluvial-plain slopes are expected at the top of the Ipu Formation in Area II (red line from 250 to 500 ka in Figure 9c), whereas much gentler slopes ( $1$  to  $2 \times 10^{-4}$ ) are expected in the coeval section in Area I (red line from 250 to 500 ka in Figure 9d), matching the inferences of Ganti *et al.* (2019) for similar systems. On the other hand, progradational fluvial systems after the Tianguá transgression display low slopes on both areas (red lines after 700 ka in Figure 9c, 9d). These predictions are in clear agreement with the observed fluvial architecture, with dominant sheet-braided deposits at the top of the Ipu Formation in Area II, most likely due to greater slopes, and the abundance of the meandering-ancestor style in the coeval Ipu Formation deposits in Area I and at the lower Jaicós Formation in both areas.

The differences in fluvial architecture between deposits of Area Ia and Area Ib are crucial to understand the effective controls on the variability of fluvial architecture. Since Area Ib is sited downstream of Area Ia, the fluvial architecture would be controlled by lower alluvial-plain slopes. Abundant deep-channel elements and preservation of fine-grained overbank and bar-top deposits related to downstream-accreting bar elements. Such documentation clearly departs from the dominant sheet-braided style for pre-vegetational settings described in the literature. Meandering river ancestors are found in the areas predicted by the numerical model results: distal environments of a larger river system (Area Ib) in the Ipu Formation and throughout the whole area in the Jaicós Formation, as the progradation of alluvial deposits after the transgressive event resulted in low-slope fluvial deposits at the observed positions for both the large river in Area I and the small river in Area II.

These findings corroborate the idea that a particular fluvial style might have populated the low-slope areas in conditions that are today dominated by meandering rivers, but that lacked the bank stabilizing effects of vegetation in areas not yet occupied by land plants as they evolved since the Silurian. Additionally, our results show that the recognized greater variability of fluvial styles than the monotonic sheet-braided model for pre-vegetation fluvial deposits might result from variations in slope and discharge in different fluvial systems and accommodation-infill histories. The



abundance of exposures of the “meandering ancestor” style, makes the study outcrops of the Ipu and Jaicós formations good candidates for reference sections of that peculiar fluvial style.

### Acknowledgments

The authors were funded by the São Paulo Research Foundation (FAPESP) through Research Grants #2016/03091-5. We are thankful to CNPq for scholarship of L. Janikian (#306513/2016-1) and for students scholarships: FAPESP (#2017-06874-3), CNPq (#142276/2015-4) and CAPES-DF (demanda social).

### References

- Almeida, R.P.A., Marconato, A., Freitas, B.T. and Turra, B.B., 2016a. The ancestors of meandering rivers. *Geology*, **44**, 203-206.
- Almeida, R.P.A., Freitas, B.T., Turra, B.B., Figueiredo, F.T., Marconato, A. and Janikian, L., 2016b. Reconstructing fluvial bar surfaces from compound cross-strata and the interpretation of bar accretion direction in large river deposits. *Sedimentology*, **63**, 609-628.
- Birdland, D.R., Bennett, J.A., McVicar-Wright, S.E. and Scrivener, R.C., 2014. Rivers through geological time: the fluvial contribution to understanding of our planet. *P. Geologist Assoc.*, **125**, 503-510.
- Button, A. and Tyler, N., 1981. The character and economic significance of Precambrian paleoweathering and erosion surfaces in Southern Africa. *Economic Geology 75th Anniversary Volume*, 686-709.
- Caputo, M.V. and Lima, E.C., 1984. Estratigrafia, idade e correlação do Grupo Serra Grande - Bacia do Parnaíba. *Anais do 33° Congresso Brasileiro de Geologia*. Rio de Janeiro, SBG, 740-753.



Casshyap, S.M., 1968. Huronian stratigraphy and paleocurrent analysis in the Espanola-Willisville area, Sudbury District, Ontario, Canada. *J. Sediment. Petrol.*, **38**, 920-942.

Cotter, E., 1978. The evolution of fluvial style, with special reference to the central Appalachian Paleozoic. In: *Fluvial Sedimentology* (A.D. Miall, ed), Cambridge University Press, United Kingdom.

CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. *Machine learning*, v. 20, n. 3, p. 273-297, 1995.

Cunha, F.M.B., 1986. Evolução paleozóica da Bacia do Parnaíba e seu arcabouço tectônico. Rio de Janeiro, UFRJ, master thesis, 118 pp.

Davies, N.S. and Gibling, M.R., 2010. Cambrian to Devonian evolution alluvial system: the sedimentological impact of the earliest land plants. *Earth Sci. Rev.*, **98**, 171-200.

Davies, N.S., Gibling, M.R. and Rygel, M.C., 2011. Alluvial facies evolution during the Paleozoic greening of the continents: case studies, conceptual models and modern analogues. *Sedimentology*, **58**, 220-258.

Fuller, A.O., 1985. A contribution to the conceptual modelling of pre-Devonian fluvial systems. *S. Afri. J. Geol.*, **88**, 189-194.

Ganti, V., Whittaker, A.C., Lamb, M.P., Fischer, W.W., 2019. Low-gradient, single-threaded rivers prior to greening of the continents. *Proc. Natl. Acad. Sci. USA*, **116**, 11652-11657.

Gibling, M. & Davies, N.S., 2012. Paleozoic landscape shaped by plant evolution. *Nature Geoscience*, **5**, 99-105.



Gibling, M.R., Davies, N.S., Falcon-Lang, H.J., Bashforth, A.R., DiMichele, W.A., Rygel, M.C. and Ielpi, A., 2014. Paleozoic co-evolution of rivers and vegetation: A synthesis of current knowledge. *P. Geologist Assoc.*, **125**, 524-533.

Grahn, Y., de Melo, J.H.G., Steemans, P., 2005. Integrated Chitinozoan and Miospore zonation of the Serra grande Group (Silurian-Lower Devonian), Parnaíba Basin, Northeast Brazil. *Rev. Esp. Micropaleontol.*, **37**, 183-204.

Ielpi, A. and Rainbird, R.H., 2015. Architecture and morphodynamics of a 1-6 a Ga fluvial sandstone: Ellice Formation of Elu Basin, Arctic Canada. *Sedimentology*, **62**, 1950-1977.

Ielpi, A., Fralick, P., Ventra, D., Ghinassi, M., Lebeau, L.E. and Marconato, A., 2018. Fluvial floodplains prior to greening of the continents: Stratigraphic record, geodynamic setting, and modern analogues. *Sediment. Geol.*, **373**, 140-172.

Kenrick, P. and Crane, P.R., 1997. The origin and early evolution of plants on land. *Nature*, **389**, 33-39.

Long, D.G.F., 1978. Proterozoic stream deposits: Some problems of recognition and interpretation of ancient sandy fluvial systems. In: *Fluvial sedimentology* (A.D. Miall, ed.), Cambridge University Press, United Kingdom.

Long, D.G.F., 2011. Architecture and depositional style of fluvial systems before land plants: A comparison of Precambrian, early Paleozoic, and modern river deposits. In: Davidson SK et al. (Eds) *From river to rock record: The preservation of fluvial sediments and their subsequent interpretation*. *SEPM Spec. Publ.*, **97**, 37-61.

McMahon, W.J. and Davies, N.S., 2012. The shortage of geological evidence for pre-vegetation meandering rivers. In: *Fluvial Meanders and their sedimentary products in the rock record* (M. Guinassi, L. Colombera, N.P. Mountney, A.J.H. Reesink and R. Bateman, eds.). Wiley-Blackwell, Hoboken.



- Miall, A.D., 1996. *The Geology of Fluvial Deposits: Sedimentary Facies, Basin Analysis and Petroleum Geology*. Springer, Berlin, pp. 668.
- Paola, C., 2000. Quantitative models of sedimentary basin filling. *Sedimentology*, **47**, 121–178.
- Pedreira, A.J., Lopes, R.C., Vasconcelos, A.M. and Bahia, R.B.C., 2003. Bacias Sedimentares Paleozóicas e Meso-Cenozóicas Interiores. In: *Geologia, Tectônica e Recursos Minerais do Brasil* (L.A. Bizzi, C. Schobbenhaus, R.M. Vidotti, J.H. Gonçalves, eds.), Brasília, 55-85.
- Pemberton, S.G., Van Wagoner, J.C. and Wach, G.D., 1992. Ichnofacies of a wave-dominated shoreline. *SEPM Spec. Pub.*, **17**, 339-382.
- Santos, M.G.M., Almeida, R.P., Godinho, L.P.S., Marconato, A. and Mountney, N.P., 2014. Distinct styles of fluvial deposition in a Cambrian rift basin. *Sedimentology*, **61**, 881-914.
- Schumm, S.A., 1969. Speculations concerning the paleo-hydraulic controls of terrestrial sedimentation. *GSA Bull.*, **79**, 1573-1588.
- Stemans, P., Herissé, A., Melvin, J., Miller, M.A., Paris, F., Verniers, J. and Wellman, C.H., 2014. Origin and Radiation of the Earliest Vascular Land Plants. *Science*, **324**, 352.
- Sweet, I.P., 1988. Early Proterozoic stream deposits: Braided or meandering - Evidence from central Australia. *Sediment. Geol.*, **58**, 277-293.
- Syvitski, J.P.M. and Milliman, J.D., 2007. Geology, geography, and humans battle for dominance over the delivery of fluvial sediment to the coastal ocean. *J. Geol.*, **115**, 1-19.
- Vaz, P.T., Rezende, N.G.A.M., Wanderley Filho, J.R. and Travassos, W.A.S., 2007. Bacia do Parnaíba. *Bol. de Geoc. da Petrobrás*, **15**, 253-263.



Figures:

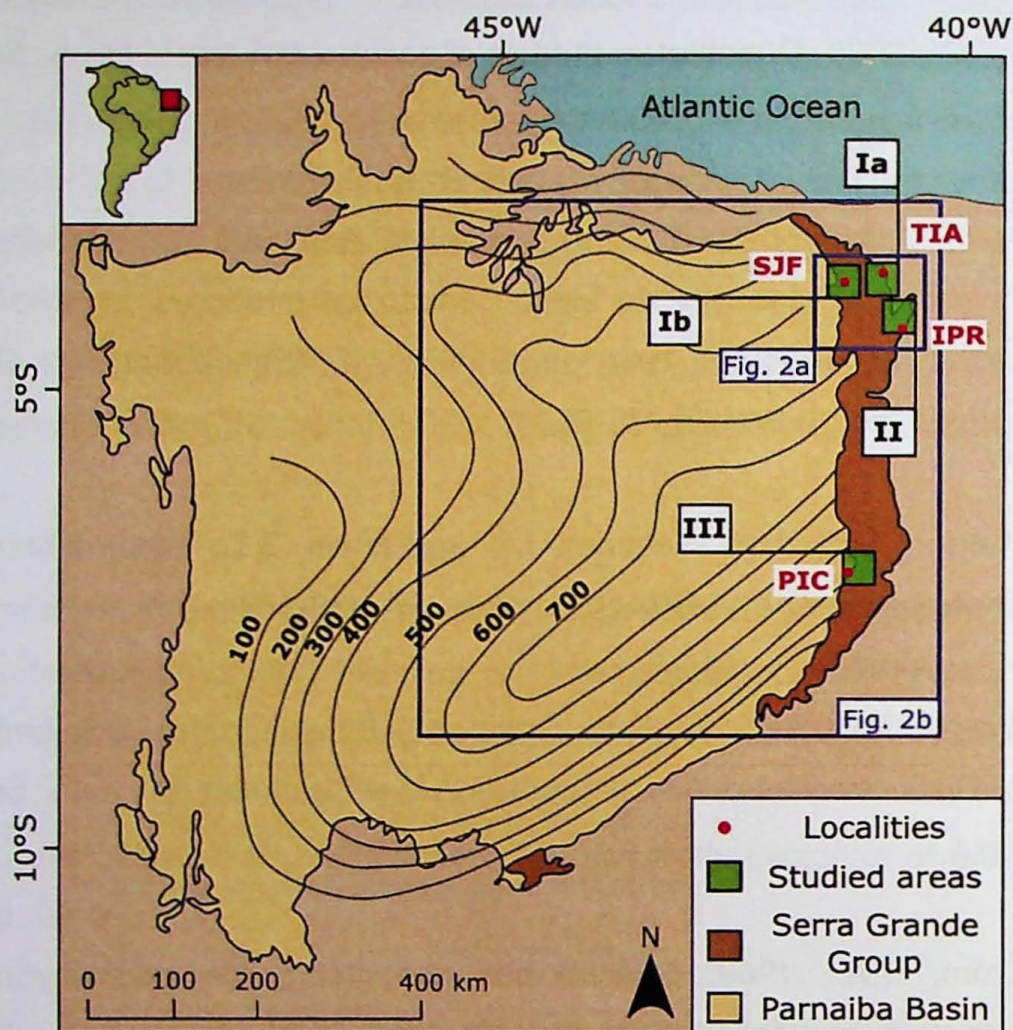


Figure 1: Location of studied areas on the eastern border of the Parnaíba Basin, displaying the area of exposure of the Silurian Serra Grande Group (in brown) and isopachs for the whole group based on Cunha (1986). Green boxes show the investigated areas. Red dots indicate the main localities for each area. SJF: São João da Fronteira; TIA: Tanguá; IPR: Ipueiras; PIC: Picos.



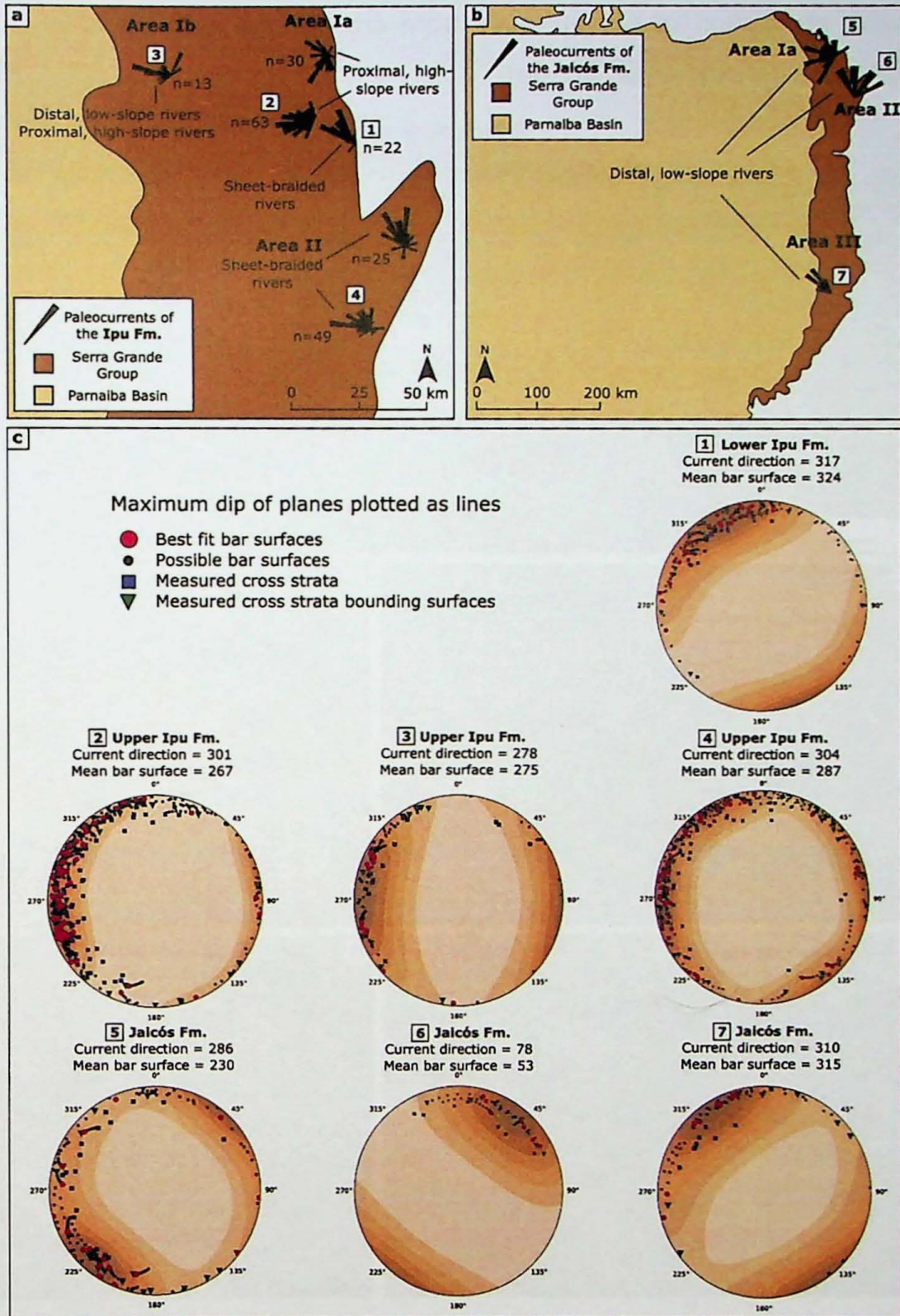


Figure 2: Paleocurrent maps for the Ipu Formation (a) and the Jaicós Formation (b), showing main flow direction towards the northeast, therefore orthogonal to the main isopach direction displayed in Figure 1. Note that Areas I and II are to the northwest of the depocenter, pointing to the progradation of fluvial clastic wedges while maintaining the northwest-dipping long profile; (c) Paleocurrents and reconstruction of bar surface orientation. Rose diagrams display azimuth of cross-strata. Lower-hemisphere equal-area stereoplots display maximum dip as lineations, reconstructing



mean orientation of bar surfaces from cross-strata and cross-strata set bounding surfaces (Almeida *et al.*, 2016b).

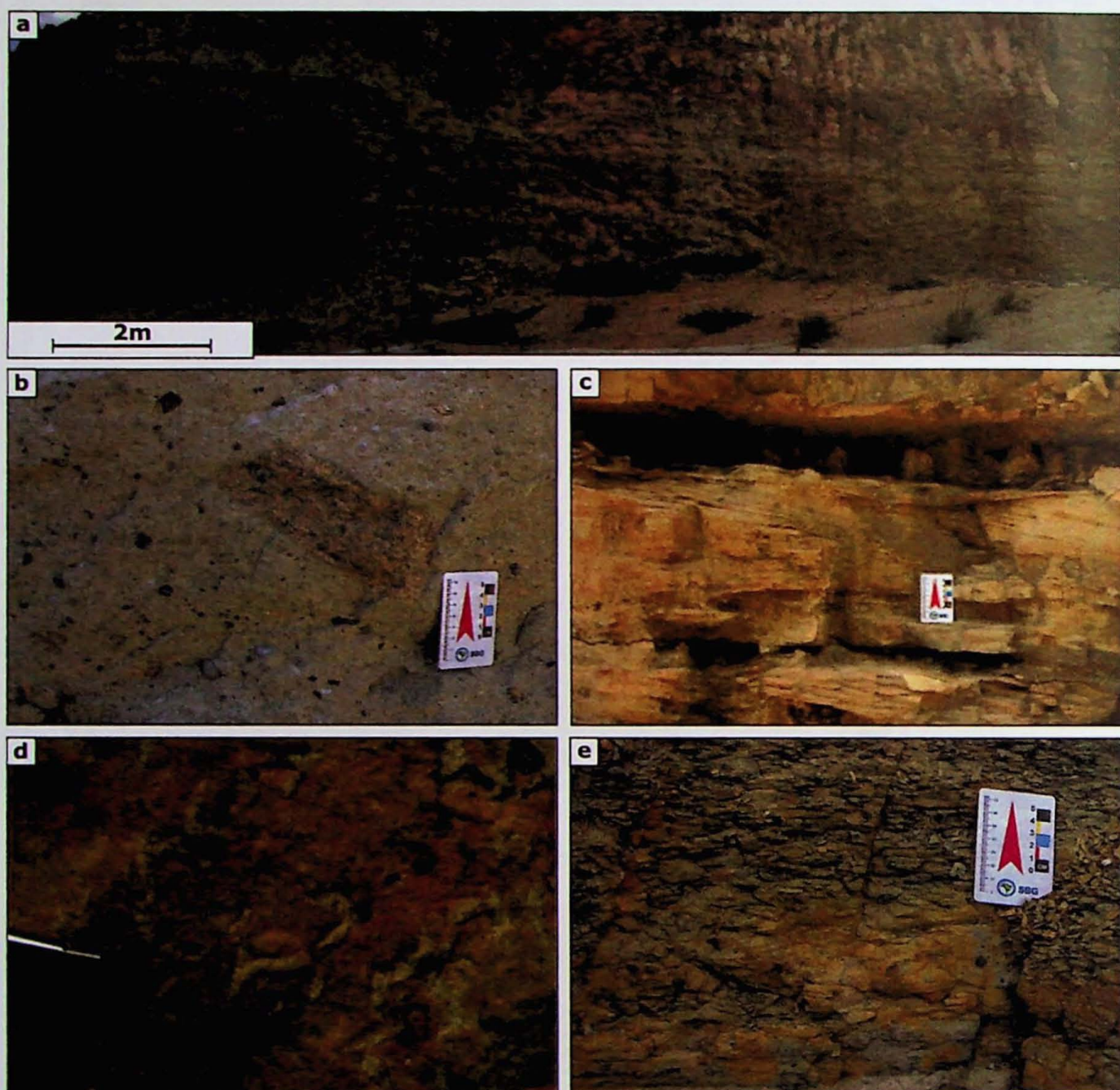


Figure 3: Examples of sedimentary facies found in the Serra Grande Group: (a) sheet-braided coarse pebbly sandstone in the Ipu Formation in Area II; (b) diamictite in the Ipu Formation in Area II (c) hummocky cross-stratification and wave ripple lamination in fine sandstone in the Tianguá Formation in Area Ia; (d) three-dimensional galleries of feeding burrows, Tianguá Formation in Area Ia; (e) marine siltstone with ice-rafted pebbles near its base in the Tianguá Formation in Area II;



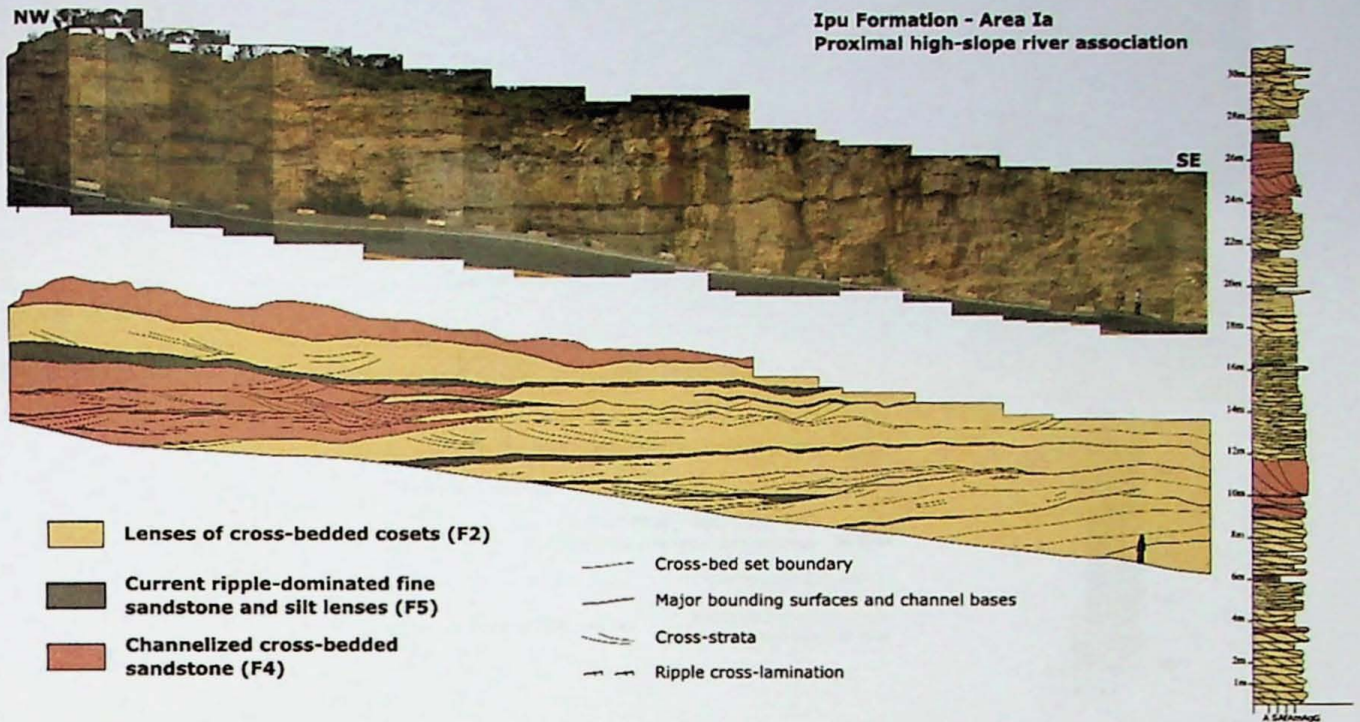


Figure 4: Example of the depositional architecture of *Proximal high-slope rivers association* in the Ipu Formation of Area Ia, typically dominated by lenses of cross-bedded cosets (F2) in metric fining-upward cycles bounded by bar-top current-ripple dominated fine sandstone and siltstone (F5) and minor occurrences of channelized cross-bedded sandstone (F4).



Ipu Formation - Area Ib

Distal low-slope rivers association

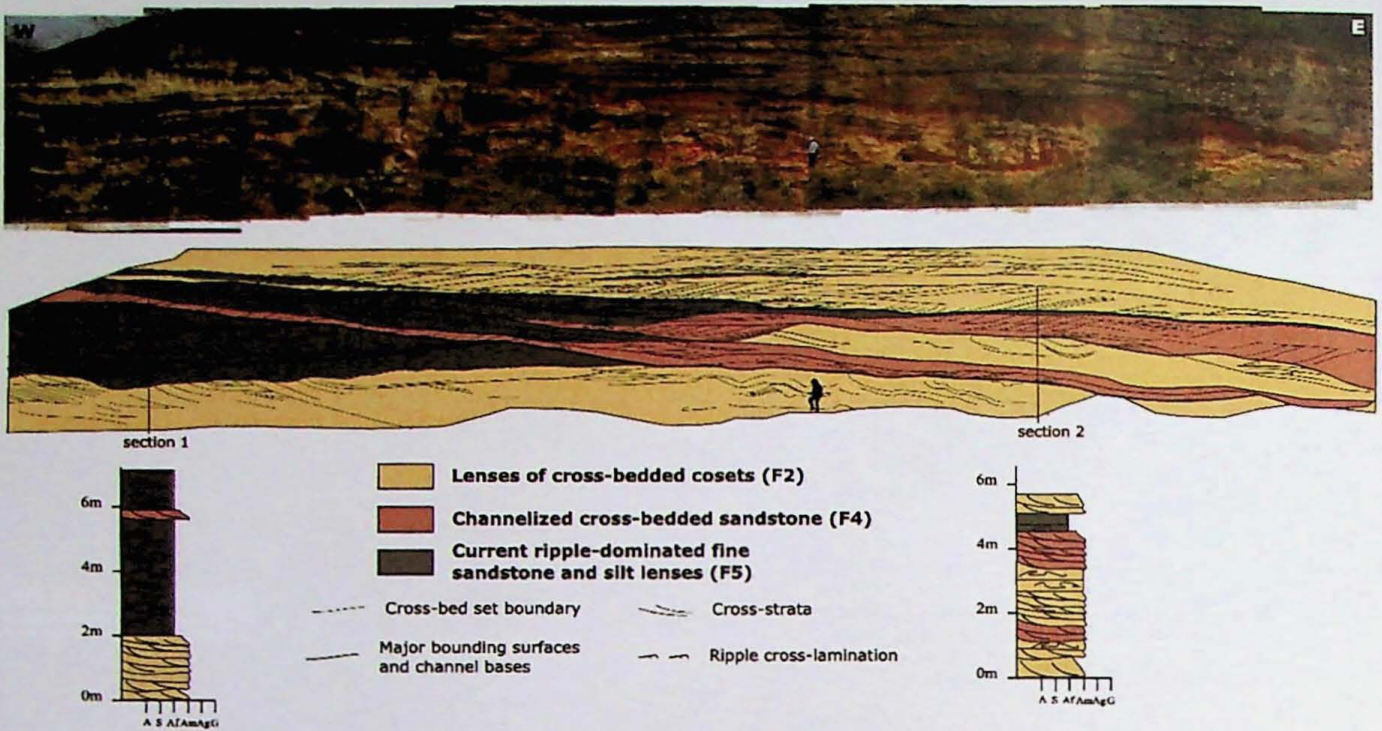


Figure 5: Example of the depositional architecture of *Distal low-slope rivers and floodplains association* in the Ipu Formation in Area Ib, typically dominated by fine sandstone in lenses of cross-bedded cosets in (F2) with abundant preserved overbank deposits of current-ripple dominated fine sandstone and siltstone (F5) and minor occurrences of channelized cross-bedded sandstone (F4). This association represents the “meandering ancestor” proposed by Almeida *et al.* (2016).

POI I6C  
01 07 20







**USP**

Universidade  
de São Paulo

Pós Graduação em Geociências  
Programa de Geoquímica e Geotectônica