



Teresa Mourão Cerqueira e Silva

***MACHINE LEARNING* APLICADO À CLASSIFICAÇÃO DE
LITOFÁCIES UTILIZANDO PERFIS DE POÇOS DA BACIA DE
SANTOS**

**Trabalho Final de Curso
(Geologia)**

UFRJ
Rio de Janeiro
2022



UFRJ

Teresa Mourão Cerqueira e Silva

***MACHINE LEARNING* APLICADO À CLASSIFICAÇÃO DE LITOFÁCIES
UTILIZANDO PERFIS DE POÇOS DA BACIA DE SANTOS**

Trabalho Final de Curso de Graduação em Geologia do Instituto de Geociências, Universidade Federal do Rio de Janeiro – UFRJ, apresentado como requisito necessário para obtenção do grau de Geólogo.

Orientador:

Prof. Dr. Leonardo Fonseca Borghi de Almeida

Coorientador:

Lucas Abreu Blanes de Oliveira, MSc

TERESA, Mourão Cerqueira e Silva

***Machine Learning* Aplicado à Classificação de Litofácies Utilizando Perfis de Poços da Bacia de Santos** / Teresa Mourão Cerqueira e Silva -- Rio de Janeiro: UFRJ / IGeo, 2022.

49 p. : il.; 30cm

Trabalho Final de Curso (Geologia) – Universidade Federal do Rio de Janeiro, Instituto de Geociências, Departamento de Geologia, ano.

Orientador: Prof. Dr. Leonardo Fonseca Borghi de Almeida

Coorientador: Lucas Abreu Blanes de Oliveira, MSc

1. Geologia. 2. IGEO – Trabalho de Conclusão de Curso. I. Leonardo Fonseca Borghi de Almeida. II. Universidade Federal do Rio de Janeiro, Instituto de Geociências, Departamento de Geologia. III. Título.

Teresa Mourão Cerqueira e Silva

MACHINE LEARNING APLICADO À CLASSIFICAÇÃO DE LITOFÁCIES
UTILIZANDO PERFIS DE POÇOS DA BACIA DE SANTOS

Trabalho Final de Curso de Graduação em
Geologia do Instituto de Geociências,
Universidade Federal do Rio de Janeiro –
UFRJ, apresentado como requisito
necessário para obtenção do grau de
Geólogo.

Orientador:

Prof. Dr. Leonardo Fonseca Borghi de
Almeida

Coorientador:

Lucas Abreu Blanes de Oliveira, MSc

Aprovada em: 29.12.2022

Por:

Orientador: Prof. Dr. Leonardo Fonseca Borghi de Almeida (UFRJ)

D.Sc. Marco Antonio da Silva Braga (UFRJ)

M.Sc. Vinícius Carneiro Santana (Lagesed)

“Escrever é fácil. Se começa com uma letra maiúscula e se termina com um ponto final. No meio colocam-se as idéias.”

Pablo Neruda

AGRADECIMENTOS

Escrever pode ser fácil, mas elaborar as ideias que vêm entre a letra maiúscula e o ponto final é um grande desafio. Como pôr em poucas palavras a tamanha gratidão que sinto às tantas pessoas e oportunidades que me fizeram chegar até aqui?

Os primeiros nomes que me vêm à mente, como é de se esperar, são os dos meus pais: Luciana e Wanderley. Afinal, foram eles quem me deram os primeiros ensinamentos, os primeiros incentivos e as primeiras oportunidades. Minha mãe é a maior inspiração que tenho, a quem eu recorro em todos os momentos. E meu pai é o embaixador do riso, me alegrou em tantos momentos difíceis da vida.

Em seguida, agradeço aos meus irmãos de sangue Tiago e Lucas, e às minhas irmãs de coração Anna e Darién. Vocês me trazem a privilegiada certeza de que estarei sempre me aventurando pela vida nas mais seguras companhias. Obrigada aos meus avós, Beto, Teca, Ceição e Genô, que fazem festa no céu, celebrando esse momento. E enfim, dedico minhas melhores palavras ao meu companheiro Ciro Rosa, que nunca me deixa esquecer do quanto sou capaz e que me apoia em todos os momentos. A vida do seu lado é sempre a que quero viver.

Gostaria de manifestar minha mais sincera gratidão aos meus orientadores, Borghi e Lucas. Sem eles a realização deste trabalho não seria possível. Sobre tudo ao Lucas, que tanto me ensinou e me incentivou durante essa etapa de conclusão de curso. Sua parceria foi fundamental, me faltam palavras para agradecer.

Não poderia deixar de fora meus amigos de turma e colegas de curso, que estiveram comigo durante todo o longo trajeto até a formatura. Agradeço especialmente às minhas amigas do grupo M. Queens, por tudo que vivemos juntas. E o mais sincero obrigada a minha amiga Sarah Aleixo que vem me acompanhando nas minhas maiores conquistas profissionais até então.

Agradeço ainda a todo corpo docente da Geologia UFRJ, um super obrigada a cada professor que contribuiu para minha formação e meus conhecimentos. Também há de se agradecer a todos os funcionários que desempenham um papel fundamental no funcionamento da faculdade e com os quais convivi por anos: Anadja e Elizabeh da secretaria, todos os funcionários da limpeza, da segurança e os motoristas de campo. Um obrigada também aos meus colegas de trabalho da ExxonMobil que muito contribuem para meu desenvolvimento profissional.

Por fim, agradeço à UFRJ, pela minha formação. Os aprendizados e as memórias que construí aqui foram meus maiores presentes da faculdade.

RESUMO

SILVA, Teresa Mourão. **Machine learning aplicado à classificação de litofácies utilizando perfis de poços da Bacia de Santos**. 2022. Trabalho Final de Curso (Geologia) – Departamento de Geologia, Instituto de Geociências, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Atualmente, é muito comum a classificação automática de litofácies por meio de sistemas especializados. Entretanto, mesmo que os algoritmos de aprendizagem de máquinas em *softwares* comerciais de avaliação petrofísica permitam uma classificação rápida, ela às vezes é feita utilizando perfis de poços inadequados para o propósito da avaliação, o que pode levar a resultados inconsistentes. A partir dessa motivação, o intuito do trabalho é investigar a capacidade do algoritmo de clusterização *k-means* de classificar litofácies a partir de perfis de poço. Dessa forma, foram utilizados intervalos siliciclásticos do reservatório do Pós-sal e carbonáticos do reservatório do Pré-sal para entender a diferença das respostas que a máquina fornece em cada um dos cenários testados. Os dados, como mineralogia da rocha, salinidade e temperatura da água, composição do fluido presente no reservatório, além dos perfis de poços (calibre do poço, Raios-gama e Raios-gama espectral, resistividade, densidade, porosidade neutrônica, porosidade total do RMN, fator fotoelétrico e vagarosidade da onda cisalhante e compressional), foram fornecidos pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP). O projeto foi desenvolvido em duas principais etapas: a primeira consistiu em gerar um modelo de subsuperfície e simular nele perfis sintéticos utilizando o *software* UTAPWeLS e compará-los com os perfis reais para validar o modelo gerado; já a segunda etapa compreendeu o processo de clusterização do modelo de subsuperfície e dos perfis reais para estimar litofácies e eletrofácies (respectivamente) através do *software* Weka. Em seguida, os perfis de poços foram removidos um a um do processo de clusterização para avaliar como a falta de informação poderia influenciar na classificação do algoritmo. Visando garantir o controle nesse processo, toda a primeira etapa levou em consideração os dados de rocha e fluido cedidos pela ANP para construir um modelo de subsuperfície representativo da área de estudo. A segunda etapa, por sua vez, teve a mesma rotina de clusterização aplicada tanto no modelo de subsuperfície quanto nos perfis de poço, a fim de tornar a comparação entre eles válida. Como resultado, foram obtidos os perfis faciológicos gerados a partir do modelo de subsuperfície e a partir dos perfis de poços, que se mostraram muito mais satisfatórios no reservatório siliciclástico do que no carbonático. Fica evidente, ao final do estudo, que o processo de classificação de litofácies por meio de perfis de poços não é algo trivial e, por mais que as máquinas auxiliem nesse processo, ele ainda demanda um conhecimento profundo dos especialistas.

PALAVRAS-CHAVE: aprendizado de máquinas; perfis de poços; classificação de litofácies; Bacia de Santos.

ABSTRACT

SILVA, Teresa Mourão. **Machine learning applied to facies classification using well logs from the Santos Basin**. 2022. Trabalho Final de Curso (Geologia) – Departamento de Geologia, Instituto de Geociências, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Currently, it is very common to perform facies classification automatically through specialized systems. However, even if machine learning algorithms in commercial petrophysical evaluation software allows a fast classification, it is sometimes done using inadequate well logs for the evaluation purpose, which can lead to inconsistent results. Based on this motivation, the aim of this study is to investigate the ability of the *k-means* clustering algorithm to classify lithofacies from well logs. Thus, siliciclastic intervals of the post-salt reservoir and carbonate intervals of the pre-salt reservoir were used to understand the difference in the responses that the machine provides in each of the tested scenarios. The data, such as rock mineralogy, salinity and water temperature, composition of the fluid present in the reservoir, plus the well logs (caliper, gamma ray and spectral gamma ray, resistivity, density, neutrons porosity, total porosity of NMR, photoelectric factor, compressional and shear wave slowness), were provided by the National Petroleum Agency (ANP). The study was developed in two main stages: the first one consisted of generating an earth model and simulating synthetic logs using the UTAPWeLS software to compare them with the real ones; the second stage comprised the clustering process of the earth model and the real logs to estimate lithofacies and electrofacies (respectively) using the Weka software. After that, the well logs were removed one by one from the clustering process to assess how the lack of information might influence in the algorithm classification. To ensure the quality control in this process, the entire first stage took into account the rock and fluid data provided by the ANP to build an earth model representative of the study area. The second stage, in turn, had the same clustering routine applied to both the earth model and the well logs, to make the comparison between them valid. As a result, the facies logs generated from the earth model and from the well logs were obtained, which were much more satisfactory in the siliciclastic reservoir than in the carbonate one. It is evident, at the end of the study, that the process of classifying facies using well logs is not trivial and, as much as machines may help in this, it still demands an in-depth knowledge of well log experts.

KEYWORDS: machine learning; well logs; facies classification; Santos Basin.

LISTA DE FIGURAS

Figura 1: (a) Mapa de localização da Bacia de Santos (modificado de Souza, 2019). (b) Algumas das principais feições de seu arcabouço estrutural (modificado de Garcia et al. 2012; retirado de Fernandez, 2017). ACF = Alto de Cabo Frio; AF = Alto de Florianópolis; BC = Bacia de Campos; BP = Bacia de Pelotas; RJ = Rio de Janeiro; SC = Santa Catarina; SP = São Paulo; PR = Paraná;

Figura 2: Produção mensal de petróleo por Bacia no Brasil no ano de 2021 (Painel Dinâmico da ANP, 2022);

Figura 3: Carta estratigráfica da Bacia de Santos (Moreira et al 2007);

Figura 4: Mapa de localização do Campo de Mexilhão (ANP, 2016);

Figura 5: Histórico de produção de gás não associado e condensado do bloco de Mexilhão, na Bacia de Santos (ANP, 2016);

Figura 6: Mapa de localização do Campo de Tupi (modificado de ANP, 2018);

Figura 7: Histórico de produção de petróleo e gás natural do bloco de Tupi, na Bacia de Santos (ANP, 2018);

Figura 8: Distribuição do espectro de radiação gama entre as respectivas janelas de energia referentes ao conteúdo de Th, U e K presentes na formação (Ellis e Singer, 2007);

Figura 9: Resposta típica de um perfil de resistividade (Glover, 2016);

Figura 10: Modelo esquemático da ferramenta de que faz a leitura da densidade da formação, correndo junto à parede do poço (Serra, 1984);

Figura 11: Gráficos indicando a relação entre (a) a razão das contagens dos detectores e a porosidade para diferentes tipos de rocha, (b) a razão das contagens dos detectores e a distância de migração e (c) a distância de migração e a porosidade para diferentes tipos de rocha. (Ellis & Singer, 2007);

Figura 12: Exemplo de posicionamento de limites de camadas (retas verticais do perfil quadratizado) em relação às variações do perfil de nêutrons adquirido em um poço (curvas preta e vermelha representando o perfil real e simulado através do modelos de subsuperfície, respectivamente). (Tittman, 1991);

Figura 13: Módulo Paleta Composicional do *software* UTAPWeLS, que permite criar formações geológicas através da escolha de componentes minerais, fluidos, porosidade, pressão e temperatura das rochas encontradas no poço;

Figura 14: Exemplo de definição de informação a priori do módulo Inversão Composicional Conjunta do *software* UTAPWeLS. Considerando o histograma acima para o componente quartzo, o processo iterativo irá convergir para uma fração de quartzo que respeite a distribuição a priori, e que pode ser calibrada através do histograma de fração de quartzo observado em análises de amostras de rocha;

Figura 15: representação esquemática de como o algoritmo *k-means* funciona, onde (a) representa a primeira etapa do processo, em que são estabelecidos os objetos que representam o centroide de cada *cluster*, (b) é o passo seguinte, onde são calculadas as distâncias dos centróides até todos os objetos; (c) definição dos *clusters* e cálculo da média de cada objeto dentro do *cluster* e, por fim; (d) representação da última etapa do processo onde os objetos de cada *cluster* são confirmados e os *clusters* são estabelecidos definitivamente. (CARDOSO et al., 2008);

Figura 16: fluxograma proposto para metodologia do trabalho;

Figura 17: Histograma de frações minerais utilizados como informação a priori do módulo Inversão Composicional Conjunta do poço do intervalo Pós-sal;

Figura 18: Histograma de frações minerais utilizados como informação a priori do módulo Inversão Composicional Conjunta do poço do intervalo Pré-sal;

Figura 19: fluxograma indicando a ordem de clusterização utilizada no presente trabalho;

Figura 20: exemplo didático de um gráfico do cotovelo utilizado para determinar a quantidade ideal de *clusters*;

Figura 21: modelo de subsuperfície gerado a partir dos dados do poço do intervalo Pós-sal;

Figura 22: modelo de subsuperfície gerado a partir dos dados do poço do intervalo Pré-sal;

Figura 23: sobreposição dos perfis adquiridos em poço e dos simulados no UTAPWeLS, referente ao intervalo Pós-Sal. REAL = perfis adquiridos no poço; SIM = perfis simulados no *software*;

Figura 24: sobreposição dos perfis adquiridos em poço e dos simulados no UTAPWeLS, referente ao intervalo Pré-Sal. REAL = perfis adquiridos no poço; SIM = perfis simulados no *software*;

Figura 25: Gráfico do cotovelo contendo todos os resultados normalizados do poço do intervalo Pós-sal;

Figura 26: Gráfico do cotovelo contendo todos os resultados normalizados do poço do intervalo Pré-sal;

Figura 27: perfil faciológico a partir do resultado de clusterização das frações minerais do modelo de subsuperfície, de todos os poços e de cada etapa de retirada de informação. Referente aos resultados do poço do intervalo Pós-sal. Cluster 1 = oliva; cluster 2 = laranja; cluster 3 = bege; cluster 4 = verde; cluster 5 = amarelo;

Figura 28: perfil faciológico a partir do resultado de clusterização das frações minerais do modelo de subsuperfície, de todos os poços e de cada etapa de retirada de informação. Referente aos resultados do poço do intervalo Pré-sal. Cluster 1 = azul; cluster 2 = amarelo; cluster 3 = vermelho; cluster 4 = celeste; cluster 5 = laranja; cluster 6 = verde.

LISTA DE TABELAS

Tabela 1: informações da base de dados do poço do intervalo Pós-sal e Pré-sal, com total de metros;

Tabela 2: valores da dispersão média para cada valor de k nos diferentes cenários avaliados;

Tabela 3: média do conteúdo mineralógico de cada *clusters*;

Tabela 4: classificação das litofácies que cada cluster representa.

SUMÁRIO

Agradecimentos	vi
Resumo.....	vii
Abstract.....	viii
Lista de Figuras.....	ix
Lista de Tabelas.....	xiii
1. INTRODUÇÃO	1
1.1 Apresentação	1
1.2 Objetivo	2
2. FUNDAMENTOS TEÓRICOS.....	4
2.1 Área de Estudo	4
2.1.1 Pré-Sal	6
2.1.2 Pós-Sal.....	8
2.2 Perfis de Poços	10
2.3 Modelo de Subsuperfície.....	17
2.4 <i>Machine Learning</i>	21
2.4.1 Clusterização.....	22
3. MATERIAIS E MÉTODOS.....	24
3.1 Base de Dados.....	25
3.2 Criação do Modelo de Subsuperfície	27
3.3 Clusterização	29
4. RESULTADOS	32
4.1 Modelo de Subsuperfície.....	32
4.2 Clusterização	36
4.2.1 Gráfico do Cotovelo.....	36
4.2.2 Análise de <i>Clusters</i>	39
5. DISCUSSÕES	44
6. CONCLUSÃO	46
7. REFERÊNCIAS BIBLIOGRÁFICAS	47

1. INTRODUÇÃO

1.1 Apresentação

O *machine learning* tem se tornado cada vez mais popular em diferentes áreas das Geociências, tanto na indústria quanto na academia (GUARIDO, 2018). Trata-se de um conjunto de análise de dados e métodos que incluem classificação, agrupamento e regressão, e que pode ser definida como a capacidade de sistemas de adaptar ou modificar suas ações continuamente em busca de um maior nível de precisão (HALL, 2016).

A vantagem de utilizar as máquinas para determinadas tarefas é que elas são capazes de acelerar processos e, muitas vezes, entregar resultados com maior acurácia. Se por um lado, há atividades na geologia em que o uso de *machine learning* (ML) já é consagrado, por outro, atividades como a classificação de litofácies ainda encontra desafios para o uso de ML, sobretudo por ser baseada em parâmetros que dependem de uma análise interpretativa humana. Portanto, ainda não é possível ter uma autonomia completa das máquinas para realizar esta atividade sem supervisão de um intérprete, pois trata-se de uma prática que demanda conhecimento tácito.

Litofácies são unidades rochosas que se caracterizam por suas propriedades composicionais e texturais. A classificação de litofácies, por sua vez, é o ato de analisar os diferentes padrões litológicos da subsuperfície, reconhecendo os limites entre uma litologia e outra e designar um nome à unidade de rocha com base em suas características litológicas. Naturalmente, a forma mais precisa para se realizar a classificação de litofácies é por meio de métodos diretos, como a análise de amostras de testemunhos. O problema, porém, está nos custos associados a esta prática, que tornam inviável realizar a classificação litológica de um poço baseando-se apenas em testemunhagem.

Nesse sentido, os métodos indiretos de avaliação se fazem necessários para que a classificação de litofácies seja factível. Diversas ferramentas e parâmetros são utilizados como soluções alternativas nesses casos, como a interpretação de perfis de poços, por exemplo. Estabelecer os limites entre uma camada e outra a partir da interpretação de perfis de poços é o processo de reconhecer as chamadas eletrofácies.

Este método é amplamente empregado na indústria de óleo e gás, e, muitas vezes, é feito de forma manual por intérpretes, o que pode tornar o processo lento.

Nesse contexto, uma opção interessante para otimizar e aprimorar a classificação faciológica, através de perfis de poços, é treinar algoritmos para que eles se tornem capazes de realizar uma interpretação inteligente e confiável, automatizando esse processo.

Atualmente, a classificação de litofácies, muitas vezes, já é feita por sistemas especialistas. No entanto, embora os algoritmos de ML, em *softwares* comerciais de avaliação de perfis, permitam uma classificação rápida, geralmente ela é feita através de métodos de ML supervisionados, podendo causar memorização (*overfitting*). Além disso, outro problema dessa questão é que nem sempre as propriedades medidas em perfis são as mais adequadas para diferenciar determinadas litofácies, sendo que raramente uma análise criteriosa dessa capacidade é feita. Assim, o uso da simulação de perfis pelo *software University of Texas at Austin Petrophysical and Well-Log Simulator* (UTAPWeLS) pode apresentar novas alternativas interessantes para contribuir com o avanço científico-tecnológico desta área de estudo.

Para permitir a testagem de ML, foram simulados dois poços com base em dados de poços do campo de Tupi e de Mexilhão, ambos na Bacia de Santos. A escolha dessa bacia justifica-se pela sua diversidade de litofácies de interesse, abrangendo desde rochas siliciclásticas até carbonatos, por ser uma grande área de potencial exploratório, além de ser, atualmente, a maior bacia em termos de produção de hidrocarbonetos do país (ANP, 2022). A partir dessas considerações iniciais, apresenta-se a seguir o objetivo geral do estudo.

1.2 Objetivo

O presente trabalho tem como objetivo investigar a capacidade do algoritmo de ML, *K-means*, de prever e classificar litofácies através de dados de perfis de poços, tanto em rochas carbonáticas do intervalo Pré-sal, quanto em rochas siliciclásticas do intervalo Pós-sal.

Buscou-se gerar um modelo de subsuperfície (*earth model*) através do *software* UTAPWeLS que honrasse os dados de perfis adquiridos em poços e disponibilizados pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP). Se validado, esse modelo de subsuperfície seria o que melhor representaria em termos composicionais as litofácies observadas nesses reservatórios, fornecendo um gabarito para se comparar com as respostas de perfis. Além disso, a proposta é

analisar a influência de determinados perfis para a classificação de litofácies, e se eles, de fato, apresentam uma boa acurácia nesse processo.

Os objetivos específicos do estudo compreendem:

- O capítulo 4.1 apresenta a geração de um modelo de subsuperfície que honra os perfis adquiridos nos poços através de simulações no software UTAPWeLS;
- No capítulo 4.3 tem-se a clusterização do modelo de subsuperfície gerando litofácies e, posteriormente, os perfis de poços reais, gerando eltrofácies. Ambos utilizando o algoritmo *k-means*;
- Ainda no capítulo 4.3, é apresentado o teste dos limites de quais perfis de poços são necessários para estabelecer uma boa classificação litológica através da remoção progressiva de perfis, durante a clusterização;
- O capítulo 5 analisa as diferenças entre os resultados obtidos em cada poço estudado.

2. FUNDAMENTOS TEÓRICOS

2.1 Área de Estudo

Para investigar a validade do uso de ML para previsão e classificação de litofácies, no contexto geológico de rochas carbonáticas do intervalo Pré-sal, e de rochas siliciclásticas do intervalo Pós-sal, foram selecionados dois poços distintos: um no Campo de Tupi e outro no Campo de Mexilhão, na Bacia de Santos.

A Bacia de Santos localiza-se no sudeste brasileiro, ocupando o litoral de quatro estados: Rio de Janeiro, São Paulo, Paraná e Santa Catarina (Figura 1). Sua área é de aproximadamente 350.000 km², e ela pode apresentar lâmina d'água de até 3.000 m. Seus limites geológicos se dão a norte com a Bacia de Campos, separada pelo alto estrutural de Cabo Frio, e a sul com a Bacia de Pelotas, pelo alto de Florianópolis (MOREIRA et al., 2007). Já a oeste da bacia, o limite é estabelecido pelos cinturões das serras costeiras, enquanto a leste, a bacia culmina na base do Platô de São Paulo (MACEDO, 1989).

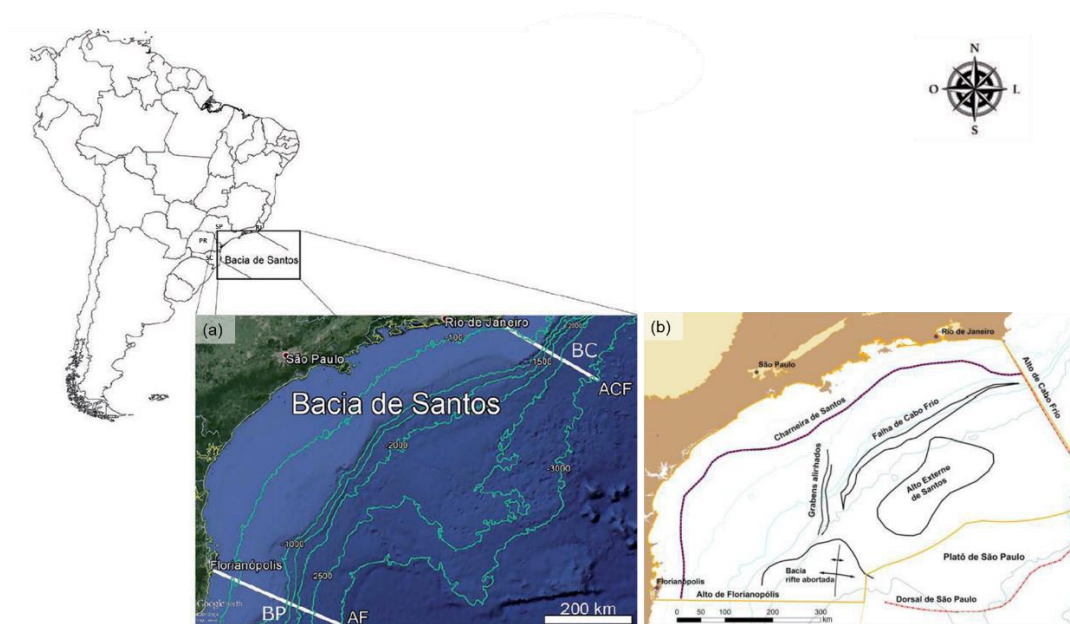


Figura 1: (a) Mapa de localização da Bacia de Santos (modificado de Souza, 2019). (b) Algumas das principais feições de seu arcabouço estrutural (modificado de Garcia et al. 2012; retirado de Fernandez, 2017). ACF = Alto de Cabo Frio; AF = Alto de Florianópolis; BC = Bacia de Campos; BP = Bacia de Pelotas; RJ = Rio de Janeiro; SC = Santa Catarina; SP = São Paulo; PR = Paraná.

A Bacia de Santos apresenta um notório potencial de exploração petrolífera. Ela lidera o *ranking* de produção de hidrocarbonetos no Brasil (Figura 2), como pode ser observado nos boletins mensais de produção da ANP.

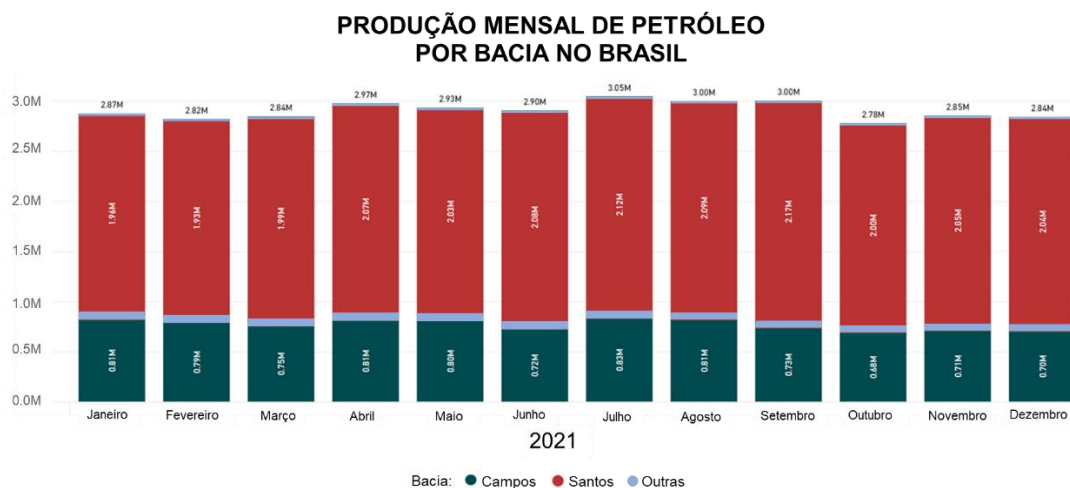


Figura 2: Produção mensal de petróleo por Bacia no Brasil no ano de 2021 (modificado do Painel Dinâmico da ANP, 2022).

A evolução geológica da Bacia de Santos se iniciou a cerca de 130 M.a., e é comumente descrita em três estágios tectônicos: Rifite, Pós-rifite e Drifte (Figura 3). O primeiro estágio foi caracterizado por esforços extensionais progressivos, durante ruptura do supercontinente Gondwana, no Eocretáceo, que permitiram a geração de espaço de acomodação sedimentar (SOUZA, 2019). Este estágio é representado estratigraficamente por três formações do Grupo Guaratiba: Camboriú, Piçarras e Itapema. Posteriormente ocorreu o estágio Pós-rifite, marcado por um aumento gradual das incursões marinhas, que levaram à formação do Oceano Atlântico Sul (SOUZA, 2019). A restrita circulação de água marinha no oceano que estava sendo formado, em conjunto ao clima árido que dominava na época (Arai, 2009) permitiu as condições ideais para sedimentação da seção evaporítica (Fm. Ariri), que se depositou sobre a seção carbonática (Fm. Barra Velha) (MOREIRA et al., 2007). Por fim, o último estágio, Drifte, é marcado pela individualização dos continentes africano e sul-americano, encerrando a seção evaporítica da bacia e dando início à sedimentação de ambiente de mar raso, e, posteriormente, mar profundo (MOREIRA et al., 2007). No que diz respeito a estratigrafia, este estágio é representado pelos grupos Camburi, Frade e Itamambuca e todas as formações pertencentes a esses grupos.

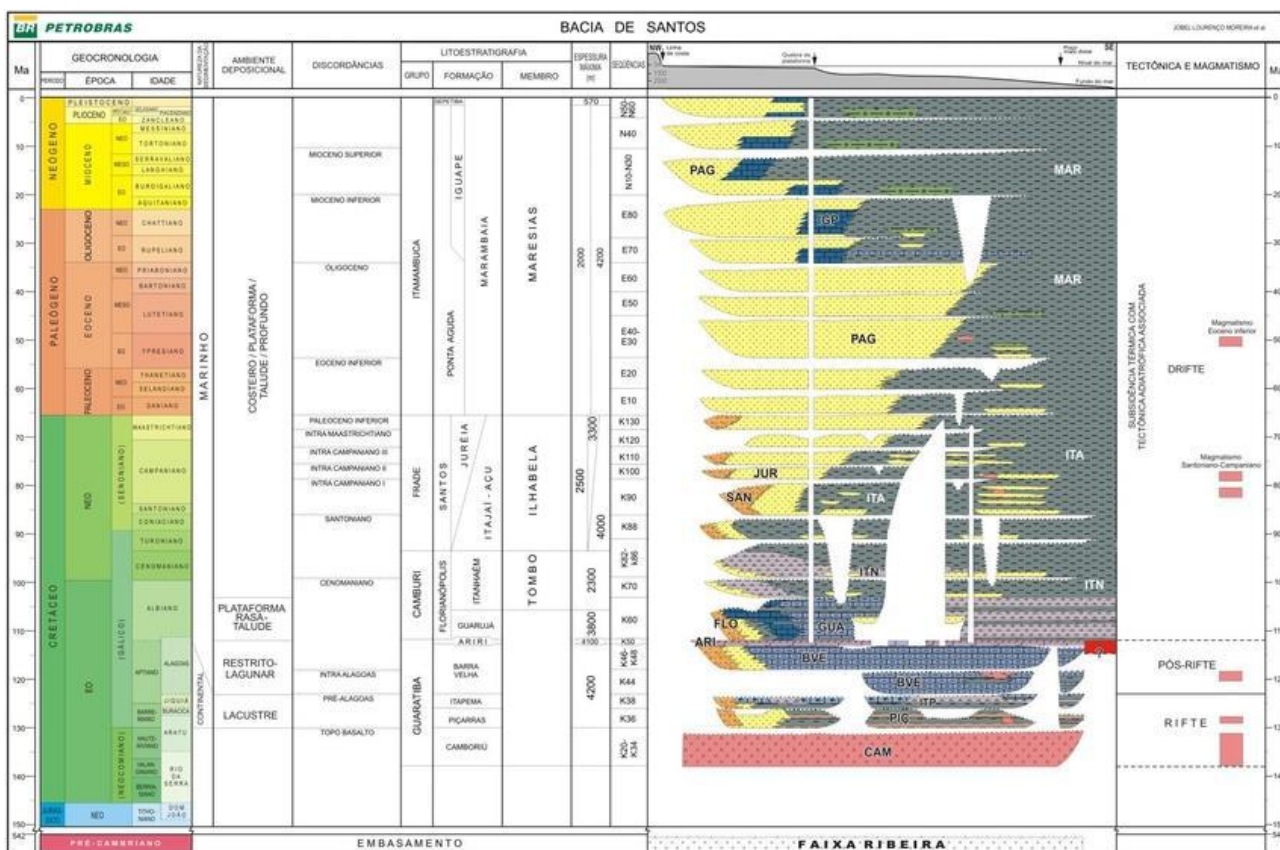


Figura 3: Carta estratigráfica da Bacia de Santos (Moreira et al 2007).

2.1.1 Pré-sal

Para avaliar a capacidade de algoritmos de ML de classificar litofácies inseridas no contexto geológico de rochas carbonáticas do intervalo Pré-sal, selecionou-se um poço do campo de Tupi, originalmente chamado de Lula.

O campo de Tupi encontra-se inserido no polígono do Pré-sal, na porção central da Bacia de Santos. Ele está dentro do estado do Rio de Janeiro, a 230 km da costa da capital carioca (Figura 6). O bloco ocupa uma área de 1.523,22 km², apresentando lâmina d'água de até 2.156 m de profundidade (ANP, 2018). Ele foi descoberto em agosto de 2006, pela Petrobras, e teve sua produção iniciada em dezembro de 2010.

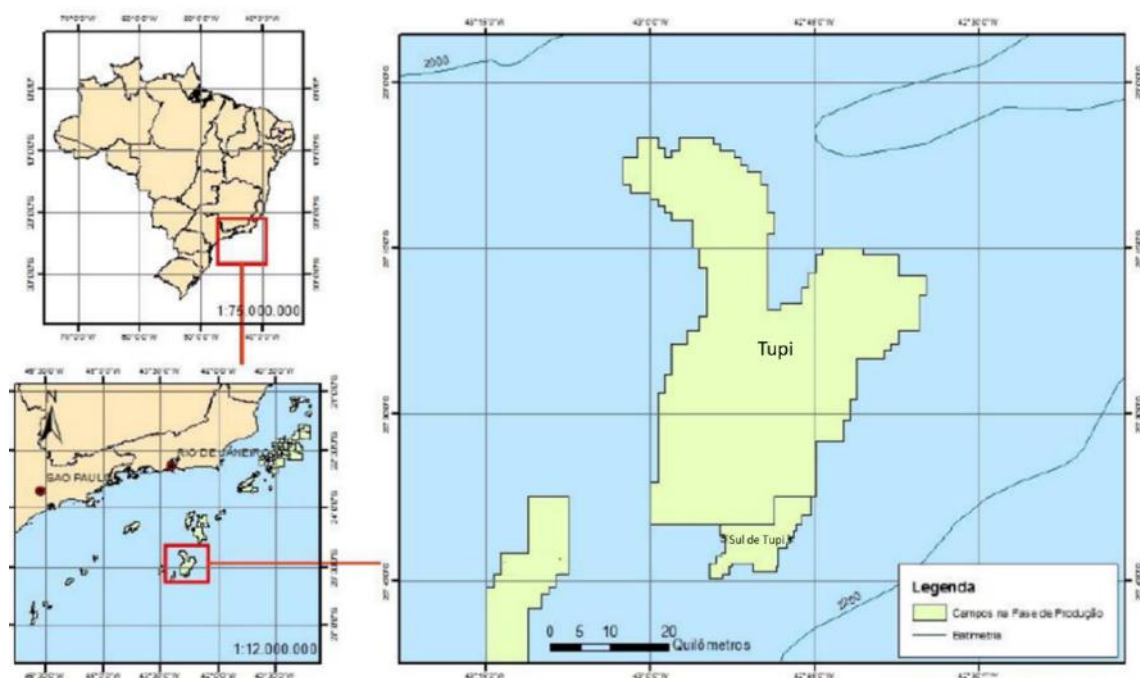


Figura 6: Mapa de localização do Campo de Tupi (modificado de ANP, 2018).

O campo de Tupi representa um enorme potencial petrolífero no Brasil, como é possível observar por meio de seu histórico de produção (Figura 7).

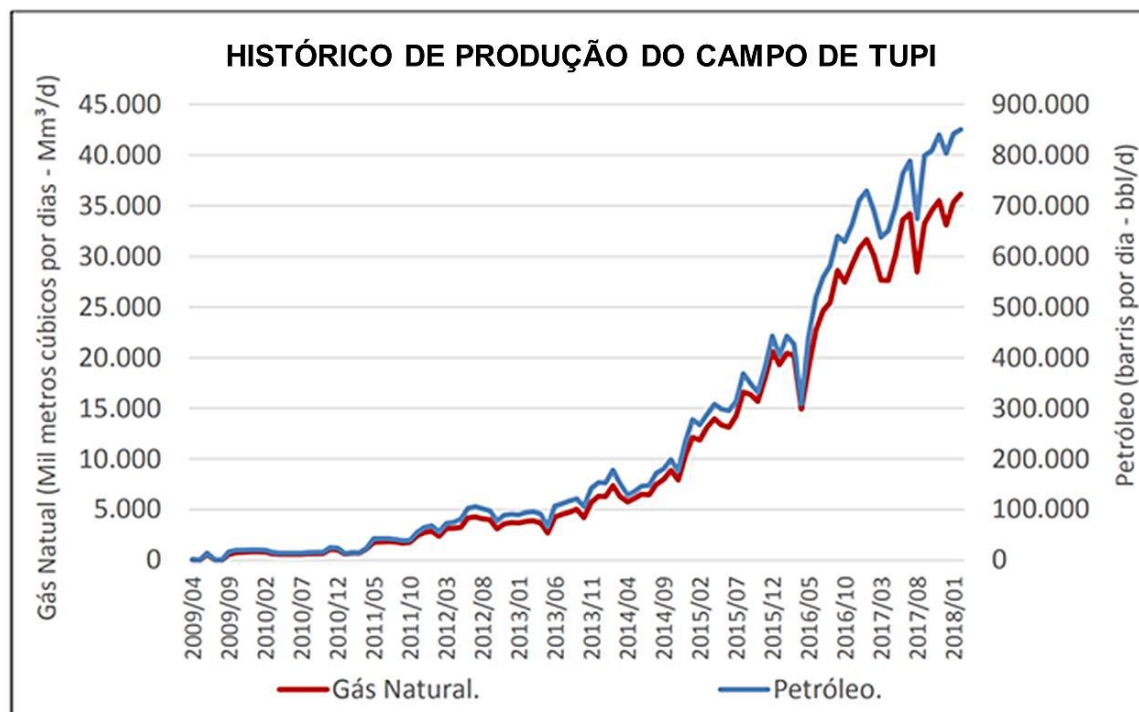


Figura 7: Histórico de produção de petróleo e gás natural do Campo de Tupi, na Bacia de Santos (ANP, 2018).

Os reservatórios desse campo integram as formações Itapema e Barra Velha. Ambas se encontram entre 4.700 e 6.000 m abaixo do nível do mar, estando sotopostos a uma espessa camada de sal, que pode chegar até 2.000 m de espessura (Fernandez, 2017). Segundo a ANP (2018), esses reservatórios caracterizam-se por apresentar alta variabilidade (tanto lateral, quanto vertical) no que diz respeito as suas propriedades permo-porosas.

A Fm. Itapema é interpretada como um ambiente de leque aluvial, tendo como produto da deposição na parte proximal conglomerados e arenitos, enquanto a porção distal apresenta intercalações de calcirruditos e folhelhos escuros (Fernandez, 2017).

Já a Fm. Barra Velha é composta por rochas carbonáticas de idade aptiana, podendo ser tanto calcários microbiais, quanto estromatólitos e laminitos. Os carbonatos que constituem os reservatórios de Tupi depositaram-se em um ambiente transicional, entre continental e marinho raso (ANP, 2018).

2.1.2 Pós-sal

Já para avaliar a classificação de litofácies inseridas em contexto geológico de rochas siliciclásticas do intervalo Pós-sal, selecionou-se um poço do campo de Mexilhão.

O campo de Mexilhão está posicionado no limite, a sudoeste, do Polígono do Pré-sal, na porção noroeste da Bacia de Santos. Ele está inserido na plataforma continental do estado de São Paulo, a 138 km do litoral paulista (Figura 4). O bloco ocupa uma área de 254 km², apresentando lâmina d'água entre 250 e 550 m de profundidade (ANP, 2016). Ele foi descoberto em fevereiro de 2003, pela Petrobras, e teve sua produção iniciada em abril de 2011.

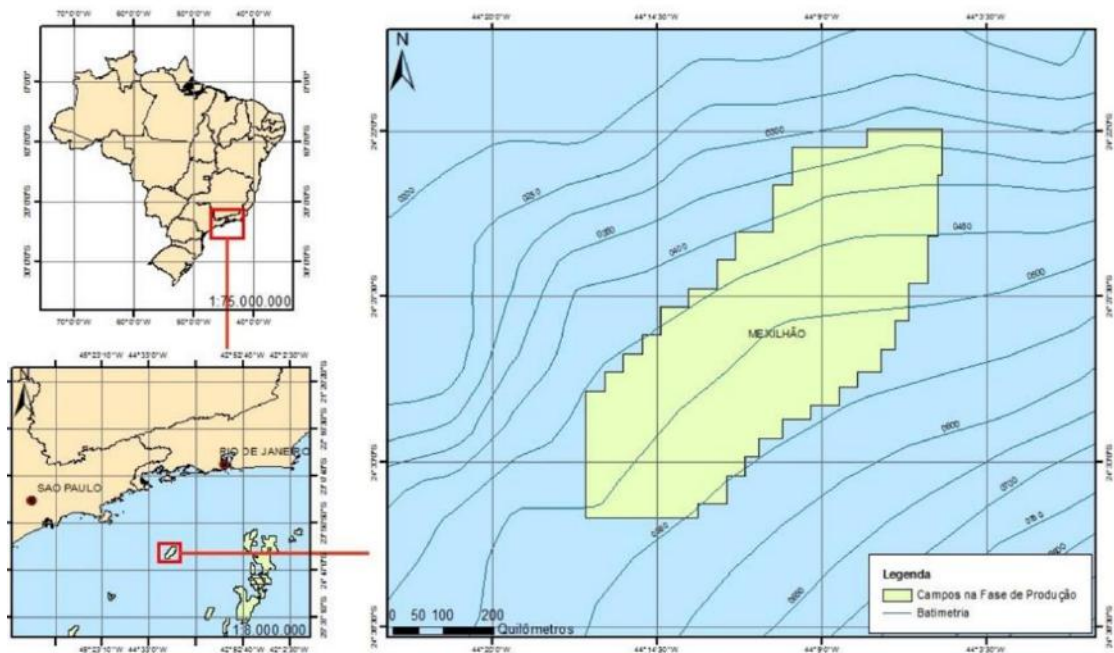


Figura 4: Mapa de localização do Campo de Mexilhão (ANP, 2016).

Esse campo destaca-se pelo seu papel fundamental no escoamento de gás dessa região da bacia. Isso porque ele apresenta uma localização estratégica, estando próximo aos maiores campos com perspectivas de reservas do país (ANP, 2016). A produção no campo de Mexilhão está focada em Gás Condensado e Gás Não Associado (Figura 5).

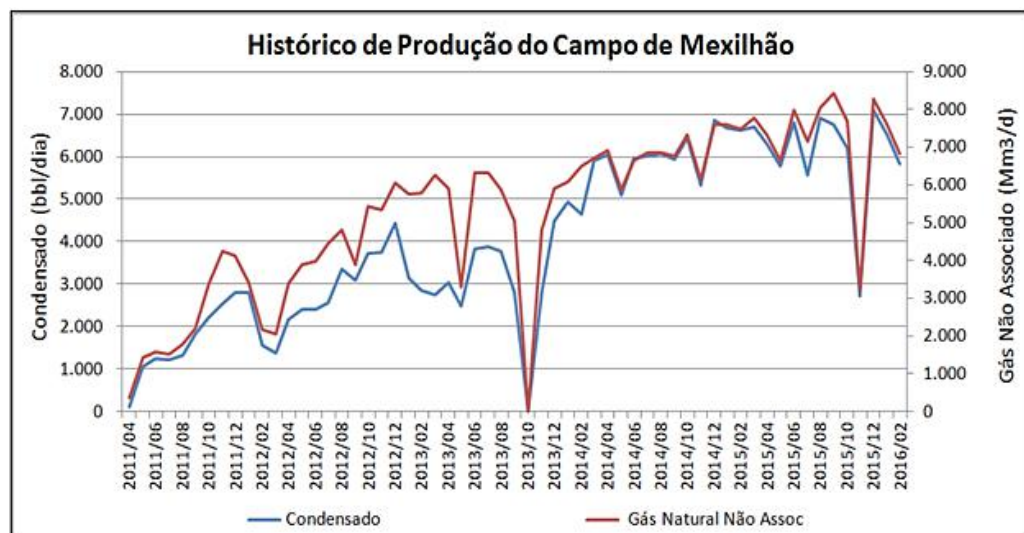


Figura 5: Histórico de produção de gás não associado e condensado do Campo de Mexilhão, na Bacia de Santos (ANP, 2016).

Os reservatórios deste campo são constituídos por arenitos turbidíticos do Membro Ilha Bela, da Formação Itajaí- Açú, de idade turoniana à maastrichtiana, encontrando-se sobrepostos a camada de sal (VIEIRA, 2009).

2.2 Perfis de Poços

Perfis de poços são curvas de registros de quaisquer características de formações rochosas, a partir do uso de ferramentas de perfilagem que realizam medidas que variam segundo a profundidade, conforme a ferramenta se desloca pelo poço (SANTOS, 2010). Geralmente, essas medidas são adquiridas através de transmissores e receptores. Nesses casos, primeiro há a emissão de algum tipo de sinal ou perturbação na rocha, interagindo com ela, e em seguida, a ferramenta faz a leitura da resposta desse sinal/perturbação. A diferença entre a informação que foi transmitida pela ferramenta e a que foi captada por ela permite ter uma noção das propriedades do meio rochoso (OLIVEIRA, 2021). Adversamente, outras ferramentas podem funcionar apenas por receptores, como será revisado neste capítulo. Por fim, destaca-se que essa aquisição de dados pode ser feita tanto ao final da perfuração do poço (perfilagem a cabo), quanto durante ela (*Logging While Drilling – LWD*).

A perfilagem de poços é uma técnica de prospecção petrofísica interessante para diversas áreas das geociências, na indústria petrolífera ela é muito útil para avaliar acumulações de hidrocarboneto. Ao longo dos anos, o papel tradicional da perfilagem foi se concentrando em duas finalidades principais: avaliação de formações e avaliação de completação (ELLIS & SINGER, 2007). Através dessa prática, é possível medir uma série de parâmetros físicos relacionados tanto às propriedades geológicas quanto às petrofísicas dos estratos perfurados.

Os produtos da perfilagem são os perfis de poços, eles podem revelar informações importantes sobre a rocha e sobre os fluidos contidos em seus poros. Encontra-se abaixo uma breve revisão das propriedades de cada perfil utilizado neste trabalho.

- Perfil de Calibre de Poço:

A medição deste perfil é feita por uma ferramenta centralizada no meio do poço com ao menos dois braços mecânicos acoplados a ela. Estes braços se ajustam

lateralmente de forma a pressionar a parede do poço, assim, eles são capazes de registrar seu diâmetro em cada profundidade amostrada.

Esse perfil é muito útil no controle das operações exploratórias, pois serve como um indicador da integridade do poço (GLOVER, 2000). Isso faz com que ele sirva, também, como controle de qualidade da perfilagem corrida, dado que a má integridade do poço frequentemente compromete a leitura de diversos perfis sensíveis ao seu diâmetro, tais como perfil sônico e perfil densidade, por exemplo.

- Perfil de Raios-gama:

Esse tipo de perfil está relacionado ao fenômeno da radioatividade, ou seja, ao decaimento espontâneo de certos isótopos presentes na rocha. Dessa forma, diferentemente da maioria dos demais perfis, o de Raios-gama (GR) lê apenas a radiação natural emitida pelo meio, sem disparar qualquer tipo de sinal para a formação. Assim, esses perfis são adquiridos a partir de ferramentas que realizam a leitura da radiação natural da rocha (SERRA, 1984). A maior causa dessa radiação é a ocorrência de três elementos abundantes na natureza que apresentam meia vida da ordem de 10^9 anos: ^{40}K , ^{232}Th e ^{238}U (SANTOS, 2010).

Em cenários siliciclásticos, os perfis de GR são amplamente empregados no reconhecimento de arenitos (potenciais reservatórios) e folhelhos (não reservatório), dado que este segundo tipo de rocha apresenta maior concentração dos elementos radioativos. Já em um contexto de ambiente carbonático, a radiação gama está mais associada à fração de matéria orgânica presente na rocha (OLIVEIRA, 2021).

Atualmente, com o avanço tecnológico, ferramentas mais modernas já são capazes de ler individualmente os diferentes espectros da radiação gama (Figura 8). Com isso, é possível saber a fração relativa ao conteúdo de Tório (Th), Urânio (U) e Potássio (K) presente na rocha, aprimorando o processo de caracterização litológica.

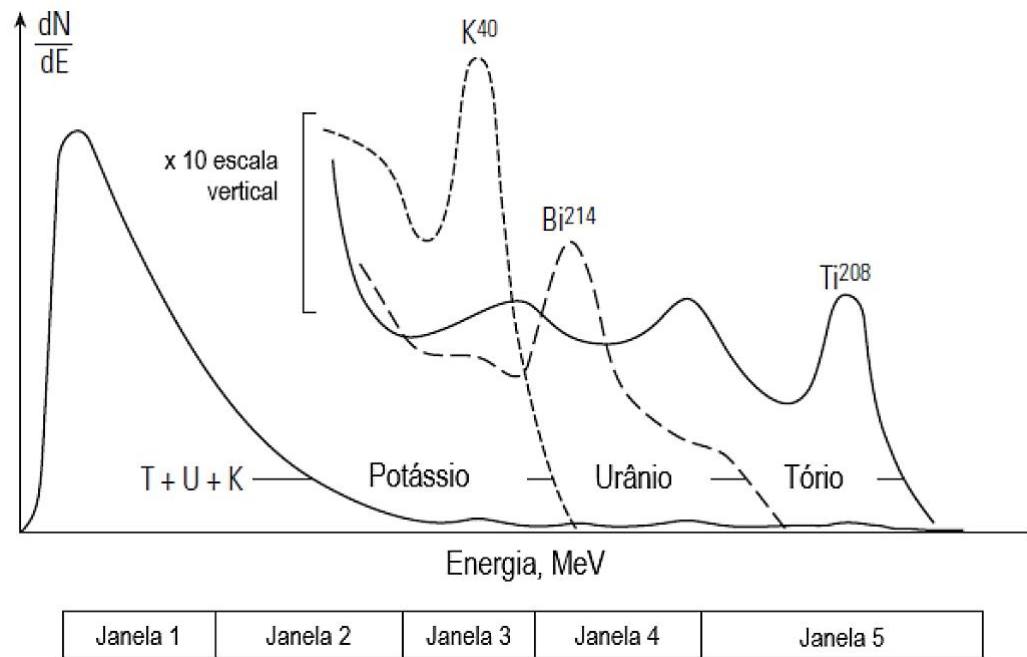


Figura 8: Distribuição do espectro de radiação gama entre as respectivas janelas de energia referentes ao conteúdo de Th, U e K presentes na formação (Ellis e Singer, 2007).

- Perfil de Resistividade:

Os perfis elétricos medem a resistividade da formação (em Ohms), e eles se relacionam com diversas características petrofísicas das rochas. Assim, para o entendimento desses perfis é fundamental ter a noção das relações identificadas por Archie (1942), representadas pela Equação 1:

$$R_t = \frac{a}{\phi_m} \frac{R_w}{S_w^n} \quad \text{Equação 1}$$

Onde R_t é a resistividade da formação (resistividade verdadeira – *true*); a é a chamada “constante litológica”, que varia de acordo com o tipo de rocha, sua granulometria, a geometria dos poros e o grau de compactação; ϕ representa a porosidade efetiva; m consiste no expoente de cimentação; R_w expressa a resistividade da água na temperatura da formação; S_w é a saturação de água e; n representa o expoente de saturação para ajustar a resistividade de acordo com o tipo de fluido presente no meio rochoso.

Segundo Archie (1942), quando um arenito não fornece uma leitura de resistividade constante, implica-se que o caráter da areia provavelmente mudou ou que o tipo/quantidade de fluido contido na areia deve ter variado. Por isso, no contexto

de exploração de petróleo, esses perfis são largamente utilizados para fornecer alguma indicação de fluido presente no reservatório. Além disso, apesar de ser uma ferramenta insuficiente para identificação litológica, seus perfis são muito usados de forma qualitativa para correlacionar as formações penetradas pela broca (GLOVER, 2016). A figura 9 mostra a típica resposta de um perfil de resistividade em uma sequência de arenitos e folhelhos, sendo que a resistividade responde a alterações em diversas variáveis, como: (i) a porosidade; (ii) a estrutura dos poros; (iii) a resistividade da água de formação; (iv) a saturações de água, óleo e gás.

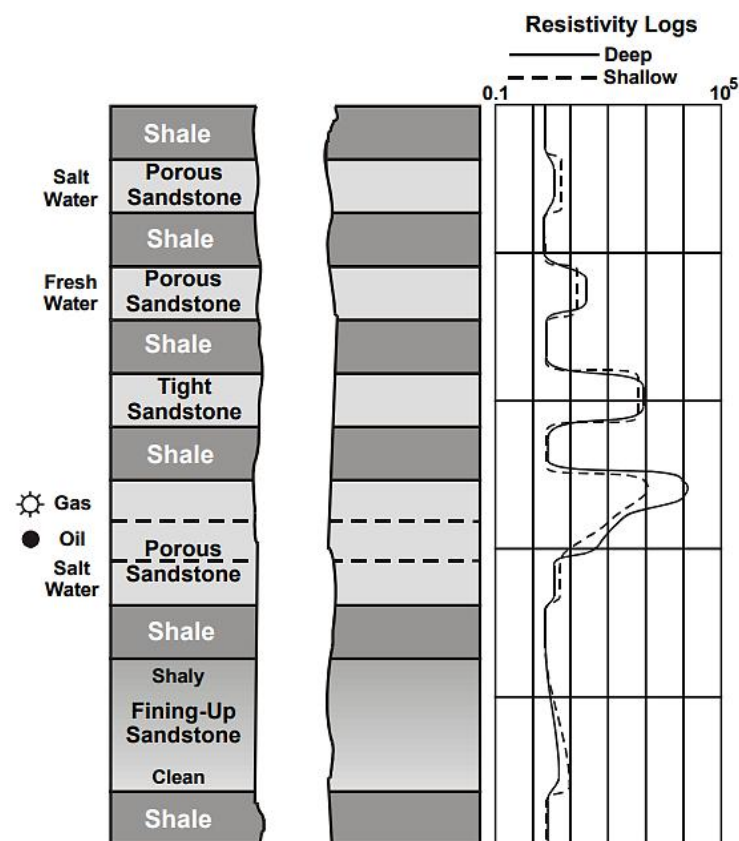


Figura 9: Resposta típica de um perfil de resistividade (Glover, 2016).

- Perfis de Densidade e Fator Fotoelétrico:

O perfil de Densidade (RHOB) registra as variações de densidade das formações cortadas pelo poço. Essa propriedade da rocha é medida a partir da emissão de raios-gama por uma fonte artificial (em geral ^{137}Cs) na ferramenta de perfilagem. Quando a interação entre os raios-gama emitidos e o meio rochoso se dá, principalmente, por espalhamento Compton, é possível determinar a densidade eletrônica da matéria em análise. E quando se estabelece, também, a medida da

atenuação dos raios-gama que atravessam a formação, torna-se possível calcular sua densidade em g/cm³. Esse cálculo pode ser feito através da Equação 2, que converte os valores da variação na intensidade do fluxo de raios-gama emitidos e captados em valores de densidade:

$$I_{captado} = I_{emitido} \cdot e^{-\rho_e \cdot \mu \cdot x} \quad \text{Equação 2}$$

Onde ρ_e é a densidade eletrônica do meio, μ é seu o coeficiente de absorção de massa e x é a sua espessura. Entretanto, a equação apresentada rege o fenômeno em nível atômico. Uma segunda equação (Equação 3) é utilizada para realizar a extrapolação aproximada do valor em nível atômico da densidade para o nível molecular. Segundo Santos (2010), esse cálculo entrega bons resultados para matrizes areníticas e carbonáticas saturadas em água:

$$\rho_b = 1,07 \cdot \rho_e - 0,1883 \quad \text{Equação 3}$$

Onde ρ_b é a densidade da formação.

Para se obter a medida de atenuação dos raios-gama, necessária para o modelo matemático, a ferramenta apresenta dois detectores: um proximal e outro distal (Figura 11). Idealmente, os dois detectores deveriam indicar o mesmo valor, entretanto, algumas influências externas (como fluido de perfuração) podem causar distorções na leitura (OLIVEIRA, 2022). O uso de dois detectores serve, justamente, para corrigir as alterações provocadas por esses fatores, permitindo chegar na real densidade da formação.

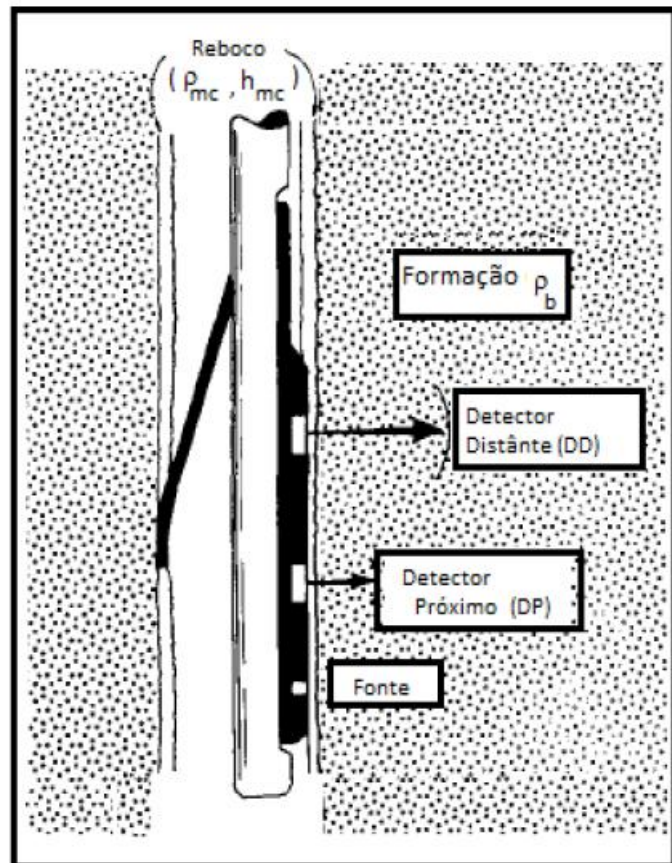


Figura 10: Modelo esquemático da ferramenta que faz a leitura da densidade da formação, correndo junto à parede do poço (Serra, 1984).

As ferramentas de densidade mais modernas são capazes de entregar também o perfil fotoelétrico (P_e), pois elas permitem a leitura da radiação gama em altas e baixas energias, tornando possível calcular o valor do fator fotoelétrico (OLIVEIRA, 2022). Esse perfil é relevante na estimativa da composição mineralógica da matriz da rocha, por ser pouco suscetível as variações na porosidade, e apresentar relação com número atômico das partículas que compõem a formação (ELLIS & SINGER, 2007).

- Perfil de Nêutrons

O perfil de nêutrons é amplamente utilizado para avaliar a porosidade da formação, dado que ele é bastante sensível a quantidade de átomos de hidrogênio, e o índice de hidrogênio tem relação direta com a porosidade da rocha (OLIVEIRA, 2022).

A ferramenta que permite a leitura deste tipo de perfil opera colada a parede do poço, a partir do bombardeamento de nêutrons de alta energia para a formação, podendo chegar até 4,5 MeV (SANTOS, 2010). Essas partículas sem carga elétrica,

e em alta energia, são capazes de penetrar profundamente na matéria, interagindo elástica e/ou inelasticamente com os núcleos dos elementos químicos que compõem a mineralogia da rocha e a composição dos fluidos. Bem como nas leituras do perfil de densidade, os perfis de nêutrons são adquiridos a partir do uso de dois detectores, que captam a resposta dessa interação, sendo um proximal e outro distal.

Conforme os nêutrons bombardeados interagem com a formação, eles vão perdendo energia para ela, gerando raios-gama de alta energia. Dessa forma, como o espalhamento dos nêutrons emitidos é mais eficiente em átomos de hidrogênio, é possível calcular a quantidade aproximada de átomos de hidrogênio na formação (GLOVER, 2016), e correlacionar este dado à porosidade da rocha, dado que este elemento está muito presente nos fluidos.

Em formações com grandes quantidades de átomos de hidrogênio, os nêutrons são desacelerados e absorvidos rapidamente, em curta distância. Enquanto em formações com uma pequena quantidade, os nêutrons viajam mais pela formação antes de serem absorvidos, logo desaceleram mais lentamente. Assim, no primeiro cenário, a taxa de contagem de captura de nêutrons lentos ou raios-gama é baixa, estando associados a rochas de alta porosidade. Já no segundo cenário, a taxa de contagem é maior, indicando rochas de baixa porosidade (GLOVER, 2016). A relação entre a distância de migração dos nêutrons, a razão das contagens dos dois detectores e a porosidade da rocha estão demonstradas na Figura 10.

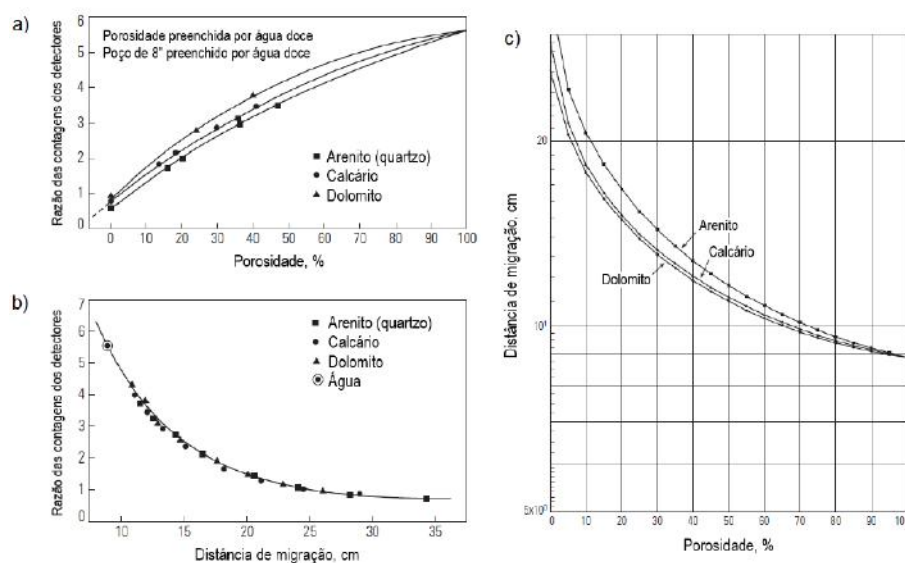


Figura 11: Gráficos indicando a relação entre (a) a razão das contagens dos detectores e a porosidade para diferentes tipos de rocha, (b) a razão das contagens dos detectores e a distância de

migração e (c) a distância de migração e a porosidade para diferentes tipos de rocha. (Ellis & Singer, 2007).

- **Perfil de Ressonância Magnética Nuclear**

Segundo Victor (2017), os perfis de ressonância magnética nuclear (RMN) têm como foco de atuação os núcleos de hidrogênio presentes nos fluidos da formação. Um fator essencial na perfilagem RMN é o tempo relaxação (T_2), medido em milissegundos. Isso porque a análise da distribuição de T_2 permite estabelecer o tamanho dos poros da rocha, inferir algumas propriedades dos fluidos que ela contém, bem como calcular sua permeabilidade absoluta de forma mais precisa. Além disso, essa ferramenta de perfilagem fornece a porosidade total da formação, independente da matriz (OLIVEIRA, 2022).

O experimento Carr-Purcell-Meiboom-Gill (CARR; PURCELL, 1954; MEIBOOM; GILL, 1958) descreve o processo de leitura da ferramenta de RMN, havendo três momentos principais: primeiramente ocorre a polarização dos átomos de hidrogênio presentes na formação por meio de um campo magnético constante (B_0); em seguida, devido a pulsos de rádio frequência, que oscilam na frequência ressonante dos átomos excitados, é gerado um campo magnético (B_1); por fim, quando há a remoção de B_1 , os átomos retornam para seu estado original. Neste momento, torna-se possível realizar a leitura do decaimento do sinal magnético emitido por esses átomos.

2.3 Modelo de Subsuperfície

O modelo de subsuperfície (*Earth Model*) é um conceito introduzido pela equipe do Departamento de Engenharia de Petróleo e Geossistemas da Universidade do Texas, em Austin. Consiste em um modelo de multicamadas que descreve detalhadamente as formações rochosas encontradas em um poço, honrando os componentes minerais, porosidade, saturação, temperatura e composição de fluidos, pressão e até mesmo geometria de camadas, no caso de poços direcionais ou horizontais. Assim, quando construído adequadamente, esse modelo pode ser empregado na modelagem direta dos perfis de poços, permitindo analisar quais características da formação impactam nas respostas dos perfis.

O *software* UTAPWeLS foi desenvolvido através de um consórcio da Universidade do Texas e empresas atuantes na indústria de óleo e gás. Ele disponibiliza uma série de módulos próprios para a criação de um bom modelo de

subsuperfície. Entretanto, a presente monografia de conclusão de curso irá descrever apenas os principais módulos utilizados neste trabalho. Infelizmente, devido a pouca informação já publicada em forma de artigos científicos acerca desse *software*, existe uma dificuldade em realizar a citação formal de trabalhos. Porém, muito do que será discutido nesse capítulo pode ser encontrado no canal do YouTube do *software*.

Para a geração de um modelo de subsuperfície o primeiro passo é estabelecer os limites que definem o topo e a base de camadas, delimitando as formações geológicas. Esses limites podem ser identificados automaticamente pelo *software*, através da análise da máxima variância de uma determinada janela de profundidade pré-estabelecida, ou manualmente pelo próprio usuário. De maneira geral, os limites entre camadas são os pontos de inflexão dos perfis entre valores máximos e mínimos de uma determinada região, conforme a figura 12.

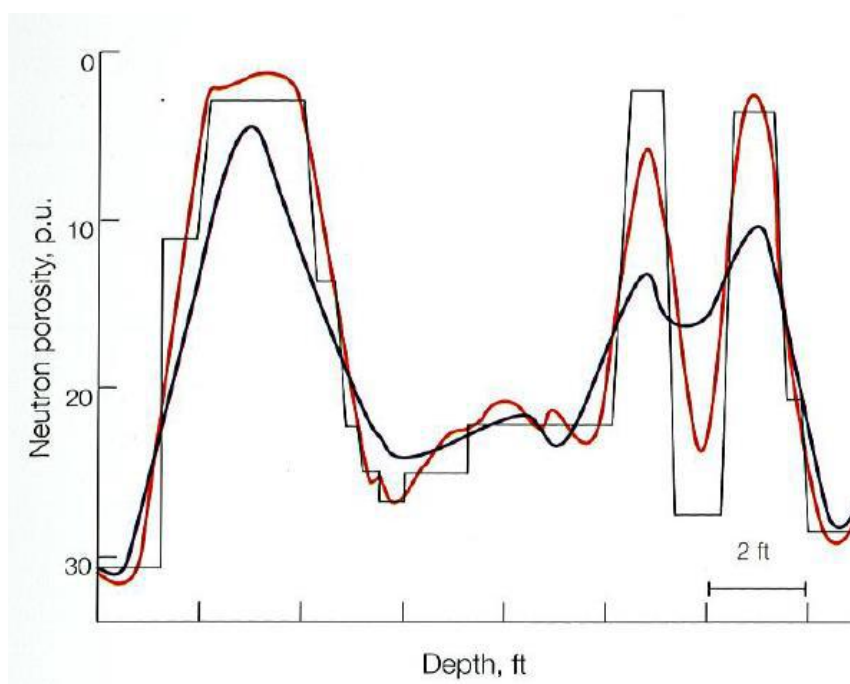


Figura 12: Exemplo de posicionamento de limites de camadas (retas verticais do perfil quadratizado) em relação às variações do perfil de nêutrons adquirido em um poço (curvas preta e vermelha representando o perfil real e simulado através dos modelos de subsuperfície, respectivamente). (Tittman, 1991).

Após as camadas terem sido devidamente delimitadas, a segunda etapa consiste na escolha dos componentes sólidos (minerais e matéria orgânica) e fluidos de cada camada, determinando também suas frações relativas, além de suas pressões e temperaturas. Para isso se utiliza o módulo Paleta Composicional (Figura

13). O ideal é que a etapa baseie-se em análises laboratoriais de amostras de rocha e fluido coletadas, preferencialmente, no poço analisado, ou ainda em poços análogos.

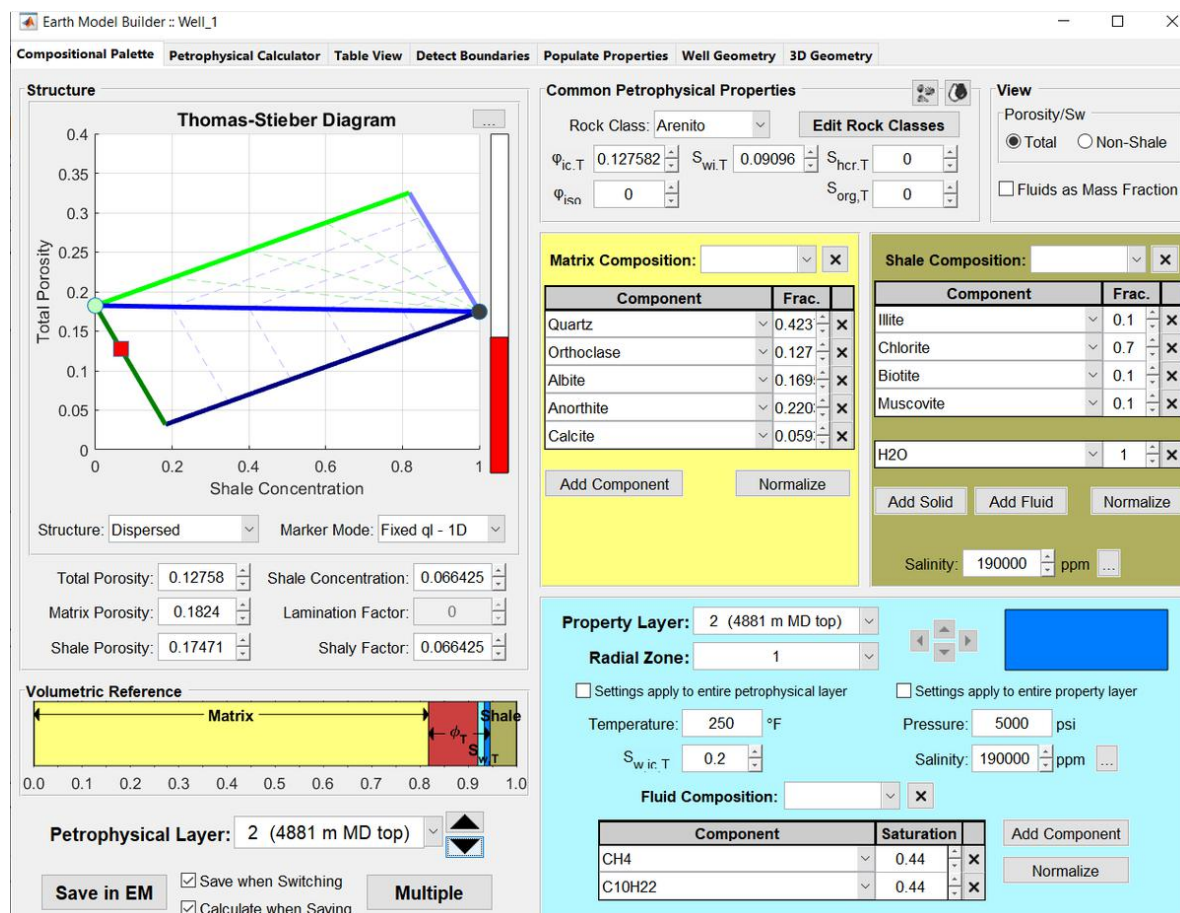


Figura 13: Módulo Paleta Composicional do *software* UTAPWeLS, que permite criar formações geológicas através da escolha de componentes minerais, fluidos, porosidade, pressão e temperatura das rochas encontradas no poço.

Naturalmente, não é esperado que a calibração manual, por meio da Paleta Composicional, gere um modelo de subsuperfície que honre completamente o contexto litológico das camadas perfuradas. Entretanto, é fundamental que essa primeira calibração seja feita cuidadosamente, pois ela é a estimativa inicial do processo iterativo que irá refinar a composição das camadas para simulação dos perfis de poços sintéticos, através do módulo de Inversão Composicional Conjunta (*Composition Joint Inversion*). Nesse sentido, quanto mais realista for o modelo de subsuperfície, mais próximo os perfis simulados estarão dos reais.

Cabe destacar que o módulo Inversão Composicional Conjunta é bastante complexo, e depende de dois principais fatores: (1) a estimativa inicial calibrada

manualmente na Paleta Composicional, e; (2) a definição a priori de frações de componentes que irão guiar o processo iterativo. Essa informação a priori consiste em histogramas de fração de componentes (Figura 14) que podem ser geradas a partir de análises laboratoriais de rocha e fluido. Além disso, caso haja grande incerteza na resposta de algum perfil, seja por problemas ambientais ou por incerteza em suas equações, o módulo permite diminuir o peso desses perfis para que eles não tenham grande influência no processo iterativo.

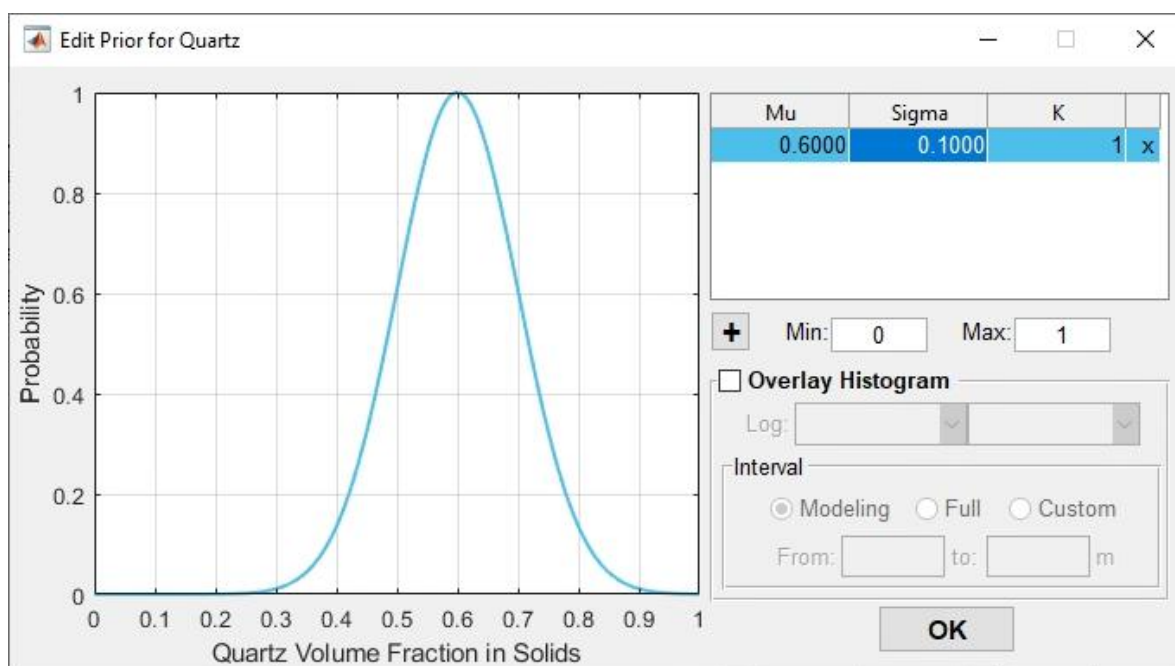


Figura 14. Exemplo de definição de informação a priori do módulo Inversão Composicional Conjunta do software UTAPWeLS. Considerando o histograma acima para o componente quartzo, o processo iterativo irá convergir para uma fração de quartzo que respeite a distribuição a priori, e que pode ser calibrada através do histograma de fração de quartzo observado em análises de amostras de rocha.

Esse último módulo faz pequenas alterações na fração dos componentes do modelo de subsuperfície prévio. Ao simular os perfis sintéticos e os comparar com os reais, o software calcula a função custo, definida pela diferença média entre os perfis, e usa o gradiente dessa função, em relação aos componentes, para estimar suas novas frações. Assim, entende-se que, ao realizar diversas iterações, a fração de componentes irá convergir para um valor com o qual os perfis simulados ficarão suficientemente próximos aos reais.

Portanto, ao final do processo iterativo, espera-se obter um modelo de subsuperfície contínuo para toda a extensão do poço, devidamente calibrado com

dados de amostras de rocha e fluido e posteriormente refinado pelo *software*. Dessa forma, este resultado gerado seria o elemento do presente trabalho que melhor honra as formações encontradas nos poços estudados.

2.4 *Machine Learning*

O aprendizado de máquinas é um campo de estudo da inteligência artificial que transformou a ciência da computação, pois representou uma quebra de paradigmas da programação tradicional (KORCHI et al., 2019). Antigamente, era necessário fornecer instruções explícitas ao computador para que a programação fosse possível. Já hoje, com o advento do ML, a programação indireta se tornou uma realidade. Agora, as máquinas são capazes de aprimorar seu próprio desempenho, sem serem explicitamente programadas para tal. Essa automatização do aprendizado computacional se dá através de algoritmos eficientes no reconhecimento de padrões e na identificação das “regras ocultas” nos dados fornecidos (MITCHELL et al., 1990).

Quando aplicado a dados de perfis de poços, os algoritmos de ML podem ser divididos em três principais tipos de aplicação: predição, interpretação e imputação de dados. O primeiro deles busca calcular propriedades petrofísicas da rocha a partir perfis de poços; o segundo visa a classificação litológica/petrofísica de um poço; o último, por fim, tem por objetivo simular dados de perfis em intervalos em que a perfilagem apresenta baixa qualidade (OLIVEIRA, 2022). Portanto, o tipo de problema de interpretação receberá maior destaque neste trabalho, por melhor atender à natureza do estudo em questão.

Chapelle et al. (2006) separaram o ML em duas principais categorias: aprendizado supervisionado e aprendizado não supervisionado. Posteriormente, surgiram novas subdivisões que expandiram essa classificação, porém no contexto de aplicação do ML em problemas da geociências as duas abordagens iniciais permaneceram sendo as mais utilizadas (LIMA, 2019).

Segundo Bishop (2006 apud Oliveira, 2022), o aprendizado é chamado de supervisionado quando o algoritmo de ML calibra uma série de parâmetros para estabelecer a maneira que as variáveis de entrada dos dados de treinamento se relacionam entre si, gerando estimativas das variáveis de saída. Já quando os dados de treinamento não possuem variáveis de saída, o aprendizado é considerado não

supervisionado. Nesse caso, o algoritmo irá, por conta própria, avaliar uma série de parâmetros dos dados de entrada para agrupá-los em conjuntos de afinidade.

2.4.1 Clusterização

Clusterização é o procedimento de agrupar diversos dados em conjuntos de familiaridade de forma não supervisionada, ou seja, sem que haja uma predefinição de como os padrões de cada grupo devem ser gerados (JAIN e DUBES, 1988). Assim, é necessário que o próprio algoritmo de clusterização analise a base de dados a fim de verificar as similaridades dos dados e identificar padrões existentes entre eles para realizar o agrupamento adequadamente. O objetivo é obter a maior homogeneidade possível dentro de cada grupo formado e a maior heterogeneidade entre os grupos (CARDOSO et al., 2008). Os grupos são comumente denominados de “*clusters*”.

Essa técnica é uma excelente ferramenta para aprimorar o entendimento de um determinado conjunto de dados, pois ela torna mais fácil de visualizar correlações interessantes entre os diversos atributos dos dados originais. Nesse sentido, a análise de *clusters* desempenha um papel muito importante no processo de organizar uma coleção de dados, e pode ser feita através do uso de diversos algoritmos de aprendizado de máquina. Para o presente trabalho, foi escolhido o método de clusterização por meio do algoritmo *k-means*.

Esse algoritmo particiona a base de dados em *k clusters*, sendo que “*k*” representa a quantidade de agrupamentos que o usuário deseja obter, portanto esse valor deve ser fornecido previamente. O *k-means* opera da seguinte maneira: primeiro, ele eleger, aleatoriamente, os objetos da base de dados que irão representar o centróide de cada *cluster*. Em seguida, é calculada a distância desses centróides para todos os demais objetos. Cole (1998) explica que quanto menor for a distância entre um objeto e outro, maior será sua similaridade. Assim, esse valor, conhecido como “Medida de Similaridade”, é utilizado para definir a qual *cluster* cada objeto pertence, de maneira que os objetos sejam designados para o *cluster* de menor distância em relação ao centroide correspondente (Figura 15).

Não há uma única função de distância que se aplique a todas as variáveis, mas geralmente as medidas de similaridade mais empregadas são: Euclidiana, Manhattan, Minkowski e Mahalanobis (CASSIANO, 2014). Neste trabalho, foi utilizada a distância Euclidiana, definida pela Equação 3:

$$D(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad \text{Equação 3}$$

Onde $D(x, y)$ é a distância do objeto x da base de dados para o objeto y , n representa o número total de atributos, x_i e y_i representam o atributo i dos objetos x e y .

Uma vez que os objetos foram alocados em seus respectivos *clusters*, calcula-se a média dos objetos pertencentes a cada grupo, esse valor é chamado de centro de gravidade do *cluster*. Então, o último passo é quando o algoritmo aplica uma estratégia de controle para determinar quais objetos deveriam mudar de *cluster*, de forma que a função objetivo venha a convergir (CASSIANO, 2014).

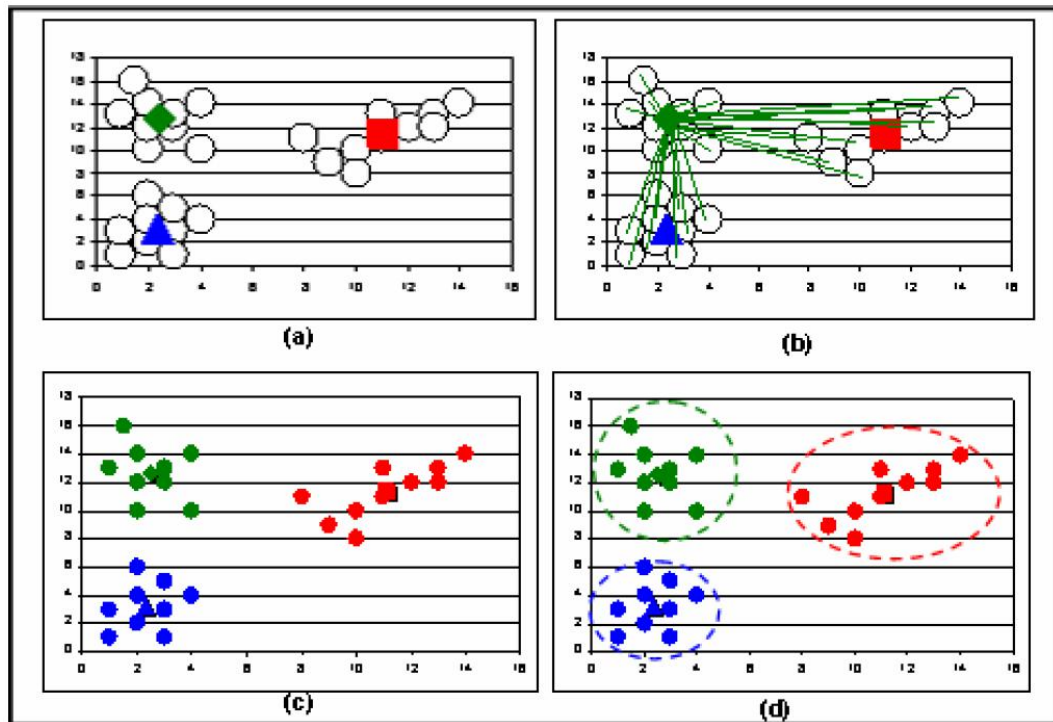


Figura 15: representação esquemática de como o algoritmo *k-means* funciona, onde (a) representa a primeira etapa do processo, em que são estabelecidos os objetos que representam o centróide de cada *cluster*; (b) é o passo seguinte, onde são calculadas as distâncias dos centróides até todos os objetos; (c) definição dos *clusters* e cálculo da média de cada objeto dentro do *cluster* e, por fim; (d) representação da última etapa do processo onde os objetos de cada *cluster* são confirmados e os *clusters* são estabelecidos definitivamente. (CARDOSO et al., 2008).

3. MATERIAIS E MÉTODOS

O estudo foi realizado em duas etapas: a primeira consistiu na geração de um modelo de subsuperfície (*earth model*) no *software* UTAPWeLS e na posterior simulação de uma corrida de perfilagem no modelo gerado para verificar sua validade; a segunda etapa, por sua vez, aproveitou o resultado gerado na primeira para realizar a clusterização das frações minerais do modelo de subsuperfície e a clusterização dos perfis de poços reais.

Abaixo, na figura 16, pode ser observado o fluxograma que resume as etapas realizadas no desenvolvimento do presente trabalho.

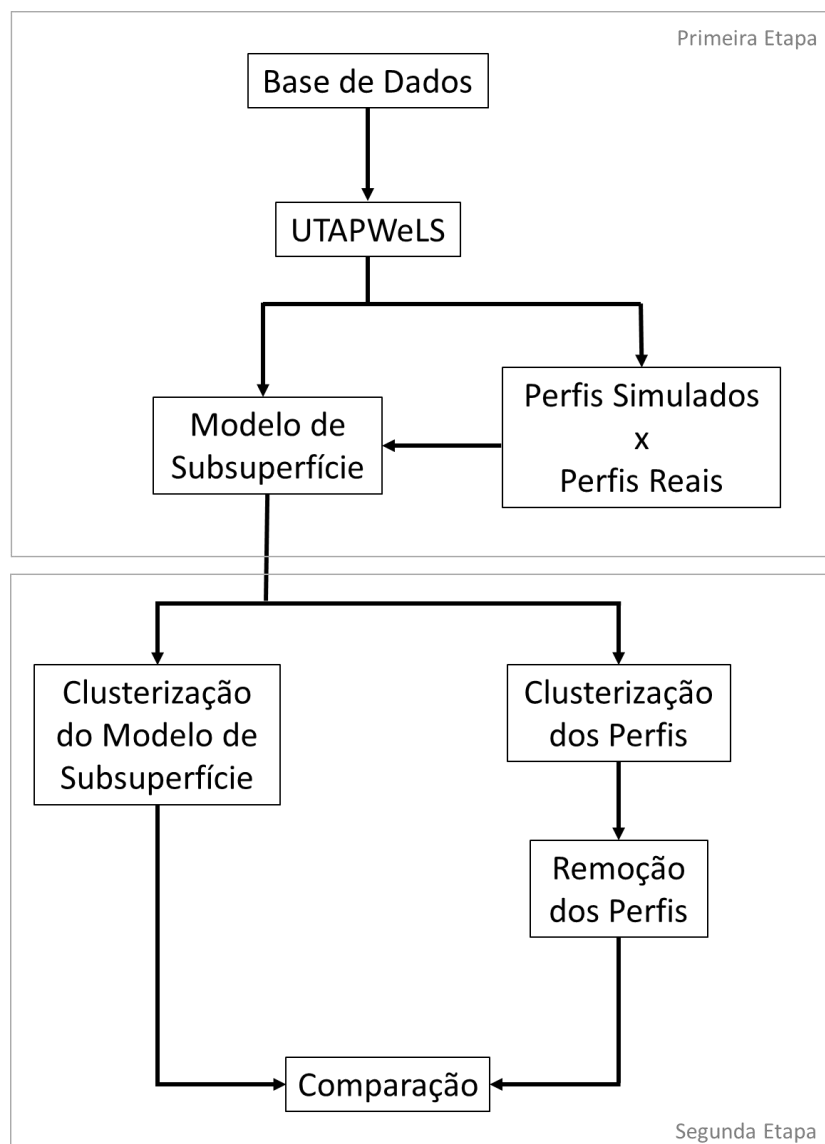


Figura 16: fluxograma proposto para metodologia do trabalho.

3.1 Base de Dados

Dentre os dados disponibilizados pela ANP constam dez perfis de poços para cada um dos poços estudados, sendo eles: calibre do poço (*caliper*), Raios-gama total e espectral (concentração de Th, U e K), resistividade, porosidade neutrônica, densidade, fator fotoelétrico, sônico, e porosidade total através de ressonância magnética nuclear (RMN). Para além dos dados de perfilagem, também foram usados neste trabalho dados de análises laboratoriais de amostras de rocha e fluido, tais como: mineralogia; salinidade e temperatura da água de formação; e composição dos fluidos.

Para realizar o processo de clusterização, foram utilizadas informações que se encontram em conjuntos de dados, chamados de base de dados (*datasets*). As quatro base de dados usadas neste estudo contam com os seguintes componentes sólidos (mineralogia e matéria orgânica) e tipos de perfis de poços (Tabela 1):

Tabela 1 – informações da base de dados do poço dos intervalos Pós e Pré-sal, com total de metros e amostras analisadas

Base de Dados	Dados Disponíveis	Tipo de Dado	Intervalo Analisado (m) / Quantidade de Amostras
Pós-Sal	Anortita	Componentes Sólidos	26
	Biotita		
	Calcita		
	Clorita		
	Ilita		
	Ortoclásio		
	Quarzo		
	Densidade (ZDEN)	Perfis de Poços	232,72
	Diâmetro de Poço (CAL)		
	Fator Fotoelétrico (PE)		
	Porosidade de Nêutrons (NPHI)		
	Porosidade Total do RMN (MPHS)		
	Raios Gama (GR)		
	Raios Gama Espectral (K, Th, U)		
	Resistividade (M2R)		
	Vagarosidade da Onda-Compressional (DTCF)		
	Vagarosidade da Onda Cisalhante-(DTSF)		
Pré-Sal	Calcita	Componentes Sólidos	82
	Dolomita		
	Matério Orgânica		
	Ortoclásio		
	Quartzo		
	Densidade (ZDEN)	Perfis de Poços	194,77
	Diâmetro de Poço (CAL)		
	Fator Fotoelétrico (PE)		
	Porosidade de Nêutrons (CNC)		
	Porosidade Total do RMN (TCMR)		
	Raios Gama (GR)		
	Raios Gama Espectral (K, Th, U)		
	Resistividade (M2R)		
	Vagarosidade da Onda Compressional (DT24)		

Muitas vezes, ao estudar uma base de dados, percebe-se a presença de dados espúrios, que prejudicam sua análise – por mais que sejam minoritários. Nesses casos, é necessário que haja uma etapa de tratamento de dados, desprezando qualquer dado espúrio na base de dados, para que ele se torne representativo no contexto do problema.

No caso do presente trabalho, dados espúrios costumam ser causados por regiões com intenso arrombamento. Sendo assim, perfis muito afetados por este

motivo, e que não foram possíveis de serem recuperados através de correções ambientais, tiveram seus pesos diminuídos na construção do modelo de subsuperfície, conforme detalhado posteriormente. Ademais, como na base de dados havia valores que indicavam erros de leitura de ferramenta, foi necessário realizar a exclusão dessas informações, de tal forma que a base de dados dispusesse apenas de dados confiáveis.

3.2 Criação do Modelo de Subsuperfície

Para gerar o modelo de subsuperfície foi utilizado o *software* UTAPWeLS. A metodologia adotada neste trabalho necessita que haja um gabarito do modelo de subsuperfície, ou seja, informações contínuas de frações minerais e de fluido e, conseqüentemente, de litofácies. Com esse gabarito do modelo de subsuperfície, é que seja possível comparar os resultados gerados e verificar sua acurácia. Dessa forma, o uso desse *software* foi uma escolha oportuna para cumprir essa proposta, pois tornou possível gerar um modelo de subsuperfície simulando as características das rochas em torno dos poços selecionados.

Para tanto, o primeiro passo da etapa de geração foi estabelecer os limites das litofácies através da interpretação dos perfis de poços. Em seguida, através do módulo Paleta Composicional, cada litofácies foi caracterizada, fornecendo ao *software* informações de mineralogia, saturação das rochas, tipo de fluido presente, temperatura e salinidade da água da formação. Após esta etapa, é realizada então a simulação dos perfis de poço sintéticos referentes às litofácies caracterizadas, que foram comparados com os reais para avaliar a representatividade das litofácies calibradas. Por último, a fim de se obter o gabarito necessário, as frações minerais e de fluido do modelo de subsuperfície foram refinadas através do módulo Inversão Composicional Conjunta, supervisionado pela informação a priori das frações minerais observadas nas análises laboratoriais das amostras de rocha (Figuras 17 e 18).

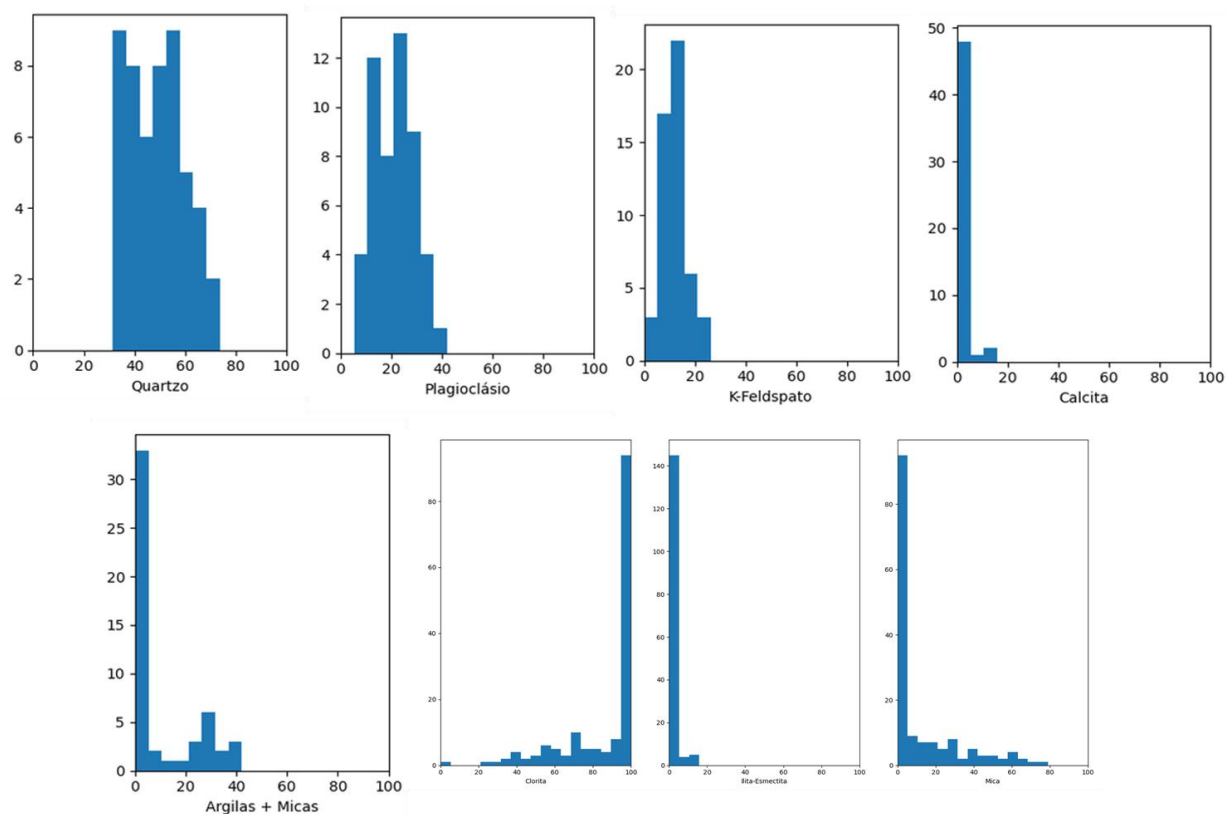


Figura 17: Histograma de frações minerais utilizados como informação a priori do módulo Inversão Composicional Conjunta do poço do intervalo Pós-sal.

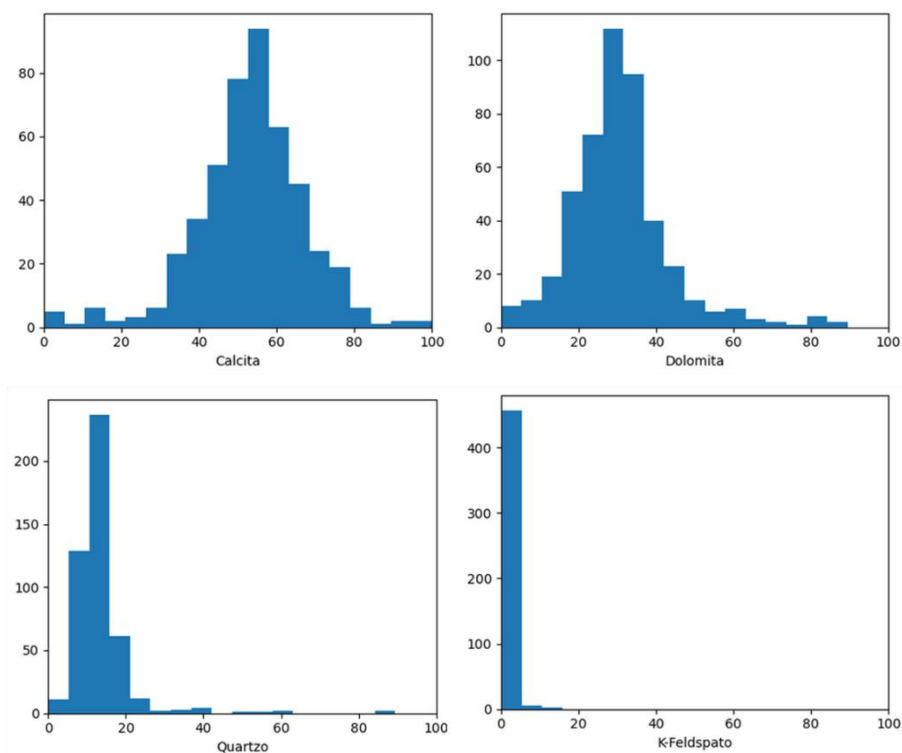


Figura 18: Histograma de frações minerais utilizados como informação a priori do módulo Inversão Composicional Conjunta do poço do intervalo Pré-sal.

Os minerais utilizados nos modelos de subsuperfície e os perfis de poços para validar esses modelos para os intervalos do Pós e Pré-sal foram os apresentados na Tabela 1 acima. Como o módulo Inversão Composicional Conjunta permite alterar os pesos que cada perfil tem no refinamento do modelo de subsuperfície, aqueles perfis que foram muito afetados por fatores ambientais, como arrombamento, tiveram seu peso diminuído.

Todos os dados disponibilizados pela ANP para realização do presente estudo foram utilizados como parâmetros para auxiliar o *software* na calibração do Inversão Composicional Conjunta. Cabe destacar que, visando garantir o controle de qualidade nesse processo, toda a etapa foi supervisionada por resultados de análises de amostras de rocha e fluido coletadas nos intervalos analisados, e disponibilizados pela ANP.

3.3 Clusterização

O processo de clusterização foi realizado através do algoritmo de aprendizado de máquina não-supervisionado *k-means*. Essa escolha se deu devido a esse algoritmo já ser considerado consagrado, e por ele ser de fácil interpretação. O *software* utilizado nessa etapa do trabalho foi o Weka. Desenvolvido pela Universidade de Waikato (Nova Zelândia), esse *software* apresenta uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados, contendo, inclusive, ferramentas para realizar clusterizações.

A primeira etapa do processo de clusterização foi realizada nos dados do modelo de subsuperfície gerado previamente, considerado o mais próximo do real. Os *clusters* definidos pela máquina indicam diferentes litofácies. Como a classificação de litofácies gerada a partir da clusterização do modelo de subsuperfície foi considerado a mais representativa das formações, ela foi definida como o “gabarito das litofácies”.

Em seguida, todos os perfis de poços foram clusterizados juntos, seguindo a mesma rotina de clusterização adotada anteriormente. Nesse momento, a máquina realiza a classificação das eletrofácies, por meio do agrupamento dos dados de perfis de poços que são semelhantes entre si.

Por fim, sabe-se que, muitas vezes, ao longo da elaboração do plano de aquisição de perfis de poços de um campo, há uma tendência de adquirir apenas as informações mais relevantes para a sua avaliação. Portanto, foi testado os limites de

até onde é possível manter boas respostas de clusterização, quantos e quais tipos de perfis de poços são realmente necessários para estabelecer uma boa interpretação.

Para cumprir esse objetivo, os perfis de poços foram sendo removidos um a um, para testar como a falta de informação poderia impactar na resposta de agrupamento do algoritmo. A ordem de clusterização seguida está indicada no fluxograma abaixo (Figura 19):

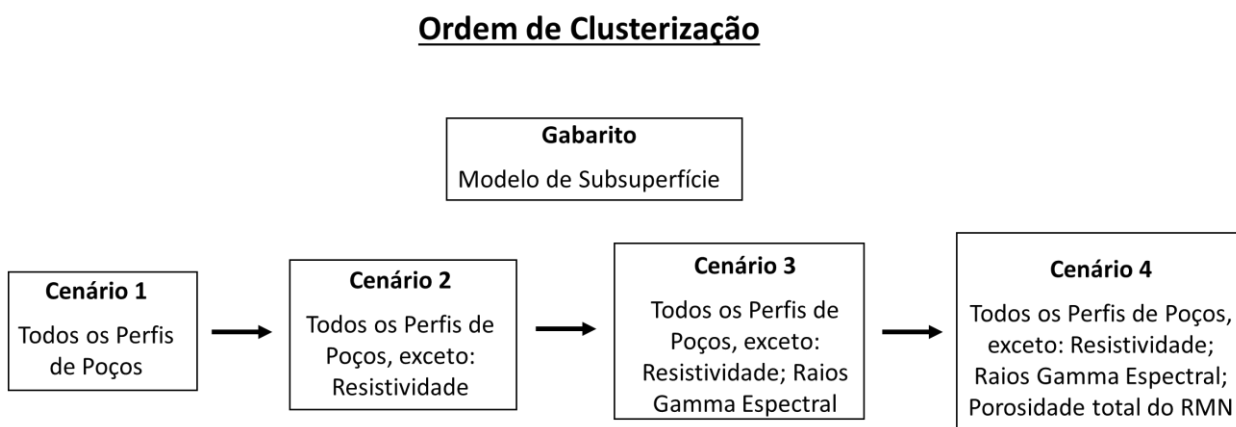


Figura 19: fluxograma indicando a ordem de clusterização utilizada no presente trabalho.

Como o interesse do estudo se voltou para a classificação de litofácies, nesse momento, a única característica litológica das rochas analisadas que foi levada em consideração para clusterização foi suas frações minerais, pois traria resultados mais representativos. Se o interesse do estudo envolvesse, por exemplo, a compreensão de saturação de fluido, uma outra metodologia teria que ser desenvolvida.

Para todas os cenários de clusterização, foram testados valores de “k” variando entre 2 e 10. Posteriormente, foi utilizada a dispersão média, calculada a partir da distância euclidiana entre os pontos de cada *cluster*, para cada valor de k testado, afim de plotar no gráfico do cotovelo (*elbow graph*) (Figura 20).

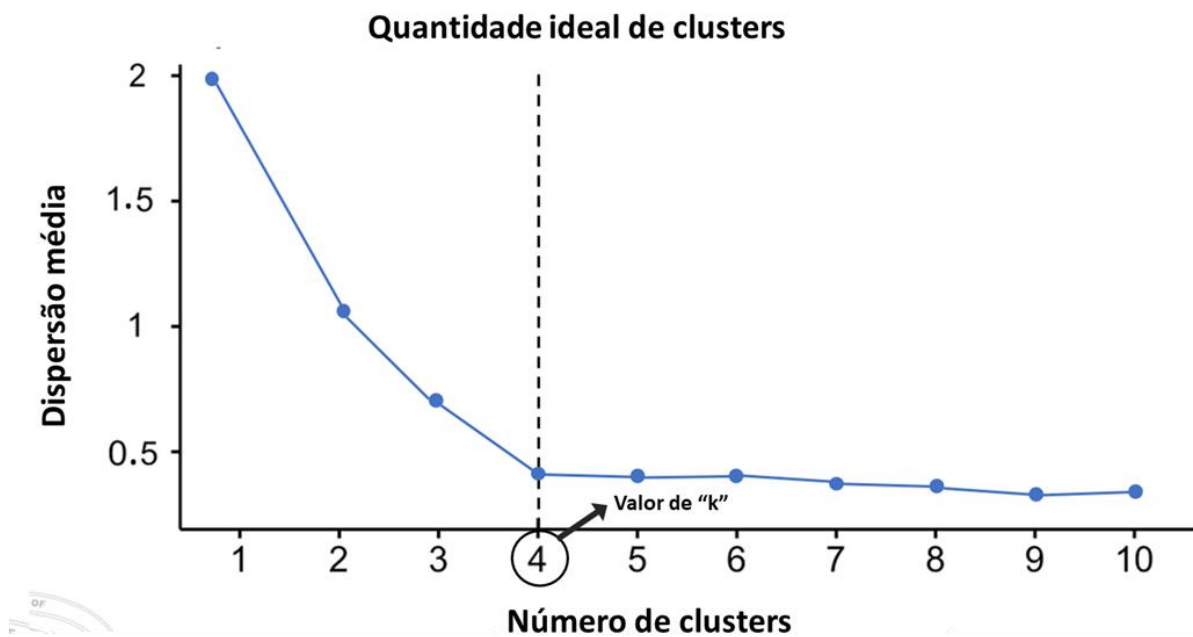


Figura 20: exemplo didático de um gráfico do cotovelo utilizado para determinar a quantidade ideal de clusters.

Este gráfico auxilia a estabelecer qual seria a quantidade mais adequada de clusters a se adotar em cada cenário. O ponto ideal, conhecido como “*elbow point*”, normalmente é selecionado onde há a uma quebra no gráfico em que, a partir deste valor, a dispersão média deixa de diminuir drasticamente a cada ponto e passa a variar pouco. Assim, do *elbow point* em diante, idealmente, o desenho do gráfico tende a assemelhar-se a uma reta horizontal (plateau).

4. RESULTADOS

Os resultados gerados ao longo deste trabalho serão apresentados no presente capítulo subdividindo em dois tópicos: (4.1) Modelo de Subsuperfície, onde será apresentado o modelo gerado e a comparação entre os perfis de poços simulados e os adquiridos para validação do modelo de subsuperfície, e; (4.2) Clusterização, que apresentará tanto os Gráficos de Cotovelo que auxiliaram a estabelecer o número ideal de *clusters* quanto a Análise de *Clusters*.

Para cada tópico será primeiro apontado os produtos provenientes dos dados do poço do intervalo Pós-sal e, em seguida, os produtos provenientes dos dados do poço do intervalo Pré-sal.

4.1 Modelo de Subsuperfície

Os primeiros produtos obtidos neste trabalho foram os modelos de subsuperfície de cada poço, calibrados e gerados no UTAPWeLS. Na última trilha da figura 21, pode-se observar o modelo de subsuperfície após a inversão composicional. O mesmo produto, obtido através dos dados do poço do intervalo Pré-sal, está disposto na última trilha da figura 22.

A presença de cada mineral no modelo de subsuperfície gerado é indicado por diferentes cores, que variam de proporção conforme a profundidade. A quantidade de cada mineral aumenta da esquerda (0%) para direita (100%), de forma cumulativa.

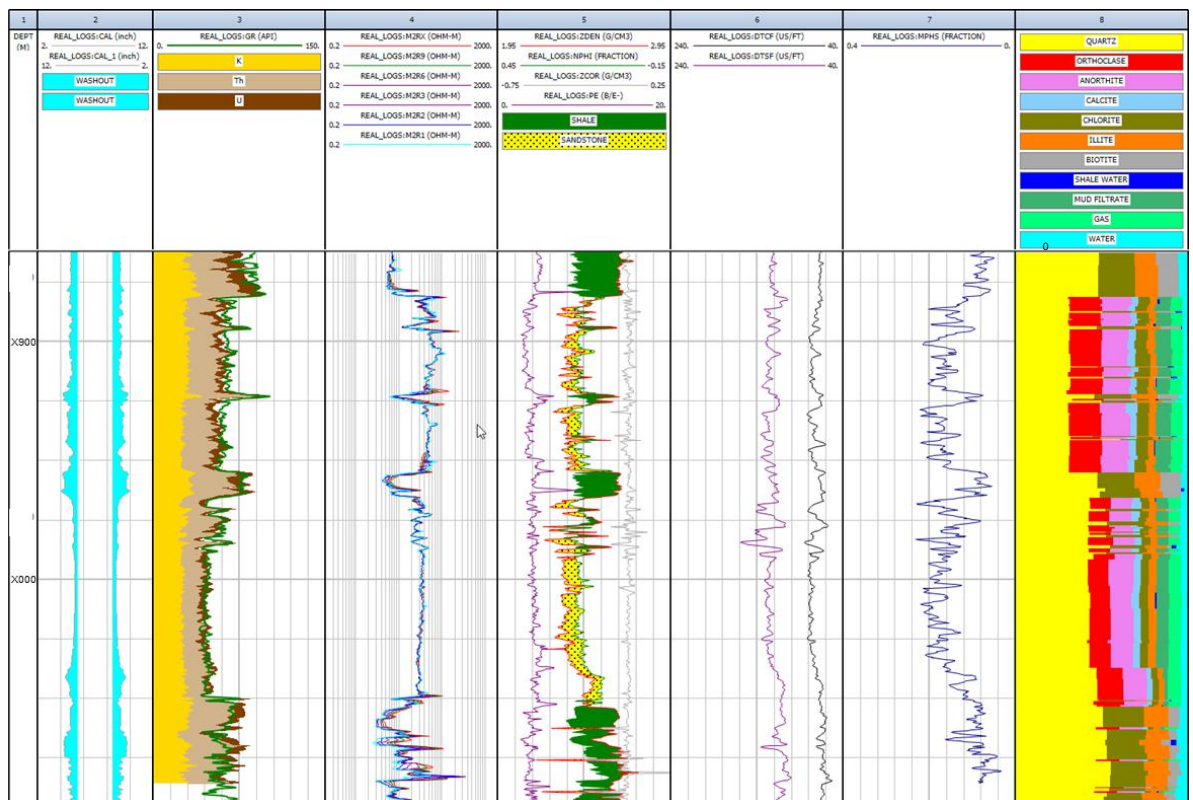


Figura 21: modelo de subsuperfície gerado a partir dos dados do poço do intervalo Pós-sal.

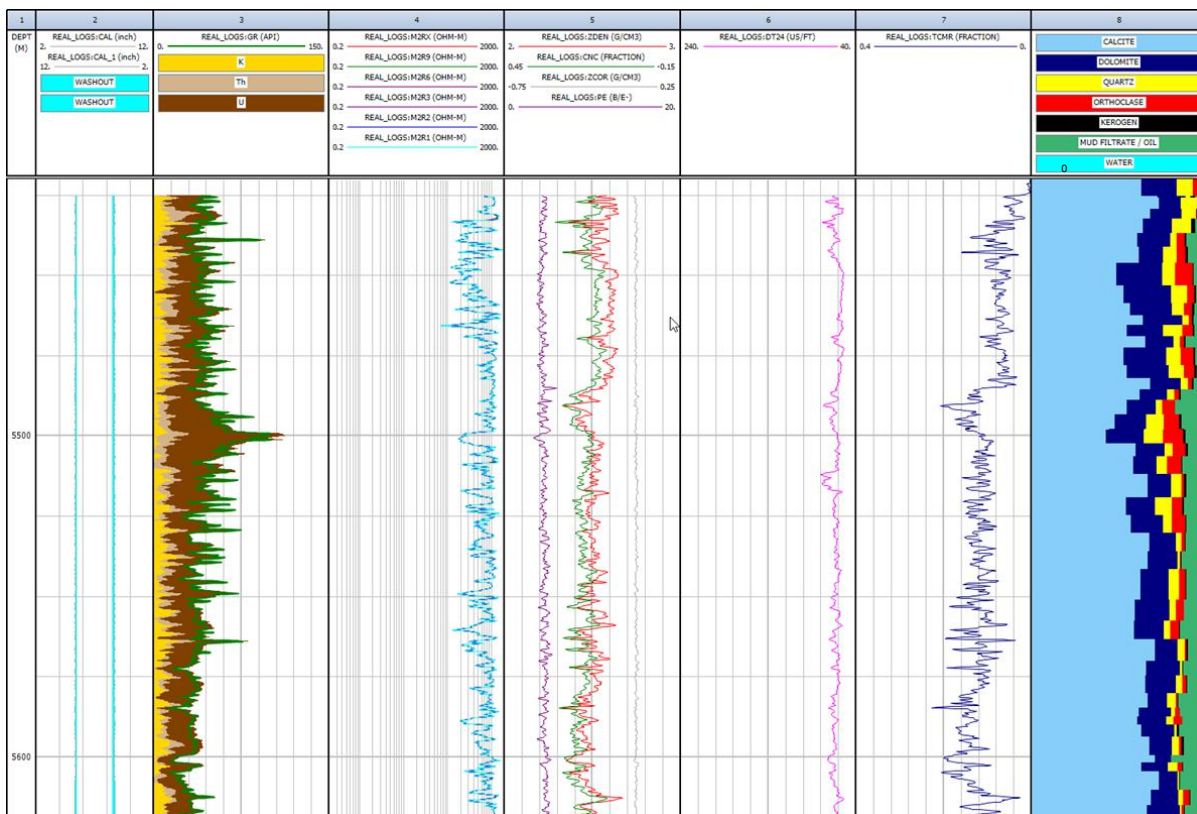


Figura 22: modelo de subsuperfície gerado a partir dos dados do poço do intervalo Pré-sal.

Para atestar a qualidade do modelo de subsuperfície gerado, os perfis reais foram comparados com perfis simulados a partir do modelo de subsuperfície. Obteve-se como resultado seis perfis de poços simulados para cada um dos poços: calibre do poço (trilha 2), Raios-gama (trilha 3), resistividade (trilha 4), densidade (trilha 5), nêutrons (trilha 6), fator fotoelétrico (trilha 7), vagarosidade das ondas compressional e cisalhante (trilha 8) – no caso do Pós-sal – e vagarosidade da onda compressional (trilha 8) – no caso do Pré-sal.

Para checar o nível de compatibilidade entre os perfis adquiridos nos poços perfurados e os perfis simulados no UTAPWeLS, todos os perfis foram plotados sobrepostos aos seus correspondentes. O produto obtido nessa etapa para o poço do intervalo Pós-sal encontra-se abaixo, na figura 23:

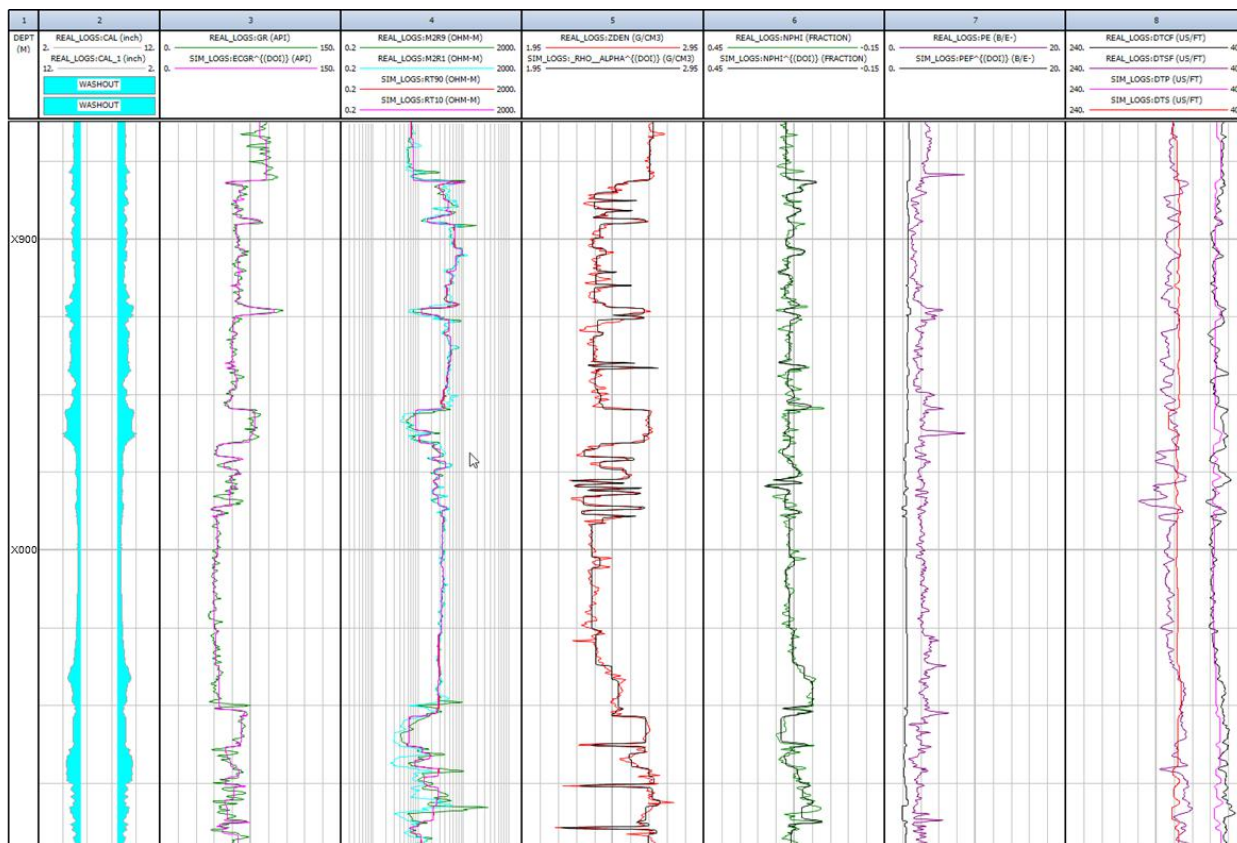


Figura 23: sobreposição dos perfis adquiridos em poço e dos simulados no UTAPWeLS, referente ao intervalo Pós-Sal. REAL = perfis adquiridos no poço; SIM = perfis simulados no *software*.

Os quatro primeiros perfis simulados, da esquerda para direita (Raios-gama, resistividade, densidade e nêutrons), tiveram um bom ajuste com os perfis reais. Por outro lado, o fator fotoelétrico e a vagarosidade da onda cisalhante não apresentaram

uma boa correspondência entre os perfis simulados pelo *software* e os fornecidos pela ANP.

No primeiro caso, é provável que o baixo ajuste entre os perfis tenha se dado pelo fato do poço estar com elevado nível de arrombamento. É possível observar isto através do perfil de calibre de poço, que indica um poço com paredes rugosas. Essa característica do poço faz com que o fluido de perfuração, rico em baritina, impacte na leitura do perfil fotoelétrico durante sua aquisição, gerando anomalias nos valores lidos pela ferramenta. Por este motivo, o peso designado a este perfil na etapa de inversão composicional foi menor que os demais perfis.

Já no segundo caso, como não foi possível calibrar adequadamente o modelo de velocidade da onda cisalhante no *software* utilizado, optou-se por desconsiderar este perfil na etapa de inversão composicional, para evitar incoerências no resultado. Essa impossibilidade de calibração se deve a presença de gás no reservatório, que demandaria um modelo calibrado considerando um cenário de arenitos com gás. Esta tarefa iria comprometer a viabilidade de execução dos resultados, pois exigiria um tempo impraticável para a proposta do trabalho.

No resultado obtido a partir do poço do intervalo Pré-sal (Figura 24) fica perceptível a dificuldade dos perfis simulados de honrarem a variação de alta frequência que é observada nos perfis de poços adquiridos. Mesmo que esse fato possa ser reparado em todos os perfis simulados, eles não deixaram de ter uma boa correlação com os perfis originais, pois ainda assim foram capazes de capturar as grandes tendências de cada curva.

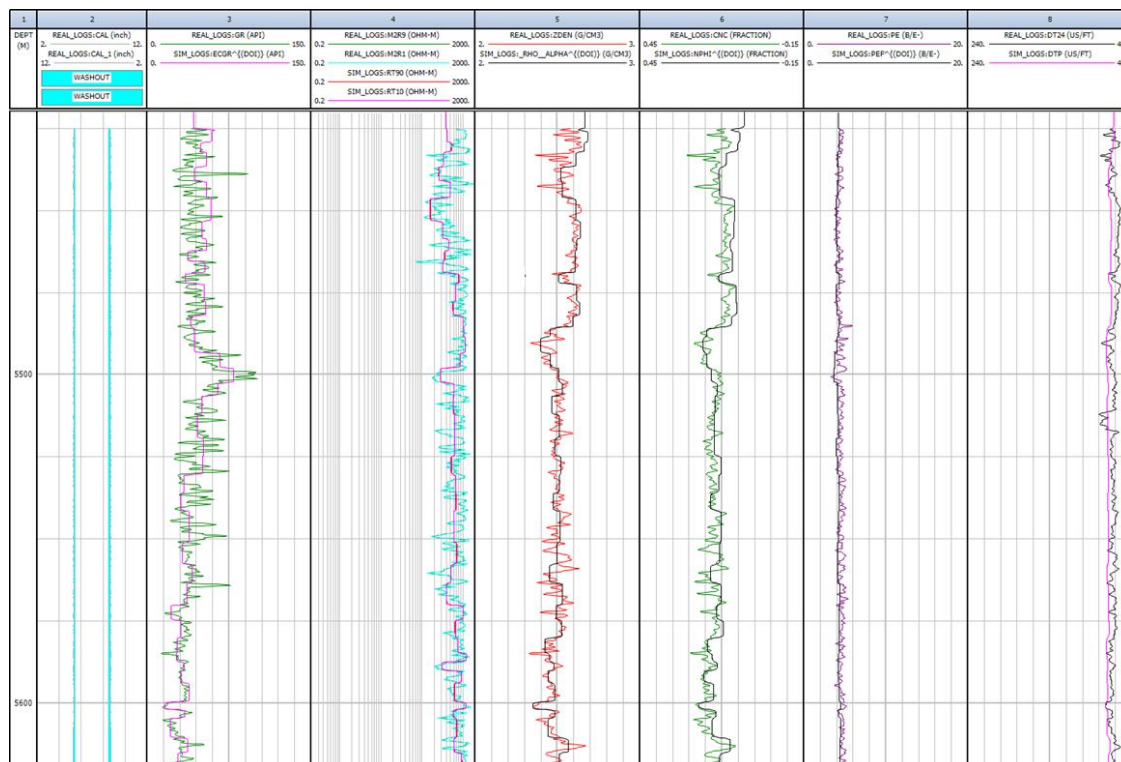


Figura 24: sobreposição dos perfis adquiridos em poço e dos simulados no UTAPWeLS, referente ao intervalo Pré-Sal. REAL = perfis adquiridos no poço; SIM = perfis simulados no *software*.

4.2 Clusterização

4.2.1 Gráficos do cotovelo

Foram gerados gráficos de cotovelo para o resultado da clusterização do modelo de subsuperfície, de todos os perfis de poços, e de cada etapa em que uma informação foi subtraída. Os valores da dispersão média de cada cenário avaliado estão resumidos na tabela 2, abaixo:

Tabela 2: valores da dispersão média para cada valor de k nos diferentes cenários avaliados

Dispersão Média		Valor de k:	2	3	4	5	6	7	8	9	10
Intervalo Pós-Sal	Modelo de subsuperfície		128,40	63,52	29,48	28,61	26,64	20,92	20,06	18,64	17,80
	Todos os perfis		330,80	285,25	234,85	223,50	200,63	190,71	181,39	174,47	170,86
	Todos os perfis, exceto: Resistividade		280,67	243,59	193,83	180,92	163,96	154,92	147,77	142,13	139,17
	Todos os perfis, exceto: Resistividade; RG Espectral		177,31	137,62	120,62	107,69	97,41	93,11	89,11	85,60	82,74
	Todos os perfis, exceto: Resistividade; RG Espectral; Porosidade Total do RMN		158,05	141,09	102,12	89,36	78,61	75,26	70,13	67,20	63,75
Intervalo Pré-Sal	Modelo de subsuperfície		215,46	193,34	167,91	123,01	90,94	84,67	78,51	64,31	60,07
	Todos os perfis		507,52	449,31	349,90	310,15	298,31	271,60	262,19	233,48	222,31
	Todos os perfis, exceto: Resistividade		324,57	200,66	179,91	155,41	144,08	138,15	132,62	120,52	115,80
	Todos os perfis, exceto: Resistividade; RG Espectral		171,42	103,93	84,62	74,44	68,85	65,21	62,49	51,99	49,06
	Todos os perfis, exceto: Resistividade; RG Espectral; Porosidade Total do RMN		131,83	83,76	68,97	58,58	54,73	50,88	47,62	39,97	37,81

Para facilitar a visualização e comparação dos gráficos de cotovelo obtidos em cada etapa, eles foram normalizados e plotados em um único gráfico, para cada poço.

No caso do intervalo Pós-sal (Figura 25), observou-se que o gráfico do cotovelo dos perfis apresenta uma dispersão média maior que a do modelo de subsuperfície. Além disso, em geral, eles não apresentam uma quebra da curva tão acentuada quanto no modelo de subsuperfície. Possivelmente isso se deve ao fato do modelo de subsuperfície gerado no UTAPWeLS, considerado o gabarito, apresentar litofácies cujo conteúdo mineralógico é melhor identificável e mais fácil de se agrupar, facilitando a clusterização. Já os clusters gerados através dos perfis de poço são afetados por diversos fatores além do conteúdo mineral, como porosidade, variação de fluidos, fatores ambientais etc.

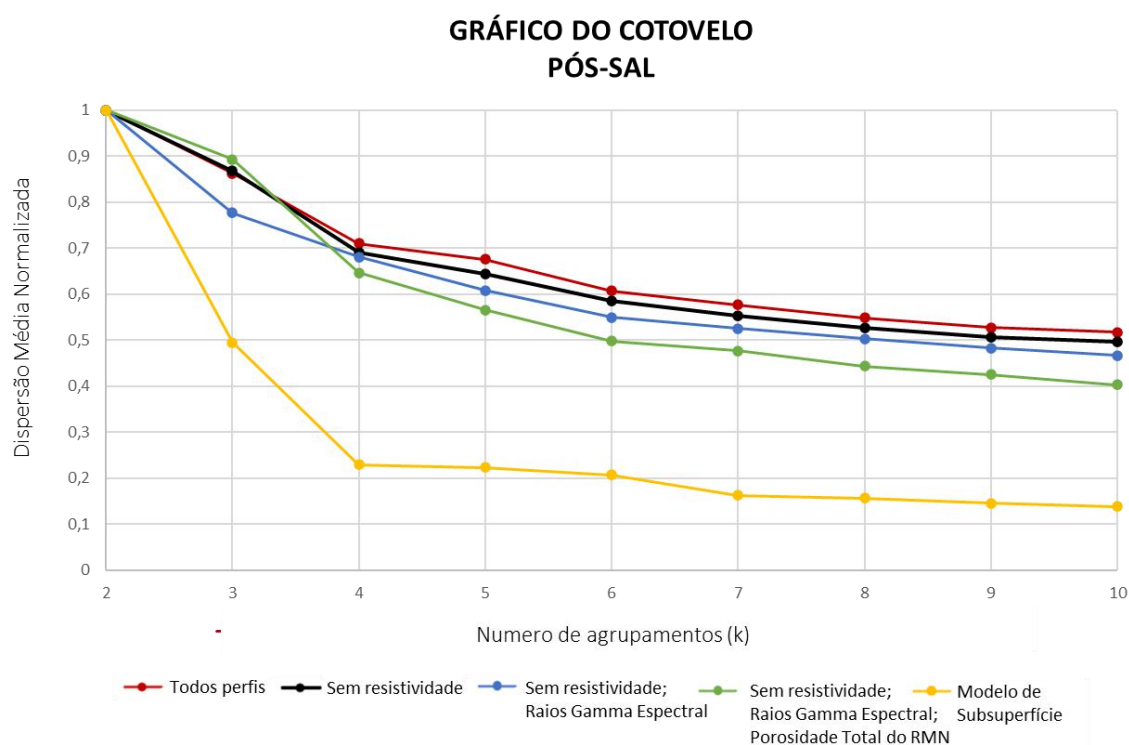


Figura 25: Gráfico do cotovelo contendo todos os resultados normalizados do poço do intervalo Pós-Sal.

Através do gráfico de cotovelo do modelo de subsuperfície, indicado pela linha amarela, notou-se que o número de *clusters* ideal varia entre quatro e seis, não apresentando alterações muito significativas a partir disso. Para ter uma variedade adequada de litologias, o valor de k foi definido como cinco.

Já para o intervalo Pré-sal (Figura 26), percebe-se que a dispersão média é mais elevada logo no modelo de subsuperfície. Assim, o ponto clássico de inflexão na dobra do cotovelo fica mais atenuado, tornando difícil determinar o valor ideal de k . De toda forma, ainda é possível notar que a curva começa a se estabilizar próxima ao número seis, e, portanto, esta foi a quantidade de *clusters* adotada nesse caso.

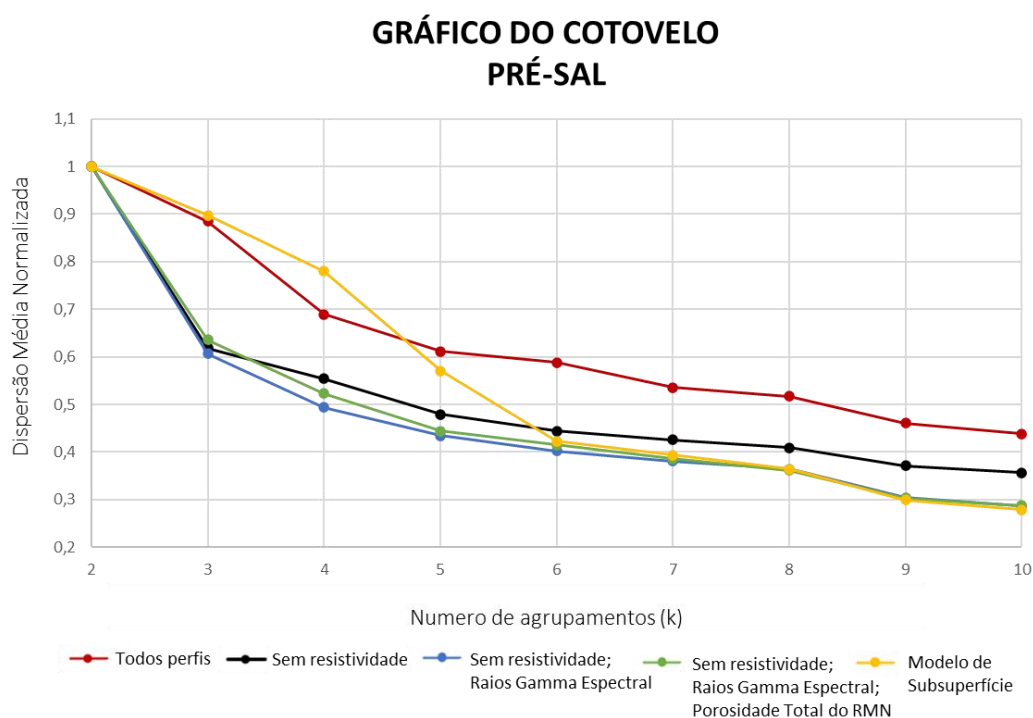


Figura 26: Gráfico do cotovelo contendo todos os resultados normalizados do poço do intervalo Pré-sal

4.2.2 Análise de *Clusters*

Cada *cluster* gerado a partir do processo de clusterização do modelo de subsuperfície indica uma diferente litofácies. E cada *cluster* gerado a partir do processo de clusterização dos dados de perfis de poços indica uma diferente eletrofácies. Como a classificação de litofácies, proveniente do modelo de subsuperfície, foi considerada como gabarito, esse capítulo irá comparar a classificação das eletrofácies com o gabarito. Podendo avaliar, portanto, a capacidade do *K-means* de classificar litofácies a partir de perfis de poços.

As médias das frações minerais de cada *cluster* do modelo de subsuperfície, que representam as diferentes litologias presentes nos poços, estão retratadas nos resultados conforme indicado na tabela 3:

Tabela 3 – média do conteúdo mineralógico de cada *cluster*

	Cluster	Quartzo	Ortoclásio	Anortita	Calcita	Ilita	Clorita	Biotita
Intervalo Pós-Sal	1	0,3305	0,1688	0,1454	0,0383	0,0521	0,0829	0,0103
	2	0,4057	0,1198	0,1119	0,0322	0,0637	0,0928	0,0247
	3	0,4710	0,0826	0,0839	0,0259	0,0763	0,1089	0,0215
	4	0,4246	0,1090	0,1018	0,0294	0,0701	0,1027	0,0213
	5	0,3160	0,1789	0,1579	0,0481	0,0422	0,0741	0,0029
	Cluster	Calcita	Dolomita	Quartzo	Ortoclásio	Querogênio	-	-
Intervalo Pré-Sal	1	0,5789	0,1882	0,0730	0,0517	0,0116	-	-
	2	0,5957	0,1806	0,0566	0,0455	0,0081	-	-
	3	0,5457	0,2156	0,0735	0,0618	0,0097	-	-
	4	0,6136	0,1730	0,0537	0,0382	0,0085	-	-
	5	0,6045	0,1773	0,0578	0,0410	0,0091	-	-
	6	0,6169	0,1661	0,0491	0,0377	0,0377	-	-

É possível observar que, para o reservatório siliciclástico, a areia com maior conteúdo quartzoso é representada pelo *cluster* 3, enquanto aquela com maior conteúdo feldspático é representada pelo *cluster* 5. No contexto do reservatório carbonático, o *cluster* 4 destaca-se por apresentar o maior conteúdo de carbonato total, mas a litologia mais dolomitizada é simbolizada pelo *cluster* 3. O *cluster* 6 se destaca por ser aquele com maior nível de querogênio presente (3,78%) em sua composição.

Assim, as litofácies, representadas por cada cluster, foram classificadas de acordo com a média do conteúdo mineralógico de cada clusterização. E encontram-se indicadas na tabela 4, abaixo:

Tabela 4 – classificação das litofácies que cada cluster representa

	Cluster	Litofácies
Intervalo Pós-Sal	1	Folhelho
	2	Arenito com argilo minerais
	3	Arenito quartzoso
	4	Folhelho com silte
	5	Arenito feldspático
Intervalo Pré-Sal	1	Carbonato silicificado
	2	Carbonato 1
	3	Carbonato dolomitizado
	4	Calcário
	5	Carbonato 2
	6	Calcário com M.O

Entretanto, vale ressaltar que, no caso do intervalo Pré-sal, as diferenças entre cada *cluster* foram muito pequenas. Assim, não existe uma variação expressiva em termos de mineralogia entre os 6 clusters obtidos. Como exemplo, a diferença entre o cluster com maior conteúdo carbonático (*cluster* 4) e o menor (*cluster* 3) é de apenas 2,53%.

O produto final do trabalho foi o perfil faciológico de cada etapa pré-estabelecida no segundo fluxograma apresentado na metodologia. Os resultados obtidos foram dispostos lado a lado para facilitar a comparação entre eles.

Para o intervalo Pós-sal (Figura 27) é perceptível que as respostas de clusterização foram bastante semelhantes entre o gabarito e os cenários testados. Isso ocorre para todos os resultados do intervalo Pós-sal, com exceção apenas do cenário 1, que apresentou algumas diferenças.

Muito provavelmente, isso se deve ao fato deste cenário ser o único que considera o perfil de resistividade para clusterização. Dessa forma, este perfil deve estar gerando uma resposta muito mais relacionada aos fluidos presentes nessas profundidades do que a litologia em si.

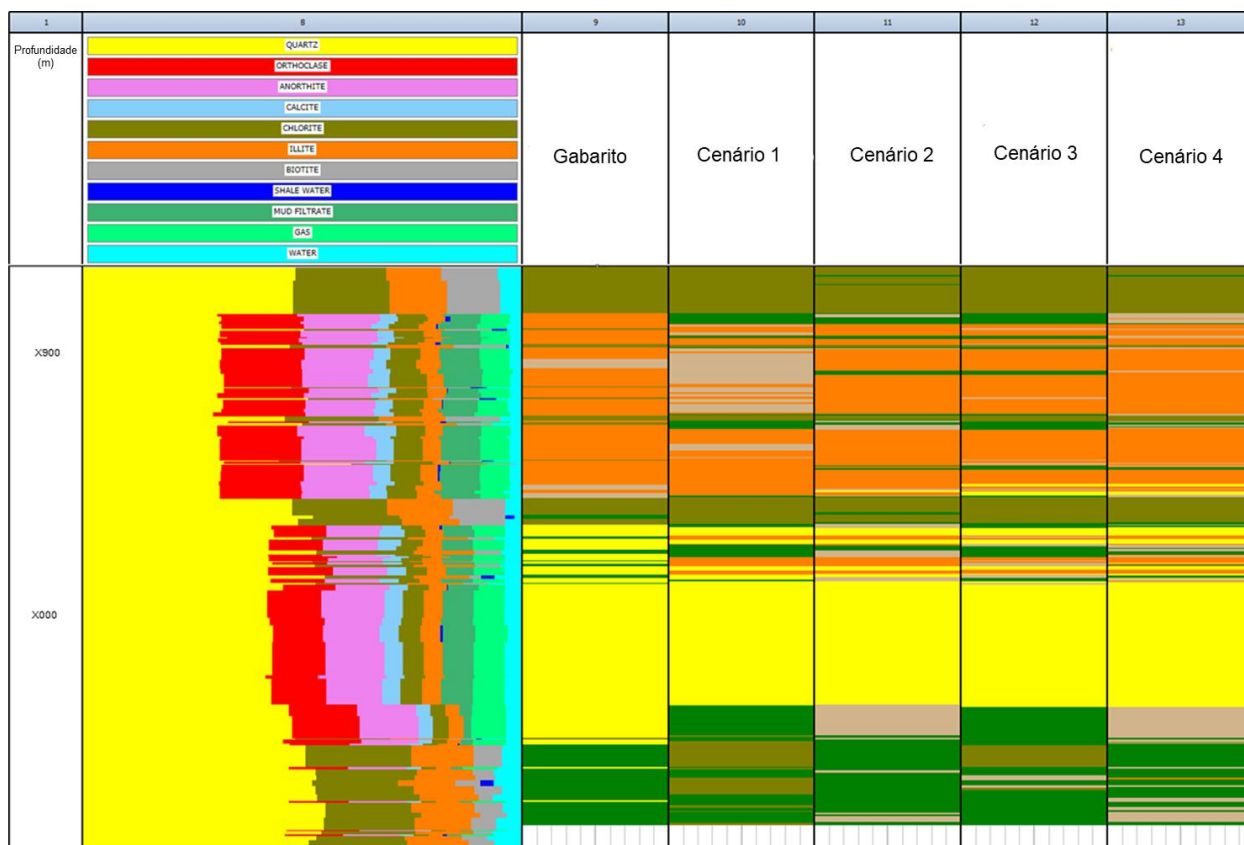


Figura 27: perfil faciológico a partir do resultado de clusterização das frações minerais do modelo de subsuperfície, de todos os poços e de cada etapa de retirada de informação. Referente aos resultados do poço do intervalo Pós-sal. Cluster 1 = oliva; cluster 2 = laranja; cluster 3 = bege; cluster 4 = verde; cluster 5 = amarelo

Além disso, pode-se observar que a alternância entre os clusters 4 e 5, observada em uma dada profundidade do gabarito, também é captada nos demais cenários. Porém, em todos os cenários testados o cluster 4 é substituído pelo cluster 2. É provável que isso ocorra devido a resolução vertical dos perfis de poços ser inferior à do modelo de subsuperfície. Assim, a clusterização dos perfis fica prejudicada, sem conseguir retratar a verdadeira variação litológica nessas profundidades.

Por outro lado, o resultado do intervalo Pré-sal (Figura 28) destaca-se por mostrar variações de alta frequência, sem um padrão claro de clusterização ao longo do poço ou entre os cenários testados. Até mesmo a classificação das litofácies já não aparenta mostrar nenhuma distinção clara das diferentes litologias. E a classificação das eletrofácies, por sua vez, não foi capaz sequer de se aproximar do observado na clusterização do modelo de subsuperfície.

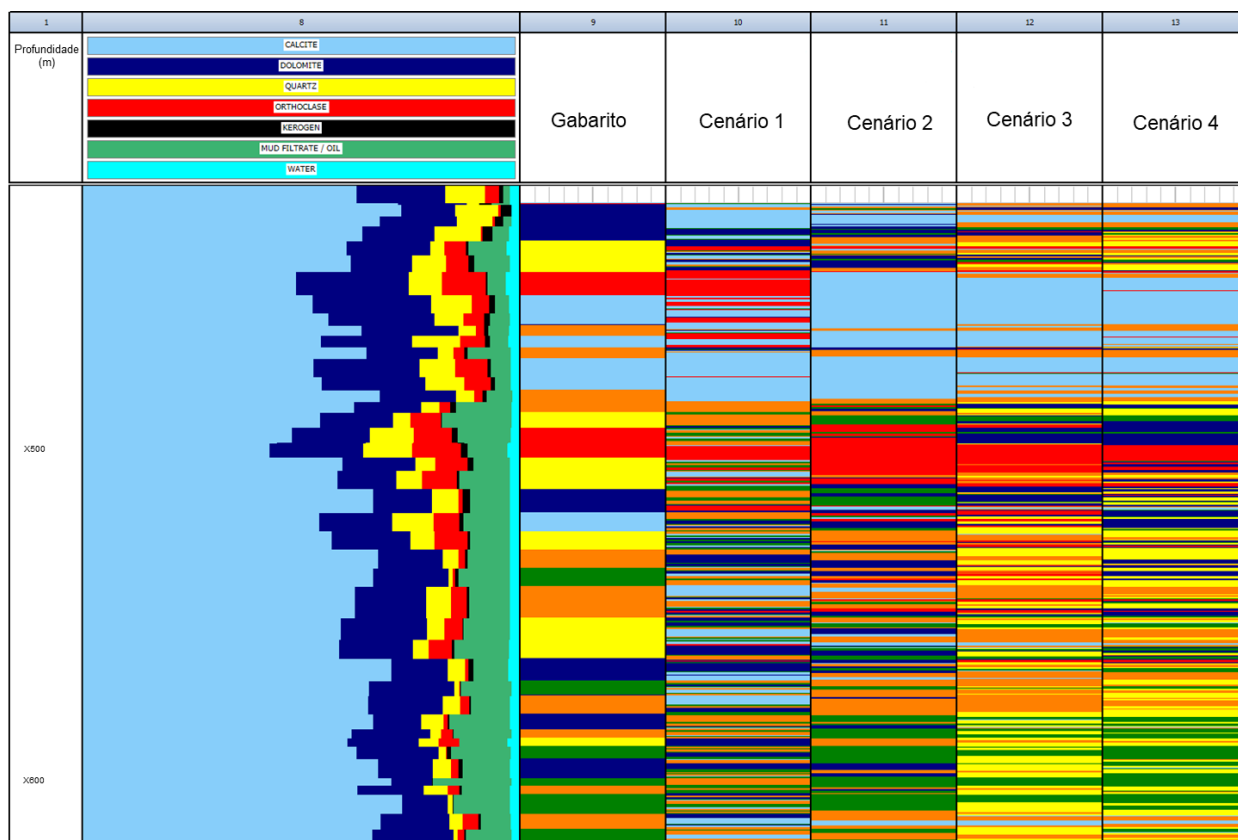


Figura 28: perfil faciológico a partir do resultado de clusterização das frações minerais do modelo de subsuperfície, de todos os poços e de cada etapa de retirada de informação. Referente aos resultados do poço do intervalo Pré-sal. Cluster 1 = azul; cluster 2 = amarelo; cluster 3 = vermelho; cluster 4 = celeste; cluster 5 = laranja; cluster 6 = verde.

Essa baixa correspondência entre as classificações sugere que aquelas variações de alta frequência observadas nos perfis faciológicos representam variações que não estão relacionadas somente a diversidade mineralógica. Ou seja, as litofácies estão relacionadas não somente as variações composicionais, mas também a outras características da rocha. Assim, para a assembleia de perfis de poços selecionados não foi possível honrar a classificação de litofácies do gabarito.

5. DISCUSSÕES

Como mencionado no capítulo 1.2, o objetivo deste estudo compreende a análise das diferenças entre os produtos obtidos para o poço de Mexilhão (intervalo Pós-sal) e de Tupi (intervalo Pré-sal). Portanto, as discussões do trabalho visam comparar os resultados de cada um deles.

Quando postos lado a lado, percebe-se facilmente uma dessemelhança nos resultados referentes a cada poço. Enquanto no contexto das rochas siliciclásticas do intervalo Pós-sal tem-se uma assembleia mineralógica muito variada – com presença de diversos argilominerais – no contexto carbonático do intervalo Pré-sal, a composição das rochas é essencialmente calcita, dolomita e quartzo. Isto evidencia a discrepância de complexidade composicional entre os poços, o que impacta diretamente na classificação de litofácies realizada pelo algoritmo.

Em relação aos diferentes cenários de clusterização, os clusters obtidos com os perfis de poço em rochas siliciclásticas apresentaram boa semelhança com os obtidos através do modelo de subsuperfície. Isso demonstra que as variações observadas nos perfis são diretamente relacionadas ao conteúdo mineralógico das litofácies em ambientes siliciclásticos. Portanto, a classificação de eletrofácies nesse caso foi capaz de honrar a classificação de litofácies.

Já nos carbonatos do intervalo Pré-sal, a clusterização do próprio modelo de subsuperfície não foi capaz de definir litofácies claras. Como consequência, os clusters obtidos com os perfis tiveram pouca representatividade, tanto quando comparados com o modelo de subsuperfície quanto quando comparados entre si. Isso é reflexo de uma mineralogia relativamente constante dessas rochas e, como os perfis respondem às variações composicionais, o processo de clusterização tem dificuldade de encontrar diferença significativa entre os clusters.

Sendo assim, para classificação de litofácies utilizando o algoritmo de ML *k-means*, a utilização dos perfis de poço não é adequada para a identificação de litofácies no ambiente carbonático do intervalo Pré-sal. Possivelmente, a inclusão de informações texturais, como as obtidas em perfis de imagem, podem auxiliar nesse objetivo.

Quando se trata da influência do perfil de resistividade no processo de clusterização, fica evidente que ele gerou uma resposta prejudicial à classificação. Essa dificuldade desse perfil em honrar a classificação de litofácies observada no

gabário ocorre devido ao fato dele ser um tipo de perfil extremamente relacionado às características do fluido presente na formação. Assim, as clusterizações que utilizaram perfil de resistividade (cenário 1 de cada poço) apresentaram um certo grau de incoerência com relação aos demais.

6. CONCLUSÃO

A partir da classificação de litofácies usando técnicas de ML através de dados de perfis de poços nos intervalos selecionados dos reservatórios siliciclásticos de Mexilhão e carbonáticos de Tupi, foram gerados e analisados perfis faciográficos de cada um desses poços. Assim, as principais conclusões do presente trabalho se encontram pontuadas abaixo:

- A classificação de litofácies usando perfis de poços através de técnicas de ML com algoritmos de clusterização, como o *k-means*, é mais representativa em formações siliciclásticas;
 - Nesses casos, a suíte básica de perfilagem (Raios-gama, resistividade, densidade, porosidade de nêutrons, fator fotoelétrico) aparenta ser suficiente para caracterizar essas formações;
- Devido à baixa variedade mineralógica das rochas carbonáticas, essas formações devem ser avaliadas com maior atenção;
 - As variações observadas nos perfis de poços não são representativas das variações nos componentes da rocha;
 - A inclusão de perfis com componente textural, como os perfis de imagem, podem auxiliar nesse processo.
- Perfis de resistividade podem comprometer a classificação de litofácies que se baseiam apenas em características mineralógicas, por serem fortemente influenciados pela presença de fluidos;
- O processo de classificação de litofácies e caracterização de poços não é trivial e exige um conhecimento aprofundado de analistas de perfis de poços/ petrofísicos;

7. REFERÊNCIAS BIBLIOGRÁFICAS

ANP (2022); Boletim Mensal da Produção de Petróleo e Gás Natural. Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. Acesso em 02 de agosto de 2022. <http://www.anp.gov.br/publicacoes/boletins-anp/2395-boletim-mensal-daproducao-de-petroleo-e-gas-natural>

ANP (2018) Plano de Desenvolvimento Aprovado – Lula, 2018. Acessado em 11 de agosto de 2022. <https://www.gov.br/anp/pt-br/assuntos/exploracao-e-producao-de-oleo-e-gas/gestao-de-contratos-de-e-p/fase-de-producao/pd/lula.pdf>

ANP (2016) Plano de Desenvolvimento Aprovado – Mexilhão, 2016. Acessado em 11 de agosto de 2022. <https://www.gov.br/anp/pt-br/assuntos/exploracao-e-producao-de-oleo-e-gas/gestao-de-contratos-de-e-p/fase-de-producao/pd/mexilhao.pdf>

ARAI, Mitsuru. Paleogeografia do Atlântico Sul no Aptiano: um novo modelo a partir de dados micropaleontológicos recentes. **Boletim de Geociências da PETROBRAS**, v. 17, n. 2, p. 331-351, 2009.

ARCHIE, Gustave E. **The electrical resistivity log as an aid in determining some reservoir characteristics**. Transactions of the AIME, v. 146, n. 01, p. 54-62, 1942.

BESTAGINI, Paolo; LIPARI, Vincenzo; TUBARO, Stefano. **A machine learning approach to facies classification using well logs**. In: Seg technical program expanded abstracts 2017. Society of Exploration Geophysicists p. 2137-2142, 2017.

CARDOSO, Gustavo do S; NASCIMENTO, Victor L. R.; NUNES, Eldman de O.; NETO, Ernesto S. M. **Clusterização K-means: Uma Proposta de Melhoria** (Workshop de Trabalhos de Iniciação Científica e Graduação Bahia, Alagoas e Sergipe (WTICG), Salvador) 2008.

CARR, Herman Y.; PURCELL, Edward M. **Effects of diffusion on free precession in nuclear magnetic resonance experiments**. Physical review, v. 94, n. 3, p. 630, 1954. ELLIS, Darwin V.; SINGER, Julian M. Well logging for earth scientists. Dordrecht: Springer, 2007.

CASSIANO, Keila Mara. **Análise de Séries Temporais Usando Análise Espectral Singular (SSA) e Clusterização de Suas Componentes Baseada em Densidade**. Pontifícia Universidade Católica do Rio de Janeiro, 2014.

CHAPELLE, O.; SCHÖLKOPF, B.; ZIEN, A. **Semi-Supervised Learning**. Cambridge, 2006.

COLE, R. M. **Clustering with genetic algorithms**. Department of Computer Science, University of Western Australia, Australia: Citeseer, 1998.

FERNANDEZ, R. O. BACIA DE SANTOS **Sumário Geológico e Setores em Oferta**. Rodada Brasil 15, Concessões de Petróleo e Gás, ANP, 2017.

GUARIDO, Marcelo. **Machine learning in geoscience: facies classification with features engineering, clustering, and gradient boosting trees**. CREWES Research Report 30: 13.1–13.23, 2018.

GLOVER, Paul WJ. **Petrophysics**. University of Aberdeen, UK, p. 246-269, 2000.

HALL, Brendon. **Facies classification using machine learning**. The Leading Edge, v. 35, n. 10, p. 906-909, 2016.

JAIN, Anil K.; DUBES, Richard C. **Algorithms for clustering data**. Prentice-Hall, Inc., 1988.

KORCHI, Adil; MESSAOUDI, Fayçal; OUGHDIR, Lahcen. **Machine Learning and Deep Learning Revolutionize Artificial Intelligence**. International Journal of Scientific & Engineering Research Volume 10, Issue 9, 2019.

LIMA, Rafael Augusto Pires de et al. **Machine learning applications for geoscience problems**. Tese de Doutorado, University of Oklahoma Graduate College, 2019.

MACEDO, J. M. **Evolução tectônica da Bacia de Santos e áreas continentais adjacentes**. Boletim de Geociências da Petrobrás, v. 3, n. 3, p. 159-173, 1989.

MEIBOOM, Saul; GILL, David. **Modified spin-echo method for measuring nuclear relaxation times**. Review of scientific instruments, v. 29, n. 8, p. 688-691, 1958.

MITCHELL, Tom et al. **Machine learning**. Annual review of computer science, v. 4, n. 1, p. 417-433, 1990.

MOREIRA, Jobel Lourenço Pinheiro et al. **Bacia de Santos**. Boletim de Geociências da PETROBRAS, v. 15, n. 2, p. 531-549, 2007.

SANTOS, Rodrigo Andrade. **Interpretação e modelagem de perfis geofísicos de poço nos campos marginais de Quiambina e Fazenda Mamoeiro**. Monografia (Graduação). Salvador, IGEO/UFBA, 2010.

SERRA, Oberto. **The fundamentals of well log interpretation. Part 1: The acquisition of logging data**, p 423. Elsevier, 1984.

DE SOUZA, Leonardo Silveira; CHAVES SGARBI, Geraldo Norberto. **Bacia de Santos no Brasil: geologia, exploração e produção de petróleo e gás natural**. Boletín de Geología, v. 41, n. 1, p. 175-195, 2019.

TITTMAN, Jay. **Vertical resolution of well logs: Recent developments**. Oilfield Review;(Netherlands), v. 3, n. 3, 1991.

VIEIRA, J. Nona Rodada de Licitações-Bacia de Santos, v. 10, 2009. Acessado em 11 de agosto de 2022. [http://rodadas.anp.gov.br/arquivos/Round9/palestras/Santos\(portugues\).pdf](http://rodadas.anp.gov.br/arquivos/Round9/palestras/Santos(portugues).pdf).

OLIVEIRA, Lucas. **Modelagem geoquímica e mineralógica dos reservatórios carbonáticos do Pré-sal da bacia de Santos através de perfis de poços e inteligência artificial**. Tese (Mestrado) – Universidade de São Paulo, 2022.

VICTOR, R. A. **Multiscale, image-based interpretation of well logs acquired in a complex, deepwater carbonate reservoir**. 240 p. Tese (Doutorado) – The University of Texas at Austin, 2017.