

ANÁLISE PETROFÍSICA E DE
AGRUPAMENTOS
FACIOLÓGICOS DO SISTEMA
AQUÍFERO SÃO SEBASTIÃO NA
PORÇÃO NORTE DA BACIA DO
RECÔNCAVO

ALEXSANDRO GUERRA CERQUEIRA

SALVADOR – BAHIA
JANEIRO – 2019

Análise Petrofísica e de Agrupamentos Faciológicos do Sistema Aquífero São Sebastião na Porção Norte da Bacia do Recôncavo

por

ALEXSANDRO GUERRA CERQUEIRA

Me. Geofísica (Universidade Federal da Bahia – 2016)

Orientador: Prof. Dr. Olivar Antônio Lima de Lima

Co-Orientador: Prof. Dr. Ricardo Araújo Rios

TESE DE DOUTORADO

Submetida em satisfação parcial dos requisitos ao grau de

DOUTOR EM CIÊNCIAS

EM

GEOFÍSICA

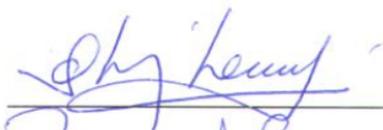
ao

Conselho Acadêmico de Ensino

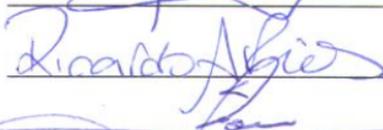
da

Universidade Federal da Bahia

Comissão Examinadora



Dr. Olivar Antônio Lima de Lima (Orientador)



Dr. Ricardo Araújo Rios (Co-Orientador)



Dr. Milton José Porsani



Dr. Michael Holz



Dr. Washington de Jesus Sant'Anna da Franca Rocha

Aprovada em 21 de janeiro de 2019

A presente pesquisa foi desenvolvida no Centro de Pesquisa em Geofísica e Geologia da UFBA, com recursos da CAPES.

C416 Cerqueira, Alexsandro Guerra,
Análise Petrofísica e de Agrupamentos Faciológicos do Sistema
Aquífero São Sebastião na Porção Norte da Bacia do Recôncavo
/ Alexsandro Guerra Cerqueira. — Salvador, 2019.
94 f.: il.

Orientador: Prof. Dr. Olivar Antônio Lima de Lima
Co-Orientador: Prof. Dr. Ricardo Araújo Rios
Tese (Doutorado) - Pós-Graduação em Geofísica. Instituto de
Geociências da Universidade Federal da Bahia, 2019.

1. Bacia do Recôncavo. 2. Aquífero São Sebastião 3. Análise
de Agrupamento 4. K-médias 5. Modelo de Misturas Gaussianas

550.3(134.3)(043)

"É fácil construir histórias para explicar o passado ou confiar em algum desdobramento futuro duvidoso. O fato de que tais empreendimentos tenham suas armadilhas não significa que não devamos levar adiante."

Leonard Mlodinow

Resumo

Nesta tese realiza-se uma análise hidrológica quantitativa do sistema aquífero Recôncavo, na porção norte da bacia do Recôncavo, utilizando conceitos de aprendizado de máquina para a interpretação de agrupamentos faciológicos e salinidade das águas subterrâneas. São analisados 101 perfis geofísicos contendo, principalmente, medidas de potencial elétrico espontâneo, resistividades elétricas rasa e profunda e contagem de radiação gama. Em poucos perfis desse conjunto usou-se medidas de densidade eletrônica e vagarosidade sônica para avaliações petrofísicas de porosidade. É realizada a distinção entre camadas permeáveis e não-permeáveis saturadas com águas de formações doces ou salgadas a partir da análise de agrupamento dos registros de potencial espontâneo, resistividade e profundidade, utilizando um algoritmo de análise de agrupamento de Modelos de Misturas Gaussianas inicializado a partir dos centroides obtidos com o método K-médias. Isto permite gerar um modelo hidro-litológico de poço interpretado usando um sistema de recomendação, de maneira rápida e eficiente, na maioria dos casos. Uma segunda análise de agrupamento é feita através do atributo da radiação gama, onde sucessivos testes estatísticos apontam uma forte hipótese bimodal das distribuições areno-argilosas, propiciando a automatização da estimativa do conteúdo de argilas. Uma versão alternativa utilizando medidas de resistividade elétrica também é apresentada. Após distinguir camadas permeáveis com água doce (arenitos aquíferos) e definir a profundidade máxima em que ocorrem nos poços, foram gerados mapas estruturais da base da água doce e de espessura acumulada de arenitos nas áreas selecionadas de Pojuca e Araçás. Estimativas de porosidade são feitas utilizando os perfis de densidade e vagarosidade. Adicionando-se, a essas análises, informações de profundidade da zona de água doce de outros 74 poços, foram construídos mapas similares para toda porção norte da bacia do Recôncavo, e estimada uma reserva hídrica potável de $2,6 \times 10^{11} \text{ m}^3$, numa extensão superficial de aproximadamente 2900 km^2 .

Palavras Chaves: Bacia do Recôncavo, Aquífero São Sebastião, Análise de Agrupamento, K-médias, Modelo de Misturas Gaussianas.

Abstract

In this thesis it is performed a quantitative hidrological analysis of the Recôncavo aquifer system, in the northern portion of the Recôncavo basin, using machining learning concepts to interpret lithological facies clusters and groundwater salinity. One hundred one well logs containing mainly electric spontaneous potential, resistivity, and gamma-ray measurements are analyzed and, for a small portion of the dataset, there are electronic density and sonic vagarosity measurements. The distinction between permeable and non-permeable layers containing fresh or saltwater is accomplished via clustering analysis of the spontaneous potential, resistivity, and depth the Gaussian mixture models algorithm initialized with the centroids obtained using the K-means method. This allows the generation of an interpreted model utilizing a recommendation system quickly and efficiently for most cases. A second clustering analysis is accomplished through the gamma-ray attribute, where successive statistical tests point out a strong bimodal hypothesis for the shaly sand distributions, propitiating an automatization of the clay content estimatives. An alternative version using resistivity measurements is also presented. After distinguishing the freshwater permeable layers and defining the maximum depth of its occurrence, structural maps of the fresh-water base and accumulated sandstone thickness are generated for Pojuca and Araçás areas. Porosity estimations are also made using the density and vagarosity logs. Summing up, to this analysis, the information of the depth of the freshwater zone from other 74 wells, similar maps are constructed for the northern portion of the Recôncavo basin and estimated a stored fresh-water volume of $2.6 \times 10^{11} \text{ m}^3$ in a surface extent of 2900 km^2 .

Key words: Recôncavo Basin, São Sebastião Aquifer, Cluster Analysis, K-means, Gaussian Mixture Models.

Índice

| | |
|---|----|
| Resumo | 4 |
| Abstract | 5 |
| Índice | 6 |
| Índice de Tabelas | 8 |
| Índice de Figuras | 10 |
| Introdução | 14 |
| 1 Geologia e Hidrogeologia da Bacia do Recôncavo | 17 |
| 1.1 Sequências Sedimentares | 17 |
| 1.1.1 Super-Sequência Pré-Rifte | 18 |
| 1.1.2 Super-Sequência Rifte | 19 |
| 1.1.3 Super-Sequência Pós-Rifte | 20 |
| 1.2 Sistema Aquífero Recôncavo | 20 |
| 2 Propriedades Físicas das rochas | 23 |
| 2.1 Propriedades Hidráulicas das Rochas | 23 |
| 2.1.1 Porosidade | 24 |
| 2.1.2 Permeabilidade | 25 |
| 2.2 Propriedades Elétricas das Rochas | 27 |
| 2.2.1 Resistividade elétrica de rochas porosas | 28 |
| 3 Perfilagem Geofísica de Poços | 31 |
| 3.1 Potencial Espontâneo (SP) | 31 |
| 3.2 Perfil de Indução Elétrica | 34 |
| 3.3 Perfil de Raios Gama | 35 |
| 3.4 Perfil Sônico | 36 |
| 3.5 Perfil de Densidade | 37 |

| | | |
|----------|--|----|
| 4 | Análise de Agrupamento | 39 |
| 4.1 | Algoritmos de Agrupamento | 39 |
| 4.1.1 | K-médias (<i>K-means</i>) | 39 |
| 4.1.2 | Modelos de Misturas Gaussianas (GMM) | 42 |
| 4.2 | Validação de Grupos | 46 |
| 4.2.1 | Índice Rand Ajustado | 46 |
| 4.2.2 | Índice de Silhueta | 47 |
| 5 | Metodologia da Análise de Agrupamento de Fácies | 49 |
| 5.1 | Pré-processamento e Edição | 49 |
| 5.2 | Análise de Agrupamentos | 51 |
| 5.2.1 | Sistema de Recomendação de Grupos | 52 |
| 5.2.2 | Validação | 55 |
| 5.3 | Cálculo da Argilosidade | 55 |
| 6 | Resultados e Interpretações | 61 |
| 6.1 | Área de Pojuca | 62 |
| 6.2 | Área de Araçás | 72 |
| 6.3 | Porção Norte da Bacia do Recôncavo | 79 |
| 7 | Conclusões | 84 |
| | Referências Bibliográficas | 87 |
| | Agradecimentos | 91 |

Índice de Tabelas

| | | |
|-----|--|----|
| 4.1 | Matriz de concordância entre a Partição P e Agrupamento C | 47 |
| 5.1 | Valor médio do índice silhueta de diferentes números de grupos para cada análise de agrupamento das medidas de raios gama de amostras pertencentes ao sistema aquífero do Recôncavo. | 55 |
| 5.2 | Média (μ) e desvio padrão (σ) do modelo de misturas gaussianas da análise de agrupamento com dois grupos. Correspondem a arenitos e folhelhos, derivados das medidas de raios gama, e os parâmetros da Eq. 3.10, dado em unidade uAPI derivados dos Modelos de Misturas Gaussianas. | 56 |
| 5.3 | Valor médio do índice silhueta de diferentes números de grupos para cada agrupamento das medidas do logaritmo da resistividade elétrica. | 57 |
| 5.4 | Média (μ) e desvio padrão (σ) do modelo de misturas gaussianas da análise de agrupamento de com dois grupos. Correspondem a arenitos e folhelhos, derivados das medidas do logaritmo da resistividade elétrica, e os parâmetros da Eq. (3.7), dado em unidade $\log(\Omega m)$ derivados dos Modelos de Misturas Gaussianas. | 58 |
| 6.1 | Valor médio do índice silhueta de diferentes números de grupos para cada análise de agrupamento das medidas de raios gama dos perfis pertencentes a área de Pojuca. | 65 |
| 6.2 | Média (μ) e desvio padrão (σ) do modelo de misturas gaussianas da análise de agrupamento com dois grupos. Correspondem a arenitos e folhelhos, obtidos a partir das medidas de raios gama, e parâmetros da Eq. 3.10, dados em unidade uAPI derivados dos Modelos de Misturas Gaussianas dos perfis pertencentes a área de Pojuca. | 65 |
| 6.3 | Valor médio do índice silhueta de diferentes números de grupos para cada análise de agrupamento das medidas de raios gama dos perfis pertencentes a área de Araçás. | 79 |

| | |
|---|----|
| 6.4 Média (μ) e desvio padrão (σ) do modelo de misturas gaussianas da análise de agrupamento com dois grupos correspondentes a arenitos e folhelhos das medidas de raios gama, e parâmetros da Eq. 3.10 dados em unidade uAPI derivados dos Modelos de Misturas Gaussianas dos perfis pertencentes a área de Araçás. | 79 |
|---|----|

Índice de Figuras

| | | |
|-----|--|----|
| 1.1 | Localização e arcabouço estrutural da Bacia do Recôncavo (Milhomem et al., (2003), apud Prates e Fernandez (2015)). | 18 |
| 1.2 | Carta estratigráfica da bacia do Recôncavo (Silva et al., 2007). | 22 |
| 2.1 | Elemento de Volume Representativo (EVR) de uma rocha de porosidade intergranular (esquerda) e representação esquemática do volume poroso e da matriz (direita). (Adaptado de Schön, 2015). | 24 |
| 2.2 | Histograma e diagrama de caixa de 1299 medidas de porosidades calculadas, a partir do perfil de densidade de arenitos da Formação São Sebastião, de um poço da área de Araças. Note que o histograma da porosidade assemelha-se a uma distribuição normal (gaussiana). | 25 |
| 2.3 | Representação do modelo de Kozeny-Carman de uma rocha porosa por tubos capilares (Schön, 2015). | 26 |
| 3.1 | Histograma de uma sequência siliciclástica Terciária. Os valores $GR_{min} = 52$ gAPI and $GR_{max} = 93$ gAPI, sugeridos por Vernik (2016), foram selecionados, visualmente, com base no histograma. | 36 |
| 4.1 | Representação gráfica de observações sintéticas referentes a dois grupos: arenitos (círculos laranjas) e folhelhos (cruzes verdes), e seus respectivos centroides. | 41 |
| 4.2 | Execução do algoritmo K-médias para subdividir um conjunto de dados em três partições (Adaptado de James et al., 2013). | 42 |
| 4.3 | Histograma e a função densidade de probabilidade de uma distribuição aproximadamente normal de uma variável x com os parâmetros: $\mu \cong 0$ e $\sigma \cong 1$ | 43 |
| 4.4 | (a) Representação visual de um modelo de misturas gaussianas (GMM) contendo três gaussianas representando a probabilidade de ocorrência de um evento e (b) a representação do valor π_k das respectivas distribuições. | 45 |

| | | |
|-----|---|----|
| 5.1 | Perfil de potencial espontâneo não corrigido (esquerda) e a correção utilizando um modelo regressivo do potencial espontâneo e a profundidade (direita). Os dados utilizados para a criação da reta de correção estão contidos no intervalos de $[-52,20]$ (mV) para os registros de SP e profundidades entre $[123,800]$ (m). | 50 |
| 5.2 | Gráficos de dispersão das medidas com ferramentas de potencial elétrico espontâneo (<i>SP</i>), logaritmo da resistividade da indução profunda (<i>LOGILD</i>) e o logaritmo da resistividade da normal curta (<i>LOGSN</i>) e seus respectivos histogramas no poço guia. | 52 |
| 5.3 | Gráficos de dispersão das medidas com ferramentas de potencial elétrico espontâneo (<i>SP</i>), logaritmo da resistividade de indução profunda (<i>LOGILD</i>) e o logaritmo da resistividade da normal curta (<i>LOGSN</i>) e seus respectivos histogramas apontando os grupos obtidos a partir do algoritmo de Modelo de Misturas Gaussianas (GMM) com centroides inicializados a partir do resultado do algoritmo K-médias, do perfil PG1. | 53 |
| 5.4 | Grupos de facies obtidos a partir do algoritmo de clusterização de Modelo de Misturas Gaussianas (GMM) com centroides inicializados a partir do resultado do algoritmo K-médias, e sistema de recomendação para correção das profundidades de zonas permeáveis e posicionamento da zona de transição do perfil PG1. | 54 |
| 5.5 | Histograma das medidas de raios gama e parâmetros obtidos a partir do Modelo de Misturas Gaussianas do poço PW 3 usados para calcular a argilosidade através da Eq. (3.10). Nota-se a tendência de uma distribuição bimodal das medidas. | 58 |
| 5.6 | Histogramas das medidas de raios gama e do logaritmo da resistividade elétrica de três perfis geofísicos diferentes. É possível notar a tendência bimodal bem definida em quase todos os histogramas, com exceção do histograma do logaritmo da resistividade do perfil PW 3. | 59 |
| 5.7 | Conteúdo de argila estimado (V_{sh}), do poço PW 9, em que $V_{sh} = 0$ é interpretado como arenito, $V_{sh} = 1$ é interpretado como folhelho e caso contrário, entende-se que a amostra avaliada trata-se de uma mistura areno-argilosa entre arenito e folhelho (siltito). | 60 |
| 6.1 | Mapa que apresenta os perfis geofísicos disponíveis para análise do sistema aquífero do Recôncavo, na porção norte. | 62 |
| 6.2 | Gráfico de dispersão da resistividade de indução <i>versus</i> o potencial elétrico espontâneo do perfil PM 21 e representação dos grupos interpretados a partir dos algoritmos de agrupamento e o sistema de recomendação. | 64 |

| | | |
|------|---|----|
| 6.3 | Interpretação dos agrupamentos obtidos no perfil PM 21 a partir do algoritmo de Modelo de Misturas Gaussianas inicializados a partir do centroide do algoritmo K-médias (terceira coluna) e a interpretação externa ao algoritmo para validação utilizando o Índice Rand Ajustado (<i>ARI</i>). | 67 |
| 6.4 | Mapa de contorno estrutural da base do reservatório de água doce na área de Pojuca. | 68 |
| 6.5 | Nível Estático do aquífero livre correspondente a área de Pojuca. | 68 |
| 6.6 | Interpretação obtida das análise de agrupamento utilizando o algoritmo de Modelo de Misturas Gaussianas inicializado a partir dos centroides do método K-médias do perfil PM 4, usando o sistema de recomendação proposto na Seção 5.2.1. | 69 |
| 6.7 | Seção longitudinal, sentido leste-oeste, apresentando quatro perfis da Área de Pojuca e suas respectivas interpretações obtidas a partir da análise de agrupamento utilizando o algoritmo de Modelos de Misturas Gaussianas inicializado a partir dos centroides do método K-médias. | 70 |
| 6.8 | Seção transversal, sentido sul-norte, apresentando quatro perfis da área de Pojuca e suas respectivas interpretações obtidas a partir da análise de agrupamento utilizando o algoritmo de Modelos de Misturas Gaussianas inicializado a partir dos centroides do método K-médias. | 71 |
| 6.9 | Mapa de espessura acumulada estimada das camadas permeáveis com água doce nos perfis disponíveis na área de Pojuca. | 72 |
| 6.10 | Histograma da porosidade calculada a partir dos registros de densidade dos perfis dos poços PM 10 e 11 (esquerda) e vagarosidade dos perfis dos poços PM 1, 8 e 9 (direita). | 72 |
| 6.11 | Interpretação dos agrupamentos obtidos no perfil PA 8 a partir do algoritmo de Modelo de Misturas Gaussianas inicializados a partir do centroide do algoritmo K-médias (terceira coluna) e a interpretação externa ao algoritmo para validação utilizando o Índice Rand Ajustado (<i>ARI</i>). | 75 |
| 6.12 | Mapa de contorno estrutural da base do reservatório de água doce na área de Araçás. | 76 |
| 6.13 | Nível Estático do aquífero livre correspondente a Área de Araçás. | 76 |
| 6.14 | Seção longitudinal, sentido sudoeste-nordeste, apresentando quatro perfis da área de Araçás e suas respectivas interpretações obtidas a partir da análise de agrupamento utilizando o algoritmo de Modelos de Misturas Gaussianas inicializado a partir dos centroides do algoritmo K-médias. | 77 |

| | | |
|------|---|----|
| 6.15 | Seção transversal, sentido sudeste-noroeste, apresentando quatro perfis da área de Araçás e suas respectivas interpretações obtidas a partir da análise de agrupamento utilizando o algoritmo de Modelos de Misturas Gaussianas inicializado a partir dos centroides do algoritmo K-médias. | 78 |
| 6.16 | Mapa de espessura acumulada estimada das camadas permeáveis com água doce nos perfis disponíveis na área de Araçás. | 80 |
| 6.17 | Histograma da porosidade calculada a partir dos registros de densidade dos perfis PA 16 e 17(esquerda) e vagarosidade dos perfis PA 2,4 e 18 (direita). | 80 |
| 6.18 | Mapa de contorno estrutural da base do reservatório de água doce da porção norte da Bacia do Recôncavo. | 82 |
| 6.19 | Mapa de espessura acumulada estimada das camadas permeáveis com água doce da porção norte do Bacia do Recôncavo. | 83 |

Introdução

O sistema aquífero Recôncavo, na bacia sedimentar homônima, constituiu-se na sua porção livre por arenitos e conglomerados das Formações Marizal e Barreiras e, na sua porção semi-confinada, por arenitos de excelente poro-permeabilidades da Formação São Sebastião. Segundo Lima (1999), o sistema dispõe de uma reserva hídrica de aproximadamente quinhentos bilhões de metros cúbicos que é utilizada para atender vilas e cidades do Recôncavo Baiano, além de complexos industriais instalados na região. Pode-se afirmar que este aquífero contém água de excelente qualidade, permitindo atribuir-lhe a classificação de água mineral, sendo considerado um dos aquíferos mais importantes do Estado da Bahia (Brandão, 2008). A quantificação de zonas permo-porosas e a determinação da qualidade das águas contidas nestas camadas permitem um melhor aproveitamento destes recursos hídricos.

Para avaliar com maior detalhe o potencial desse sistema aquífero em áreas da porção norte do Recôncavo, de forma mais automática, foram utilizados dados de perfis geofísicos de poços de exploração de petróleo e água subterrânea. Foi utilizado algoritmos de aprendizado de máquinas para examinar perfis geofísicos de 101 poços liberados pela PETROBRAS, ANP e CERB para a realização de estudos hidrogeológicos no CPGG/UFBA (Brandão, 2008; Mariano, 2013).

O uso de algoritmos inteligentes, também conhecidos como algoritmos de aprendizado de máquina (AM), fornece suporte para a determinação de parâmetros físicos das rochas, bem como permite uma análise e interpretação de dados geofísicos mais acuradas de regiões com poucas informações *a priori*, auxiliando na eliminação de ruídos e ambiguidades em modelos geofísicos (Mukerji et al., 2001; Dubois et al., 2007; Torghabeh et al., 2014; Caté et al., 2017; Xia et al., 2018). Em Aprendizado de Máquina, os algoritmos são programados para aprender com experiências passadas (Faceli et al., 2011) e segundo Mitchell (1997) preocupa-se com questões de como construir programas de computador que automaticamente melhorem o desempenho na realização de tarefas com base em experiências conhecidas previamente. A abordagem de aprendizado de máquina pode ser subdividida em dois principais grupos: supervisionada e não-supervisionada. Aplicações que consistem na criação de modelos matemáticos a partir de dados de entrada que contêm suas respectivas respostas, estão inseri-

dos na abordagem supervisionada (Ex.: métodos dos mínimos quadrados, k-vizinhos mais próximos, máquina de vetores de suporte e árvores de decisão). Para problemas de reconhecimento de padrões em que não há respostas (dados observados), ou estas observações encontram-se inconsistentes, pode ser utilizada a abordagem não-supervisionada. Estas técnicas não-supervisionadas são capazes de descobrir grupos, também chamados de "*clusters*", que possuem similaridades entre as instâncias pertencentes a cada um deles, de acordo com algum critério definido (Bishop, 2006).

A análise em medidas de perfilagens geofísicas realizadas em aquíferos tem como objetivo distinguir zonas permeáveis das não permeáveis, além de separar camadas permo-porosas cujas águas de formação sejam potável daquelas que contém águas salobras. Neste contexto, o método de agrupamento K-médias (*K-means*), é um algoritmo não-supervisionado capaz de selecionar grupos que compartilham semelhanças com um determinado centroide, em um conjunto de dados não rotulados (Bishop, 2006; James et al., 2013; Hastie et al., 2016). Esse algoritmo é utilizado com o objetivo de obter agrupamentos que definam os grupos descritos a partir registros em poços de potencial espontâneo e resistividade, considerados atributos relevantes para definições de zonas permeáveis e que são sensíveis aos fluidos presentes na formação. Para análises mais acuradas, os agrupamentos obtidos com o algoritmo K-médias é submetido como modelo inicial à técnica não-supervisionada de Modelo de Misturas Gaussianas (*Gaussian mixture models* - GMM) com o objetivo de atenuar ambiguidades entre os agrupamentos obtidos. Os agrupamentos gerados por essas análises não-supervisionadas são interpretados através de um sistema de recomendação. O sucesso das interpretações está relacionado a um conjunto de medidas confiável que, em alguns casos, requerem correções.

Uma segunda abordagem não paramétrica é feita com o objetivo de utilizar as medidas de raios gama ou de resistividade elétrica para estimar o conteúdo de argilas nos arenitos do sistema aquífero do Recôncavo, a partir da análise de agrupamentos. De acordo com Vernik (2016) a estimativa da argilosidade exige alguns parâmetros que são obtidos de testemunhos de sondagens ou através de algum critério estabelecido por um interprete. Considerando que a contagem de radiação gama de formações areno-argilosas apresentam uma distribuição bimodal, estes parâmetros foram obtidos a partir de técnicas de aprendizado de máquina usando como critérios estatísticos a média e o desvio padrão obtidos a partir de um Modelo de Misturas Gaussianas, eliminando o critério da subjetividade na obtenção dos mesmos, o que permite automatizar o calculo da argilosidade.

Neste trabalho são analisados 101 perfis geofísicos cedidos pela Companhia de Engenharia e Recursos Hídricos da Bahia (CERB), Agência Nacional de Petróleo (ANP) e Petróleo Brasileiro S.A (Petrobras), que perpassam principalmente a unidade aquífero São Sebastião. Estes

dados permitiram a implementação de metodologias inéditas de interpretações automáticas utilizando aprendizado máquina, todas elas implementadas na linguagem de programação *python 3* e foi utilizada a biblioteca *scikit-learn*. As análises oriundas destas interpretações automáticas são somadas às informações sobre a base da zona de água doce de outros 74 perfis já interpretados num trabalho inédito realizado para a Petrobras por Neto (2006). Os resultados das análises de agrupamentos faciológicas de arenitos saturados com água doce, folhelhos, siltitos argilosos e arenitos saturados com águas salobras estão apresentados na forma de seções correlativas. Dados petrofísicos derivados de perfis, tais como argilosidade e porosidade efetiva foram usados para estimar as reservas de água doce armazenadas em áreas menores na região de Pojuca e Araças e, posteriormente, em toda porção norte da Bacia do Recôncavo.

1

Geologia e Hidrogeologia da Bacia do Recôncavo

A Bacia do Recôncavo está localizada no centro-leste do Estado da Bahia, Nordeste do Brasil, e está disposta em uma área de aproximadamente 12.000 km² (Silva et al., 2007; Prates e Fernandez, 2015). Sua porção norte ocupa uma área de aproximadamente 5.500 km² e é limitada ao norte e noroeste pelo Alto de Aporá, ao leste pela Falha de Salvador e ao oeste pela falha de Mata-Catu (Figura 1.1). Esta sub-bacia sedimentar faz parte de um sistema de riftes assimétricos (o aulacógeno Recôncavo-Tucano-Jatobá), que compreende os estados da Bahia, Sergipe e Pernambuco, e foi preenchido por sedimentos clásticos continentais, com idades variando do Jurássico até o Cretáceo (Lima, 1999).

A configuração estrutural do aulacógeno é do tipo meio-gráben de direção NNE-SSW, com a borda falhada a sudeste e a flexural a oeste. O seu arcabouço tectônico é composto de falhas sintéticas e antitéticas paralelas à falha principal (Falha de Salvador) na borda SE, e zonas de transferência NW-SE que acomodam o deslocamento lateral entre blocos crustais sob os esforços distensivos da separação continental Brasil-África (Aragão, 1993).

1.1 Sequências Sedimentares

As sequências deposicionais ligadas ao preenchimento da Bacia relacionam-se a três principais estágios: pré-rifte, rifte e pós-rifte (Figura 1.2). A espessura máxima sedimentar é superior a 6.500m e é verificada no Baixo de Camaçari (Aragão, 1994).



Figura 1.1: Localização e arcabouço estrutural da Bacia do Recôncavo (Milhomem et al., (2003), apud Prates e Fernandez (2015)).

1.1.1 Super-Sequência Pré-Rifte

Esta Super-Sequência está relacionada a flexura da crosta formada no estágio inicial de distensão que originou o sistema de riftes do Eocretáceo. São verificados três grandes ciclos flúvio-eólicos representados, da base para o topo, pelo Membro Boipeba da Formação Aliança e pelas formações Sergi e Água Grande (Silva et al., 2007).

A Formação Aliança abrange os membros Boipeba que consiste de arcóseos finos a médios, vermelhos e marrons, com estratificação cruzada; e Capianga, tipicamente caracterizado por folhelhos vermelho-tijolo. Já a Formação Sergi é constituída por arenitos finos a conglomeráticos, cinza-esverdeados, com estratificação cruzada acanalada (Caixeta et al., 1994).

1.1.2 Super-Sequência Rifte

A fase Rifte foi preenchida por dois principais sistemas: o primeiro flúvio-deltáico passando a lacustre, representado pelas formações Candeias e Maracangalha, pelo Grupo Ilhas e pela Formação São-Sebastião; e o segundo sistema que consiste de fan-deltas derivados da borda falhada, com conglomerados proximais (Formação Salvador) e turbiditos mediais distais (da Silva et al., 2003).

A Formação Candeias se divide nos membros Tauá e Gomo, tem um contexto lacustre que prevalecia no início da fase rifte. O membro Tauá se caracteriza por folhelhos escuros, físseis, com partição acicular, enquanto que o Membro Gomo é formado por folhelhos cinza-esverdeados, laminados, com delgadas camadas de calcários e arenitos finos, cinza-claros. Perfis elétricos indicam uma maior resistividade no Membro Gomo. Estas rochas foram depositadas em ambiente lacustre, com rápida subsidência e com grande aporte sedimentar (Caixeta et al., 1994).

Em seguida têm-se a Formação Maracangalha que é constituída por folhelhos cinza-esverdeados e cinza-escuros, cujos valores de resistividade são inferiores se comparado aos folhelhos do Membro Gomo. O Membro Caruaçu apresenta camadas lenticulares de arenito fino e médio, com estratificações cruzadas tangenciais e estratificação plano-paralela, laminações por correntes e convolutas. Arenitos de granulometria fina, lamosos e maciços constituem o Membro Pitanga (Caixeta et al., 1994).

O Grupo Ilhas, composto pelas Formações Marfim e Pojuca, designa uma seção areno-argilosa situada entre os folhelhos Candeias e os arenitos do Grupo Massacará. A seção basal da Formação Marfim, depositou-se *onlap*¹ sobre as áreas plataformais e consiste de arenitos finos a médios intercalados por delgadas camadas de folhelhos (Campinho, 2002 apud Santos et al. (2007)) depositados em clima mais seco. Sob um contexto de reduzidas taxas de subsidência e baixos gradientes deposicionais, desenvolveram-se os registros estratigráficos da Formação Pojuca, caracterizada por uma sucessão cíclica de fácies deltaico e sequências pelíticas/carbonáticas lacustres (Santos et al., 2007). Segundo Santos e Braga (1990), foram desenvolvidos, neste grupo, marcos elétrico-radioativos correlacionáveis por toda bacia.

Na Formação Salvador são registrados conglomerados e arenitos na borda leste da Bacia. Estes conglomerados foram formados em leques aluviais sintectônicos e marcam a atuação da falha de borda que caracteriza a porção rifte desta bacia.

Por fim, os depósitos do Grupo Massacará (Formação São Sebastião) constituído por depósitos

¹De acordo com Holz (2012), é uma terminação de estratos de baixo ângulo contra uma superfície mais inclinada, normalmente marcando a terminação lateral proximal de unidades sedimentares.

fluviais. Viana et al. (1971) definiram esta formação como constituída por arenitos grossos, amarelo-avermelhados, com intercalações de argilas sílticas, depositados por sistemas fluviais atuantes do Berriasiano ao Eoaptiano, divididos em três Membros: Mb. Paciência constituído por arenitos finos e grossos e folhelhos pretos, Mb. Passagem dos Teixeiras que apresenta arenitos finos a médios e folhelhos cinzas e o Mb. Rio Joanes com arenitos finos e argilas avermelhadas (Caixeta et al., 1994).

1.1.3 Super-Sequência Pós-Rifte

Pertencente ao Cretáceo Superior, resultado das últimas atividades do rifte, é representada pela Formação Marizal. Esta sequência tem uma espessura média de 50m (da Silva et al., 2003) e é composta por clastos grossos (conglomerados e arenitos), folhelhos e calcários. Repousa discordantemente sobre a Formação São Sebastião (Murphy e Schlanger, 1963).

1.2 Sistema Aquífero Recôncavo

Segundo Brandão (2008), a água subterrânea do sistema Aquífero Recôncavo é de origem pluvial e se infiltra em solos compostos por sedimentos arenosos a cascalhosos, onde cerca de 40 a 50% das águas da chuva consegue penetrar, e através de sedimentos pouco porosos, cuja taxa de infiltração pode ser reduzida a valores de 5 a 20%. Além disso, as águas pluviais podem também infiltrar-se em rochas aflorantes.

De acordo com Lima (1999), Brandão (2008) e Mariano (2013) o sistema aquífero Recôncavo é constituído por três Formações:

- **Formação Barreiras** que encontra-se principalmente na região norte da bacia e é classificado como um aquífero livre²;
- **Formação Marizal** que ocorre nas regiões mais elevadas da bacia e é classificado como misto por conter unidades de recarga livres e confinadas; e
- **Formação São Sebastião** que pode ser classificado como livre ou confinado³, porém sua maior porção de destaque é a confinada. É constituída por camadas de arenitos com espessuras que variam de 20 a mais de 100 m, confinados por camadas de folhelhos.

²Aquíferos livres são aqueles cujo limite superior é a própria superfície de saturação (ou freático) e a água está armazenada à pressão atmosférica.

³Aquíferos confinados são aqueles contidos entre camadas impermeáveis, armazenados a pressões superiores a atmosférica.

Quanto ao armazenamento de recursos hídricos (água potável), evidencia-se a porção confinada do aquífero São Sebastião que, em sua maior parte, contém água doce classificada como mineral.

Classificando o aquífero em termos das componentes livres e confinadas, sabe-se que o componente livre, constituído pelo acoplamento de arenitos das Formações Marizal e São Sebastião, forma um extenso reservatório com espessuras médias da ordem de 80 m, alcançando espessuras superiores a 200 m nos baixos estruturais de Alagoinhas e Camaçari (Lima, 1999). A porção confinada é composta predominantemente por arenitos e folhelhos da Formação São Sebastião depositados em regimes fluviais, cujas espessuras podem ultrapassar 1000 m. Eventualmente, pode ocorrer acúmulo de águas doces na Formação Pojuca em baixos estruturais.

Moreira et al. (2016) afirmam que a base do reservatório de água doce da Bacia do Recôncavo varia de 0 a 1600 m de profundidade, com maiores espessuras nas regiões de Alagoinhas, Miranga e Camaçari. Lima (1999) previu uma reserva de água potável de aproximadamente $5,0 \times 10^{11} \text{ m}^3$. Em um estudo de uma área de 152 km^2 , na região de Araçás, realizado por Brandão (2008), foi estimado um volume hídrico da ordem de $9,8 \times 10^9 \text{ m}^3$.

Lima (1999) apresentou em seu trabalho dados a cerca da produtividade de poços no aquífero São Sebastião na região de Camaçari - Dias D'Ávila, em profundidades no intervalo de 100 a 380 m, que abrangem um intervalo de vazão entre 20 a $200 \text{ m}^3/\text{h}$.

Foi observada uma forte correlação entre o nível estático dos aquíferos da bacia do Recôncavo (Z_{ne}) e a elevação dos poços (Z_p), por Lima (1999), que desenvolveu a seguinte relação empírica entre essas variáveis, dada por:

$$Z_{ne} = 0,819012 \cdot Z_p + 4,75295, \quad (1.1)$$

em que Z_{ne} e Z_p são dados em metros.

Apesar do Sistema Aquífero Recôncavo ser composto pelas Formações São Sebastião, Marizal e Barreiras, praticamente a maioria absoluta das perfilagens geofísicas de poços se dão na Formação São Sebastião, neste trabalho.

2

Propriedades Físicas das rochas

O estudo das propriedades físicas das rochas e suas interações com os fluidos: gases, hidrocarbonetos líquidos e soluções aquosas é denominado **petrofísica** (Peters, 2012). É necessário que essas rochas reservatórios (ou aquíferos) contenham porosidade, permeabilidade, e algumas outras propriedades adequadas para a extração de água, óleo ou gás. Na exploração petrolífera sabe-se que os principais reservatórios são rochas sedimentares (arenitos e carbonatos), podendo também haver armazenamento em rochas ígneas e metamórficas fraturadas e/ou fissuradas.

2.1 Propriedades Hidráulicas das Rochas

As propriedades do espaço poroso são importantes para a descrição e caracterização do volume do poro e o comportamento do fluxo do fluido de reservatórios (Schön, 2015). Segundo Schön (2015), as propriedades fundamentais do espaço poroso de um reservatório podem ser descritas por:

- Fração de volume dos fluidos (porosidade e saturação);
- Propriedades de controle de distribuição de fluidos no espaço poroso (pressão capilar, superfície interna específica e molhabilidade); e
- Propriedades de controle do fluxo de fluidos sob influência de um gradiente de pressão (permeabilidade).

Esta seção se limitará a discussão da porosidade e da permeabilidade.

2.1.1 Porosidade

A porosidade é uma propriedade fundamental das rochas e dos meios porosos em geral e é responsável por caracterizar a quantidade de interstícios ou vazios (poros, fraturas e fissuras) de um meio poroso rochoso (Lima, 2014). Esta propriedade influencia na maioria das propriedades físicas das rochas, por exemplo: na velocidade de ondas elásticas, resistividade elétrica e densidade das formações (Schön, 2015). Em resumo, pode-se dizer que a porosidade é uma propriedade adimensional, descrita em termos fracionais ou de porcentagem, sendo definida como o volume de todos os poros (vazios, fraturas e fissuras), V_p , dividido pelo volume total (V_t) de uma dada amostra:

$$\phi = \frac{V_p}{V_t} = \frac{V_t - V_m}{V_t}, \quad (2.1)$$

onde V_m representa o volume da matriz de sólidos. A Figura 2.1 ilustra os elementos apresentados na Eq. (2.1).

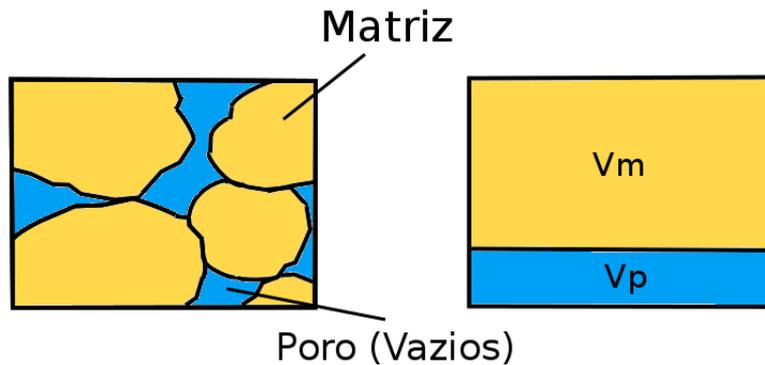


Figura 2.1: Elemento de Volume Representativo (EVR) de uma rocha de porosidade intergranular (esquerda) e representação esquemática do volume poroso e da matriz (direita). (Adaptado de Schön, 2015).

Essa definição de porosidade total leva em consideração poros e gargantas não-conectados (V_{nc}), por exemplo: bolhas, porosidade móldica, etc. Por isso, é necessário a introdução do conceito de porosidade efetiva de uma rocha (ϕ_e) que é a razão entre o volume poroso interconectado (V_i) e o volume total da rocha (Eq. 2.2), excluindo o efeito de poros não conectados.

$$\phi_e = \frac{V_i}{V_t} = \frac{V_t - V_m - V_{nc}}{V_t}. \quad (2.2)$$

A porosidade das unidades areno-argilosas contidas na Formação São Sebastião é controlada por fatores pré-diagenéticos, tais como: tamanho, distribuição e empacotamento dos grãos

e a forma média das partículas (porosidade primária). O histograma e o diagrama de caixa apresentados na Figura 2.2 ilustram a distribuição de porosidade dos arenitos da Formação São Sebastião calculada a partir de curvas de perfis geofísicos em um dos poços localizado no município de Araças. O diagrama de caixa, também conhecido como diagrama de extremos e quartis, é uma ferramenta gráfica que aponta a variação de dados observados de uma variável numérica através dos quartis. Neste caso, o segundo quartil aponta a mediana da porosidade das amostras, que é aproximadamente igual a 28,4%. A esquerda do limite inferior é possível verificar a ocorrência de amostras discrepantes com porosidades inferiores a 22%.

Segundo Lima (1999) os arenitos da Formação São Sebastião, obtidos em testemunhos de um poço estratigráfico, apresentam elevadas porosidades que variam entre 24% e 32,7%, compatíveis com o histograma apresentado na Figura 2.2.

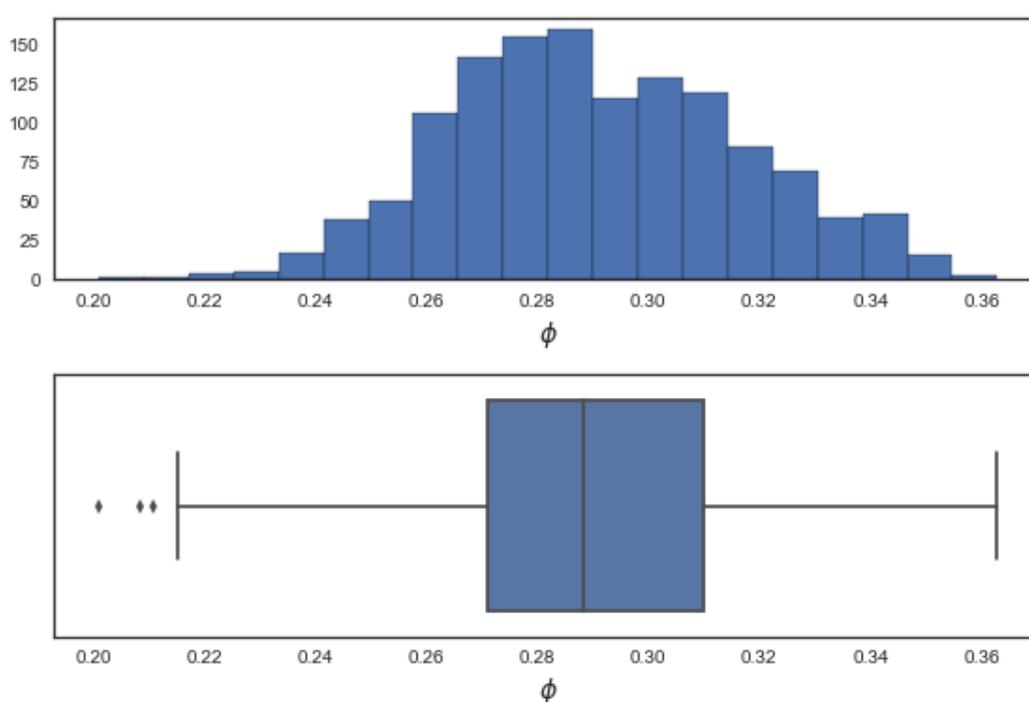


Figura 2.2: Histograma e diagrama de caixa de 1299 medidas de porosidades calculadas, a partir do perfil de densidade de arenitos da Formação São Sebastião, de um poço da área de Araças. Note que o histograma da porosidade assemelha-se a uma distribuição normal (gaussiana).

2.1.2 Permeabilidade

A permeabilidade intrínseca (k) é um parâmetro tensorial (Schön, 2015) e é definida como a medida da capacidade de um meio transmitir fluidos (óleo, água e gás) sob algum diferencial de pressão específico (Yang e WEI, 2017). Este parâmetro depende da porosidade efetiva, da

estrutura interna do espaço poroso (superfície específica e tortuosidade) e de propriedades físicas do fluido de saturação (Lima, 2014). Sua definição matemática é dada pela Eq. 2.3:

$$k = \eta \cdot \frac{\mathbf{u}}{\nabla P}, \quad (2.3)$$

em que \mathbf{u} é definido como o volume que atravessa uma área de seção transversal unitária, em um determinado tempo t - ou vetor fluxo volumétrico de um fluido;

η é a viscosidade dinâmica do fluido; e

$\nabla P = \left(\frac{\partial P}{\partial x}, \frac{\partial P}{\partial y}, \frac{\partial P}{\partial z} \right)$ é o gradiente de pressão macroscópico de um fluido.

Este ente petrofísico apresenta-se em um intervalo de várias ordens de magnitude nas rochas. Tomando-se como parâmetro a produção de óleo, a permeabilidade mínima necessária para a extração deste recurso mineral gira em torno de 0,1 md (Lucia, 1999).

Existem diversas expressões empíricas que tentam estimar a permeabilidade através de observações de parâmetros físicos das rochas, tais como: argilosidade (V_{sh}), porosidade (ϕ) e resistividade (ρ). Uma das mais populares é a expressão para o modelo de um tubo capilar, conhecida como a equação de Kozeny-Carman. Este modelo consiste em simular o comportamento hidráulico de rochas granulares através de arranjos tridimensionais de tubos capilares, similar a Figura 2.3. A Eq. (2.4) mostra a relação matemática relacionada a este modelo:

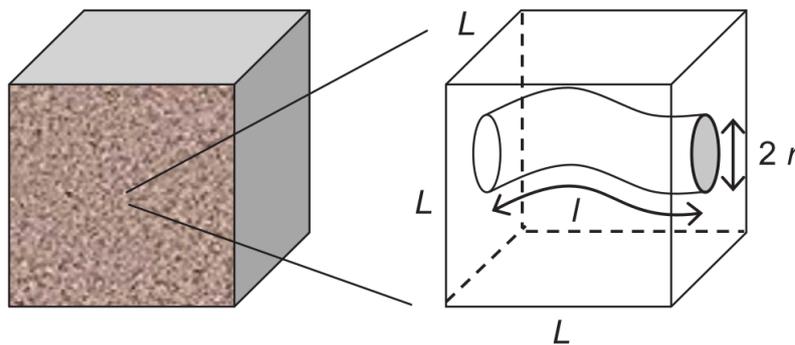


Figura 2.3: Representação do modelo de Kozeny-Carman de uma rocha porosa por tubos capilares (Schön, 2015).

$$k = \frac{1}{8} \cdot \phi \cdot \frac{r^2}{\tau^2}, \quad (2.4)$$

onde: r^2 é o raio da seção transversal ao capilar e $\tau = l/L$ é a tortuosidade média do espaço poroso.

Em se tratando de um meio poroso granular, outras muitas relações se destacam, entre elas a posposta por Lima (1995) que derivou a Eq. (2.5) a partir da teoria de Kozeny-Carman, que descreve a permeabilidade de formações areno-argilosas cujo os grãos de areia são revestidos por argila:

$$k = \alpha_0 \left(\frac{\phi^{(m-\frac{1}{2})}}{1 + \delta \cdot Q_v} \right)^q, \quad (2.5)$$

onde m é o expoente de cimentação da equação Archie (abordada na seção 2.2 - Propriedades Elétricas das Rochas), q é um expoente empírico (para poros de grãos esféricos $q = 2$), δ é um parâmetro controlado pelo tamanho médio dos grãos de areia e tipo de argila, Q_v é a concentração de troca catiônica e α_0 é um fator que depende da geometria porosa do arenito.

Schön e Georgi (2003) também desenvolveram um modelo para o cálculo da permeabilidade de formações areno-argilosas ($k_{sh|sd}$) que é descrito em função da permeabilidade da areia limpa k_{sd} e do teor de folhelho disperso V_{sh} , dada pela Eq. (2.6):

$$k_{sh|sd} = k_{sd} \left(1 - \alpha \cdot \frac{V_{sh}}{\phi} \right)^2 \quad (2.6)$$

onde α é um parâmetro empírico e ϕ a porosidade do arenito. Detalhes sobre α desta equação podem ser obtidos através do artigo escrito por Schön e Georgi (2003) listado nas referências bibliográficas deste trabalho.

2.2 Propriedades Elétricas das Rochas

As propriedades que definem o comportamento elétrico das rochas sob a ação de campos eletromagnéticos são: a resistividade elétrica ρ , a condutividade elétrica σ que é a inversa da resistividade ($\sigma = 1/\rho$), a permitividade dielétrica ϵ e a permeabilidade magnética μ (Lima, 2014). Para medições de resistividade mais comuns, a lei de Ohm, determinada pela Eq. (2.7), dá a base para a determinação da resistência (R) de um meio a qual é relacionada com a resistividade elétrica (Schön, 2015).

$$R = \rho \frac{l}{A}. \quad (2.7)$$

Ao aplicar um campo elétrico \vec{E} a uma rocha, pode-se observar dois tipos de fenômenos: condução elétrica (Eq. (2.8)) e um processo de polarização dielétrica (Eq. (2.9)), onde ambas as correntes podem ser descritas como:

$$\vec{j}_C = \sigma \vec{E} \text{ (corrente de condução),} \quad (2.8)$$

$$\vec{j}_D = \epsilon \frac{\partial \vec{E}}{\partial t} \text{ (corrente de deslocamento).} \quad (2.9)$$

Os tópicos a seguir discutem a resistividade elétrica e a sua inversa, condutividade, que são fenômenos de baixa frequência relacionados a corrente de condução.

2.2.1 Resistividade elétrica de rochas porosas

A resistividade elétrica das rochas varia em muitas ordens de magnitude e isso deve-se às variações na matriz, presença de argila, porosidade, conectividade entre os poros e conteúdo de fluido. Em 1942, Archie, embasado em diversos dados experimentais obtidos a partir de formações arenosas limpas, propôs uma relação empírica para a resistividade elétrica de uma formação. Ele sugeriu que a resistividade de uma rocha saturada (ρ_o) com uma determinada água de resistividade (ρ_w) são proporcionais através de um termo denominado fator de formação de resistividade F , como mostra a Eq. (2.10):

$$\rho_o = F \cdot \rho_w. \quad (2.10)$$

Considerando que a única forma de condução elétrica se dá através dos poros da rocha, ou seja, não há presença de minerais condutores, é possível observar que há uma relação entre o fator de formação e a porosidade:

$$F = \frac{1}{\phi^m} = \frac{\rho_o}{\rho_w}, \quad (2.11)$$

em que m é uma constante empírica relacionada ao grau de compactação e cimentação da rocha e foi chamada de "expoente de cimentação". Nota-se que esta constante tem valores aproximados de 1,3 para areias não consolidadas e entre 1,8 - 2,0 para arenitos muito consolidados.

Acrescentando-se um novo parâmetro " a " que expressa a geometria de um canal poroso à relação apontada na Eq. (2.11), pode-se escrever a seguinte relação:

$$\rho_o = \frac{a \cdot \rho_w}{\phi^m}. \quad (2.12)$$

A Eq. (2.12) pode ser generalizada se levado em consideração a saturação de água junto a outros fluidos (gás ou petróleo) no mesmo meio poroso através da Eq. (2.13) :

$$S_w = \left(\frac{\rho_w a}{\rho_t \phi^m} \right)^{\frac{1}{n}}, \quad (2.13)$$

em que ρ_t é a resistividade da rocha e n é o expoente de saturação.

Algumas formulações foram propostas com o objetivo de inserir a contribuição da argila em um meio poroso. Vale ressaltar que a presença de argilas em reservatórios de hidrocarbonetos e aquíferos diminui a qualidade em termos das propriedades hidráulicas (porosidade efetiva e permeabilidade), além de criar uma componente adicional de condutividade ao meio.

Poupon (1954) derivou uma equação para a resistividade de rochas arenosas que apresentam folhelhos laminados em sua composição, que mostra uma semelhança a um circuito de resistividades em paralelo:

$$\frac{1}{\rho_t} = \frac{1 - V_{sh}}{\rho_{sd}} + \frac{V_{sh}}{\rho_{sh}}, \quad (2.14)$$

e levando-se em consideração a presença de mais de um fluido no meio poroso, tem-se:

$$\frac{1}{\rho_t} = \frac{1 - V_{sh}}{\rho_w} \left(\phi^m \cdot S_w^m + \frac{V_{sh}}{\rho_{sh}} \right), \quad (2.15)$$

ρ e ρ_{sh} são as resistividades da matriz arenosa e da argila da formação respectivamente.

Waxman e Smits (1967) criaram uma equação levando em consideração a presença de argila dispersa com base em processos de troca catiônica na interface argila-água:

$$\frac{1}{\rho_o} = \frac{1}{F^*} \left(\frac{1}{\rho_w} + B \cdot Q_v \right), \quad (2.16)$$

onde

F^* é o fator de formação do arenito argiloso;

B é a condutividade equivalente de cátions de troca sódio-argila (em $S.cm^2.meq^{-1}$);

Q_v é um fator de argilosidade (proporcional a capacidade de troca catiônica (CEC) por unidade de volume); e

Considerando ρ_m a densidade de uma substância mineral em $g.cm^{-2}$, Q_v é expressa pela Eq. (2.17):

$$Q_v = CEC \frac{1 - \phi}{\phi} \rho_m \quad (2.17)$$

Simandoux (1963) propôs uma equação que leva em consideração um modelo areno-argiloso contendo argilas estruturais e dispersas:

$$\frac{1}{\rho_t} = \frac{\phi^m}{a \cdot \rho_w} S_w^n + \frac{V_{sh} \cdot S_w}{\rho_{sh}}, \quad (2.18)$$

com os parâmetros já especificados.

3

Perfilagem Geofísica de Poços

Segundo Nery (2013) a perfuração de poços constitui a última etapa da prospecção, seja para água subterrânea ou quaisquer outros alvos como hidrocarbonetos, minerais ou estudos geotécnicos e ambientais. Sondagens são introduzidas num furo aberto, por meio de um cabo e conexões elétricas, com o objetivo de coletar, discretamente, informações acerca de propriedades físicas das formações geológicas em subsuperfície, tais como: resistividade, potencial elétrico espontâneo, contagem de radiação gama, densidade eletrônica, porosidade neutrônica, entre outras, permitindo o cálculo de propriedades indiretas (que não são obtidas diretamente a partir dos sensores) bem como a correlação litológica com auxílio de amostras de calhas e testemunhos.

O objetivo dessa seção é apresentar as bases teóricas dos perfis utilizados para a interpretação geofísica do sistema aquífero Recôncavo.

3.1 Potencial Espontâneo (SP)

Segundo Serra (1984), Schlumberger descobriu, em 1928, que existia uma diferença de potencial elétrico, espontaneamente, entre um eletrodo posicionado dentro de um poço e um eletrodo remoto situado na superfície, o que gerou posteriormente o perfil de potencial espontâneo. Esse potencial pode ser atribuído a três processos envolvendo movimentos ordenados dos íons das soluções aquosas (Serra, 1984):

- **Potencial de junção líquida (E_j)** - este potencial é estabelecido no contato direto entre a lama e a água nativa contida na formação invadida. Íons de Na^+ e Cl^- , nas soluções, têm diferentes mobilidades, e, portanto, não se difundem com as mesmas

taxas entre os dois eletrólitos. Geralmente os cátions são menos móveis por causa de seu tamanho maior e uma afinidade com moléculas de água. Conseqüentemente ocorre uma separação de ânions na solução mais diluída, e cátions na solução mais concentrada. O potencial que se estabelece nessa junção é:

$$E_j = 2,3 \frac{v-u}{v+u} \frac{RT}{F} \log \left(\frac{a_w}{a_{mf}} \right), \quad (3.1)$$

onde:

- R - Constante universal dos gases (8,314 joules/ $K - mol$);
 - T - Temperatura absoluta (em *Kelvin*);
 - F - Constante de Faraday (96.489 coulombs/ mol);
 - a_w - Atividade iônica da água de formação;
 - a_{mf} - Atividade iônica do filtrado de lama;
 - v - Mobilidade dos ânions ($67,6 \times 10^{-5} cm^2/sV$)
 - u - Mobilidade dos cátions ($45,6 \times 10^{-5} cm^2/sV$)
 - $c_1 = 2,3 \frac{v-u}{v+u} \frac{RT}{F} = 11,6 mV$ a $25^\circ C$ ($77^\circ F$).
- **Potencial de membrana (E_m)** - As argilas são filosilicatos cujos defeitos em suas redes cristalinas (normalmente a substituição de Si por Al em tetraedros SiO_4) criam uma carga superficial negativa em suas superfícies folheadas. Como resultado, íons de sinais contrários (chamados contra-íons) compensam essas cargas superficiais através de estruturas de duplas-camadas (Lima, 2014). Quando partículas de argilas com soluções eletrolíticas são aproximadas por compactação, suas duplas camadas se interpenetram formando uma membrana filtradora de íons que permite a passagem de cátions mas impedem a passagem dos ânions. O potencial de membrana é dado por:

$$E_m = 2,3 \frac{RT}{F} \cdot \log \left(\frac{a_w}{a_{mf}} \right), \quad (3.2)$$

onde a constante $c_2 = 2,3RT/F$ é igual a 59,1 mV na temperatura de $25^\circ C$ ($77^\circ F$).

- **Potencial eletrocínético (E_k)** - aparece quando o filtrado, a partir da lama de perfuração, é forçado a invadir as formações pela pressão diferencial entre a coluna de lama e a formação. Este fluxo ocorre: (a) através do reboco em frente a formação permeável; (b) através da unidade permeável sendo invadida e (c) através de camadas de folhelhos.

Geralmente o potencial eletrocinético através do reboco é compensado pelo potencial eletrocinético no folhelho. Por isso, assume-se que o potencial medido num poço é apenas relacionado aos potenciais eletroquímicos (E_m e E_j), exceto em algumas situações, segundo Serra (1984), tais como: formações depletadas, lamas especiais, lamas pesadas e formações de baixa permeabilidade.

A ferramenta do SP constitui-se de um circuito simples onde um milivoltímetro é interconectado por cabo a um eletrodo posicionado dentro do poço e a outro fixo na superfície. As diferenças de potenciais medidas são da ordem de dezenas a centenas milivolts. No gráfico do potencial espontâneo *versus* a profundidade ocorrem deflexões negativas e positivas nas zonas permeáveis, levando-se em consideração a linha base de folhelhos, que depende da razão entre a resistividade do filtrado (R_{mf}) e a resistividade do fluido da formação (R_f).

Quantitativamente, quanto maior a deflexão de SP maior o contraste de salinidade entre R_{mf} e R_w (Rider, 2002). Em resumo, quando ocorre uma deflexão positiva $R_{mf} < R_f$, analogamente, quando há uma deflexão negativa $R_{mf} > R_f$. Esta relação permite utilizar uma expressão matemática que relaciona a amplitude da deflexão do SP ($SSP = E_j + E_m$) expressa em termos da resistividade da água de formação da seguinte forma:

$$SSP = -K \log \left(\frac{R_{mf}}{R_w} \right), \quad (3.3)$$

onde K é um coeficiente dependente dos sais em solução e da temperatura. Para soluções de $NaCl$ $K = 65 + 0,24 \text{ } ^\circ C$. A partir da Eq. (3.3), podemos obter a resistividade da água de formação (R_w) como:

$$R_w = R_{mf} \times 10^{-SSP/K}. \quad (3.4)$$

O R_{mf} normalmente é expresso em relação à temperatura de superfície e calculado para a temperatura na profundidade da formação que se deseja determinar R_w (Nery, 2013).

Considerando que a deflexão do SP aponta se uma formação é permeável ou não, é possível calcular a argilosidade (V_{shSP}) utilizando a seguinte fórmula:

$$V_{shSP}(\%) = \left(1 - \frac{PSP}{SSP} \right) \times 100. \quad (3.5)$$

PSP é o potencial medido no meio argiloso e SSP é o potencial estático no arenito mais limpo da seção atravessada no poço. Ambos (PSP e SSP) tem a mesma unidade de medida de SSP (mV).

3.2 Perfil de Indução Elétrica

Segundo Rider (2002), os perfis de resistividade, inicialmente desenvolvidos na pesquisa de hidrocarbonetos, são fundamentais para a estimativa direta da resistividade verdadeira da formação e da zona invadida (R_t e R_{xo}), e servem como parâmetros de uma série de fórmulas petrofísicas para obtenção de propriedades físicas indiretas.

O perfil de indução elétrica, considerado um perfil de corrente focalizada, se sobressai em relação aos arranjos convencionais (normal e lateral) devido a sua maior resolução vertical (detecção de camadas pouco espessas) e por ser menos sensível aos efeitos de poço (lama e zona invadida).

É uma ferramenta de princípio eletromagnético composta por bobinas emissoras e receptoras. Um oscilador produz uma corrente alternada de alta frequência resultando em um campo eletromagnético que induzirá uma corrente secundária na formação (efeito de Foucault). Este processo gera um campo eletromagnético secundário que é detectado por uma bobina receptora. Uma vez que a corrente da bobina emissora tem frequências e amplitudes constantes, a intensidade das correntes induzidas de Foucault na formação é proporcional a condutividade ($\sigma_t = 1/R_t$) do meio, permitindo assim a obtenção desta propriedade.

É importante ressaltar o efeito denominado efeito pelicular (*skin effect*), cuja profundidade de penetração do campo eletromagnético é dada por:

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}}, \quad (3.6)$$

onde δ é a profundidade (m) na qual 63% da amplitude do sinal é atenuada, ω é a frequência angular do sinal, μ é a permeabilidade magnética e σ é a condutividade. A correção para o efeito pelicular é automaticamente realizada na atividade da perfilagem, com exceções de ferramentas muito antigas, onde a correção é realizada através de cartas de correção (Serra, 1984).

Segundo (Nery, 2013) esta ferramenta é indicada para poços com lamas não condutivas, mas apresenta problemas relacionados a poços que são perfilados com lamas muito condutivas (salinidade superior a 30000 ppm de NaCl) e/ou formações que apresentem resistividades muito altas.

A resistividade medida em distribuições areno-argilosas mostram uma forte correlação com as medidas de raios gama total (GR) quando a resistividade da água de formação é relativamente constante. Isto torna possível calcular a argilosidade em formações areno-argilosas a partir da Eq. (3.7):

$$V_{shR} = \left(\frac{\log(R_{max}) - \log(R_m)}{\log(R_{max}) - \log(R_{min})} \right), \quad (3.7)$$

Os valores de R_{max} e R_{min} dependem de uma constante n , onde para $n = 1$:

$$V_{shR} = \left(\frac{\log(R_s) - \log(R_m)}{\log(R_s) - \log(R_{sh})} \right), \quad (3.8)$$

caso contrário, $n = -1$:

$$V_{shR} = \left(\frac{\log(R_{sh}) - \log(R_m)}{\log(R_{sh}) - \log(R_s)} \right), \quad (3.9)$$

onde R_s é a resistividade do arenito limpo, R_m é a resistividade medida e R_{sh} é a resistividade do folhelho. Quando $n = 1$ significa que o valor médio da resistividade dos arenitos é maior que o do folhelho, caso contrário, $n = -1$. A definição de n pode ser realizada com o auxílio de perfis litológicos. Utiliza-se o logaritmo da resistividade para estimar a argilosidade devido ao comportamento log-normal da resistividade elétrica nas formações.

3.3 Perfil de Raios Gama

O perfil de Raios Gama é uma ferramenta tradicionalmente empregada para analisar formações areno-argilosas (Ellis e Singer, 2007). Esta ferramenta faz o registro da radiação gama natural total, normalmente proveniente de três principais tipos de fontes radioativas: potássio (^{40}K), Tório (^{232}Th) e Urânio (^{238}U) presentes nos minerais das rochas (Serra, 1984; Schön, 2015). Devido a maior concentração de potássio nos folhelhos, se comparado a arenitos, carbonatos e sais não potássicos, sua principal aplicação é na distinção de folhelhos e não-folhelhos e por consequência é utilizado na estimativa do conteúdo de argila, também conhecido como argilosidade (V_{sh}).

A estimativa da argilosidade é realizada através do índice de Raios Gama, Eq.(3.10), e consiste em uma normalização da medida sendo dada por:

$$I_{GR} = \frac{GR_m - GR_{min}}{GR_{max} - GR_{min}}, \quad (3.10)$$

onde GR_m é o valor medido pelo instrumento, GR_{min} e GR_{max} são os valores mínimos e máximos registrados ou obtidos por algum critério. Segundo Vernik (2008), os parâmetros GR_{min} e GR_{max} podem ser obtidos de forma acurada através da análise de testemunhos e,

na ausência destes, Vernik (2016) estabeleceu que os parâmetros para uma formação areno-argilosa podem ser estimados a partir de um histograma em que a distribuição das medidas de raios gama se comportam de forma bimodal, apresentando duas gaussianas bem definidas em um formato de "sino", similar a Figura (3.1).

Larionov (1969) e Thomas e Stieber (1975) proporam correções para o calculo da argilosidade levando-se em consideração a idade das rochas.

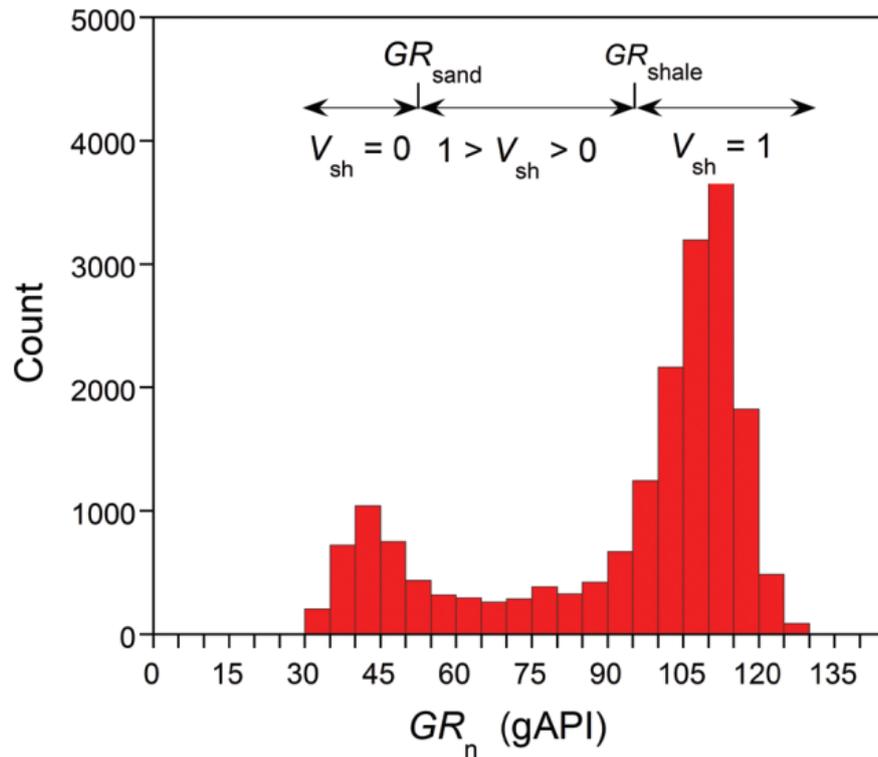


Figura 3.1: Histograma de uma sequência siliciclástica Terciária. Os valores $GR_{min} = 52$ gAPI and $GR_{max} = 93$ gAPI, sugeridos por Vernik (2016), foram selecionados, visualmente, com base no histograma.

3.4 Perfil Sônico

O uso da energia acústica para produzir uma imagem de subsuperfície tem uma longa história. Prospecção sísmica de reflexão consiste em usar uma fonte acústica de baixa frequência na superfície para criar pulsos de energia que são parcialmente refletidos aproximadamente diretamente abaixo da fonte. Entretanto foi a necessidade de acessar a velocidades das camadas que impulsionou a criação de ferramentas sônicas para poços (Ellis e Singer, 2007).

A ferramenta mais simples do perfil Sônico usa um transmissor de frequência constante, ultrassônica baixa e dois receptores (Nery, 2013). O funcionamento deste aparato se dá da seguinte maneira: uma onda elástica de alta amplitude é enviada radialmente pelo trans-

missor e se espalha pela rocha até acionar dois receptores posicionados a distâncias fixas conhecidas.

A velocidade da onda acústica varia de acordo com meio em que ela se propaga, sendo mais rápida em meios sólidos, cuja densidade é maior, do que nos fluidos. Em vista disso, como a velocidade e o tempo de propagação têm uma relação inversa, quanto maior a presença de fluidos na rocha, maior o tempo de propagação de uma onda sônica.

Devido a relação existente entre velocidade da onda e proporção de fluidos e sólidos em uma camada, pode-se estimar a porosidade da rocha através de uma simples relação envolvendo tempos de trânsito. Na Eq (3.11) a porosidade calculada pelo perfil Sônico (ϕ_S) é função do tempo de propagação na matriz sólida (Δt_m), do tempo de propagação no fluido (Δt_f) e do tempo registrado no ponto de interesse (Δt).

$$\phi_S = \frac{\Delta t - \Delta t_m}{\Delta t_f - \Delta t_m} \quad (3.11)$$

A Eq. (3.11) é derivada da equação do tempo médio, proposta Wyllie et al. (1956), e apenas calcula porosidades realistas quando $S_w = 1$ e $V_{sh} = 0$.

Para rochas não compactadas, é necessário aplicar uma correção devido à atenuação em forma de leve estiramento (*stretch*) que as ondas sofrem ao se propagarem por elas. Essa correção é executada quando o folhelho sotoposto à camada de interesse tem tempo médio (Δt_{sh}) superior a 100 $\mu s/pé$. A correção é dada pela Eq. (3.12):

$$\phi_{Scor} = \phi_S \cdot \frac{1}{C_p} \cdot \frac{100}{\Delta t_{sh}} \quad (3.12)$$

onde ϕ_{Scor} é a porosidade do perfil Sônico corrigida para a ausência de compactação e C_p é uma constante.

3.5 Perfil de Densidade

O perfil de Densidade, que tem princípio radioativo assim como o perfil de Raios Gama, registra as variações de densidade das camadas, geralmente, em g/cm^3 . Para efetuar a medição, a ferramenta é pressionada contra a parede do poço com um patim metálico portando uma fonte radioativa, o Césio 137, com nível energético da ordem de 0,622 MeV (Serra, 1984; Ellis e Singer, 2007).

De acordo com Serra (1984), existem três tipos de interação entre os raios gamas emitido

pela ferramenta e a matéria, que dependem tanto do material em questão quanto da energia do fóton:

- **Efeito Fotoelétrico** - choque entre um fóton de baixa energia (< 75 keV) e um átomo;
- **Efeito de Produção de Par** - choque de um fóton de alta energia ($> 1,02$ MeV) na vizinhança de um núcleo; e
- **Efeito Compton** - choque entre um fóton de média energia e um átomo qualquer.

Dentre os efeitos citados acima, o Compton é a interação primacial entre os raios gama e as rochas, pois a fonte utilizada está dentro de um nível energético compatível.

De acordo com Nery (2013), um feixe monoenergético de raios gama, de intensidade fixa, ao sair da fonte, choca-se sucessivamente com os elétrons da formação por meio do efeito Compton. À proporção que os raios gama vão se dispersando ou sendo absorvidos, a intensidade do feixe inicial diminui. Essa diminuição de densidade, que é uma função da mudança de densidade eletrônica do meio, é então medida pelo detector. Assim, quanto mais densa for a rocha, menor a intensidade da radiação no detector, e vice-versa.

Semelhante ao perfil Sônico, a porosidade da formação também pode ser estimada através do perfil de Densidade. A expressão usada para tal é dada pela Eq. 3.13:

$$\phi_D = \frac{\rho_m - \rho_B}{\rho_m - \rho_f} \quad (3.13)$$

em que ϕ_D é a porosidade calculada, ρ_m é a densidade da matriz da rocha, ρ_f é a densidade do fluido da zona investigada e ρ_B é a densidade registrada pela ferramenta.

4

Análise de Agrupamento

Dados de diferentes tipos (medidas e observações) descrevem as características de uma determinada espécie, resultados de um experimentos científico, diagnósticos de doenças, entre outros eventos. Essas informações são responsáveis por prover a base para analisar e entender todos os tipos de objetos ou fenômenos (Xu e Wunsch, 2009). Em geofísica, estes dados podem ser vistos como medidas relativas a propriedades físicas das rochas que intrinsecamente revelam características petrofísicas de facies, formação, ou uma sequência sedimentar.

A análise de agrupamento (popularmente conhecida como *Cluster Analysis*) tem como objetivo encontrar estruturas onde os dados pertencentes a cada grupo (*cluster*) compartilham alguma(s) característica(s) relevante(s) para o domínio do problema em estudo, com base em alguns critérios estatísticos (Faceli et al., 2011).

Este capítulo aborda os algoritmos de agrupamentos utilizados para as análises de propriedades físicas mensuradas por meio de perfilagens geofísicas de poços que perpassam os arenitos e folhelhos contidos na Formação São Sebastião e em alguns casos na Marizal, que constitui o sistema aquífero Recôncavo.

4.1 Algoritmos de Agrupamento

4.1.1 K-médias (*K-means*)

O método K-médias é designado para identificar grupos que não contêm nenhum tipo de rótulo, sendo considerado uma técnica de aprendizado de máquina não-supervisionada. Uma vez designado o número de grupos que se deseja encontrar, o algoritmo realiza uma série de

iterações de forma a agrupar medidas que são similares entre si relacionadas a um centroide. A ideia por trás deste algoritmo é que bons grupos são aqueles que tem a menor variância possível intra-grupo (James et al., 2013).

Suponha que exista um conjunto de dados designados pela matriz das observações \mathbf{X} :

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1D} \\ x_{21} & x_{22} & \cdots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{ND} \end{bmatrix} \quad (4.1)$$

onde D é o número de atributos (neste trabalho é o número de ferramentas utilizadas na perfilagem geofísica) e N é o número de observações (medidas). O objetivo é particionar este dado em um certo número K de grupos ou fácies. Cada grupo compartilha características em comum a um dado centroide, $\boldsymbol{\mu}_k$, que é capaz de minimizar a variância intra-cluster das amostras pertencentes a um mesmo conjunto. O centroide é definido como um vetor que contém o valor médio das medidas de um determinado grupo k , com $k = 1, \dots, K$, representando assim o centroide de cada grupo (Bishop, 2006). A Figura 4.1 mostra uma representação gráfica para duas distribuições (sintéticas) de amostras de facies arenitos e folhelhos apontando os respectivos *clusters* e centroides para observações realizadas com perfis de raios gama e resistividade elétrica indutiva.

A variância acumulada intra-grupo para os K grupos é definida por:

$$Var = \frac{1}{N} \sum_{\substack{n=1 \\ \forall x_n \in K}}^{N_k} \sum_{k=1}^K \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2, \quad (4.2)$$

em outras palavras, pode ser vista como o somatório da distância euclideana, para toda amostra \mathbf{x}_n pertencente ao grupo k , entre cada componente de um grupo e seu centroide, dividido pelo número total de observações. Outros tipos de métricas de distâncias podem ser utilizadas, tais como as de Mahalanobis e Manhattan.

O problema de otimização a ser obtido pelo algoritmo de agrupamento K-médias para um número de agrupamentos K , em que C_k designa o agrupamento k , é estabelecido por (Eq. 4.3):

$$\underset{C_1, \dots, C_K}{\text{minimizar}} \left(Var = \frac{1}{N} \sum_{\substack{n=1 \\ \forall x_n \in K}}^{N_k} \sum_{k=1}^K \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2 \right). \quad (4.3)$$

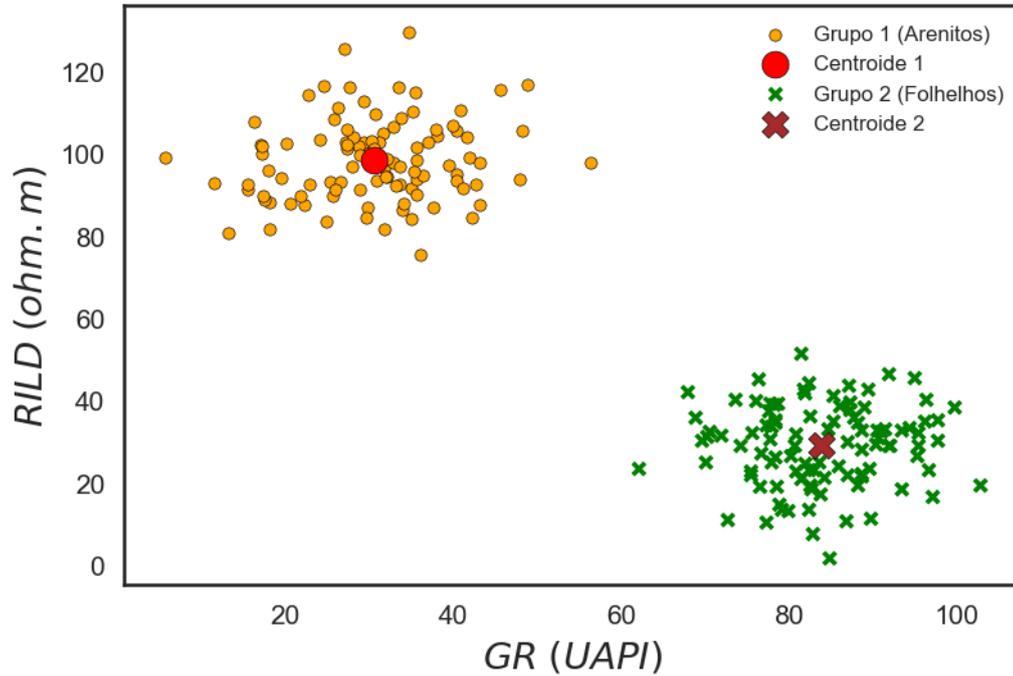


Figura 4.1: Representação gráfica de observações sintéticas referentes a dois grupos: arenitos (círculos laranjas) e folhelhos (cruzes verdes), e seus respectivos centroides.

Esta técnica de agrupamento pode ser descrita através dos passos enumerados a seguir e pela Figura 4.2.

1. Inicializa-se o algoritmo atribuindo rótulos numéricos aleatórios para cada medida, simbolizando, inicialmente, o grupo que cada medida ou amostra representa;
2. Calcula-se os centroides, μ_k , de todos esses grupos através da expressão:

$$\mu_k = \frac{1}{N_k} \sum_{\substack{j=1 \\ \forall x_n \in K}}^{N_k} \mathbf{x}_j, \quad (4.4)$$

3. Obtém-se a variância desse agrupamento inicial ($Var(i)$);
4. Atribui-se a cada medida o rótulo numérico referente ao grupo de centroide mais próximo;
5. Recalcula-se os valores dos centroides, com base nos novos grupos, utilizando a Eq. (4.4), e obtém-se a variância do novo agrupamento ($Var(i+1)$);
6. Se $Var(i+1) < Var(i)$: atualiza-se as variâncias e os centroides e retorna-se para o passo 4. Caso contrário, assume-se o ultimo agrupamento como o resultado final.

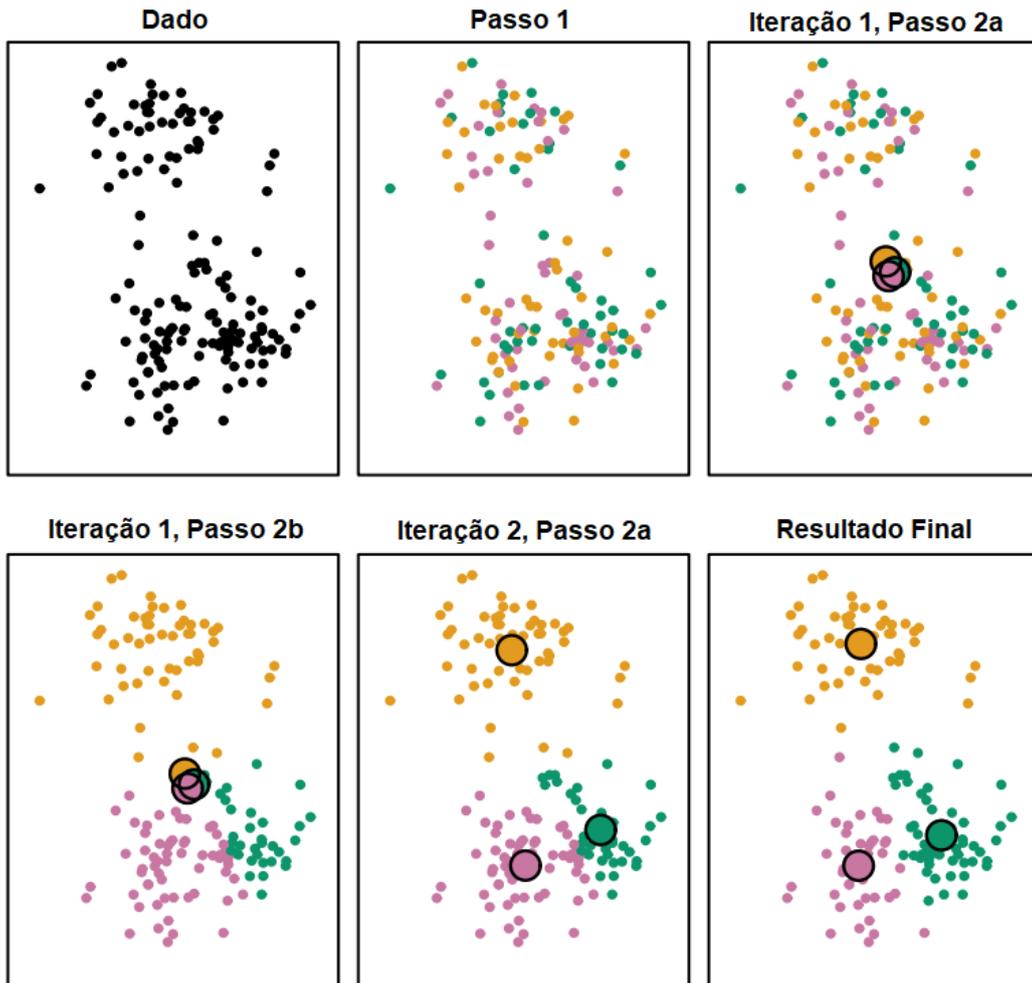


Figura 4.2: Execução do algoritmo K-médias para subdividir um conjunto de dados em três partições (Adaptado de James et al., 2013).

Devido a inicialização aleatória, faz-se necessária a execução do algoritmo diversas vezes visando a obtenção de grupos que contabilizem a menor variância intra-cluster possível.

4.1.2 Modelos de Misturas Gaussianas (GMM)

Distribuição Gaussiana

A gaussiana, também conhecida como distribuição normal, é largamente utilizada em modelos de distribuição de variáveis contínuas (Bishop, 2006). Considerando uma única variável x , pode-se escrever a equação da gaussiana da seguinte forma:

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\}, \quad (4.5)$$

onde μ é a média e σ^2 é a variância que é igual ao desvio padrão ao quadrado da distribuição de probabilidade $\mathcal{N}(x|\mu, \sigma^2)$. A Figura (4.3) mostra um histograma de distribuição em frequência e a função densidade de probabilidade da variável x , descrita por uma distribuição gaussiana.

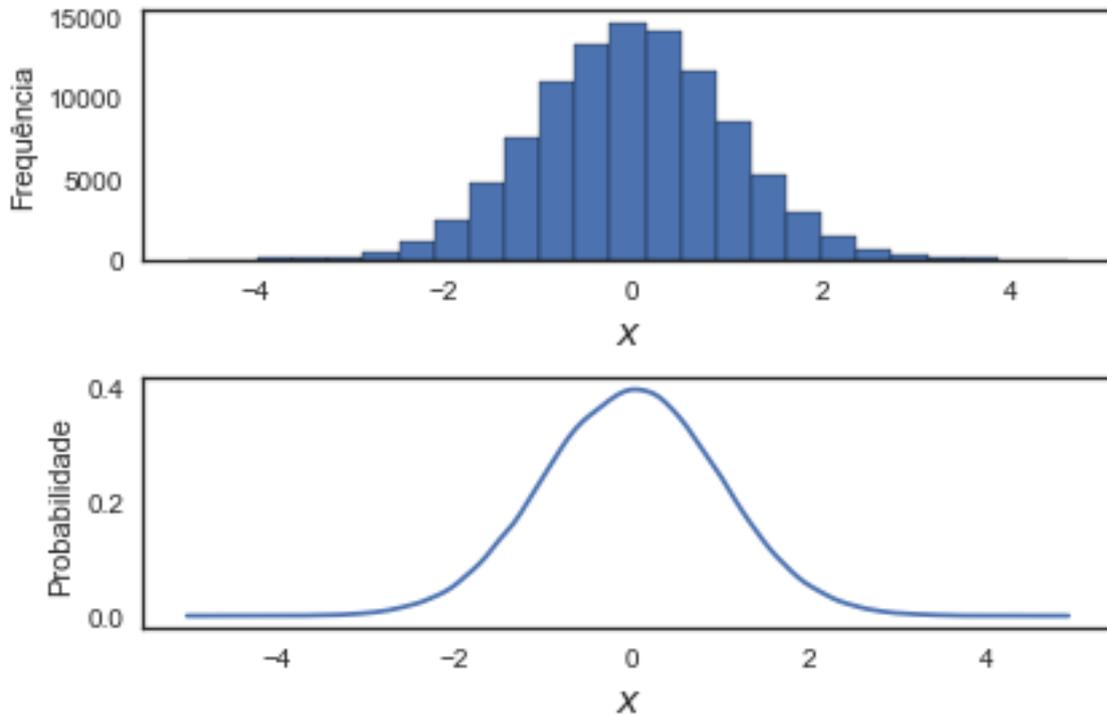


Figura 4.3: Histograma e a função densidade de probabilidade de uma distribuição aproximadamente normal de uma variável x com os parâmetros: $\mu \cong 0$ e $\sigma \cong 1$.

Considerando novamente a matriz de observações \mathbf{X} (Eq. (4.1)), a distribuição normal deste conjunto de dados é obtida por:

$$\mathcal{N}(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})\right\}, \quad (4.6)$$

onde $\boldsymbol{\mu}$ é a média D -dimensional da matriz \mathbf{X} , $\boldsymbol{\Sigma}$ é a matriz de covariância da distribuição e $|\boldsymbol{\Sigma}|$ denota o determinante da matriz de covariância.

A matriz de covariância $\boldsymbol{\Sigma}$, também conhecida como matriz de dispersão, generaliza a noção de variância para múltiplas dimensões. Definindo-se E como um operador que denota a média esperada para um respectivo elemento, a covariância pode ser escrita como:

$$\boldsymbol{\Sigma}_{ij} = \Sigma(\mathbf{x}_i, \mathbf{x}_j) = E[(\mathbf{x}_i - \mu_i)(\mathbf{x}_j - \mu_j)] = E[\mathbf{x}_i \mathbf{x}_j] - \mu_i \mu_j, \quad (4.7)$$

e a matriz de covariância é dada por:

$$\Sigma = \begin{bmatrix} E[(\mathbf{x}_1 - \mu_1)(\mathbf{x}_1 - \mu_1)] & E[(\mathbf{x}_1 - \mu_1)(\mathbf{x}_2 - \mu_2)] & \cdots & E[(\mathbf{x}_1 - \mu_1)(\mathbf{x}_D - \mu_D)] \\ E[(\mathbf{x}_2 - \mu_2)(\mathbf{x}_1 - \mu_1)] & E[(\mathbf{x}_2 - \mu_2)(\mathbf{x}_2 - \mu_2)] & \cdots & E[(\mathbf{x}_2 - \mu_2)(\mathbf{x}_D - \mu_D)] \\ \vdots & \vdots & \ddots & \vdots \\ E[(\mathbf{x}_D - \mu_D)(\mathbf{x}_1 - \mu_1)] & E[(\mathbf{x}_D - \mu_D)(\mathbf{x}_2 - \mu_2)] & \cdots & E[(\mathbf{x}_D - \mu_D)(\mathbf{x}_D - \mu_D)] \end{bmatrix}. \quad (4.8)$$

Algoritmo GMM

Suponha um evento constituído por várias funções probabilidades, composto por K gaussianas (distribuição de misturas gaussianas), que pode ser escrito como uma superposição linear de gaussianas (função densidade de probabilidade conjunta) na forma:

$$p(\mathbf{X}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{X} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (4.9)$$

onde π_k é um peso atribuído para ocorrência de cada distribuição normal, e atende os seguintes critérios:

$$0 \leq \pi_k \leq 1, \quad (4.10)$$

e

$$\sum_{k=1}^K \pi_k = 1. \quad (4.11)$$

A Figura (4.4) ilustra uma distribuição de misturas de três gaussianas unidimensionais e o valor corresponde de π para cada função probabilidade.

É possível calcular a probabilidade de que uma determinada amostra pertença a uma das distribuições normais da mistura através da responsabilidade (r_k):

$$r_k = \frac{\pi_k \mathcal{N}(\mathbf{X} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{X} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}, \quad (4.12)$$

Dada a função de máxima verossimilhança (*likelihood*):

$$\log p(\mathbf{X} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^N \log \left(\sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right), \quad (4.13)$$

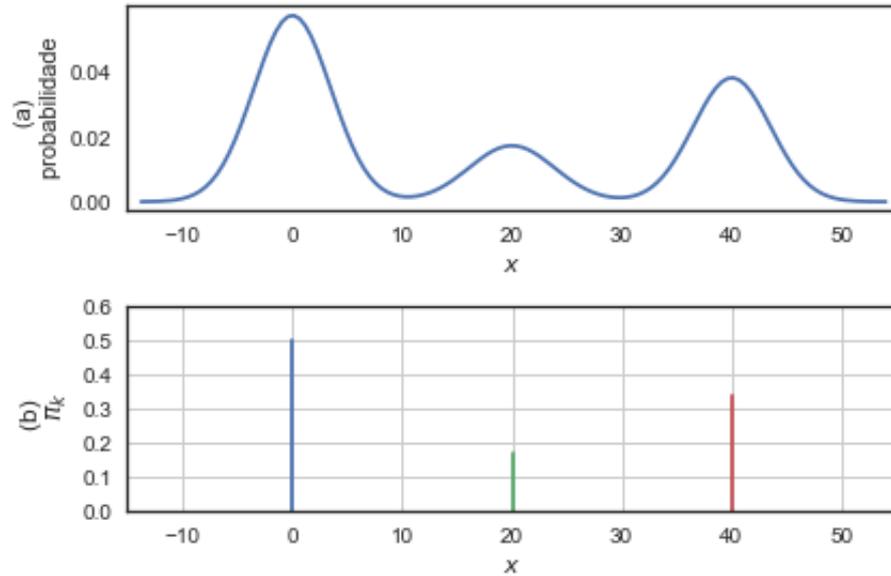


Figura 4.4: (a) Representação visual de um modelo de misturas gaussianas (GMM) contendo três gaussianas representando a probabilidade de ocorrência de um evento e (b) a representação do valor π_k das respectivas distribuições.

o objetivo deste algoritmo é maximizar esta função a cada iteração.

Utilizando, novamente, um conjunto de dados similar a matriz \mathbf{X} (Eq. (4.1)) onde D é o número de atributos contendo N amostras, o algoritmo utilizado neste trabalho pode ser dividido nos seguintes passos:

1. Os parâmetros do modelo inicial π_k , $\boldsymbol{\mu}_k$ e $\boldsymbol{\Sigma}$ são obtidos a partir do algoritmo k-médias;
2. (início do laço) Avalia-se as responsabilidades utilizando os parâmetros obtidos a partir do algoritmo k-médias:

$$r_k = \frac{\pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \quad (4.14)$$

3. Em seguida atualiza-se os parâmetros π_k , $\boldsymbol{\mu}_k$ e $\boldsymbol{\Sigma}_k$ das distribuições normais através das expressões:

$$\pi_k = \frac{1}{N} \sum_{n=1}^N r_k = \frac{N_k}{N}, \quad (4.15)$$

$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^N r_{n,k} \cdot \mathbf{x}_n}{N_k}, \quad (4.16)$$

e

$$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^N r_{n,k} (\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^T. \quad (4.17)$$

4. Por fim, avalia-se se a diferença entre a função máxima verossimilhança na iteração i e $i - 1$ é menor que uma determinada tolerância (ϵ) que é definida ao iniciar o algoritmo:

$$\log P_i - \log P_{i-1} < \epsilon, \quad (4.18)$$

se esta condição for atendida, o algoritmo é finalizado, sendo possível estimar os grupos a partir das responsabilidades obtidas na última iteração.

4.2 Validação de Grupos

A avaliação dos grupos obtidos em uma análise de agrupamento é o principal escopo da atividade. Neste processo as seguintes propriedades podem ser examinadas: densidade, tamanho, forma e separação dos grupos e robustez na implementação de sistemas de recomendação. As subseções a seguir apresentam os índices de validação utilizados para o julgamento dos agrupamentos obtidos.

4.2.1 Índice Rand Ajustado

O Índice Rand Ajustado (*ARI*) é um critério de validação externo que permite levar em consideração informações *a priori* sobre os grupos interpretados. Dada uma partição \mathbf{P} de um dado conjunto de dados (informações reais e/ou dados interpretados por um especialista) e o resultado do algoritmo que é um agrupamento, \mathbf{C} , interpretado, pode-se avaliar o desempenho do agrupamento comparando \mathbf{C} com \mathbf{P} (Xu e Wunsch, 2009). Considere que a partição (\mathbf{P}) cedida tem uma quantidade de grupos r ($\mathbf{P} = (P_1, P_2, \dots, P_r)$) e o agrupamento (\mathbf{C}) contém s grupos ($\mathbf{C} = C_1, C_2, \dots, C_s$). Pode-se, então, criar uma tabela de concordância entre a partição e o agrupamento:

| | C_1 | C_2 | ... | C_s | a_r |
|----------|-----------------------------|-----------------------------|----------|-----------------------------|-----------------------------|
| P_1 | n_{11} | n_{12} | \vdots | n_{1s} | $a_1 = \sum_{j=1}^s n_{1j}$ |
| P_2 | n_{21} | n_{22} | \vdots | n_{2s} | $a_2 = \sum_{j=1}^s n_{2j}$ |
| \vdots | \vdots | \vdots | \ddots | \vdots | \vdots |
| P_r | n_{r1} | n_{r2} | ... | n_{rs} | $a_r = \sum_{j=1}^s n_{rj}$ |
| b_s | $b_1 = \sum_{i=1}^r n_{i1}$ | $b_2 = \sum_{i=1}^r n_{i2}$ | \vdots | $b_s = \sum_{i=1}^r n_{is}$ | |

Tabela 4.1: Matriz de concordância entre a Partição \mathbf{P} e Agrupamento \mathbf{C} .

onde b_s e a_r são os somatórios das linhas e colunas da matriz de concordância.

O índice Rand ajustado (ARI) é calculado usando a seguinte expressão:

$$ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}] / \binom{n}{2}}{\frac{1}{2}[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}] - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}] / \binom{n}{2}} \quad (4.19)$$

ou de maneira mais simplificada, pode-se escrever:

$$ARI = \frac{I - I_e}{\max(I) - I_e}, \quad (4.20)$$

onde I é o índice indicado pelo agrupamento e I_e o índice esperado. Os valores do ARI estão entre 0 e 1, tal que quanto mais próximo de 1, melhor é o agrupamento obtido pelo algoritmo de clusterização.

4.2.2 Índice de Silhueta

O índice (ou coeficiente) silhueta é uma técnica de validação utilizada para extrair padrões de algoritmos de agrupamento. O principal propósito deste índice é selecionar um número apropriado de grupos a partir de um conjunto de dados quando não há nenhuma informação *a priori* e/ou comparar diferentes métodos de agrupamento e parâmetros utilizados (Rousseeuw, 1987). O coeficiente silhueta é calculado da seguinte forma:

$$s(\mathbf{x}_i) = \frac{b(\mathbf{x}_i) - a(\mathbf{x}_i, C_i)}{\max\{a(\mathbf{x}_i, C_i), b(\mathbf{x}_i)\}} \quad (4.21)$$

onde \mathbf{x}_i é uma amostra do conjunto de dados. A função $a(\mathbf{x}_i, C_i)$ calcula a distância média intra-grupos que é a distância média entre \mathbf{x}_i e todos os outros objetos colocados no mesmo grupo C_i , como é apresentado na Eq. (4.22):

$$a(\mathbf{x}_i, C_i) = \frac{1}{|C_i|} \sum_{\mathbf{x}_i, \mathbf{x}_j \in C_i, \mathbf{x}_i \neq \mathbf{x}_j} d(\mathbf{x}_i, \mathbf{x}_j), \quad (4.22)$$

em que $d(\mathbf{x}_i, \mathbf{x}_j)$ é a métrica de distância adotada para calcular a dissimilaridade entre os pares de objetos.

Na Eq. 4.21 a função $b(\mathbf{x}_i)$ é utilizada para calcular a menor distância média entre \mathbf{x}_i e todos os outros grupos, sem considerar o grupo o qual a amostra pertence. Esta função é definida pela Eq. (4.23):

$$b(\mathbf{x}_i) = \min_{\forall C_j, \mathbf{x}_i \notin C_j} a(\mathbf{x}_i, C_j). \quad (4.23)$$

O coeficiente silhueta varia no intervalo de $[-1, 1]$, tal que quanto maior o coeficiente melhor é o agrupamento (Rousseeuw, 1987). Em resumo, um coeficiente silhueta com valor próximo a 1 implica que a amostra \mathbf{x}_i foi colocada no grupo correto. Entretanto, quando o índice se aproxima de 0 mostra que \mathbf{x}_i está no limite entre dois grupos, e caso o valor seja negativo entende-se que a amostra foi colocada no grupo incorreto. O valor médio da silhueta é calculado a partir da Eq. (4.24):

$$s_\mu = \frac{1}{N} \sum_{i=1}^N s(x_i). \quad (4.24)$$

A métrica de distância adotada nos cálculos das silhuetas foi a distância de Mahalanobis, definida pela Eq. (4.25), onde Σ é a matriz de covariância. Para um conjunto de dados unidimensional, pode-se escrever a distância de Mahalanobis como uma função do desvio padrão (σ), resultando na distância Euclideana normalizada.

$$d(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{(\mathbf{x}_1 - \mathbf{x}_2)^T \Sigma^{-1} (\mathbf{x}_1 - \mathbf{x}_2)}. \quad (4.25)$$

5

Metodologia da Análise de Agrupamento de Fácies

Esta seção tem como objetivo demonstrar a metodologia que é utilizada para a obtenção dos agrupamentos de fácies da Formação São Sebastião, além de apresentar como os algoritmos de análise de agrupamento são utilizados para auxiliar as interpretações de zonas permeosas e a caracterização das águas subterrâneas. A distinção entre a Fm. São Sebastião e as Formações subjacentes, quando necessária, é realizada levando-se em consideração a estratigrafia de sequência das camadas depositadas através da análise das assinaturas de perfis geofísicos (litológicos e de resistividade) e perfis compostos.

5.1 Pré-processamento e Edição

Com base em consultas bibliográficas, em dados de perfis compostos e análises estratigráficas, verificou-se que a porção de interesse no conjunto de medidas não ultrapassa a profundidade de 1500 m nos conjuntos de dados analisados, permitindo excluir todas as medidas realizadas a profundidades superiores a mencionada.

A separação das zonas de água doce e água salgada foi feita utilizando medidas de potencial espontâneo e de resistividade indutiva das camadas. Notou-se, em alguns perfis de poços, que a linha base das camadas impermeáveis, nos registros de potencial espontâneo, não era regular (Figura 5.1), o que dificulta a distinção dos grupos de camadas permeáveis com conteúdo de água potável e salobra. Para contornar este problema, foi usada uma correção a partir de um modelo de regressão linear obtido entre a relação do potencial espontâneo e a profundidade.

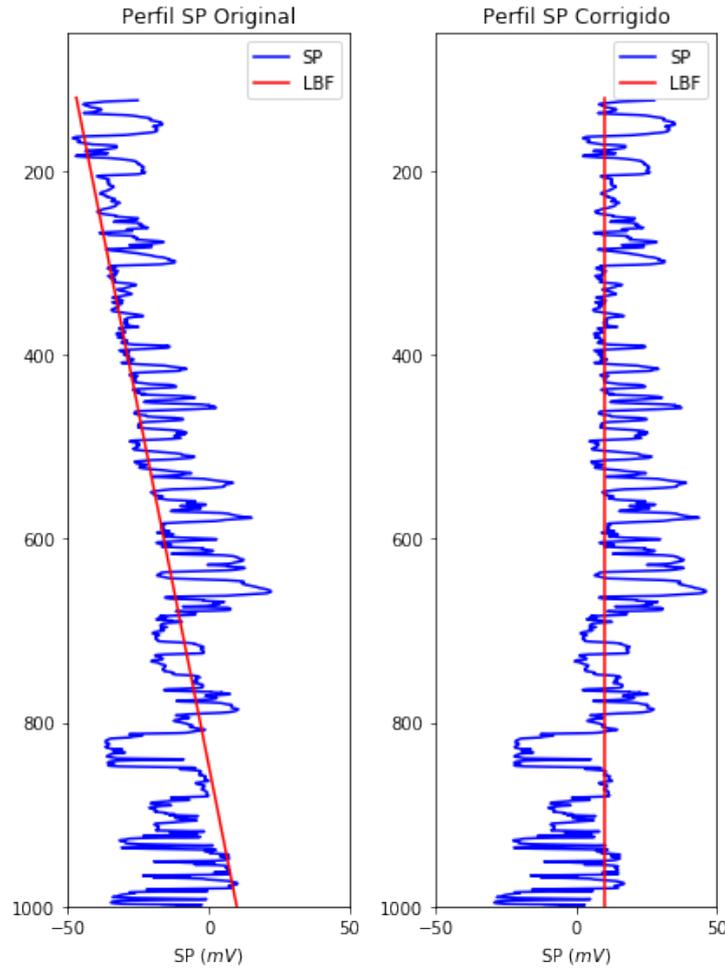


Figura 5.1: Perfil de potencial espontâneo não corrigido (esquerda) e a correção utilizando um modelo regressivo do potencial espontâneo e a profundidade (direita). Os dados utilizados para a criação da reta de correção estão contidos no intervalos de $[-52,20]$ (mV) para os registros de SP e profundidades entre $[123,800]$ (m).

As medidas de profundidade e potencial espontâneo, utilizadas na regressão linear, são limitadas por um critério empírico, estabelecido pelo interprete. Esta restrição tem como objetivo gerar a melhor reta de correção da linha base. Considerando que o modelo obtido é dado por:

$$SP_{calc}(DEPT) = a \cdot DEPT - b \quad (5.1)$$

em que $DEPT$ é a profundidade da leitura e a e b são as constantes da regressão. Têm-se que quando o coeficiente a é positivo o potencial espontâneo corrigido (SP_{cor}) é dado por:

$$SP_{cor} = SP_m - SP_{calc}, \quad (5.2)$$

e quando a for negativo:

$$SP_{cor} = SP_m + SP_{calc}. \quad (5.3)$$

SP_m é o potencial medido e SP_{calc} é estimado pelo modelo de regressão. A Figura 5.1 mostra o exemplo para o caso em que o coeficiente a , da Eq. (5.1), é positivo.

5.2 Análise de Agrupamentos

O poço denominado PG-1 (Poço Guia 1 - Figura 5.4) é aqui utilizado como referência para ilustrar a obtenção dos parâmetros e a execução e interpretação dos algoritmos não-supervisionados a partir de um conjunto de medidas de perfilagens geofísicas de poços.

Na análise de agrupamentos em perfis geofísicos foram utilizadas as curvas de potencial elétrico espontâneo e resistividade elétrica (indução profunda) para a separação de intervalos permeáveis e não-permeáveis e a tipificação dos fluidos contidos nas zonas permeáveis (água doce ou salgada). Na ausência de registros de resistividade de indução profunda (*RILD*), quando disponível, foram utilizadas medidas de indução rasa ou normal curta (*RILS* ou *SN*). Vale a pena ressaltar que os agrupamentos são realizados perfil por perfil e levando-se em consideração as leituras de profundidade como um atributo para a realização da clusterização dos parâmetros físicos medidos.

Tomando os dados do perfil guia da área de Pojuca como exemplo e apresentando estes dados na forma de gráficos de dispersão entre as propriedades físicas das rochas (*crossplot*) e histogramas, observa-se um comportamento característico, dos atributos, semelhantes a distribuições normais (Figura 5.2).

Para análise das medidas nos perfis é esperado obter-se, por um interprete, três grupos a partir do algoritmo de Modelos de Misturas Gaussianas (GMM), inicializado a partir dos centroides obtidos na técnica K-médias, que são interpretados como: zonas permeáveis com água doce (arenitos), zonas permeáveis com água salgada e zonas não-permeáveis (folhelhos e eventualmente carbonatos) (Figura 5.3). Notou-se, em alguns perfis do conjunto de poços analisados, após a análise de agrupamento, a ocorrência de zonas permeáveis com água doce em profundidades superiores às camadas de água salgada e vice-versa. Isso deveu-se às mudanças nos valores do potencial espontâneo, em linhas bases que não foram totalmente corrigidas, variações anômalas de resistividades, desmoronamentos, ou até mesmo alterações de fluidos dentro da formação. Devido a tais ocorrências de amostras em grupos indevidos, foi necessário adicionar um sistema de recomendação para posicionar essas zonas nas

profundidades adequadas (Seção 5.2.1).

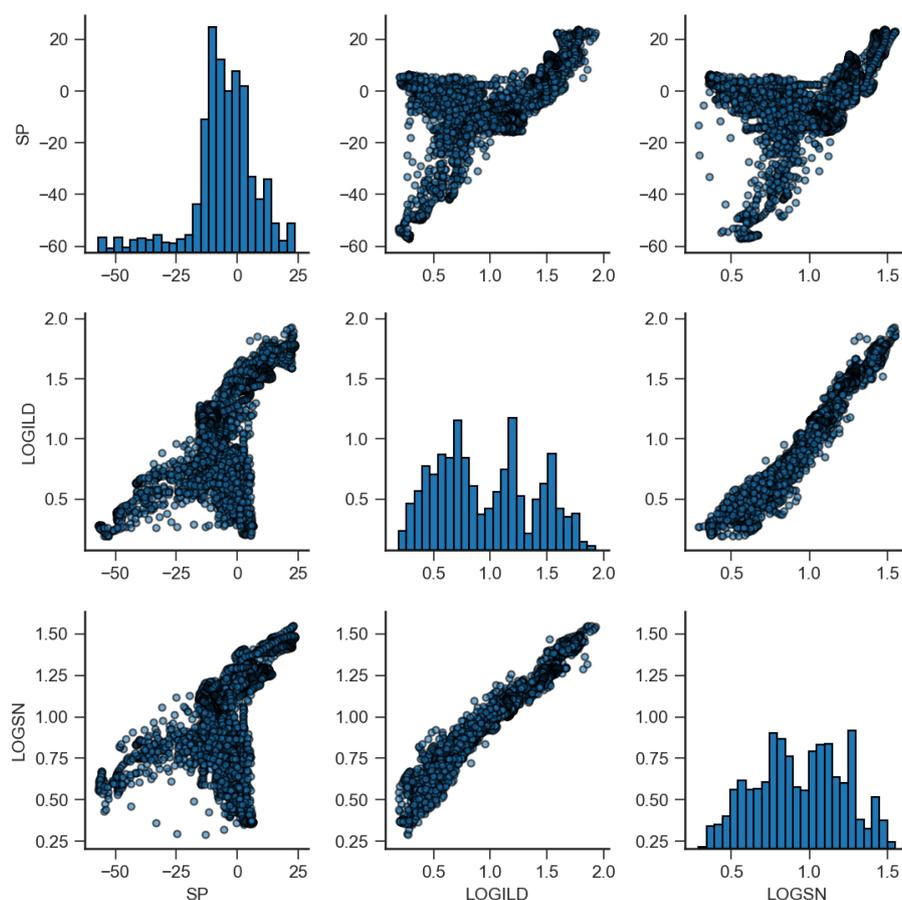


Figura 5.2: Gráficos de dispersão das medidas com ferramentas de potencial elétrico espontâneo (SP), logaritmo da resistividade da indução profunda ($LOGILD$) e o logaritmo da resistividade da normal curta ($LOGSN$) e seus respectivos histogramas no poço guia.

5.2.1 Sistema de Recomendação de Grupos

Foram verificadas ocorrências dos grupos interpretados como zonas permeáveis de água doce e salgada em profundidades incompatíveis. Para solução desse problema, foi utilizado um recurso estatístico denominado Percentil. **Percentis** são medidas que dividem as amostras de um dado atributo, em ordem crescente, em 100 partes. Por exemplo: o 5º percentil determina os 5% menores valores dos dados.

A correção dos grupos de arenitos com água doce é feita da seguinte forma, assume-se que a profundidade máxima que pode haver água doce são as profundidades equivalentes ao 99º percentil. Para profundidades superiores a desse percentil, todas as amostras interpretadas como zonas permeáveis com conteúdo de água doce, são consideradas como uma zona de

transição, pois não é possível distinguir camadas permeáveis das não-permeáveis. A segunda correção que se faz necessária diz respeito aos grupos interpretados como zonas permeáveis contendo água salgada. Para isto, foi atribuído que essas amostras com valores de profundidades inferiores ao 1º percentil fossem consideradas, também, como uma zona de transição. Quando recomendações errôneas eram mantidas na zona de água salgada, determinava-se que todas as amostras rotuladas como zona permeável de água doce fossem alteradas para não-permeável (Folhelho). Figura 5.4 ilustra o produto final gerado pelo sistema de recomendação para o perfil PG1.

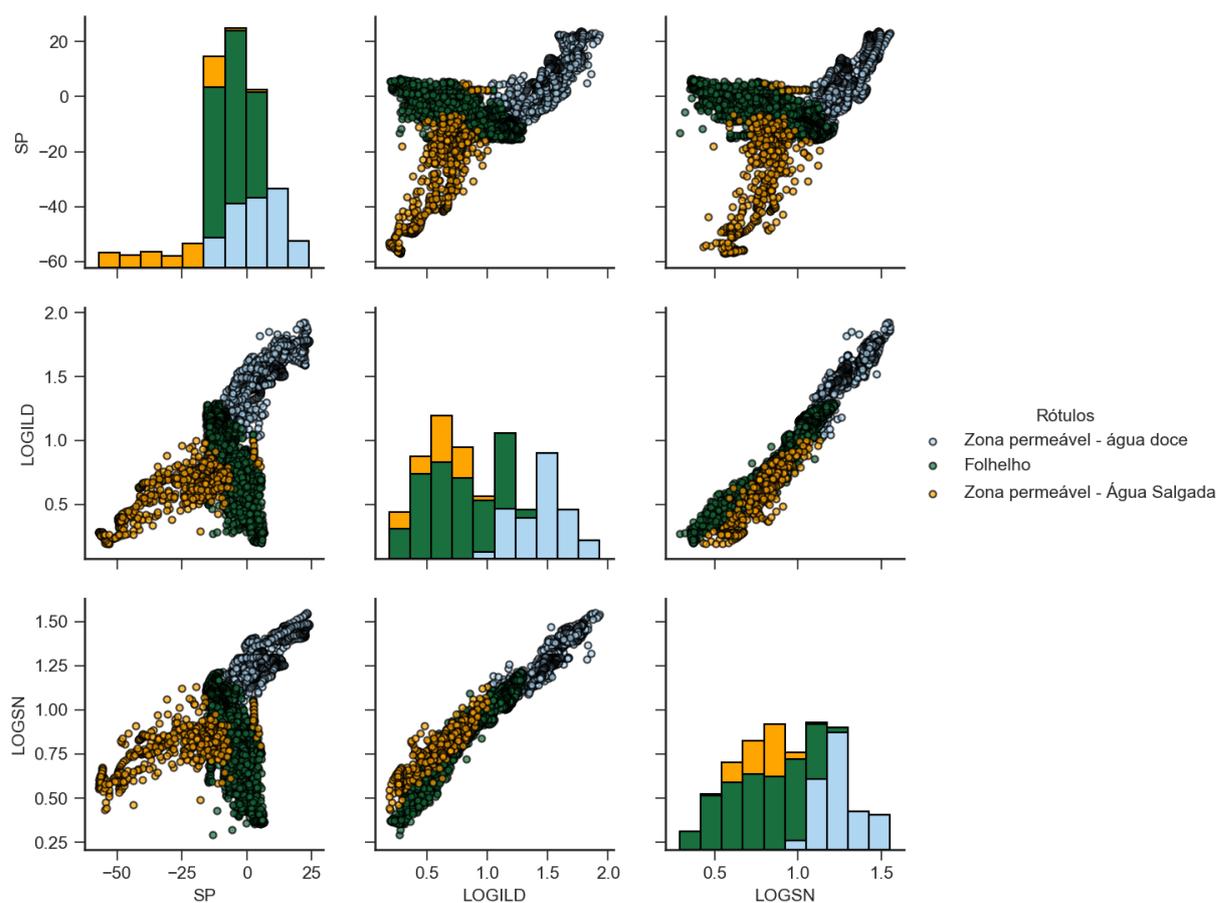


Figura 5.3: Gráficos de dispersão das medidas com ferramentas de potencial elétrico espontâneo (SP), logaritmo da resistividade de indução profunda ($LOGILD$) e o logaritmo da resistividade da normal curta ($LOGSN$) e seus respectivos histogramas apontando os grupos obtidos a partir do algoritmo de Modelo de Misturas Gaussianas (GMM) com centroides inicializados a partir do resultado do algoritmo K-médias, do perfil PG1.

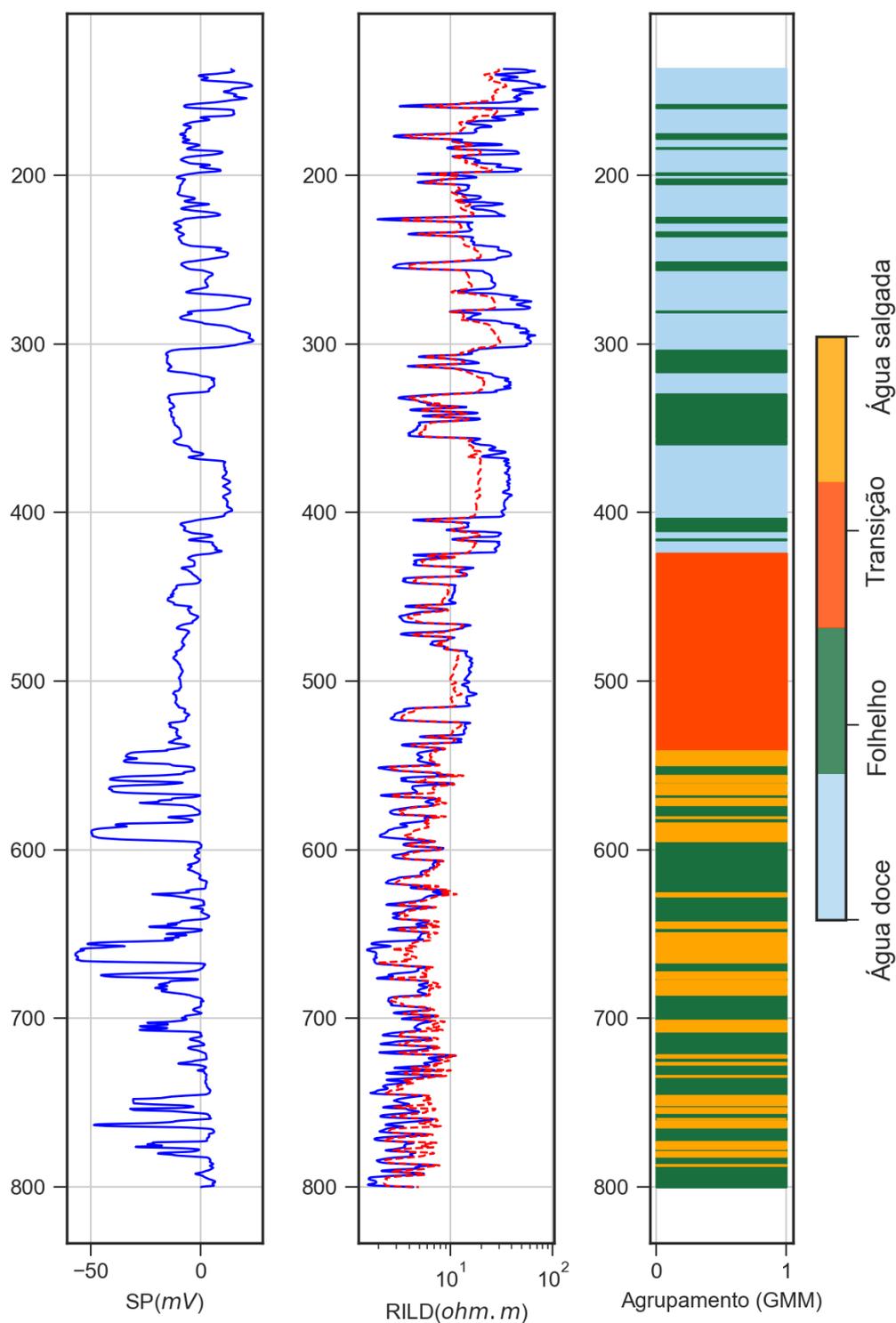


Figura 5.4: Grupos de fácies obtidos a partir do algoritmo de clusterização de Modelo de Misturas Gaussianas (GMM) com centroides inicializados a partir do resultado do algoritmo K-médias, e sistema de recomendação para correção das profundidades de zonas permeáveis e posicionamento da zona de transição do perfil PG1.

5.2.2 Validação

Depois de interpretados os grupos obtidos com o algoritmo de agrupamento, fazendo uso do sistema de recomendação, foi gerada uma partição com uma interpretação convencional para ser comparada aos agrupamentos interpretados utilizando o índice estatístico Rand Ajustado (*ARI*), apresentado na Seção 4.2.1. Para esta tarefa, foi selecionado um perfil com bom comportamento dos registros de potencial espontâneo para serem validados. Concluídas as validações, foram construídos mapas estruturais da base da água doce, do topo da água salgada e espessura acumulada das camadas permeáveis com água doce.

5.3 Cálculo da Argilosidade

A estimativa da argilosidade (V_{sh}) foi feita a partir de uma adaptação do modelo proposto por Vernik (2016). Para demonstrar a metodologia adotada, foram utilizados, nessa seção, 11 perfisagens geofísicas de poços de água, cedidos pela Companhia de Engenharia e Recursos Hídricos da Bahia (CERB), que perpassam o sistema aquífero do Recôncavo composto pelas Formações São Sebastião e Marizal, batizado como de PW 1 a PW 11.

Vernik (2016) apontou que medidas de raios gama de formações areno argilosas se comportam como uma distribuição bimodal. Para confirmar tal hipótese, foram feitos diversos agrupamentos com diferentes números de grupos, utilizando a técnica de Modelo de Misturas Gaussianas, inicializada a partir do modelo gerado pelo algoritmo K-médias, para verificar se a hipótese formulada para dois grupos é de fato a mais forte nos perfis utilizados. Empregando o índice de silhueta para comprovação do melhor número de grupos, pôde-se perceber que os valores médios mais altos dos coeficientes silhuetas obtidos foram para os agrupamentos com o número de grupos igual a 2, como mostra a Tabela 5.1.

| N. Grupos | PW 1 | PW 2 | PW 3 | PW 4 | PW 5 | PW 6 | PW 7 | PW 8 | PW 9 | PW 10 | PW 11 |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 2 | 0.698 | 0.730 | 0.789 | 0.701 | 0.784 | 0.763 | 0.786 | 0.782 | 0.720 | 0.684 | 0.634 |
| 3 | 0.624 | 0.700 | 0.690 | 0.660 | 0.700 | 0.630 | 0.681 | 0.706 | 0.676 | 0.607 | 0.621 |
| 4 | 0.614 | 0.648 | 0.626 | 0.561 | 0.635 | 0.602 | 0.653 | 0.616 | 0.634 | 0.574 | 0.593 |
| 5 | 0.585 | 0.579 | 0.560 | 0.562 | 0.602 | 0.590 | 0.591 | 0.570 | 0.521 | 0.552 | 0.572 |
| 6 | 0.570 | 0.565 | 0.534 | 0.553 | 0.572 | 0.586 | 0.570 | 0.547 | 0.538 | 0.542 | 0.570 |
| 7 | 0.540 | 0.560 | 0.549 | 0.536 | 0.553 | 0.593 | 0.569 | 0.578 | 0.548 | 0.527 | 0.577 |

Tabela 5.1: Valor médio do índice silhueta de diferentes números de grupos para cada análise de agrupamento das medidas de raios gama de amostras pertencentes ao sistema aquífero do Recôncavo.

Após confirmada a hipótese bimodal, os parâmetros GR_{min} e GR_{max} da Eq. (3.10) são derivados a partir do Modelo de Misturas Gaussianas da seguinte forma:

$$Gr_{min} = \mu_s + \sigma_s \quad (5.4)$$

e

$$Gr_{max} = \mu_{sh} - \sigma_{sh}, \quad (5.5)$$

onde μ_s e μ_{sh} são os valores de raios gama médio do Modelo de Mistura de Gaussianas do arenito e do folhelho respectivamente e σ_s e σ_{sh} são os desvios padrões correspondentes. Sabe-se que numa distribuição areno-argilosa bimodal, os valores médios das medidas de raio gama são maiores para os folhelhos. Os valores das constantes obtidas a partir dos modelos estão mostradas na Tabela 5.2

| POÇO | μ_s | σ_s | μ_{sh} | σ_{sh} | Gr_{min} | Gr_{max} |
|-------|---------|------------|------------|---------------|------------|------------|
| PW 1 | 43.03 | 8.02 | 81.50 | 10.08 | 51 | 71 |
| PW 2 | 19.80 | 6.04 | 70.66 | 19.14 | 25 | 51 |
| PW 3 | 39.18 | 10.52 | 105.69 | 13.11 | 49 | 92 |
| PW 4 | 32.61 | 8.54 | 77.59 | 15.42 | 41 | 62 |
| PW 5 | 28.97 | 10.01 | 100.73 | 19.51 | 38 | 81 |
| PW 6 | 23.30 | 10.09 | 103.70 | 20.94 | 33 | 82 |
| PW 7 | 29.46 | 8.58 | 98.57 | 19.75 | 38 | 78 |
| PW 8 | 26.45 | 9.97 | 98.27 | 15.15 | 36 | 83 |
| PW 9 | 27.08 | 7.10 | 83.41 | 28.17 | 34 | 55 |
| PW 10 | 42.63 | 10.49 | 85.75 | 13.47 | 53 | 72 |
| PW 11 | 23.77 | 4.13 | 51.45 | 10.04 | 27 | 41 |

Tabela 5.2: Média (μ) e desvio padrão (σ) do modelo de misturas gaussianas da análise de agrupamento com dois grupos. Correspondem a arenitos e folhelhos, derivados das medidas de raios gama, e os parâmetros da Eq. 3.10, dado em unidade uAPI derivados dos Modelos de Misturas Gaussianas.

A Figura 5.5 mostra o posicionamento dos parâmetros obtidos no histograma das medidas de GR do poço PW 3. Por inspeção visual, pode-se notar que as constantes obtidas tem posicionamentos semelhantes àquelas propostas por Vernik (2016), apresentado na Figura 3.1.

Um procedimento similar foi criado utilizando o logaritmo da resistividade de indução. Percebe-se nos histogramas dos registros dos poços que o formato de "sino" é bem definido para alguns perfis geofísicos (Figura 5.6). Ao realizar os agrupamentos e calcular os valores médios das silhuetas, tem-se que a hipótese bimodal é a mais forte para a maioria dos conjuntos de dados, exceto para os perfis PW 5 e PW 11, como é mostrada na Tabela 5.3.

| N. Grupos | PW 1 | PW 2 | PW 3 | PW 4 | PW 5 | PW 6 | PW 7 | PW 8 | PW 9 | PW 10 | PW 11 |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 2 | 0.720 | 0.631 | 0.629 | 0.598 | 0.600 | 0.761 | 0.647 | 0.819 | 0.710 | 0.699 | 0.598 |
| 3 | 0.634 | 0.525 | 0.610 | 0.553 | 0.627 | 0.724 | 0.453 | 0.731 | 0.579 | 0.613 | 0.636 |
| 4 | 0.603 | 0.586 | 0.581 | 0.560 | 0.649 | 0.667 | 0.668 | 0.659 | 0.563 | 0.583 | 0.604 |
| 5 | 0.579 | 0.591 | 0.568 | 0.567 | 0.642 | 0.596 | 0.637 | 0.641 | 0.552 | 0.538 | 0.471 |
| 6 | 0.554 | 0.557 | 0.569 | 0.575 | 0.618 | 0.623 | 0.579 | 0.607 | 0.551 | 0.541 | 0.514 |
| 7 | 0.546 | 0.575 | 0.565 | 0.578 | 0.579 | 0.596 | 0.583 | 0.622 | 0.553 | 0.521 | 0.553 |

Tabela 5.3: Valor médio do índice silhueta de diferentes números de grupos para cada agrupamento das medidas do logaritmo da resistividade elétrica.

Considerando o logaritmo da resistividade, os valores $\log(R_{max})$ e $\log(R_{min})$ dos parâmetros usados na Eq. (3.7) também podem ser determinados por modelos de misturas gaussianas da seguinte forma:

$$\log(R_{min}) = \mu_{sh} + \sigma_{sh} \quad (5.6)$$

e

$$\log(R_{max}) = \mu_s - \sigma_s, \quad (5.7)$$

para $n = -1$, caso contrário:

$$\log(R_{min}) = \mu_{sh} - \sigma_{sh} \quad (5.8)$$

e

$$\log(R_{max}) = \mu_s + \sigma_s. \quad (5.9)$$

Nas Eq.s (5.6), (5.7), (5.8) e (5.9) μ e σ representam os valores médios e os desvios padrões do logaritmo da resistividade. Os parâmetros obtidos utilizando as medidas de resistividade são apresentados na Tabela 5.4.

Obtido os parâmetros necessários das Eq.s (3.7) e (3.10), pode-se estimar de forma automática o conteúdo de argila dos aquíferos. A Figura 5.7 ilustra o conteúdo de argila calculado, a partir da metodologia proposta, do perfil PW 9.

| POÇO | μ_s | σ_s | μ_{sh} | σ_s | $\log(R_{min})$ | $\log(R_{max})$ |
|-------|---------|------------|------------|------------|-----------------|-----------------|
| PW 1 | 3.68 | 0.43 | 2.11 | 0.34 | 2.45 | 3.25 |
| PW 2 | 4.25 | 0.73 | 2.14 | 0.30 | 2.44 | 3.52 |
| PW 3 | 3.52 | 0.47 | 2.20 | 0.40 | 2.60 | 3.05 |
| PW 4 | 3.94 | 0.90 | 1.74 | 1.20 | 2.94 | 3.04 |
| PW 5 | 4.77 | 0.74 | 2.76 | 0.78 | 3.54 | 4.03 |
| PW 6 | 4.67 | 0.69 | 1.83 | 0.33 | 2.16 | 3.98 |
| PW 7 | 4.76 | 0.77 | 2.39 | 0.55 | 2.94 | 3.99 |
| PW 8 | 4.68 | 0.48 | 1.96 | 0.37 | 2.33 | 4.20 |
| PW 9 | 4.66 | 0.42 | 2.82 | 0.52 | 3.34 | 4.24 |
| PW 10 | 0.83 | 0.26 | 2.04 | 0.41 | 1.09 | 1.63 |
| PW 11 | 3.48 | 0.69 | 1.70 | 0.70 | 2.40 | 2.79 |

Tabela 5.4: Média (μ) e desvio padrão (σ) do modelo de misturas gaussianas da análise de agrupamento de com dois grupos. Correspondem a arenitos e folhelhos, derivados das medidas do logaritmo da resistividade elétrica, e os parâmetros da Eq. (3.7), dado em unidade $\log(\Omega m)$ derivados dos Modelos de Misturas Gaussianas.

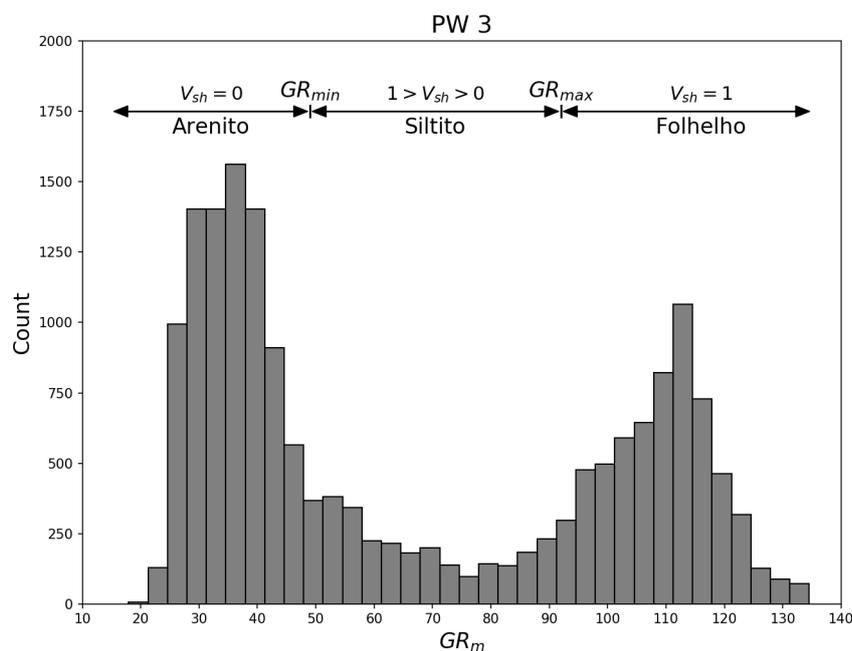


Figura 5.5: Histograma das medidas de raios gama e parâmetros obtidos a partir do Modelo de Misturas Gaussianas do poço PW 3 usados para calcular a argilosidade através da Eq. (3.10). Nota-se a tendência de uma distribuição bimodal das medidas.

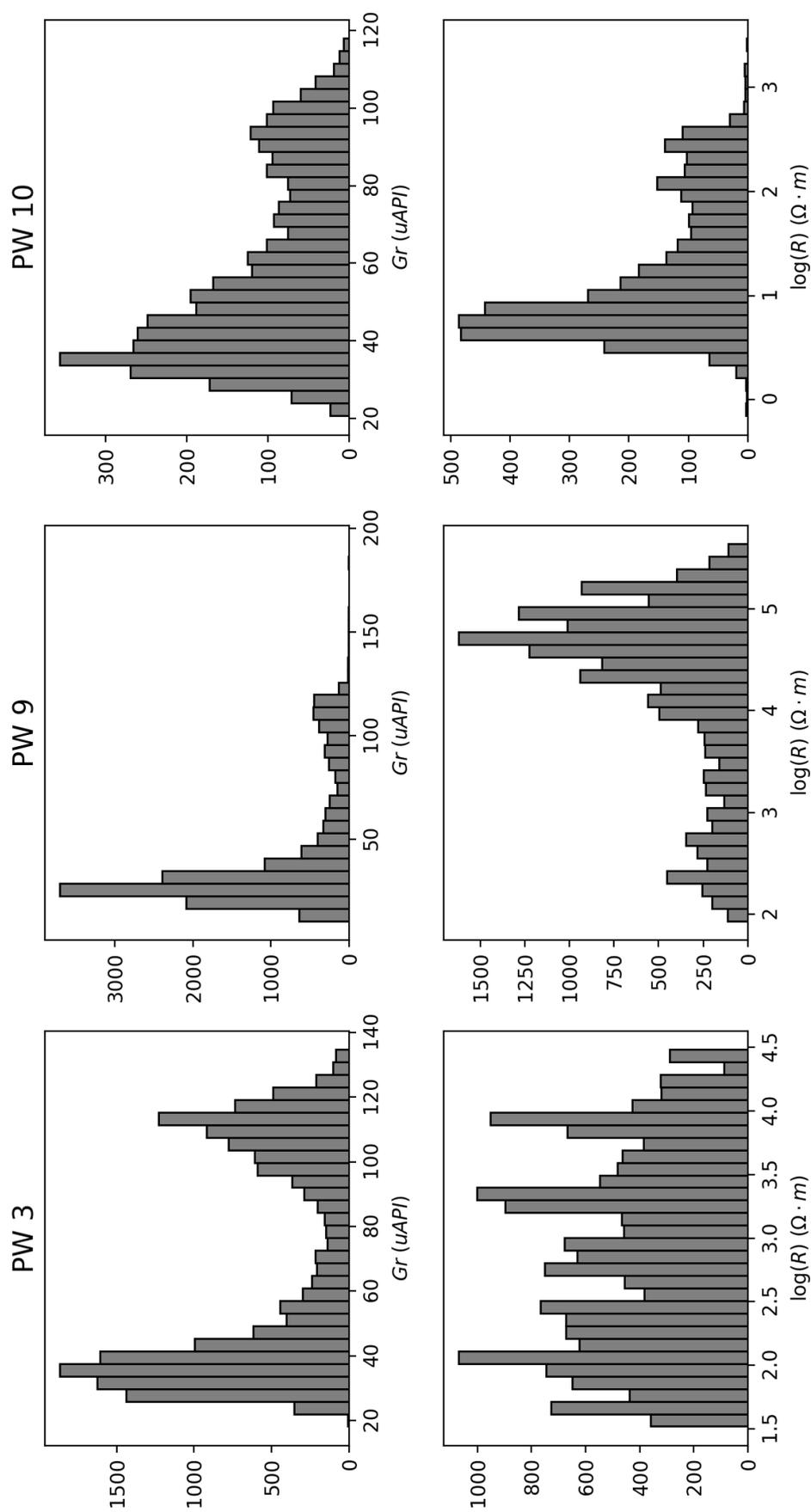


Figura 5.6: Histogramas das medidas de raios gama e do logaritmo da resistividade elétrica de três perfis geofísicos diferentes. É possível notar a tendência bimodal bem definida em quase todos os histogramas, com exceção do histograma do logaritmo da resistividade do perfil PW 3.

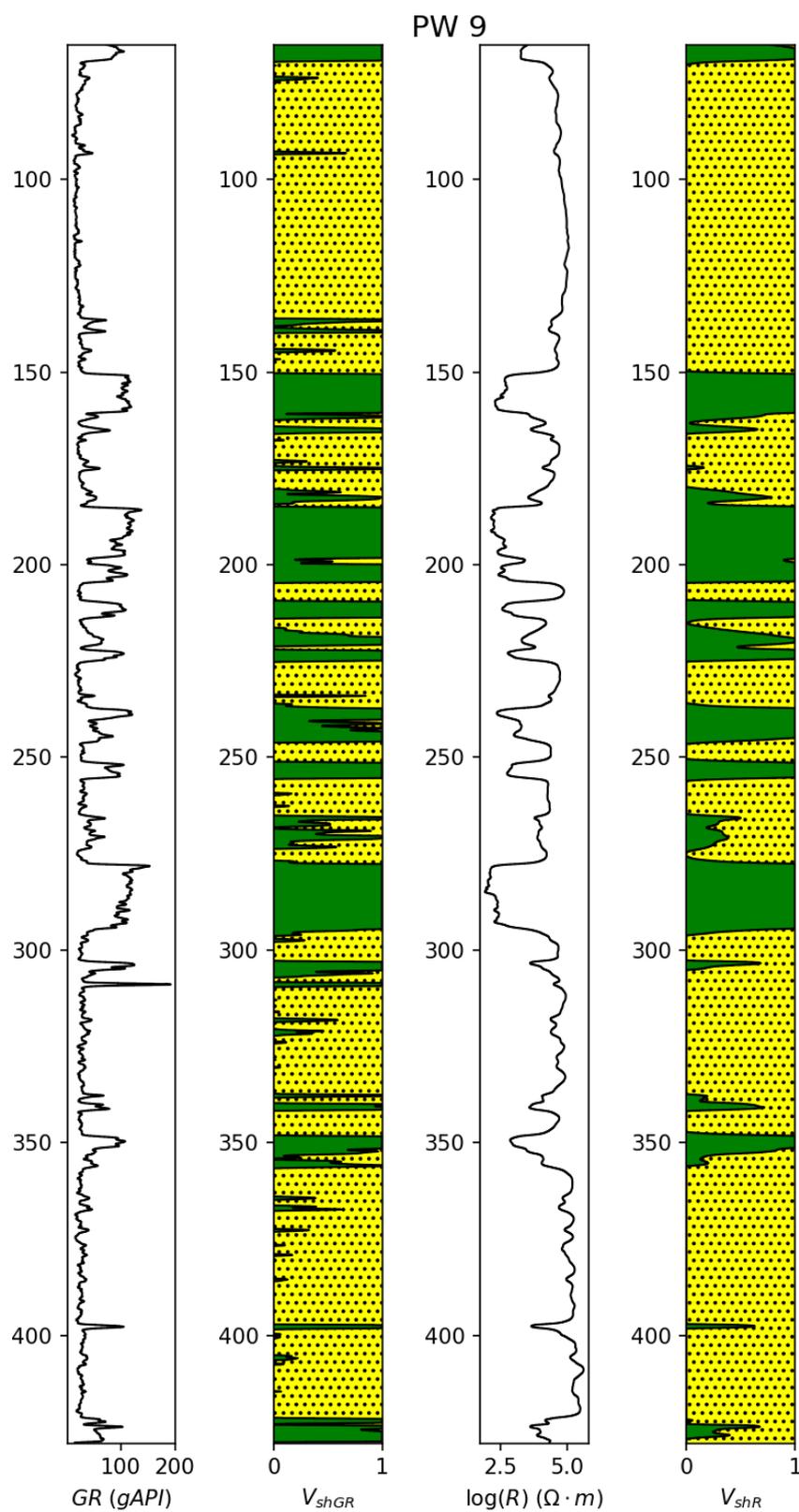


Figura 5.7: Conteúdo de argila estimado (V_{sh}), do poço PW 9, em que $V_{sh} = 0$ é interpretado como arenito, $V_{sh} = 1$ é interpretado como folhelho e caso contrário, entende-se que a amostra avaliada trata-se de uma mistura areno-argilosa entre arenito e folhelho (siltito).

6

Resultados e Interpretações

A área estudada compreende a porção norte da Bacia do Recôncavo e foram disponibilizadas informações de 164 poços que compõem o conjunto de dados, apresentados na Figura 6.1. Somente perfilagens geofísicas de 90 poços (além daquelas demonstradas na metodologia) foram disponibilizadas para análises de agrupamento, cujo os resultados serão discutidos nas seções a seguir.

Para facilitar a análise dos resultados obtidos, estes estão apresentados, com detalhes, em duas áreas representativas da ocorrência do sistema aquífero Recôncavo, designados como áreas de Pojuca e Araças, nas quais se encontram boa parcela dos perfis examinados. Numa integração para a porção norte da bacia foram usadas as seguintes informações adicionais:

1. Dados de 30 perfis de poços examinados com a mesma metodologia de automação. Esses poços distribuem-se, principalmente, entre os municípios de Catu e Alagoinhas mas, infelizmente, os dados foram liberados até a máxima profundidade de 1000m;
2. Para adensar as informações em mapa foram incorporados 74 informações sobre a profundidade de água doce em poços da área, obtidas num trabalho hidrogeológico inédito realizado por Neto (2006) para a Petrobras, nomeados PP (Figura 6.1).

Algumas dessas informações foram verificadas corretas por comparação com as análises realizadas.

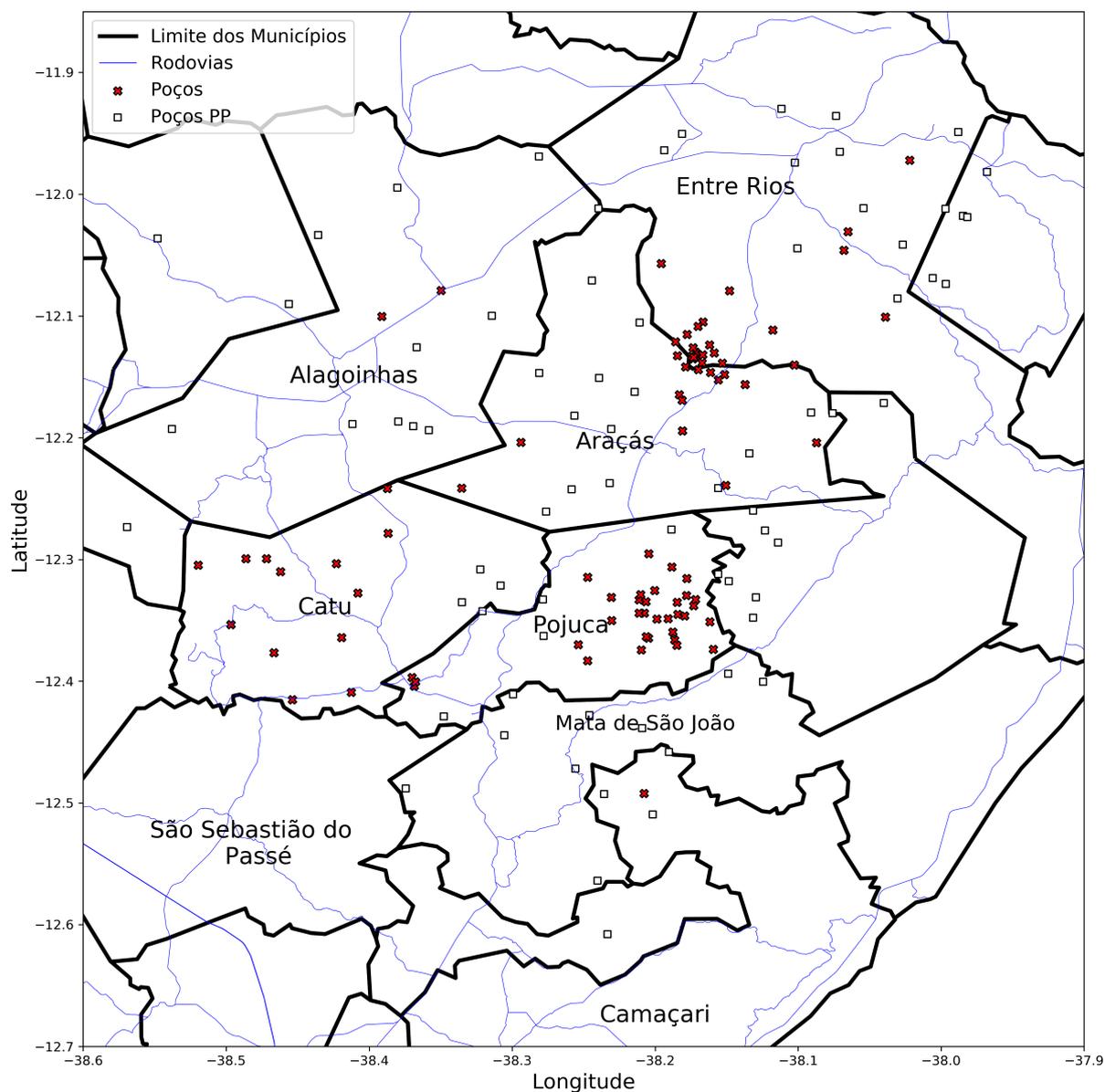


Figura 6.1: Mapa que apresenta os perfis geofísicos disponíveis para análise do sistema aquífero do Recôncavo, na porção norte.

6.1 Área de Pojuca

A área estudada em Pojuca possui uma extensão de aproximadamente 105km^2 e localiza-se na região Centro-Leste da Bacia do Recôncavo Norte, mais especificamente no município de Pojuca, a 80 km de Salvador. Os principais acessos a área são através das rodovias BA-093, BA-507 e BA-504. As principais sedes municipais circunvizinhas são Araçás, Catu e Alagoinhas ao norte e Mata de São João e São Sebastião do Passé ao sul (Mariano, 2013).

O local corresponde a um dos principais depocentros da Bacia do Recôncavo, denominado Baixo de Miranga. O Plano de Desenvolvimento da Agência Nacional de Petróleo, Resolução

nº509/2015, aponta a Formação Pojuca como o reservatório mais raso de hidrocarbonetos, e perfis compostos de poços de petróleo dessa área mostram que a Formação São Sebastião repousa discordantemente sobre esta unidade.

O conjunto de dados analisados é formado por 33 perfis geofísicos de poços cedidos pela PETROBRAS e pela Agência Nacional de Petróleo (ANP) para fins de estudos hidrogeológicos. Os registros de potencial espontâneo e resistividade elétrica (indução profunda ou normal curta) foram utilizados para realizar a separação dos horizontes ou camadas permeáveis e impermeáveis, assim como tipificar os fluidos contidos nas zonas permeáveis (água doce ou salgada). Vale ressaltar que os agrupamentos foram realizados perfil a perfil, e as leituras de profundidade foram levadas em consideração como atributo das análises de agrupamentos nas propriedades físicas.

Para agrupar os dados, foram fixados o número de grupos e os parâmetros iniciais de peso, para o algoritmo GMM, inicializados a partir da técnica K-médias. O modelo inicial proposto é composto por um grupo que contém 10% das amostras ($\pi_1 = 0,1$), um com 40% das amostras ($\pi_2 = 0,4$) e outro com 50% ($\pi_3 = 0,5$). Com isto, foram obtidos os seguintes grupos interpretados: zona impermeável, zona permeável com água doce e zona permeável com água salgada. Especialmente, em um gráfico de dispersão entre as medidas de potencial e o logaritmo da resistividade, espera-se que os grupos se apresentem conforme mostrado na Figura 5.3.

O sistema de recomendação de grupos foi implementado com o objetivo de corrigir valores anômalos presentes nos agrupamentos obtidos e sugerir uma zona de transição entre camadas permeáveis com conteúdos de água doce e de água salgada. A Figura 6.2 mostra o gráfico de dispersão correspondente à análise de agrupamento com suas respectivas interpretações para o conjunto de dados análogo ao perfil PM 21. Neste caso, é possível notar que a zona de transição sugerida situa-se espacialmente próxima ao grupo das zonas impermeáveis. Este fato é esperado, pois não há deflexão nas medidas do potencial espontâneo devido à semelhança entre a resistividade elétrica do filtrado (lama) e a resistividade da água de formação, o que dificulta a detecção de camadas permeáveis através da análise de agrupamento neste trecho do perfil.

Realizada as interpretações dos agrupamentos de todos os perfis da área que continham medidas de potencial espontâneo e resistividade elétrica, utilizou-se uma partição interpretada do perfil PM 21 advinda de uma análise faciológica convencional para comparar através do índice Rand ajustado se havia coerência nas interpretações obtidas em relação aos grupos esperados. O valor *ARI* entre a partição interpretada e o agrupamento é aproximadamente igual 0,67, indicando que o agrupamento obtido concorda significativamente com a partição

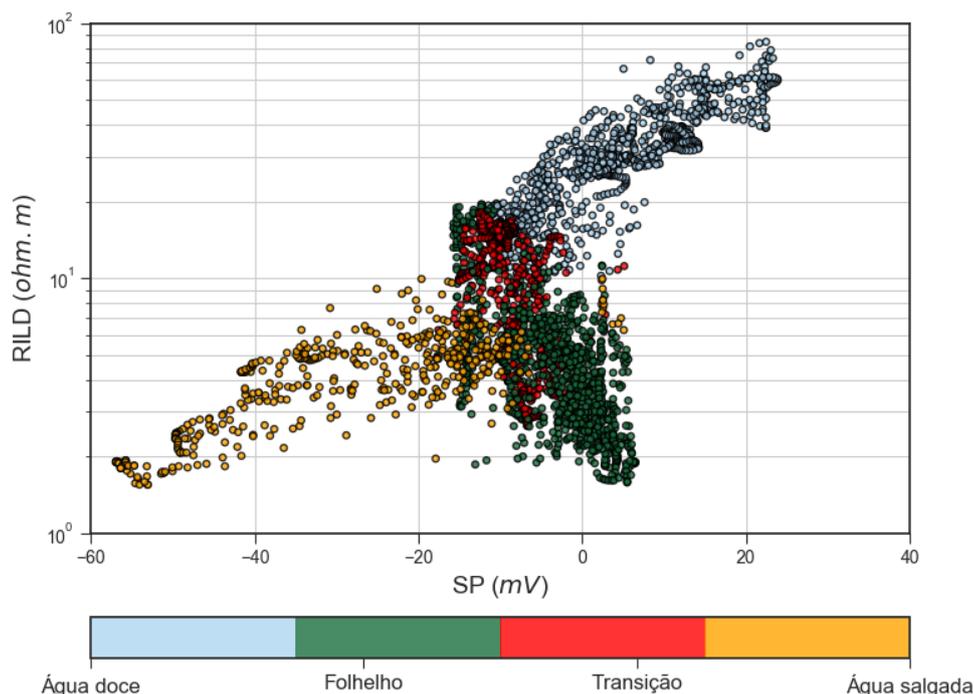


Figura 6.2: Gráfico de dispersão da resistividade de indução *versus* o potencial elétrico espontâneo do perfil PM 21 e representação dos grupos interpretados a partir dos algoritmos de agrupamento e o sistema de recomendação.

esperada. A Figura 6.3 mostra as partições externas e os agrupamentos interpretados utilizados para a validação dos grupos. Se verificada apenas a porção correspondente as zonas permeáveis contendo água doce, o valor obtido do índice *ARI* é aproximadamente igual a 0,77.

Após validado o agrupamento, foi gerado um mapa de contorno estrutural da profundidade da base da água doce, ou topo da zona de transição, mostrado na Figura 6.4. As informações sobre a base dos arenitos saturados com água doce dos perfis 1, 8, 9, 10, 11, 18 e 33 foram obtidas a partir de perfis compostos, pois estes não continham as medidas necessárias para a realização do agrupamento faciológico utilizando a metodologia da Seção 5.2. Os perfis PM 2, 4 e 14 foram excluídos por não haver medidas de potencial espontâneo e resistividade adequadas, nem informações de perfis compostos. É possível notar que os poços localizados na porção sudeste da extensão possuem a profundidade da base da água doce mais rasa, entre 300 a 500 m de profundidade, enquanto aqueles incluídos no extremo norte têm a zona de transição a profundidades superiores a 760 m. No perfil PM 4 (não apresentando no mapa interpolado), localizado a aproximadamente 4km ao norte do perfil PM 10, apresenta a base do aquífero a uma profundidade de aproximadamente 1090 m (Figura 6.6). O nível estático foi estimado a partir da Eq. (1.1), e o mapa interpolado correspondente a este é apresentado

na Figura 6.5.

Duas seções entrecruzadas, uma longitudinal (Figura 6.7) e uma transversal (Figura 6.8), foram geradas com o intuito de correlacionar os agrupamentos obtidos. É observado, na maioria dos perfis apresentados nas Figuras 6.7 e 6.8, que o algoritmo tem dificuldades de distinguir, com base em dados de resistividade, zonas permeáveis com água salgada. As zonas de interesse (água doce), porém, foram interpretadas de maneira correta.

Utilizando as medidas de raios gama dos perfis PM 1, 8, 9, 10, 11 e 18, aplicou-se a metodologia proposta na Seção 5.3. Foi necessário repartir as medidas de raios gama do perfil PM 18 em dois conjuntos de dados (18a e 18b), pois houveram, no mesmo poço, medidas realizadas em tempos diferentes. Como se sabe, a contagem de raios gama é uma contagem normalizada que pode ser alterada pelo ambiente da perfilagem. A hipótese de distribuição bimodal é confirmada para quase todos os perfis, exceto para os perfis PM 10 e 18a conforme mostra a Tabela 6.1. Os parâmetros da Eq. (3.10), apresentados na Tabela 6.2, foram utilizados para estimar a argilosidade e, por consequência, permitiram distinguir arenitos limpos e folhelhos e siltitos argilosos.

| N. Cluster | PM 1 | PM 8 | PM 9 | PM 10 | PM 11 | PM 18a | PM 18b |
|------------|-------|-------|-------|-------|-------|--------|--------|
| 2 | 0.656 | 0.696 | 0.671 | 0.577 | 0.637 | 0.602 | 0.683 |
| 3 | 0.623 | 0.598 | 0.595 | 0.625 | 0.573 | 0.624 | 0.601 |
| 4 | 0.589 | 0.560 | 0.576 | 0.594 | 0.562 | 0.585 | 0.556 |
| 5 | 0.572 | 0.539 | 0.573 | 0.568 | 0.553 | 0.578 | 0.552 |
| 6 | 0.561 | 0.520 | 0.560 | 0.546 | 0.536 | 0.549 | 0.541 |

Tabela 6.1: Valor médio do índice silhueta de diferentes números de grupos para cada análise de agrupamento das medidas de raios gama dos perfis pertencentes a área de Pojuca.

| Poço | μ_s | σ_s | μ_{sh} | σ_{sh} | Gr_{min} | Gr_{max} |
|--------|---------|------------|------------|---------------|------------|------------|
| PM 1 | 24.51 | 4.48 | 51.29 | 9.56 | 28 | 41 |
| PM 8 | 28.41 | 5.23 | 53.04 | 7.14 | 33 | 45 |
| PM 9 | 42.86 | 7.96 | 76.38 | 10.77 | 50 | 65 |
| PM 10 | 28.61 | 4.65 | 60.90 | 15.85 | 33 | 45 |
| PM 11 | 25.10 | 3.59 | 39.11 | 5.60 | 28 | 33 |
| PM 18a | 28.03 | 2.49 | 43.49 | 6.67 | 30 | 36 |
| PM 18b | 47.19 | 7.99 | 79.70 | 9.08 | 55 | 70 |

Tabela 6.2: Média (μ) e desvio padrão (σ) do modelo de misturas gaussianas da análise de agrupamento com dois grupos. Correspondem a arenitos e folhelhos, obtidos a partir das medidas de raios gama, e parâmetros da Eq. 3.10, dados em unidade uAPI derivados dos Modelos de Misturas Gaussianas dos perfis pertencentes a área de Pojuca.

Aplicadas as metodologias das Seções 5.2 e 5.3, foi possível distinguir as zonas permeáveis

com água doce das demais, permitindo estimar a espessura das camadas de arenitos e a proporção de arenito, que é dada pela razão entre a espessura total e a espessura acumulada dos arenitos das camadas perfiladas. Como não é de interesse perfilar as camadas mais rasas em poços de hidrocarbonetos, é comum não haver medidas correspondentes aos aquíferos mais superficiais. Sendo assim, considerou-se que a proporção de arenito era mantida, o que permitiu estimar as espessuras totais dos aquíferos levando em consideração o nível estático da água subterrânea, apresentado na Figura 6.9.

Uma análise sobre as porosidades derivadas dos perfis sônico e de densidade eletrônica foi feita nos perfis PM 1,8,9,10 e 11. Considerou-se que a matriz da rocha é composta por grãos de quartzo, ou seja: $\rho_{\text{quartzo}} = 2,65 \text{ g/cm}^3$ e $\Delta t_{\text{quartzo}} = 55,5 \text{ } \mu\text{s/pé}$, e assim verificou-se que a porosidade sônica calculada esperada para os perfis PM 10 e 11 foi de $27,38 \pm 3,75\%$ e para os registros de densidade dos perfis PM 1, 8 e 9 foi de $29,37 \pm 1,72\%$, ambos valores próximos aos obtidos por Mariano (2013). A Figura 6.10 apresenta os histogramas relativos as porosidades obtidas.

Estimado o nível estático, as espessuras de camadas permeáveis e porosidades médias a partir dos registros de densidade e vagarosidade, foi estimado o volume de água presente na área de Pojuca. Foi calculada uma reserva hídrica de $7,83 \cdot 10^9 \text{ m}^3$, quantidade aferida utilizando o valor médio da porosidade sônica. Usando a porosidade obtida a partir do perfil de densidade, estima-se um volume de água potável de $8,40 \cdot 10^9 \text{ m}^3$.

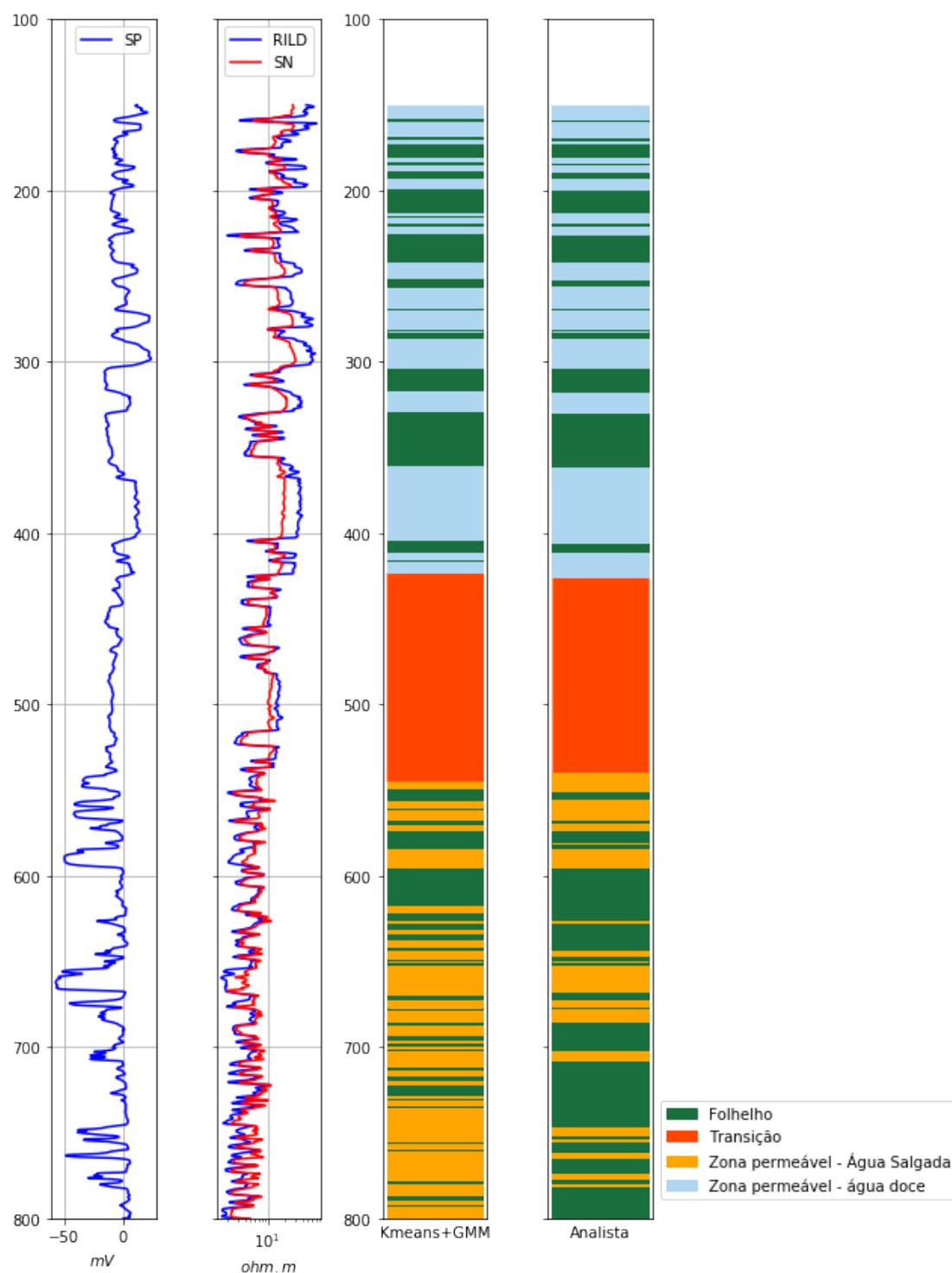


Figura 6.3: Interpretação dos agrupamentos obtidos no perfil PM 21 a partir do algoritmo de Modelo de Misturas Gaussianas inicializados a partir do centroide do algoritmo K-médias (terceira coluna) e a interpretação externa ao algoritmo para validação utilizando o Índice Rand Ajustado (*ARI*).

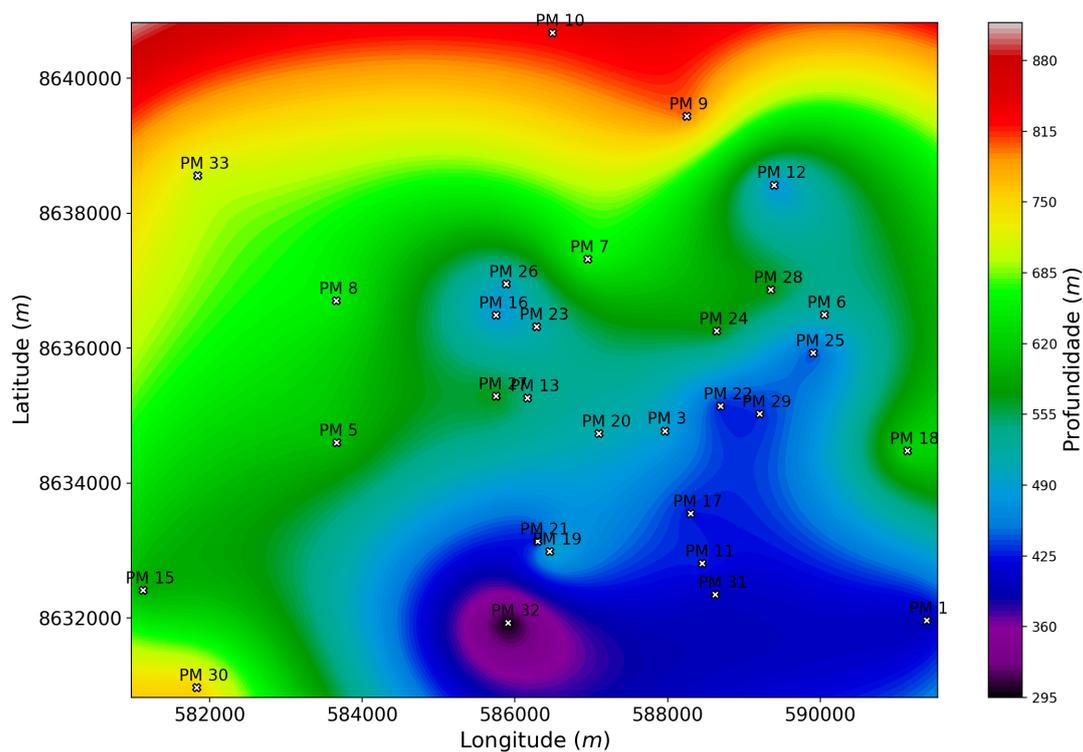


Figura 6.4: Mapa de contorno estrutural da base do reservatório de água doce na área de Pojuca.

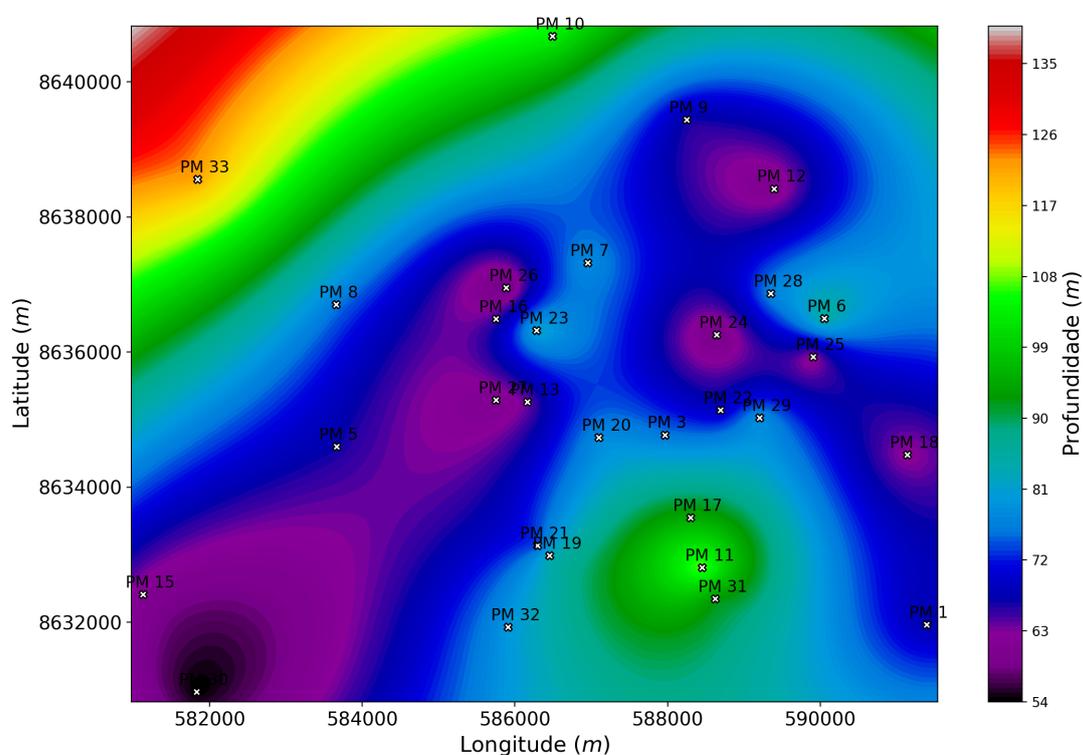


Figura 6.5: Nível Estático do aquífero livre correspondente a área de Pojuca.

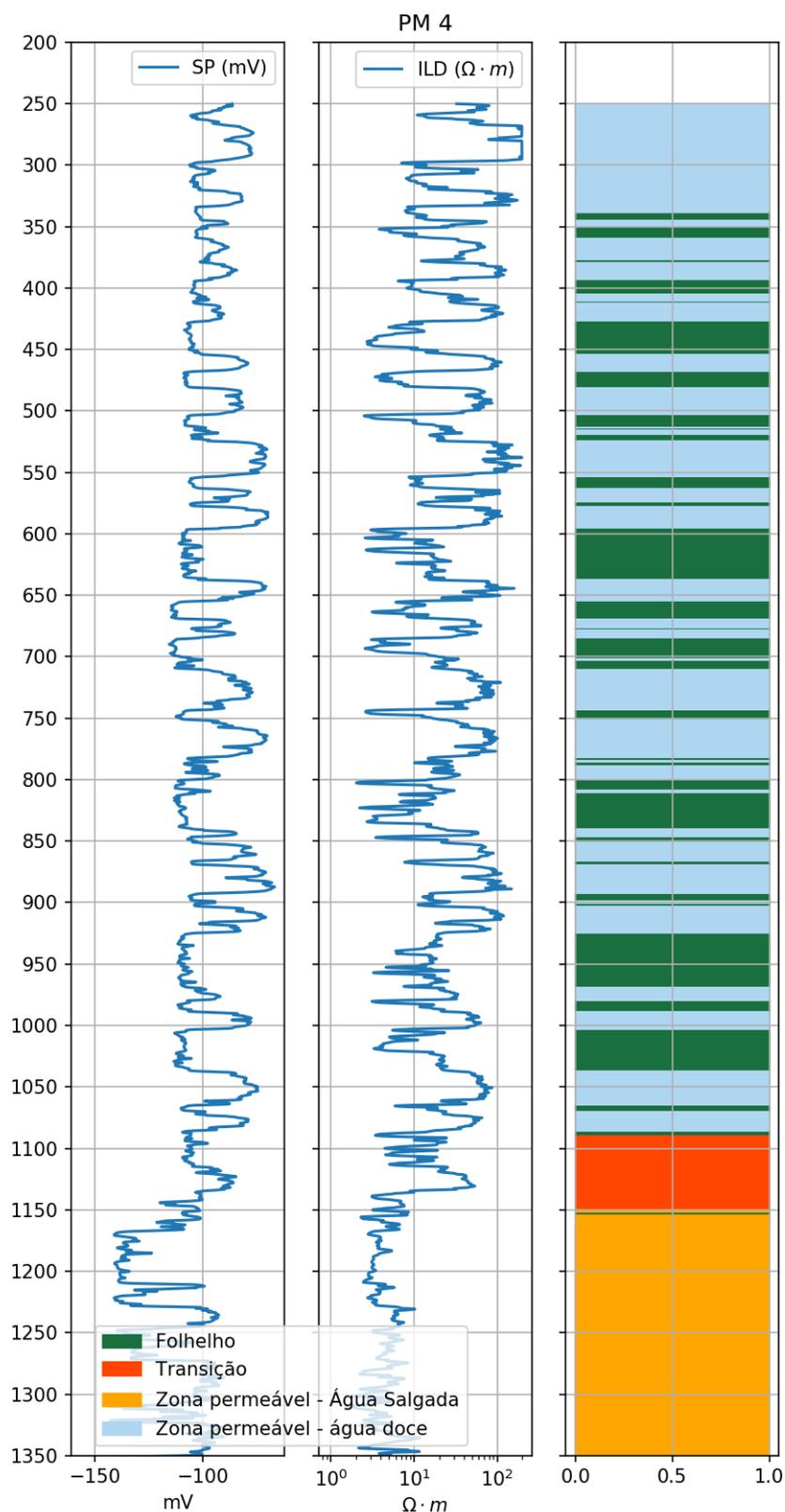


Figura 6.6: Interpretação obtida das análise de agrupamento utilizando o algoritmo de Modelo de Misturas Gaussianas inicializado a partir dos centroides do método K-médias do perfil PM 4, usando o sistema de recomendação proposto na Seção 5.2.1.

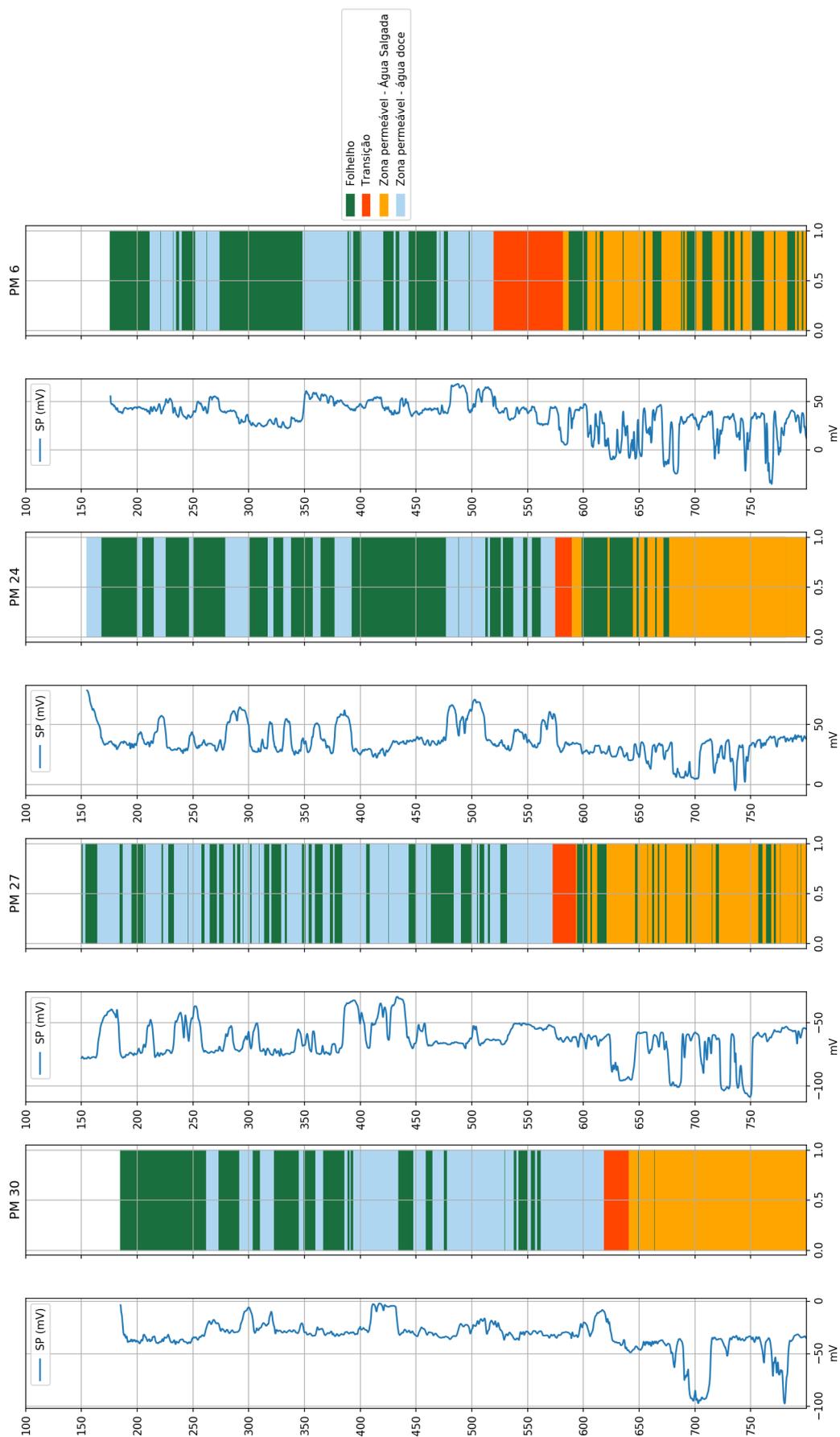


Figura 6.7: Seção longitudinal, sentido leste-oeste, apresentando quatro perfis da Área de Pojuca e suas respectivas interpretações obtidas a partir da análise de agrupamento utilizando o algoritmo de Modelos de Misturas Gaussianas inicializado a partir dos centroides do método K-médias.

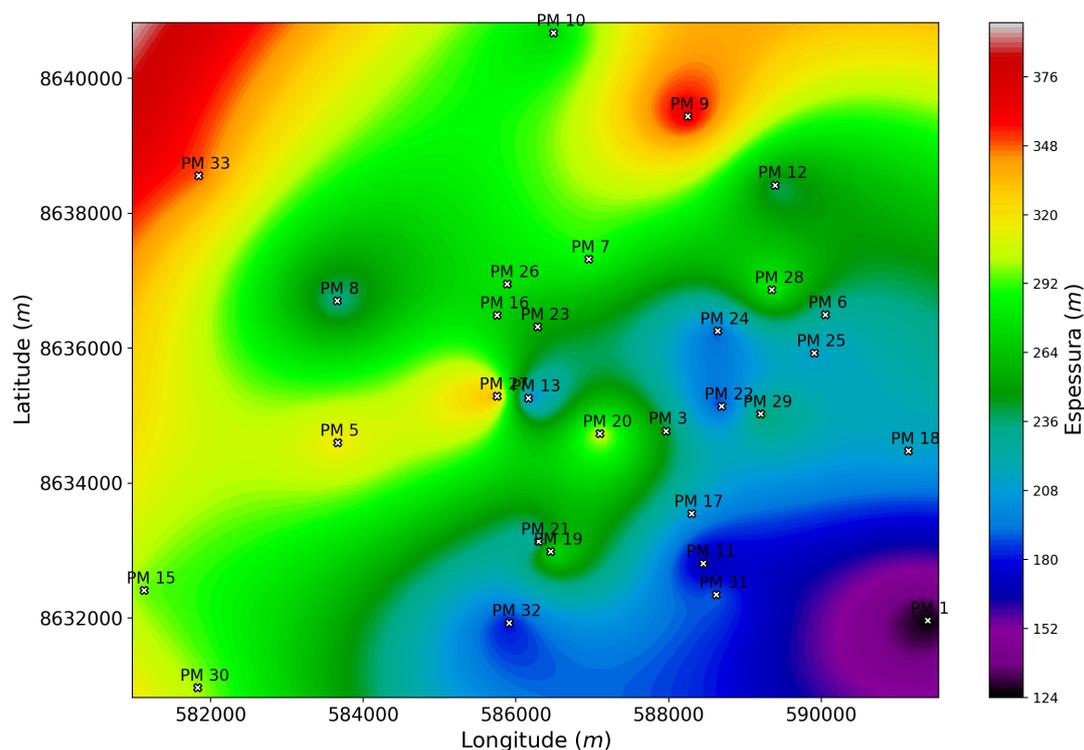


Figura 6.9: Mapa de espessura acumulada estimada das camadas permeáveis com água doce nos perfis disponíveis na área de Pojuca.

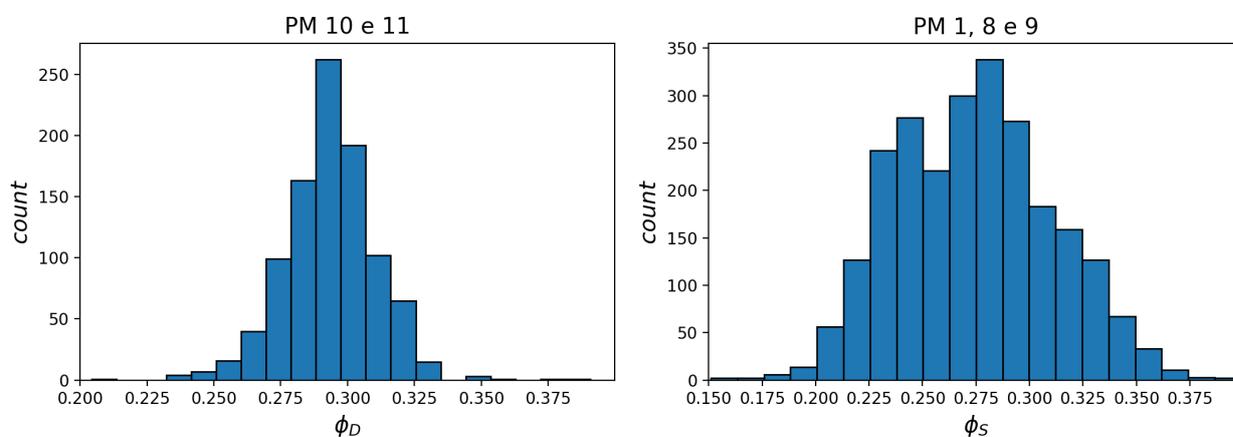


Figura 6.10: Histograma da porosidade calculada a partir dos registros de densidade dos perfis dos poços PM 10 e 11 (esquerda) e vazarosidade dos perfis dos poços PM 1, 8 e 9 (direita).

6.2 Área de Araçás

A área de estudo em Araçás possui aproximadamente 42 km² e localiza-se entre os municípios de Araçás e Entre Rios. Segundo o Plano de Desenvolvimento da Agência Nacional de Petró-

leo, Resolução nº509/2015, os reservatórios produtores de óleo nessa região são subdivididos em rasos e profundos. Os reservatórios rasos, com as melhores características hidráulicas, tem profundidades variando entre 500 a 1400 m e constituem-se de arenitos pertencentes às Formações Marfim e Pojuca, unidades sobre as quais a Formação São Sebastião repousa discordantemente.

Foram analisados 27 perfis geofísicos pertencentes a área de Araçás, e o primeiro passo foi a realização da análise de agrupamento a partir da metodologia apresentada na Seção 5.2 para separar as camadas permeáveis das impermeáveis e delimitar a base dos aquíferos potáveis. A análise de agrupamentos foi feita utilizando os mesmos parâmetros da Seção 6.1. Após interpretar os grupos utilizando o sistema de recomendação, usou-se o índice *ARI* para verificar a coerência entre análise de agrupamento do perfil PA 8 e uma partição externa obtida através de uma interpretação convencional. O índice *ARI* foi de aproximadamente 0,88. Quando verificada apenas a porção correspondente as zonas permeáveis com água doce obtém-se um valor aproximado de 0,87. A Figura 6.11 apresenta o agrupamento obtido e a partição utilizada na avaliação dos dados interpretados pelo algoritmo.

Após verificar o desempenho do algoritmo, gerou-se o mapa de contorno estrutural da profundidade da base da água doce da área de Araçás com interpretações de 18 perfis geofísicos, apresentado na Figura 6.12. É perceptível que a distribuição da base do aquífero assemelhasse a um antiforme, com ápice nas coordenadas UTM (590500,8657000) zona 24S, cuja a profundidade das zonas permeáveis com água doce é de cerca de 530 m e as zonas mais profundas tem suas bases entre 800 a 900 m. O nível estático, estimado a partir da Eq. (1.1), é apresentado na Figura 6.13. O perfil PA 27 foi eliminado, nesta etapa, para não subestimar ou superestimar os cálculos do nível estático e mais adiante, as espessuras calculadas. Com isto, a área foi reduzida para aproximadamente 30 km².

Duas seções esquemáticas, entrecruzadas, são apresentadas nas Figuras 6.14 e 6.15 com o objetivo de correlacionar os agrupamentos obtidos a partir do sistema de recomendações implementado. Novamente, é possível notar que o algoritmo apresenta dificuldades em distinguir zonas impermeáveis e zonas permeáveis com água salobra.

Para a estimativa do conteúdo de argila, 9 (nove) perfis geofísicos foram agrupados com diversos números de grupo, tomando-se como atributo a contagem de raios gama e, em seguida, calculou-se o coeficiente silhueta médio para cada um deles para verificar se a hipótese bimodal era a mais forte. A Tabela 6.3 mostra que a hipótese de dois grupos é a mais forte para todos os conjuntos de dados apresentados. Após verificada a hipótese bimodal, foram obtidos os parâmetros da Eq. (3.10) mostrados na Tabela 6.4, que permite estimar o conteúdo de argila e, conseqüentemente, distinguir arenitos, folhelhos e misturas

argilosas.

Estimada a argilosidade a partir da Eq. (3.10), e utilizando as interpretações obtidas do sistema de recomendação, foram gerados mapas de espessura acumulada das camadas permeáveis com água potável da área de Araçás (Figura 6.16). Devido a ausência de medidas de propriedades físicas em porções mais rasas do poço, foram mantidas as proporções de arenito para estimar a espessura total entre a base do aquífero com água doce e o nível estático. Para as análises de agrupamentos que não conseguiram distinguir de maneira adequada as zonas permeáveis das impermeáveis, foi utilizado o valor médio das proporções de arenito para se estimar a espessura total efetiva do aquífero. Nota-se um comportamento que remete ao antiforne da base do aquífero de água potável.

Foram calculadas as porosidades para os conjuntos de dados que haviam registros de densidade eletrônica e de vagarosidade sônica. Novamente, é considerada que a matriz é composta por quartzo. As análises para as medidas de densidade, realizadas nos perfis PA 16 e 17, apontaram um valor esperado para a porosidade calculada de $28,16 \pm 2,8\%$, enquanto as análises feitas sob os perfis PA 2, 4 e 18, usando as medidas de vagarosidade, mostraram um valor esperado de $24,32 \pm 4,08\%$: valores inferiores aos obtidos nos perfis correspondentes a área de Pojuca. A Figura 6.17 mostra a distribuição de porosidade obtida para as amostras de arenitos ($V_{sh} \leq 0$).

Depois de estimadas as espessuras e porosidades obtidas a partir da análise dos perfis de raios gama, potencial espontâneo, resistividade, sônico e densidade, calculou-se um volume de água de $2,72 \cdot 10^9 \text{ m}^3$ utilizando o valor médio da porosidade sônica e de $3,15 \cdot 10^9 \text{ m}^3$ quando utilizada a porosidade média calculada a partir das medidas de densidade. Resultados proporcionalmente compatíveis com os apresentados por Brandão (2008).

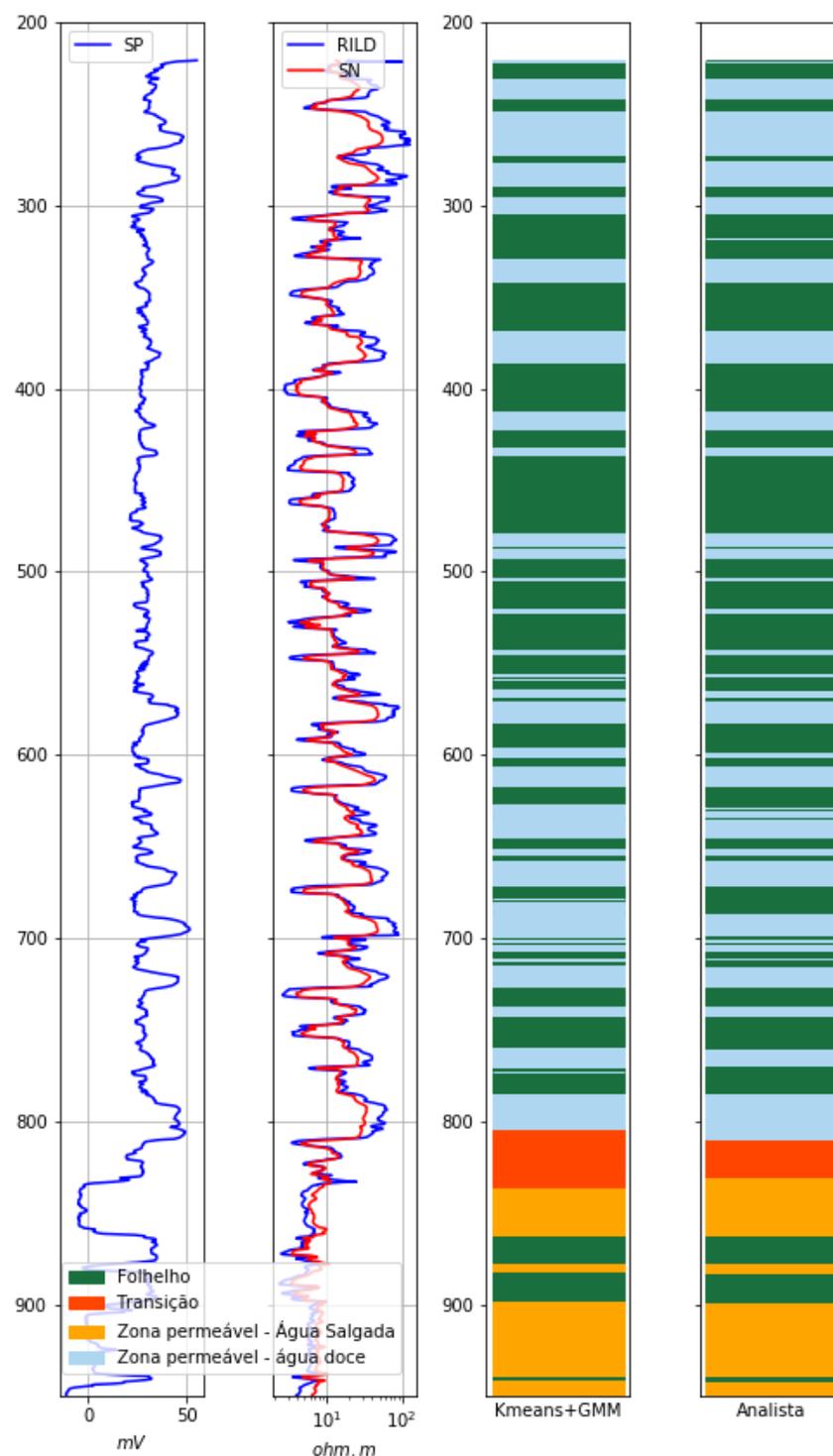


Figura 6.11: Interpretação dos agrupamentos obtidos no perfil PA 8 a partir do algoritmo de Modelo de Misturas Gaussianas inicializados a partir do centroide do algoritmo K-médias (terceira coluna) e a interpretação externa ao algoritmo para validação utilizando o Índice Rand Ajustado (*ARI*).

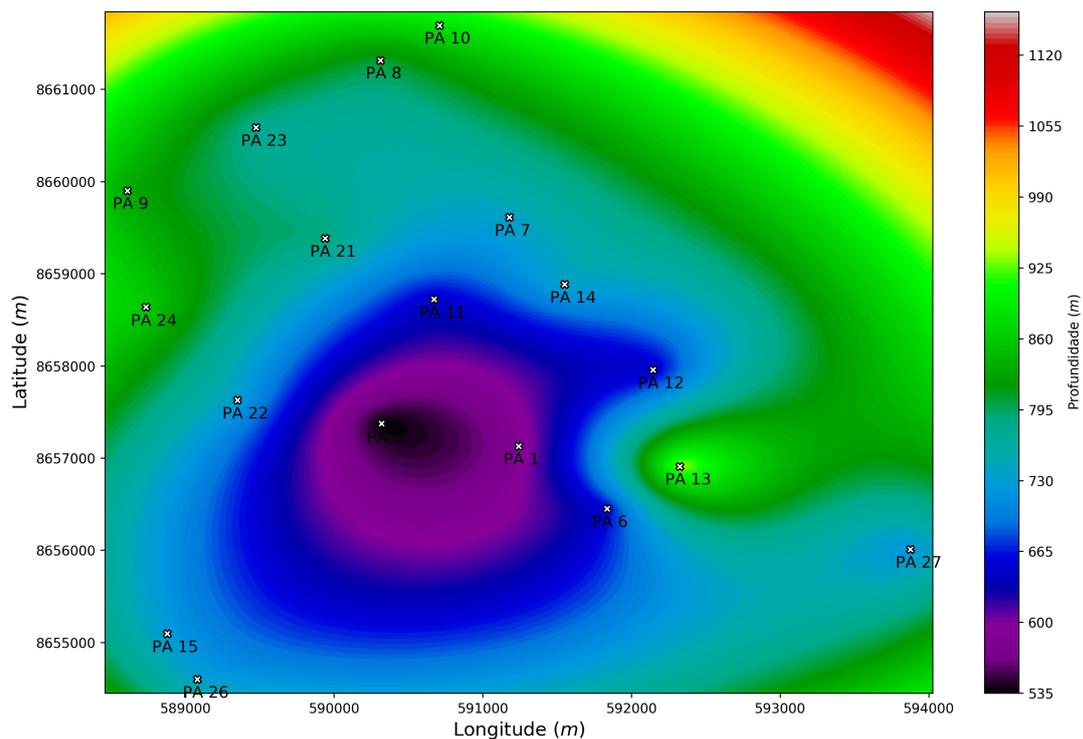


Figura 6.12: Mapa de contorno estrutural da base do reservatório de água doce na área de Araçás.

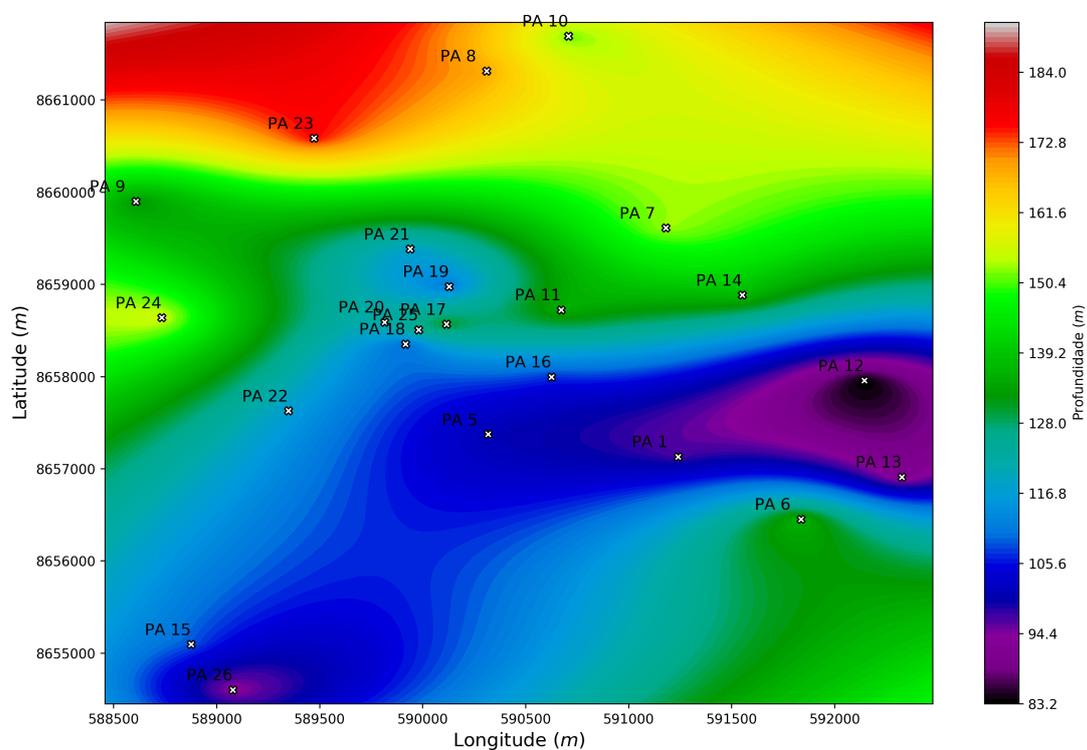


Figura 6.13: Nível Estático do aquífero livre correspondente a Área de Araçás.

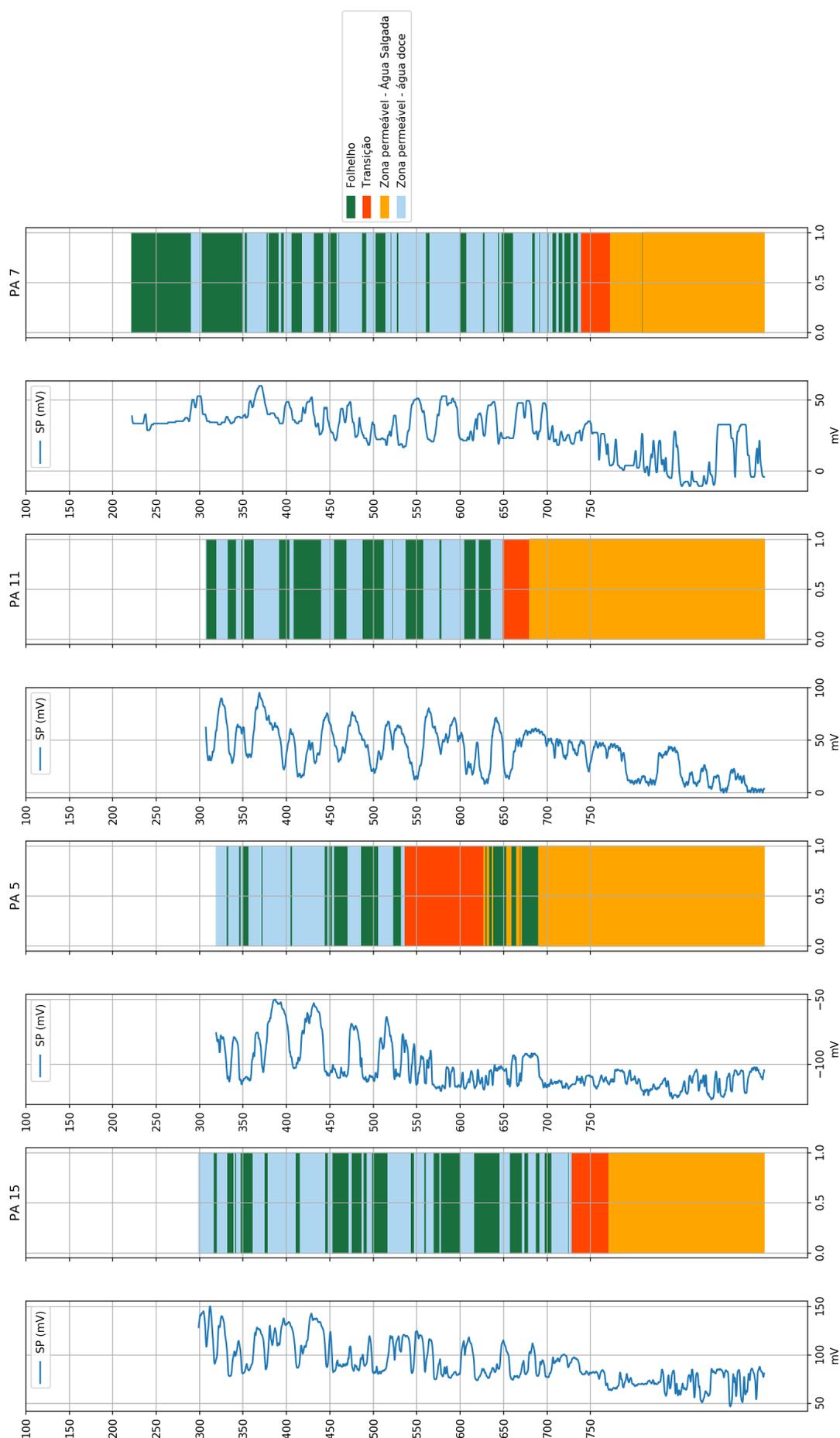


Figura 6.14: Seção longitudinal, sentido sudoeste-nordeste, apresentando quatro perfis da área de Araçás e suas respectivas interpretações obtidas a partir da análise de agrupamento utilizando o algoritmo de Modelos de Misturas Gaussianas inicializado a partir dos centroides do algoritmo K-médias.

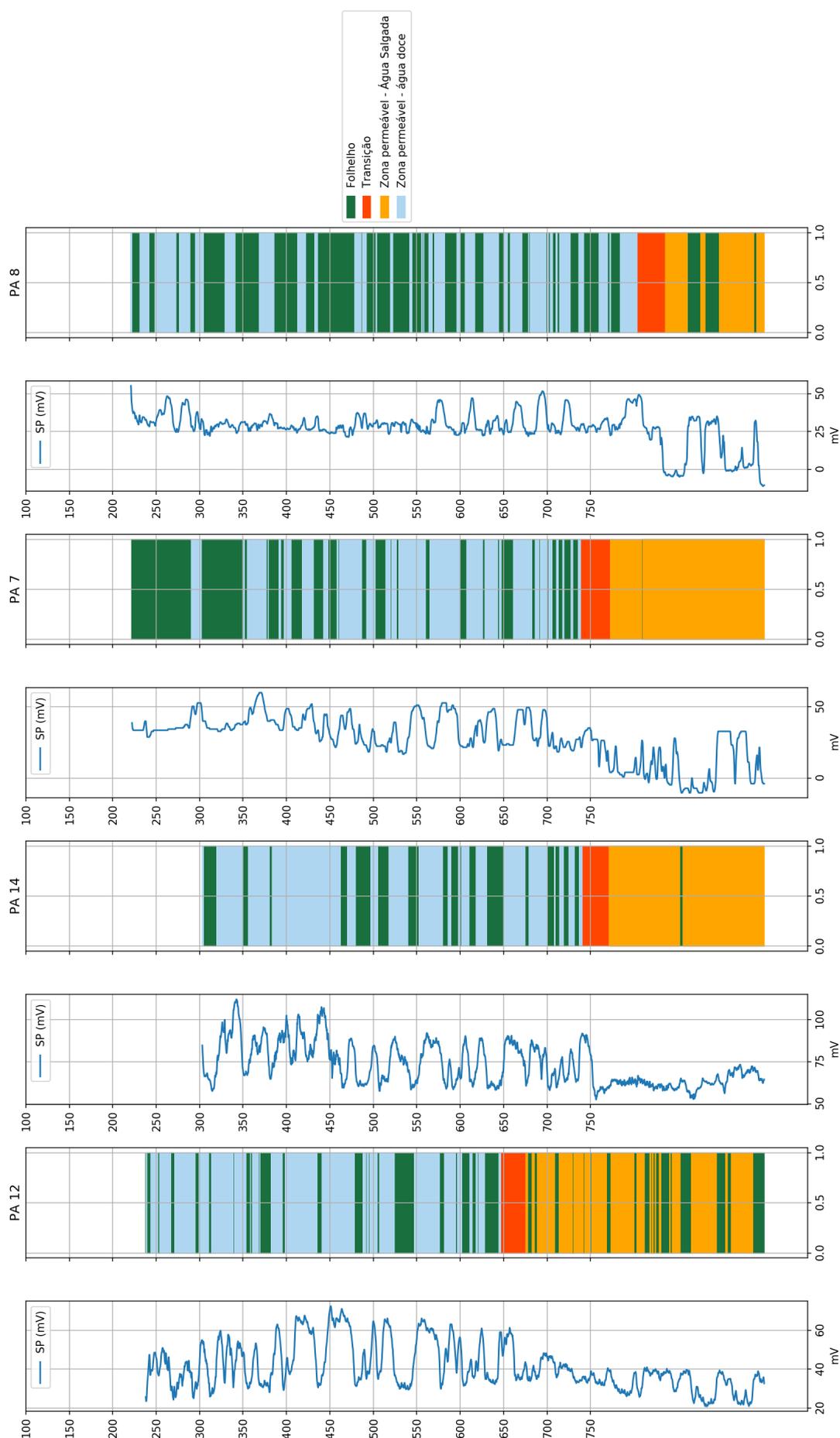


Figura 6.15: Seção transversal, sentido sudeste-noroeste, apresentando quatro perfis da área de Araçás e suas respectivas interpretações obtidas a partir da análise de agrupamento utilizando o algoritmo de Modelos de Misturas Gaussianas inicializado a partir dos centroides do algoritmo K-médias.

| N. Cluster | PA 2 | PA 4 | PA 16 | PA 17 | PA 18 | PA 19 | PA 20 | PA 21 | PA 27 |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 2 | 0.666 | 0.666 | 0.676 | 0.678 | 0.665 | 0.687 | 0.682 | 0.682 | 0.657 |
| 3 | 0.595 | 0.546 | 0.614 | 0.627 | 0.607 | 0.609 | 0.602 | 0.605 | 0.601 |
| 4 | 0.564 | 0.552 | 0.587 | 0.585 | 0.583 | 0.569 | 0.571 | 0.579 | 0.569 |
| 5 | 0.538 | 0.516 | 0.559 | 0.565 | 0.533 | 0.555 | 0.546 | 0.551 | 0.556 |
| 6 | 0.532 | 0.535 | 0.527 | 0.545 | 0.533 | 0.540 | 0.530 | 0.542 | 0.547 |

Tabela 6.3: Valor médio do índice silhueta de diferentes números de grupos para cada análise de agrupamento das medidas de raios gama dos perfis pertencentes a área de Araçás.

| Poço | μ_s | σ_s | μ_{sh} | σ_{sh} | Gr_{min} | Gr_{max} |
|-------|---------|------------|------------|---------------|------------|------------|
| PA 2 | 41.29 | 8.29 | 72.82 | 9.57 | 49 | 63 |
| PA 4 | 48.40 | 11.35 | 87.32 | 11.51 | 59 | 75 |
| PA 16 | 37.32 | 10.41 | 80.13 | 12.97 | 47 | 67 |
| PA 17 | 33.15 | 9.27 | 80.53 | 17.91 | 42 | 62 |
| PA 18 | 45.62 | 12.99 | 94.75 | 16.20 | 58 | 78 |
| PA 19 | 45.29 | 11.79 | 95.41 | 15.32 | 57 | 80 |
| PA 20 | 45.44 | 12.46 | 93.64 | 13.29 | 57 | 80 |
| PA 21 | 42.65 | 10.77 | 86.06 | 13.35 | 53 | 72 |
| PA 27 | 53.96 | 7.98 | 82.81 | 8.73 | 61 | 74 |

Tabela 6.4: Média (μ) e desvio padrão (σ) do modelo de misturas gaussianas da análise de agrupamento com dois grupos correspondentes a arenitos e folhelhos das medidas de raios gama, e parâmetros da Eq. 3.10 dados em unidade uAPI derivados dos Modelos de Misturas Gaussianas dos perfis pertencentes a área de Araçás.

6.3 Porção Norte da Bacia do Recôncavo

Informações da profundidade da base do aquífero de água potável de 74 poços foram adicionadas as análises de agrupamento dos 90 perfis examinados para estimar a reserva hídrica da porção norte da bacia do Recôncavo, compondo um total de 164 poços. O mapa de contorno estrutural da base do reservatório de água doce na porção norte da bacia do Recôncavo é apresentado na Figura 6.18. É possível notar que as regiões mais espessas do aquífero ocorrem no baixo de Alagoinhas, onde a base do reservatório atinge profundidades superiores a 1400 m. O aquífero mostra-se ainda potente em áreas dos municípios de Alagoinhas, Entre Rios e Araçás, onde alcança espessuras totais acima de 1000 m. As porções mais rasas correspondem a patamares e as regiões proximais das bordas flexural e falhada da bacia do Recôncavo. Na cunha sedimentar nordeste, à norte da Falha de Itanagra, a Formação São Sebastião esta ausente.

Dos poços adicionados não se dispunha de perfilagens geofísicas e portanto, não haviam

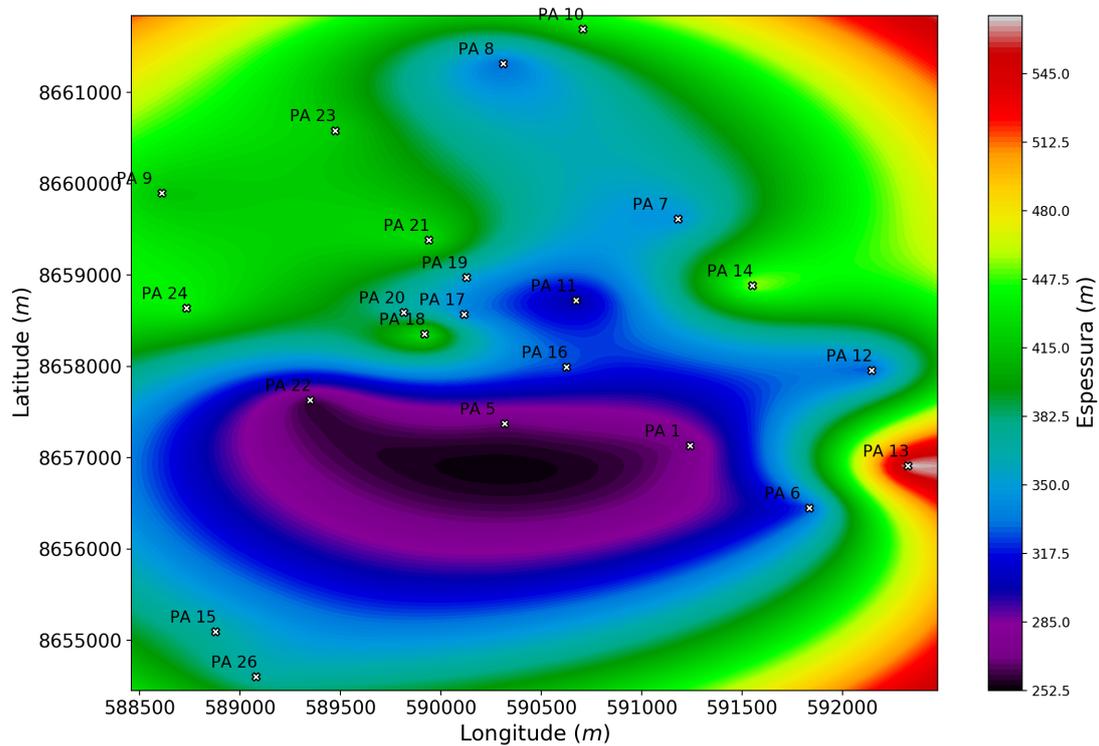


Figura 6.16: Mapa de espessura acumulada estimada das camadas permeáveis com água doce nos perfis disponíveis na área de Araçás.

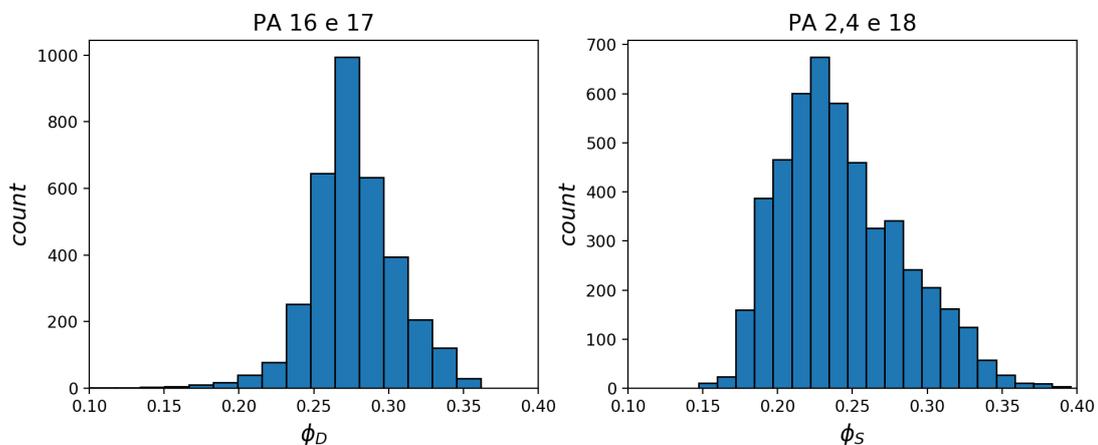


Figura 6.17: Histograma da porosidade calculada a partir dos registros de densidade dos perfis PA 16 e 17(esquerda) e vazarosidade dos perfis PA 2,4 e 18 (direita).

informações sobre as espessuras efetivas das zonas permeáveis com água potável. Por isso, espessuras das zonas permeáveis com água doce foram estimadas por comparação com a base desenvolvida. Isto é, tem aproximadamente as mesmas proporções de arenito dos poços mais próximos que foram avaliados por meio das análises de agrupamento. A Figura 6.19 mostra o mapa de espessura acumulada das camadas permeáveis indicando que as

regiões de Alagoinhas e Araças têm maiores volumes de reservatórios se comparadas com os outros municípios. Para o cálculo do volume de água potável, considerou-se que as unidades permeáveis tem uma porosidade média de 27%. Foi estimada uma reserva hídrica de $2,6 \times 10^{11} \text{ m}^3$ em uma área de 2895 km^2 , delimitada pelo polígono cuja arestas são os poços.

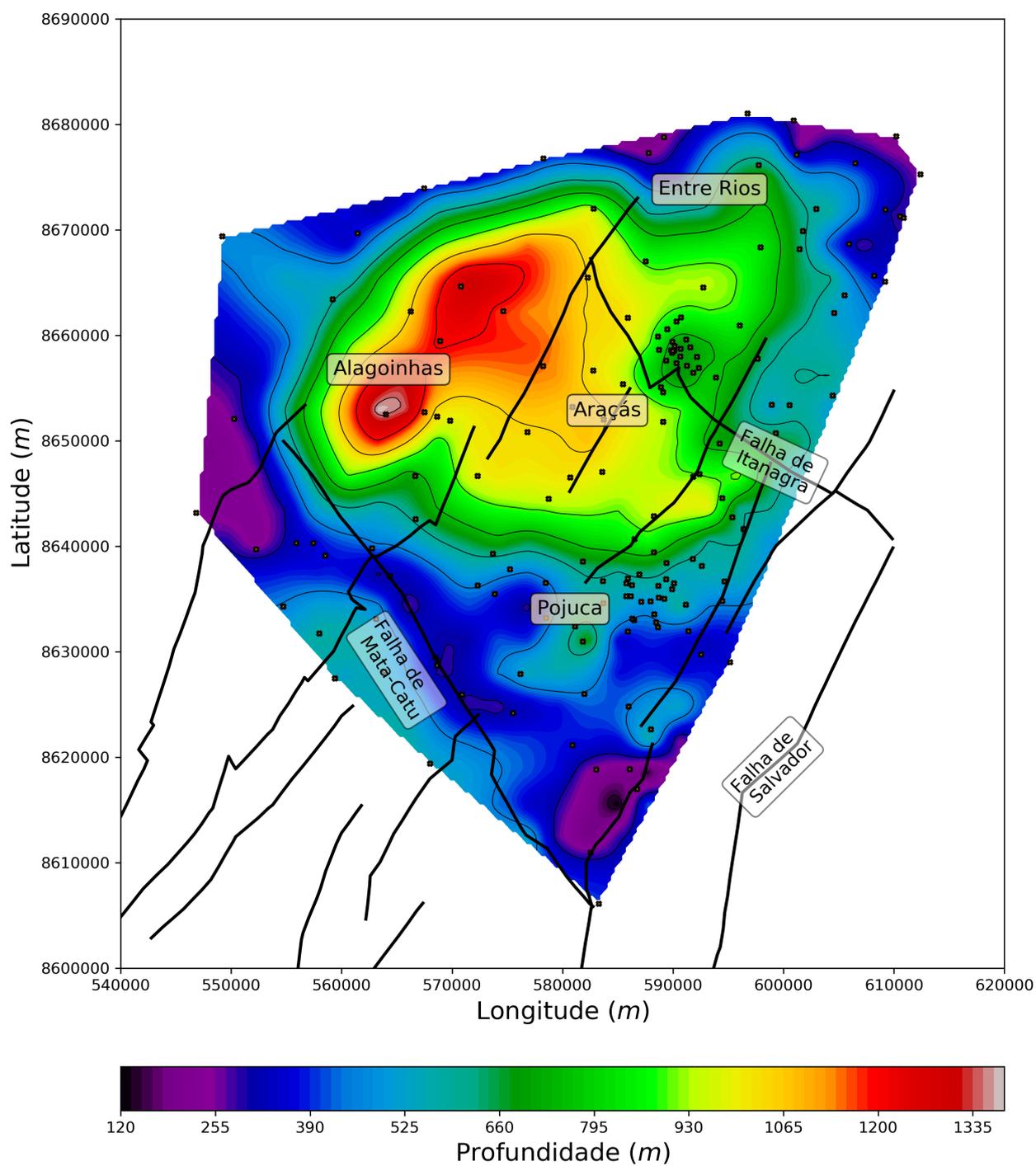


Figura 6.18: Mapa de contorno estrutural da base do reservatório de água doce da porção norte da Bacia do Recôncavo.

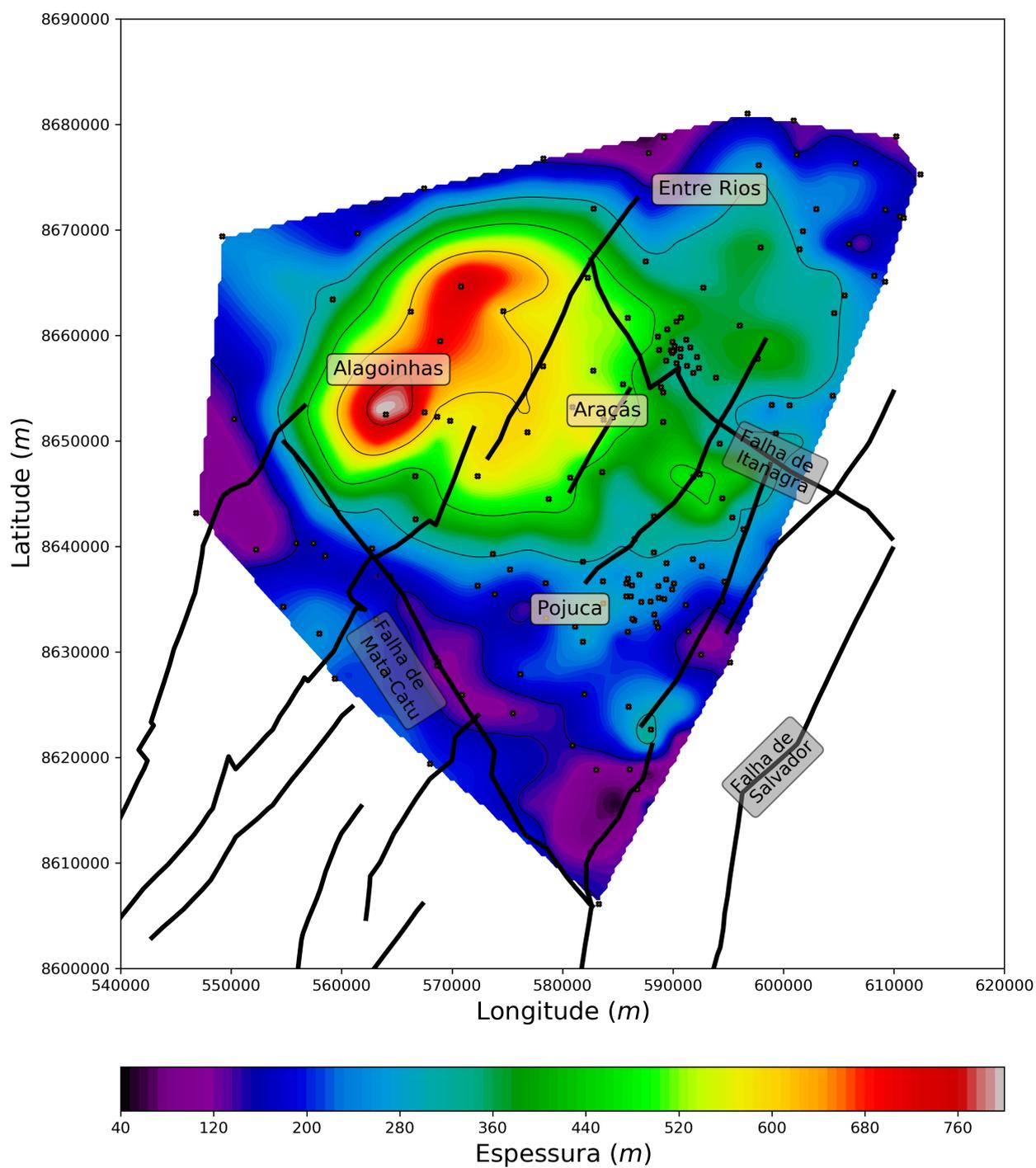


Figura 6.19: Mapa de espessura acumulada estimada das camadas permeáveis com água doce da porção norte do Bacia do Recôncavo.

7

Conclusões

Análises de 101 perfis geofísicos pertencentes a bacia do Recôncavo, foram utilizadas para desenvolver metodologias que permitiram estimar a espessura acumulada de arenitos com água potável de poços pertencentes a porção norte da Bacia. Foi observado, nas análises de agrupamentos realizadas utilizando a metodologia apresentada na Seção 5.2, a distinção de três principais grupos nos agrupamentos obtidos a partir do algoritmo de Modelo Misturas Gaussianas inicializado a partir dos centroides do método K-médias. Nos conjuntos de dados que utilizaram medidas de potencial espontâneo, resistividade e profundidade como atributos, foi constatada uma ineficiência do algoritmo em identificar zonas permeáveis em regiões de transição entre a água doce e salgada, o que comumente ocorre quando a resistividade do filtrado é aproximadamente igual a do fluido de formação, obrigando a elaboração de um sistema de recomendação para detectar tais zonas.

Obtidos os resultados das análises de agrupamento e sugerida a interpretação a partir do sistema de recomendação, utilizou-se interpretações auxiliares convencionais dos conjuntos de dados para a validação do algoritmo através do índice Rand ajustado. Observa-se uma forte coerência entre a interpretação convencional e as interpretações produzidas pelo algoritmo. Foi verificado um valor do índice *ARI* superior a 0,77 em zonas que contemplavam somente camadas permeáveis com água doce e não permeáveis.

Uma segunda análise foi feita em poços onde apenas haviam medidas de raios gama. Conforme mostrado por Vernik (2016), em uma distribuição areno-argilosa, espera-se distinguir, de maneira visual, duas gaussianas em formatos de sino correspondentes aos arenitos limpos e aos folhelhos, o que sugere uma metodologia, não-automática, de obtenção dos parâmetros requeridos para o cálculo da argilosidade. Partindo desta hipótese, foram feitos diversos

agrupamentos e, no geral, o índice de silhueta médio apontou um forte indício da existência de apenas dois grupos apresentando valores superiores 0,65, o que possibilitou introduzir uma metodologia para automatizar a obtenção dos parâmetros do índice normalizado de raios gama. A automatização do cálculo da argilosidade permitiu dividir as amostras conforme o conteúdo de argila em arenitos, folhelhos e siltitos. A metodologia foi expandida para medidas de resistividade elétrica que se comportam como distribuições bimodais, que em alguns casos, demonstra-se tão precisa quanto a análise utilizando a radiação gama.

As investigações feitas no mapa estrutural de base de água doce apontaram faixas variáveis entre 300 a 850 metros para a área de Pojuca e apresentou espessuras acumuladas de zonas permeáveis com água potável no intervalo de 125 a 400 m. Na região de Araçás é verificado um padrão estrutural de antifórme na base do aquífero, em conformidade com as análises realizadas por Brandão (2008); as localidades mais rasas estimadas foram de aproximadamente 550 m e as mais profundas em torno de 950 m. Mapas de espessura acumulada do arenito na área de Araçás apontaram regiões com espessuras úteis de arenitos superiores a 500 m.

Estimativas de porosidade empregando os perfis geofísicos que tinham medidas de densidade e vagarosidade possibilitaram constatar que o percentual de vazios é superior a 24% para os arenitos do sistema aquífero Recôncavo: em média 28,37% para a área de Pojuca e de 26,24% para a área de Araçás: valores compatíveis com aqueles disponíveis em literatura. Foi estimada uma reserva hídrica entre 7,83 e $8,4 \times 10^9$ m³ de água em uma área aproximada de 100,5 km² em Pojuca e em Araçás de 2,72 a $3,15 \times 10^9$ m³ em uma área de 30 km².

Por fim, foram combinadas informações de 164 poços, contando com aqueles utilizados na análise de agrupamento das Seções 6.1 e 6.2, com a finalidade de calcular a reserva hídrica disponível de toda a porção norte da Bacia do Recôncavo. Estimou-se um volume de água potável de $2,6 \times 10^{11}$ m³ em uma área de 2900 km². O mapa de espessura acumulada dos aquíferos mostrado na Figura 6.19 indica que os reservatórios de maiores volumes de água pertencem aos municípios de Alagoinhas e Araçás.

Pode-se afirmar, portanto, que metodologias baseadas em métodos de aprendizado de máquina aceleram o processo de interpretação e, frequentemente, produzem modelos suficientemente compatíveis com aqueles formulados por interpretações convencionais. Neste trabalho, dos 101 perfis disponíveis, pôde-se extrair informações a partir de análises de agrupamento. A utilização do algoritmo de modelos de misturas Gaussianas inicializado a partir dos centroides do método K-médias possibilitou subdividir o conjunto de dados em zonas permeáveis e não permeáveis, além de avaliar a profundidade da base da água doce através do sistema de recomendações implementado que tem como informações análises de agrupamento

dos atributos: potencial espontâneo, resistividade e profundidade. A segunda metodologia proposta, que segue as premissas que a distribuição de raios gama de uma formação areno-argilosa assemelha-se a uma distribuição bimodal, viabilizou a automatização da obtenção dos parâmetros necessários para o cálculo do conteúdo de argila e da argilosidade a partir dos desvios padrões e médias advindas dos modelos de misturas Gaussianas compostos por duas distribuições normais. Esta última abordagem propiciou a divisão das amostras em três principais grupos: Arenitos, folhelhos e misturas areno-argilosas.

Referências Bibliográficas

- Aragão, M. A. N. F. (1993) Arquitetura da Bacia do Recôncavo, 3º Congresso Internacional da Sociedade Brasileira de Geofísica, CisBGF, Resumo Expandido, p. 143-145, Rio de Janeiro, RJ, Brasil.
- Bishop, C. M. (2006) Pattern recognition and machine learning, springer.
- Brandão, A. T. (2008) Análise Estratigráfica e Hidrogeológica da Formação São Sebastião no Campo de Araças, Bacia do Recôncavo - Bahia, Dissertação de Mestrado, Centro de Pesquisa em Geofísica e Geologia CPGG Salvador, BA, Brasil.
- Caixeta, J. M.; Bueno, G. V.; Magnavita, L. P. e Feijó, F. J. (1994) Bacias do Recôncavo, Tucano e Jatobá, Boletim de Geociências, Petrobras, vol. 8, n.1, p. 163-172, Rio de Janeiro, RJ, Brasil.
- Caté, A.; Perozzi, L.; Gloaguen, E. e Blouin, M. (2017) Machine learning as a tool for geologists, The Leading Edge, **36**:215–219.
- Dubois, M. K.; Bohling, G. C. e Chakrabarti, S. (2007) Comparison of four approaches to a rock facies classification problem, Computers and Geosciences, **33**:599–617.
- Ellis, D. e Singer, J. (2007) Well Logging for Earth Scientists, Springer Netherlands, ISBN 9781402046025.
- Faceli, K.; Lorena, A. C.; Gama, J. e Carvalho, A. (2011) Inteligência artificial: Uma abordagem de aprendizado de máquina, Livros Técnicos e Científicos.
- Hastie, T.; Tibshirani, R. e Friedman, J. (2016) The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics), Springer, ISBN 0387848576.
- Holz, M. (2012) Estratigrafia de Sequencias: Historico, Principios e Aplicacoes, Interciência, ISBN 8571932905.
- James, G.; Witten, D.; Hastie, T. e Tibshirani, R. (2013) An introduction to statistical learning, vol. 112, Springer.
- Larionov, V. V. (1969) Radiometry of boreholes (in russian), Nedra, Moscow.

- Lima, A. L. L. (1999) Caracterização Hidráulica e Padrões de Poluição no Aquífero Recôncavo na Região de Camaçari - Dias D'Avila, Tese de Professor Titular, Centro de Pesquisa em Geofísica e Geologia, CPGG, Salvador, BA, Brasil.
- Lima, O. A. L. d. (1995) Water Saturation and Permeability from Resistivity, Dielectric and Porosity Logs, *Geophysics*, vol. 60, pp 1756-1764.
- Lima, O. A. L. d. (2014) Propriedades físicas das rochas. Bases da geofísica aplicada, Sociedade Brasileira de Geofísica (SBGF) e coeditoração do Instituto Nacional de Ciência e Tecnologia de Geofísica do Petróleo (INCT-GP).
- Lucia, F. J. (1999) *Carbonate Reservoir Characterization*, Springer.
- Mariano, A. H. (2013) Análise Hidrogeológica e Petrofísica do Aquífero São Sebastião no Campo de Miranga, Bacia do Recôncavo-BA, Curso de Pós-Graduação em Geologia. Instituto de Geociências - Universidade Federal da Bahia, Salvador, Bahia, Brasil.
- Mitchell, T. M. (1997) *Machine Learning*, McGraw-Hill Education, ISBN 0070428077.
- Moreira, F. S.; Garcias, E. C. O. e Soule, G. S. (2016) The Importance of Ground Water Aquifer Identification on Well Planning: An Example from the São Sebastião Aquifer, Northeast Brazil., Annual Convention and Exhibition, American Association of Petroleum Geologists (AAPG)., Calgary, Alberta, Canada., 2^o edic..
- Mukerji, T.; Avsethm, P. Mavko, G.; Takahashi, I. e González, E. (2001) Statistical rock physics: Combining rock physics, information theory, and geostatistics to reduce uncertainty in seismic reservoir characterization, *The Leading Edge*, Vol 20, pp 313-319, <http://dx.doi.org/10.1190/1.1438938>.
- Murphy, M. A. e Schlanger, S. O. (1963) Estruturas Sedimentares nas Formações Ilhas e São Sebastião, Bacia do Recôncavo, Brasil, *Boletim Técnico, Petrobras*, vol. 6, n.3/4, p. 215-258, Rio de Janeiro, RJ, Brasil.
- Nery, G. G. (2013) Perfilagem Geofísica de Poços. Fundamentos básicos com ênfase em petróleo, Sociedade Brasileira de Geofísica (SBGF) e coeditoração do Instituto Nacional de Ciência e Tecnologia de Geofísica do Petróleo (INCT-GP).
- Neto, O. V. (2006) Mapeamento da interface água doce – água salgada no aquífero são sebastião e apresentação de mapas com zoneamento para orientar o uso racional da água pela petrobras., Relatório do Núcleo de Estudos Hidrogeológicos e do Meio Ambiente (NEHMA), CDROM, Salvador.
- Peters, E. (2012) *Advanced Petrophysics: Geology, porosity, absolute permeability, heterogeneity, and geostatistics*, Advanced Petrophysics, Live Oak Book Company, ISBN 9781936909445.

- Prates, I. e Fernandez, R. (2015) *Bacia do Recôncavo: Sumário Geológico e Setores em Oferta*, Agência Nacional do Petróleo (ANP), 13^o edic..
- Rider, M. (2002) *The Geological Interpretation of Well Logs*, Rider-French Consulting Ltd., 2^o edic..
- Rousseeuw, P. J. (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics*, **20**:53–65.
- Santos, B. T.; Brandão, G. M. S.; Gomes, L. C.; Erthal, M. M. e Hatushika, R. S. (2007) *Estágios Evolutivos da Tectônica Rifte nas Bacias do Recôncavo-Tucano e Sergipe-Alagoas*, Curso de Especialização em Projetos de Análise de Bacias, Universidade do Estado do Rio de Janeiro, UERJ, Rio de Janeiro, RJ, Brasil.
- Santos, C. F. e Braga, J. A. E. (1990) O "Estado da Arte" da Bacia do Recôncavo, *Boletim de Geociências*, Petrobras, vol. 4, n.1, p. 35-43, Rio de Janeiro, RJ, Brasil.
- Schön, J. H. (2015) *Physical properties of rocks: Fundamentals and principles of petrophysics*, vol. 65, Elsevier.
- Schön, J. e Georgi, D. (2003) Dispersed shale, shaly-sand permeability - a hydraulic analog to the waxman-smits equation.
- Serra, O. (1984) *Fundamentals of Well-log Interpretation: The interpretation of logging data*, n^o v. 2 In *Developments in Petroleum Science*, Elsevier, ISBN 9780444426208.
- da Silva, A. J.; Lopes, R. C.; Vasconcelos, A. M. e Bahia, R. B. C. (2003) *Geologia, Tectônica e Recursos Minerais do Brasil*, Companhia de Pesquisa de Recursos Minerais, CPRM, cap. II, p.76, Brasília, DF, Brasil.
- Silva, O. B., C.; M, J. e Milhomem, P. S. (2007) *Bacia do Recôncavo*, *Boletim de Geociências*, Petrobras, vol. 15, n.2, p. 423-431. Rio de Janeiro, RJ, Brasil.
- Thomas, E. C. e Stieber, S. J. (1975) The distribution of shale in sandstones and its effect upon porosity., *SPWLA 16th Annual Logging Symposium*, New Orleans.
- Torghabeh, A.; Rezaee, R.; Moussavi-Harami, R.; Pradhan, B.; Kamali, M. e Kadkhodaie-Ilkhchi, A. (2014) Electrofacies in gas shale from well log data via cluster analysis: A case study of the perth basin, western australia, *Open Geosciences*, **6**(3).
- Vernik, L. (2008) Anisotropic correction of sonic logs in wells with large relative dip, *GEOPHYSICS*, **73**(1):E1–E5.
- Vernik, L. (2016) *Petrophysics of Siliciclastic Rocks*, Society of Exploration Geophysicists.
- Viana, C. F.; Gama Jr, E. G.; Simões, I. A.; Fonseca, J. R. e Alves, R. J. (1971) *Revisão Estratigráfica da Bacia do Recôncavo/Tucano*, *Boletim Técnico*, Petrobras, vol. 14, n.3/4, p. 157-192, Rio de Janeiro, RJ, Brasil.

- Waxman, M. H. e Smits, L. J. M. (1967) Electrical Conductivities in Oil-Bearing Shaly Sands, 42nd Annual Fall Meeting: Society of Petroleum Engineers, SPE, **8**:107–122.
- Wyllie, M.; Gregory, A. e Gardner, L. W. (1956) Elastic wave velocities in heterogeneous and porous media., *Geophysics*, **21**:41–70.
- Xia, K.; Hilterman, F. e Hu, H. (2018) Unsupervised machine learning algorithm for detecting and outlining surface waves on seismic shot gathers, *Journal of Applied Geophysics*, **157**:73–86.
- Xu, R. e Wunsch, D. (2009) *Clustering*, Wiley-IEEE Press, ISBN 9780470276808.
- Yang, S. e WEI, J. (2017) *Fundamentals of Petrophysics*, Springer.

Agradecimentos

Ao meu orientador Professor Olivar Antônio Lima de Lima por toda paciência, críticas positivas, ensinamentos transmitidos e incentivos que viabilizaram o surgimento e a finalização deste trabalho.

Ao meu coorientador Professor Ricardo Araújo Rios, por enxergar potencial nas aplicações de aprendizado de máquina desenvolvidas neste trabalho e por me fornecer o suporte necessário para compreender melhor este universo.

Ao meu primeiro orientador acadêmico, Wilson Mouzer Figueiró, por toda paciência e dedicação ao me iniciar como pesquisador. Seus ensinamentos foram fundamentais para o meu crescimento.

A todos os professores por cada conhecimento compartilhado e por cada desafio imposto que foram responsáveis pela construção do meu perfil acadêmico.

A minha família, meus pais: Moises de Cerqueira e Sandra da Silva Guerra, meus irmãos: Alexsandra, Alisson e Rafael e meu primo Alã.

A minha namorada, Betina Sodré, por ter me apoiado e por ter me incentivado nos momentos mais difíceis, tornando minha jornada mais leve e animada.

Aos meus sogros, Arlete Sodré e Luiz.

Aos meus amigos de graduação: Adriano, Daniel, Diego, Ildeson, Ítalo, Magnum e Yves, por todos os incentivos nos momentos em que pensei em desistir da geofísica.

Aos meus amigos de mestrado: Marcus Melo e Vitor Serravale.

Agradeço ao Centro de Pesquisa em Geofísica e Geologia (CPGG/UFBA) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte acadêmico, apoio financeiro e logístico que tornou possível a realização deste trabalho.

Agradeço a Petróleo Brasileiro S.A (Petrobras), Companhia de Engenharia e Recursos Hídricos da Bahia (CERB) e a Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) pelo fornecimento dos dados que viabilizaram esta pesquisa.